Research & Development

2010

Mechanical Engineering Letters, Szent István University

Annual Technical-Scientific Journal of the Mechanical Engineering Faculty,
Szent István University, Gödöllő, Hungary

HU ISSN 2060-3789

Páter K. u. 1., Gödöllő, H-2103 Hungary
dekan@gek.szie.hu, www.gek.szie.hu,

Volume 4 (2010)

Selected Collection from the Research Results of Year 2010

Contents

Institute for Mathematics and Informatics

Professor dr. Sándor MOLNÁR
Director of the Institute

Dear Reader,

Allow me to give an overview of our institute's research and development activity in 2010. The Institute for Mathematics and Informatics consists of two departments**:** *Department of Mathematics* (head: Prof. dr. Zoltán Varga) and *Department of Informatics* (head: Prof. dr. Sándor Molnár).

Department of Mathematics

In 2010, the main part of the research activity in the Department was dedicated to different problems concerning models of Applied Population Dynamics. Research has been carried out in fields ranging from different monitoring problems of ecology to dynamic models in fisheries, from demographic applications of age-structured population models to the dynamics of insect populations. A new three-year research project entitled Dynamic Modelling of Stage-Structured Multi-Species Population Systems, financed by the Hungarian Scientific Research Fund has also been launched. The Department's article in this issue is an overview of recent results obtained in the methodological development of the dynamics of age-structured and spatially structured populations, including different applications.

Department of Informatics:

The Department's research activities focus on the following fields

- Application of enviromental informatics and modeling methodology. The main task in 2010 was to develop a comparative model for the full costs of power generation in order to make alternative capacity enlargement scenarios and to get a precise evaluation of real costs of power generation on the national scale. Another topic of research was the application of artificial neural networks in geostatistics.
- Embedded systems and intelligent metering. Embedded systems are capable of signal processing, detection, filtering and system modeling and controlling thus providing a methodology for state monitoring of complex non-fully observable systems in every timepoint and for the solution of specific decision problems.
- Systems theory – hierarchical switched systems. In earlier research we have examined switched linear systems from the aspect of controllability and observability and we provided sufficient and necessary differential algebraic conditions for these qualities. The present research focuses on Lie-algebraic conditions on observability and controllability for non-linear systems gained by switching on and off subsystems of the original switched linear system.

Application of Geostatistics in a Numerical Ground Flow Model using a Neural Network[1]

Sándor MOLNÁR[1], Ferenc SZIDAROVSZKY[2], Márk MOLNÁR[3], Emery A. COPPOLA Jr.[4]

[1]Department of Informatics, Institute for Mathematics and Informatics, SZIE

[2] Department of Systems and Industrial Engineering, University of Arizona, Tucson, USA

[3] Department of Economical Methodology, SZIE

[4]Department of Mining and Geological Engineering, University of Arizona, Tucson, Arizona

Abstract

Numerical modeling is the most advanced physical-based method for simulating complex ground water flow systems. Spatial and temporal variability of aquifer parameters, boundary conditions, and initial conditions (for transient simulations) can be explicitly represented across the space and time domains of numerical models. While this constitutes a powerful modeling advantage, it also presents the formidable challenge of overcoming model parameter uncertainty, which, to date, has not been satisfactorily resolved. In this paper, a new modeling paradigm is proposed based on explicitly combining numerical methods with artificial neural networks (ANN) predictive capability. Mathematical theory is presented, with a relatively simple two-dimensional test case used for proof of concept and to help overview relevant mathematical and modeling issues. This hybrid approach is not limited to ANN technology, but could use other approaches as well, such as regression or support vector machines.

Keywords

Numerical ground water modeling, artificial neural networks, inverse modeling, ground water management

1. Introduction

Numerical ground water flow models are used for a variety of water resources prediction and management problems, ranging from safe yield analysis to saltwater intrusion. Because of the difficulty in developing accurate numerical ground water models, significant research has been devoted to improving development methodologies for improved simulation and prediction capability.

To attempt to circumvent parameter uncertainty and simplifying mathematical and physical simplifications inherent to numerical models, which can often

[1]This work was supported by the OTKA 68187 project.

challenge our „ssumption that reality can be nearly expressed as a model" (Porter and Anderson, 2005), alternative ground water modeling and prediction approaches have been sought. Coppola and others (2003a,b; 2005a,b) in previous work demonstrate that artificial neural networks (ANNs) can achieve highly accurate ground water state predictions, even outperforming numerical models for site specific (e.g. monitoring well) prediction problems over discrete management periods. While ANN's may often provide superior real-time prediction capability for site-specific locations (e.g. monitoring wells) in the field over discrete management periods, numerical models will often be the superior approach for predicting responses at any locations of interest as pre-defined by a model grid superimposed across the field study area.

In this paper, a new methodology is proposed where ANN location- and time-specific prediction capability is explicitly integrated with the equation set of a companion numerical model, constituting a so-called hybrid ANN-Numerical model. This methodology is demonstrated for a steady-state ground water level prediction problem, but in theory may be applied to other types of problems like water quality, as well as transient problems. In the general approach, ANN's are developed with real-world (i.e. field) data to provide accurate state variable predictions at select field locations.

As numerical modeling is the most accepted advanced method for accurately predicting head and other ground-water responses over space and time, and the power of ANN technology for providing real-time prediction capability in the real-world system have only recently been investigated, the combination of the two methods as proposed here has until recently been neither relevant or applied in the ground-water field. While methods for combining ANN technology with numerical models have been investigated and implemented in hydrology and other fields (e.g. developing ANNs from numerical models to serve as more efficient "metamodels" for simulation and optimization, Rogers and Dowla., 1994; Coppola et al., 2006; analyzing and reducing output prediction errors from a numerical model with ANNs, Babovic et al. 2001), based upon a literature review, the method presented here has not been proposed in any other physical or engineering science.

The paper is outlined as follows; after overviewing the mathematical basis for numerical modeling and ANNs, the new hybrid approach is introduced from both mathematical and physical modeling viewpoints, and then illustrated with a relatively simple two-dimensional steady-state flow problem. This is followed by a discussion and conclusions, which present the numerical and modeling issues associated with this proposed hybrid modeling method, as well as important issues and proposed future research directions.

2. Numerical Ground Water Flow Models

The power of the numerical approach is predicated on its ability to explicitly represent via its equations spatial (heterogeneity), directional (anisotropy),

and/or temporal variability of aquifer properties and conditions at any space and/or time scale of interest. Numerical models are approximations to partial differential equations that represent the physics of the system. In ground water, the partial differential equation developed for two-dimensional, steady-state ground-water flow in an anisotropic heterogeneous confined aquifer is:

$$T_x(x,y)\frac{\partial^2 h}{\partial x^2} + T_y(x,y)\frac{\partial^2 h}{\partial y^2} = w(x,y) \tag{1}$$

where h represents the dependent variable head, which varies by location (x, y) across the model domain. T_x and T_y represent aquifer transmissivity terms in the x and y directions, respectively, and each will also vary over space in the two-dimensional head field. The variable w is a sink or source term, which also may vary over space, and for a confined aquifer, generally includes some combination of well extraction (i.e. pumping) and well recharge. If the w term is assumed to represent vertical leakage into or of out the aquifer across the upper and/or lower confining aquitard or low permeability layers, the aquifer can be considered a leaky confined aquifer.

Both T_x and T_y, which are the mathematical product of hydraulic conductivity and the saturated aquifer thickness in their respective directions, can be difficult to estimate accurately, with an order of magnitude discrepancy or more possible between reasonable estimates. G. de Marsily et al. (2000) assert that „transmissivity can vary by several orders of magnitude inside the same aquifer…near the Waste Isolation Pilot Plant (WIPP), the transmissivity of the regional aquifer varies by five orders of magnitude." Inherent inaccuracies and/or smoothing of these values over the space model domain contribute to inherent prediction errors or biases which can be significant.

The partial differential equation (Equation 1) representing the two-dimensional leaky confined ground water system under steady-state condition can be approximated by a numerical finite difference equation. By using numerical differentiation (Szidarovszky and Yakowitz, 1978), equation (1) has the second order approximation:

$$T_x(x_i,y_j)\frac{h_{i,j+1} - 2h_{ij} + h_{i,j-1}}{(\Delta x)^2} + T_y(x_i,y_j)\frac{h_{i+1,j} - 2h_{ij} + h_{i-1,j}}{(\Delta y)^2} = w(x_i,y_j) \tag{2}$$

for all cells, which is a system of linear algebraic equations.

The obvious difficulty posed by Equation 2 is assigning representative aquifer properties across the model grid based upon relatively few measured values, with inevitable errors (G. de Marsily et al., 2000). Numerical ground water models typically have thousands or tens of thousands of numerical equations like that of Equation 2, one corresponding to each active cell location, each requiring

model input parameters like transmissivity. For more complex problems, like transient simulations, or inclusion of surface water systems, the number of input parameters can increase significantly. Unavoidable discrepancies between parameter estimates and real-world values inevitably produce discrepancies or errors between model predictions and system behavior. It is this type of parameter uncertainty and associated prediction error that has led researchers to pursue other modeling methods or techniques for improving prediction accuracy.

3. Artificial Neural Networks in Ground Water Modeling

Artificial neural networks are „learning" based models that through processing of relevant historical data obtain a functional mapping between input predictor and output prediction variables. Rather than explicitly using physical laws and equations to model and simulate the system of interest, they utilize an empirical-based approach which often eliminates the need to include difficult to estimate hydrogeologic parameters. A more detailed overview of the ANN technology and its mathematical basis may be found in the work of Coppola et al. (2005a,b).

Coppola and others (2003a,b, 2005a,b) have applied ANN technology to water level and water quality prediction problems in real-world ground water systems, achieving high accuracy. The mean absolute error achieved by the ANN model for eight monitoring well locations was 0.16 meters (m) feet over a 71-day validation period, using ten consecutive 7-day stress periods, compared to 0.78 m feet achieved by an extensively calibrated numerical flow model developed by the water utility. The mean absolute error for the ANN model increased to 0.18 m when it reinitialized itself for each subsequent stress period prediction, rather than being initialized by the water levels measured in the field for the corresponding stress period. Thus, for this particular problem, the ANN model's mean absolute prediction error was on average about one-quarter to one-fifth of the error of the numerical model for the site specific prediction locations (i.e. monitoring wells) over the randomly selected validation period.

As mentioned previously, a fundamental advantage of ANNs is that they can learn system behavior of interest from variables that are more accessible and less uncertain. Therefore, instead of requiring highly uncertain material properties and conditions that are required for numerical model inputs, such as transmissivity, the ANN model uses input predictor variables that are much easier to measure with much lower uncertainty. Typical inputs include water levels, pumping rates, weather variables like precipitation and temperature.

4. Hybrid Modeling Approach

A numerical ground-water flow model, by explicitly representing physical ground-water laws (e.g. conservation of mass and momentum) within its set of finite difference or finite element equations, can predict the hydraulic head field in

response to ground-water conditions and stresses. For this prediction problem, each equation represents a single unknown head value at an individual cell or element within the model domain. The general form of the numerical equation embodies the estimated ground-water system properties, boundary conditions, and initial conditions (if transient). A ground-water flow model discretized into n active cells or elements consists of n independent equations, which are solved simultaneously to achieve the model solution of final heads for each cell or element. Formally, this set of numerical equations can be written in matrix form as:

$$Ah = b \tag{3}$$

where A represents the matrix of coefficients pertaining to the aquifer properties (e.g. T_x and T_y) and model domain dimensions, h represents the vector of unknown head values, and b represents boundary conditions (and initial conditions if transient).

The objective of geostatistical techniques and inverse modeling within the context of numerical ground-water modeling is to identify the aquifer parameters and conditions, as represented by the elements of the A matrix and b vector, that minimizes the error between the predicted (i.e. computed) h vector elements, h_1, h_2, ...h_n, and the real-world head values. By using ANN, some of the components of vector h are estimated more accurately than the numerical model solutions. Appending these to the numerical equations produces an overdetermined system.

With an overdetermined system of equations, the generated numerical solution cannot be exact, but rather, is approximate, which fits best the given equations. The actual solution method depends on the selected measure of fit. In our approach we use least squares and linear programming. The accuracy of the overdetermined solution for the ground-water problem depends upon the accuracy of the A matrix and b vector, as well as the number, accuracy, and distribution of the „known" head values. In more intuitive terms, the remaining n-m head values must „equilibrate" to the "known" head values, converging to a solution that is balanced between the pre-assigned known head values and the imposed physical conditions of the model. Three particular methods will be introduced and illustrated. Each method is based on a particular way of requiring the fulfillment of the additional equations provided by the more accurate ANN predictions. Method 1 requires that the additional equations with more accurate ANN prediction values will be completely satisfied, with these values substituted into the numerical equations. The resulting overdetermined system is solved by the least squares method. Imposing the more accurate „known values", however, could produce solutions which are not consistent with the physical system.

For illustration we provide the following simple example, where a head surface is correctly modeled as relatively flat (i.e. low hydraulic gradient), but incorrectly, with each cell's head predicted approximately 50 meters ***below*** mean sea level. Assume an alternative model correctly predicts a head value for one

cell location, 50 meters ***above*** mean sea level. While assigning this one head value may force the other water levels to approach their true elevations, it could also produce an inaccurate and unrealistic ground-water mound. Thus, the mean absolute error of the simulated surface may be reduced, but the relative contours of this surface completely distorted. In Methods 2 and 3, two versions of the optimal balance are introduced between the numerical model equations and the additional equations with the more accurate ANN predictions. Least squares and linear programming are the methodological basis for these procedures.

5. Twin Model Method Approach

The feasibility of improving numerical model predictions with ANN-model generated predictions was assessed using a so-called twin modeling approach. A heterogeneous model, representing the „real" aquifer system, designated Model R, was developed for a hypothetical ground water system. A second homogenous model, designated Model S, represented a simplified version of the real aquifer, which a modeler invariably constructs. In reality, the differences between Model R and Model S are not representative of discrepancies that would occur between a real-world system and a typical model representing such a system. Nevertheless, the concept and potential utility of the new modeling paradigm is illustrated with this simple synthetic twin model approach.

Figures 1a and 1b below depict the constant head boundaries and material properties for the twin heterogeneous and homogenous models, respectively.

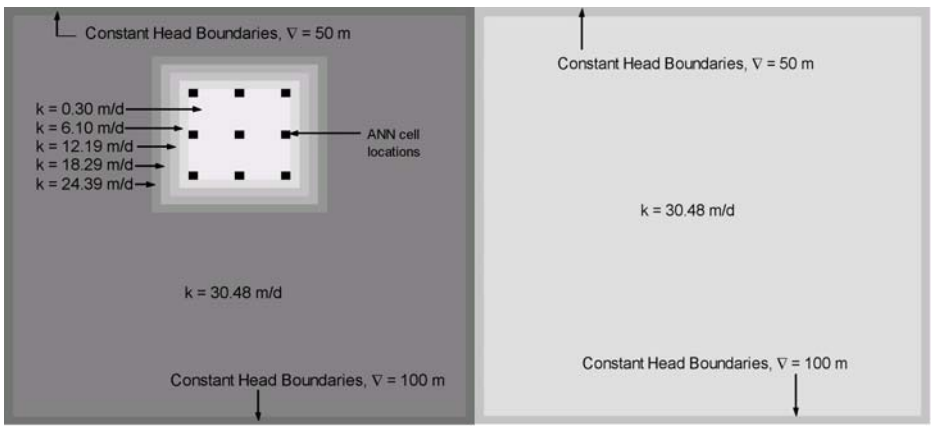


Figure 1a-b. Heterogeneous Hydraulic Conductivity Field and Boundary Conditions (Model R) and Homogeneous Hydraulic Conductivity Field and Boundary Conditions (Model S)

The domain is discretized into 51 rows and 51 columns for a total of 2,601 hundred cells, each measuring 100 by 100 meters. The upper and lower boundaries for both models were assigned the same constant head values (100

and 50 meters above mean sea level, respectively), with linearly declining constant head values assigned along the connecting boundaries. A constant recharge (i.e. leakage) value of 0.0023 meters/day was uniformly assigned across both model domains. The only difference between the two models is the transmissivity values assigned to the interior domain. Heterogeneous Model R contains regions of different transmissivity values, with distinct changes occurring within the shaded region, where progressively smaller transmissivity values transition into the interior of this region. Hydrogeologically, this region could represent transition into different sedimentary materials, from a coarser grained material like sandy gravel, where hydraulic conductivity is 30.5 m/day, to progressively finer grained material like silty sand, where hydraulic conductivity is 0.30 m/day. Model S, in contrast, is modeled as a homogenous aquifer, which even for a larger-scale ground water study under significant public scrutiny, is not unrealistic (Coppola and others, 2006).

Nine interior domain cell locations, depicted in Figure 1, were selected as the hypothetical ANN point locations, and were biased in the numerical domain region of higher errors, which corresponds to large hydraulic conductivity (i.e. transmissivity) discrepancies between twin models. The nine locations could pertain, for example, to monitoring well locations, for which ANNs have been trained to predict water levels/water level changes. Nine locations for a basin is a conservative number, where significantly more monitoring wells are typically available for ANN development. However, it provides a scenario that borders on worst case, and given the relatively small number of active cells (49 x 49) in the example numerical domain, as well as the relatively small differences in the head field for the twin models, provides a reasonable test case. To put further into perspective, the number of hypothetical ANN cell locations constitutes 0.37 *percent* of the total number of active numerical cells (2,401). Thus, for a model measuring 200 by 200 active cells, that percentage coverage value computes to 150 ANN locations for the 40,000 active cells. In a real-world setting, the ANN locations could similarly be biased towards numerical model regions prone to higher prediction errors and/or real-world locations of higher interest (e.g. closer to a contaminant plume, environmentally sensitive area, etc.).

For this twin modeling analysis, it was assumed that the ANN predictions exceed that of the numerical model by varying degrees of accuracy. Specifically, selected error ratios were 0.2, 0.4, 0.6, 0.8, and, for comparison, 1.0 (same error as generated by the numerical model). For example, assume the numerical model for a particular cell overpredicts the final head by 1.0 meters, for a final head of 33 meters above mean sea level. An error improvement ratio of 0.4 meters would predict a final head value of 32.4 meters above mean sea level. Because the homogeneous model uniformly predicted higher head values than the „true" (i.e. heterogeneous case), the assigned head values (i.e. hypothetical ANN values) with the lower prediction errors had uniformly lower values than what would be predicted by the homogenous model. For ease of implementation and analysis, the selected error ratios were uniformly assigned to all nine cell

locations for each „error" case. Thus, all nine cells were assigned a relative error of 0.2 for that error case, with no „mixing" of relative errors among the nine cells. The overdetermined system of equations was then solved using three different methods, which are presented in detail below.

6. Mathematical Methods

For the sake of mathematical simplicity assume that the last *m* components of the unknown vector *h* are known, and vector *h* is decomposed as

$$h = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}$$

where h_1 has $n-m$ elements and h_2 has *m* components. If matrix *A* and vector *b* are also decomposed accordingly, then equation (6.1) can be rewritten as follows:

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}. \tag{4}$$

Let h_2^* denote the vector of the known head values. The corresponding improved h_1 vector can be then obtained by using any one of the following methods.

Method 1

If we substitute h_2^* into equation (5) and rearrange the terms we get the following over determined system for h_1:

$$\begin{aligned} A_{11}h_1 &= b_1 - A_{12}h_2^* \\ A_{21}h_1 &= b_2 - A_{22}h_2^* \end{aligned} \tag{5}$$

The coefficient matrix and right hand side vector are

$$\begin{pmatrix} A_{11} \\ A_{21} \end{pmatrix} \text{ and } \begin{pmatrix} b_1 - A_{12}h_2^* \\ b_2 - A_{22}h_2^* \end{pmatrix}.$$

The number of unknowns is *n-m*, and we have *n* equations. A standard least squares method can be used to compute vector h_1.

Method 2
Equation (6) can be combined with the selected known head values as the appended system

$$\begin{aligned} A\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} &= b \\ h_2 &= h_2^* \end{aligned} \tag{6}$$

Here we have n unknowns and $n+m$ equations resulting again in an overdetermined system. The accuracy of h_2^* is much better than the accuracy of the coefficients of matrix A and vector b, therefore we want to satisfy the given values of h_2 more accurately than the numerical equations. A weighted least squares method can be used for this task.

Let w>1 be a parameter showing that the second equation of (6.5) has to be satisfied more accurately than the first equation, and consider the modified system

$$\begin{aligned} A\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} &= b \\ wh_2 &= wh_2^* \end{aligned} \tag{7}$$

This is again an overdetermined system with coefficient matrix and right hand side vector

$$\begin{pmatrix} A \\ 0 \quad wI \end{pmatrix} \text{ and } \begin{pmatrix} b \\ wh_2^* \end{pmatrix}$$

where I is the $m \times m$ identity matrix. A standard least squares method can be applied to find h_1 and h_2 that give the best fit of the equation system.

Method 3
The best fit of equation (6.6) can be also obtained by using linear programming. In the least squares method, the sum of the squares of the discrepancies between the left and right hand sides of equation (6.6) is minimized. In the case of linear programming the maximum discrepancy is minimized. Notice that squaring small numbers makes them even smaller and squaring numbers, which are larger than unity, makes them even larger. In the case of minimizing the largest discrepancy we do not need to square these values, their original values are used.

Notice that

$$\begin{pmatrix} A \\ 0 \quad wI \end{pmatrix} = \begin{pmatrix} a_{11} \ldots a_{1,n-m} & a_{1,n-m+1} \ldots a_{1n} \\ a_{21} \ldots a_{2,n-m} & a_{2,n-m+1} \ldots a_{2n} \\ \vdots & \\ a_{n1} \ldots a_{n,n-m} & a_{n,n-m+1} \ldots a_{nn} \\ 0 \quad \ldots \quad 0 & w \quad \ldots \quad 0 \\ \vdots & \\ 0 \quad \ldots \quad 0 & 0 \quad \ldots \quad w \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} b \\ wh_2^* \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \\ wh_{21}^* \\ \vdots \\ wh_{2m}^* \end{pmatrix}.$$

If ε denotes the maximum discrepancy between the left and right hand sides of equation (7), then

$$\left| \sum_{j=1}^{n-m} a_{ij} h_{1j} + \sum_{j=1}^{m} a_{i,n-m+j} h_{2j} - b_i \right| \le \varepsilon \tag{8}$$

for $i = 1,2,\ldots,n$, and

$$\left| wh_{2i} - wh_{2i}^* \right| \le \varepsilon \tag{9}$$

for $i = 1,2,\ldots,m$. Minimizing ε subject to constraints (8) and (9) can be rewritten as the following linear programming problem:

minimize ε ! subject to

$$\left.\begin{aligned} \sum_{j=1}^{n-m} a_{ij} h_{1j} + \sum_{j=1}^{m} a_{i,n-m+j} h_{2j} - \varepsilon \le b_i \\ \sum_{j=1}^{n-m} a_{ij} h_{1j} + \sum_{j=1}^{m} a_{i,n-m+j} h_{2j} + \varepsilon \ge b_i \end{aligned}\right\} (1 \le i \le n)$$

$$\left.\begin{aligned} wh_{2i} - \varepsilon \le wh_{2i}^* \\ wh_{2i} + \varepsilon \ge wh_{2i}^* \end{aligned}\right\} (1 \le i \le m)$$

where the unknowns are $h_{11},\cdots,h_{1,n-m},h_{21},\ldots,h_{2m}$ and ε. The simplex method and standard computer software can be used to solve this problem.

7. Modeling Results

Table 1 below compares the absolute mean errors and maximum errors for each cell that occurred for the homogeneous model relative to the heterogeneous

„real-world" system, for the three solution methods and the five error cases assigned to the nine cell locations in the domain.

Table 1. Comparison of Errors for All Active Cells (meters)

Case		Method 1	Method 2	Method 2	Method 3	Method 3
	Mean Absolute	Maximum	Mean Absolute	Maximum	Mean Absolute	Maximum
Normal	0.22	0.43	0.22	0.43	0.22	0.43
Error =0.2	0.04	0.1	0.04	0.1	0.15	0.22
Error = 0.4	0.13	0.24	0.13	0.24	0.23	0.33
Error = 0.6	0.22	0.37	0.21	0.37	0.35	0.48
Error = 0.8	0.3	0.51	0.3	0.51	0.48	0.64
Error = 1.0	0.39	0.65	0.39	0.65	0.61	0.79

As Table 1 shows, for the homogeneous and heterogeneous case prior to application of the three methods with assigned head values, the mean absolute error for all cells was 0.22 meters, with a maximum error of 0.43 meters.

Method 1 with ANN assigned errors of 0.6 for the nine ANN locations reduced the maximum error and matched the mean error for the uncorrected case. Method 2 reduced the overall error of the numerical model when the nine cell locations are assigned head values with an error at least 0.6 of what the numerical model would generate. Method 3 requires an error less than 0.4 to improve the overall solution in terms of mean absolute error, but the maximum error is lower for this case than without the ANN „corrected" values. What is interesting is that while lower errors assigned to the nine ANN locations, extrapolating to the 0.5 error or less on average, improve the overall solution, errors above this threshold but still below the values achieved by the numerical model increased the overall solution error. Thus, for this simple example, reducing the error approximately by half at the nine cell locations produces overall more accurate solutions, while errors above would increase the overall solution error, even when they remain below 1.0.

8. Discussion

Based upon the results obtained from the relatively simple two-dimensional steady-state numerical flow simulation cases presented here, an overdetermined system of equations can decrease the overall predictive error (e.g. mean absolute error) both within the areas in close proximity to the ANN assigned values and more distant locations. However, it has also been shown that the overall solution can degrade, even when the ANN assigned values exceed the accuracy of the „uncorrected" numerical model. This occurred for the higher error cases, which indicates that the magnitude of improvement achieved by the ANN models will partly determine whether the overall solution can be improved.

Intuitively, it may be expected that the cells closest to the ANN assigned values will show the greatest improvement. What is important to distinguish here, however, is an improvement between mean absolute predictive error and the simulated flow field. In some cases, an overall improvement in predicted head values, with some diminishment at other locations, may produce a flow field that is not consistent with the real-world system (i.e. flow paths), which the original numerical model may have re-produced relatively accurately. Therefore, in addition to considering the reduction in mean absolute error, the modeler must also assess how the overall system behavior is affected by the ANN values.

From mathematical theory, we know that in order to achieve relatively accurate predictions with the overdetermined system (i.e. hybrid numerical-ANN model), the physical-based model (i.e. numerical model) must be relatively accurate. Specifically, the elements of the A matrix and b vector must be representative of the real-world conditions in order to achieve an acceptable solution with the overdetermined system of equations. In addition, under these conditions, the solution computed with the overdetermined system of equations may be superior to a solution computed with a standard numerical model consisting of n equations in n unknowns. It can be proved mathematically, that if matrix A and vector b have no errors, and the additional results are exact, then the solution of the overdetermined system is also the true solution of the numerical model. If matrix A and vector b are incorrect, then the hybrid method attempts to preserve the very accurate known discrete head values and at the same time attempts to preserve the structure of the entire head surface.

It is postulated that the accuracy of the physical-based numerical model, in combination with the number, spatial distribution, and accuracy of the ANN prediction values will determine the degree of accuracy of the overdetermined solution. Assuming a reasonably accurate numerical model, as the accuracy and spatial distribution of the ANN predicted values increases, it is expected that the overall predictive performance will increase. If true, this could have implications when designing a monitoring network in a watershed or study area of interest. If data can be collected for ANN development at a discrete number of locations (i.e. monitoring wells), a trade-off in these locations should be sought, balancing increased predictive accuracy at locations of interest (e.g. plume boundary) without serious diminishment in predictive accuracy at more distant locations. It may also be possible to interpolate among ANN predicted values to increase the number of additional values assigned to the numerical model. There may also be optimal methods for locating the ANN points that bias improved accuracy to areas of higher interest and/or higher error, while not neglecting locations further away from these regions to preserve their accuracy and the integrity of the overall solution.

9. Conclusion

In this study, a hybrid numerical-ANN modeling approach is proposed for improving predictive capability for complex ground-water systems. Numerical

models and ANNs within the context of ground-water modeling and prediction each have relative advantages. By combining the two methods, it is believed that for well posed problems, the ANN predictions can be used to improve the overall accuracy of the numerical model solution. What makes this approach even more attractive is that the ANN predictions can still be considered independent of the hybrid model, enhancing real-time management capability at site specific locations for which the ANN model was developed. That these predictions can be further utilized by potentially increasing the accuracy of a companion numerical model adds further value, and would allow decision-makers to more effectively quantify the response of the system at locations or even state variables (e.g. stream stage elevations) that the ANN does not explicitly model.

For demonstration, a relatively simple two-dimensional steady-state test case was used to illustrate the mathematical and modeling approach of this method, where only the transmissivity field was different between twin models. Differences in boundary conditions between the real-world and the actual model introduce another source of prediction error, and may have to be treated differently by this hybrid method. For example, using the ANN model to predict relative state changes (e.g. drawdown), and then converting them to corresponding state values (e.g. head) in the numerical model relative to its state datum, rather than relative to the field. Still, in accordance with mathematical theory, it is expected that the findings here would be consistent for larger and more complex modeling problems, such as an unconfined aquifer under transient conditions. However, much more research is needed to investigate the strengths and limitations of this proposed modeling methodology, and the conditions under which it is best applied. Only through rigorous testing of this proposed modeling methodology under a variety of conditions will the potential merits of this hybrid modeling method be established. Based upon the results obtained, the hybrid method may hold promise as a powerful and flexible modeling approach and certainly warrants future investigation.

References

Babovic, V., R. Canizares, H. Rene Jensen, and A. Klinting. 2001. Neural Networks as Routine for Error Updating of Numerical Models. Journal of Hydraulic Engineering, 127, no. 3: 181-193.

Coppola, E., M. Poulton, E. Charles, J. Dustman, and F. Szidarovszky. 2003a. Application of Artificial Neural Networks to Complex Groundwater Management Problems. Journal of Natural Resources Research, 12, no. 4: 303-320.

Coppola, E., F. Szidarovszky, M. Poulton, and E. Charles. 2003b. Artificial Neural Network Approach for Predicting Transient Water Levels in a Multilayered Groundwater System Under Variable State, Pumping, and Climate Conditions. Journal of Hydrologic Engineering, 8, no. 6: 348-359.

Coppola, E. C. McLane, M. Poulton, F. Szidarovszky, and R. Magelky. 2005a. Predicting Conductance Due To Upconing Using Neural Networks, Journal of Ground Water, 43, no 6: 827-836.

Coppola, E., A. Rana, M. Poulton, F. Szidarovszky, and V. Uhl. 2005b. A Neural Network Model for Predicting Water Table Elevations. Journal of Ground Water, 43, no 2: 231-241.

Coppola, E. F. Szidarovszky, D. Davis, S. Spayd, M. Poulton, and E. Roman. Multiobjective Analysis of a Public Wellfield Using Artificial Neural Networks. (2006). Journal of Ground Water, in press.

Marsily, G. de, J.P. Delhomme, A. Coudrain-Ribstein, and A.M. Lavenue. 2000 Four decades of inverse problems in hydrogeology. Geological Society of America, Special Paper 348, 1-17.

Oreskes, N., K. Shrader-Frechette, and K. Belitz (1994). Verification, validation, and confirmation of numerical models in the earth sciences. Science, 263, no 5147: 641-646.

Ovaska, S. J. (ed.) (2005) Computationally Intelligent Hybrid Systems. Wiley-IEEE Press, Piscataway, NJ.

Porter, Eileen and D. Anderson. 2005 Multimodel Ranking and Inference in Ground Water Modeling. Journal of Ground Water, 43, no 4: 597-605.

Rogers, L. L., and Dowla, F. U. 1994. "Optimization of groundwater remediation using artificial neural networks with parallel solute transport modelling." J. Water Resources Research, 30, no. 2: 457-481.

Szidarovszky, F. and Yakowitz, S. 1978. Principles and Procedures of Numerical Analysis. Plenum Press, New York, New York.

Models of Applied Population Dynamics [2]

Zoltán VARGA[1], Zoltán SEBESTYÉN[1], Manuel GÁMEZ[2],
Tomás CABELLO[3], Anna ATTIAS[4]
[1]Department of Mathemtaics, Institute for Mathematics and Informatics,
[2]Department of Statistics and Applied Mathematics, University of Almería
[3]Research Center in Agro-Alimentary Biotechnology, University of Almería
[4]Department of Methods and Models for Economics, Territory and Finance,
Sapienza University of Rome

Abstract

The paper is an overview of our recent results achieved with different coauthors, concerning several research lines of applied population dynamics we initiated some years ago, and mostly published in 2010. First, based on the classical Leslie population model, a dynamic demographic model including controlled immigration is recalled, and applying the Perron-Frobenius theory of nonnegative matrices, a convergent optimal control algorithm is given, in order to address the sustainability problem of the considered pay-as-you-go pension system. Then, concerning the monitoring problem of particular population systems, observer design for certain trophic chains is presented. Furthermore, considering a fish population with a reserve area, we deal with the problem of steering the population into a desired new equilibrium, applying a time-dependent fishing effort as a control function. To this end, an optimal control problem is set up, which is numerically solved using an optimal control toolbox developed for MatLab. Finally, for the analysis of the efficiency of certain agents to be applied in biological pest control, corresponding mathematical models are presented that describe the dynamics of the interaction between predator, parasitoid and pest insect populations.

1. Introduction

In the paper our results concerning the methodological development in different fields of applied population dynamics are summarized. In Section 2, we present the analysis of the demographic background of the pay-as-you-go pension system. From Angrisani et al. (2004), we recall an age-structured two-sex demographic model and the existence of an asymptotic demographic equilibrium. For the analysis of the effect of a controlled immigration, in Angrisani et al. (2010), we developed an optimal control algorithm that steers

[2] The research has been supported by the Scientific and Technological Innovation Fund (of Hungary, ES-17/2008), the Ministry of Education and Sciences (of Spain, HH2008-0023) and the Hungarian Scientific Research Fund OTKA (K62000, K68187, K81279).

the population towards a demographic equilibrium, and at the same time minimizes the yearly immigration. Here we shortly summarize some of the results obtained for the Italian demographic data.

Section 3 is devoted to an illustrative example of a monitoring problem of population ecology studied in Varga et al. (2010). We shortly recall the observer design for certain trophic chains of the type *resource – producer – primary consumer*. This observer system makes it possible to effectively estimate the whole state process from the observation of one of the components of the chain.

In Section 4, we consider a problem of conservation ecology: controlling a spatially structured population back to an initial equilibrium state or to a required new equilibrium by an abiotic human intervention. The results presented here are recalled from Gámez et al. (2011), and deal with a fish population present in a free fishing area and in a reserve area, with migration between them. In order to deal with the problem of steering the population to equilibrium, an optimal control problem is set up, which is then numerically solved using an optimal control toolbox developed for MatLab.

Section 5 summarizes some recent methodological developments we achieved in application of population dynamics models to biological control of insect pests, cf. Gámez et al. (2010a). Based on density-dependent models of interacting populations, the effect of the release of the predator and parasitoid agents on the pest density is analyzed.

Finally, Section 6 relates the presented material to other results of the involved authors, also discussing some further possible extensions of the presented methods.

2. Demographic model for the pay-as you-go pension system

The sustainability of a pay-as-you-go pension system strongly depends on the underlying demographic process determining the proportion of the active subpopulation to pensioners, considered as a sustainability index. In our paper Angrisani et al. (2004), based on the classical Leslie population dynamics, a two-sex demographic model has been set up.

Let N be the maximal age for each sex in a human population (in our simulations N=110); for age groups defined on yearly basis (for i = 0, 1,…, N−1), let $x^F(t)$ and $x^M(t)$ be the N-dimensional population vectors for the female and the male sub-populations, respectively, α_i^F the female reproduction rates (counting all newborns of both sexes, i =15,…,50), ω_i^F and ω_i^M the female and male survival rates, respectively (i = 0,1,…, N−1), $\varphi := x_0^F(0)/x_0(0)$ the sex ratio. Then for the corresponding $N\times N$ Leslie matrices and for the dynamics of the female and male sub-populations, with the N-dimensional basic vector $e_1 = [1, 0, ..., 0]^T$, we have

$$L^F := \begin{bmatrix} 0 & 0 & \dots & \alpha^F{}_{15} & \dots & \alpha^F{}_{50} & 0 & \dots & 0 & 0 \\ \omega_0^F & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 & 0 \\ 0 & \omega_1^F & \dots & 0 & \dots & 0 & 0 & \dots & 0 & 0 \\ . & . & \dots & . & \dots & . & . & \dots & . & . \\ 0 & 0 & \dots & 0 & \dots & 0 & 0 & 0 & \omega_{N-2}^F & 0 \end{bmatrix}, \quad L^M := \begin{bmatrix} 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 & 0 \\ \omega_0^M & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 & 0 \\ 0 & \omega_1^M & \dots & 0 & \dots & 0 & 0 & \dots & 0 & 0 \\ . & . & \dots & . & \dots & . & . & \dots & . & . \\ 0 & 0 & \dots & 0 & \dots & 0 & 0 & 0 & \omega_{N-2}^M & 0 \end{bmatrix}$$

$$x^F(t+1) = L^F x^F(t), \; x^M(t+1) = L^M x^M(t) + \frac{(1-\varphi)}{\varphi} e_1 \circ L^F x^F(t) \qquad (t = 0, 1, 2, \dots),$$

where $\circ$ denotes the Hadamard product of vectors, $a \circ b = [a_1 b_1, a_2 b_2, \dots, a_n b_n]$. For the whole population, with state vector $x = [x^F \;\; x^M]^T$, we have the following dynamics:

$$x(t+1) = \Lambda x(t) \quad (t = 0, 1, 2, \dots),$$

where the $2N \times 2N$ block matrix Λ is defined as

$$\Lambda := \begin{bmatrix} L^F & O \\ d & 0 \\ O & L_1^M \end{bmatrix},$$

in which d and L_1^M denote the $(N-1)$-dimensional row vector

$$d := \begin{bmatrix} 0 & \dots & 0 & \frac{1-\varphi}{\varphi}\alpha_{15}^F & \dots & \frac{1-\varphi}{\varphi}\alpha_{50}^F & 0 & \dots & 0 \end{bmatrix}$$

and the $(N-1) \times N$ matrix

$$L_1^M := \begin{bmatrix} \omega_0^M & 0 & . & . & . & 0 & 0 \\ 0 & \omega_1^M & . & . & . & 0 & 0 \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 0 & 0 & . & . & . & \omega_{N-2}^M & 0 \end{bmatrix},$$

respectively. In Angrisani (2004), with the application of the Perron-Frobenius theory of nonnegative matrices, the problem of demographic equilibrium (i.e. the equilibrium age distribution) has been considered. Although the standard condition of irreducibility of matrix Λ does not hold, we have managed to proof the existence of a strictly dominant positive eigenvalue and a corresponding

positive eigenvector. Hence we have also proved that the age distribution of the population converges to a demographic equilibrium, where the population growth is proportional, with rate equal to the dominant eigenvalue. As a consequence of the demographic dynamics to the pay-as-you-go pension system, we have obtained a necessary and sufficient condition for the existence of a financial equilibrium at the long term demographic equilibrium.

Recently, in Angrisani et al. (2010), the above dynamic model is modified to include a controlled immigration. A convergent algorithm is given which steers the population towards a demographic equilibrium, and at the same time minimizes the yearly immigration. In the framework of this model, choosing a parameter, the consequences of different decision strategies can be numerically analyzed in terms of the *inverse old-age dependency ratio*, considered as *sustainability index*. The results are also compared with the projections calculated under the hypotheses that the yearly immigration is kept at different constant values. The simulation analysis for the *different immigration scenarios* is carried out with the Italian data of 2006 provided by the Italian Institute for Statistics. *Scenario A*: prescribed constant yearly immigration. In our case, according to the Italian law in force, this figure is yearly set at 180.000. *Scenario B*: the *de facto* present yearly immigration, kept constant, set at 500.000. *Scenario C*: immigration controlled by our model.

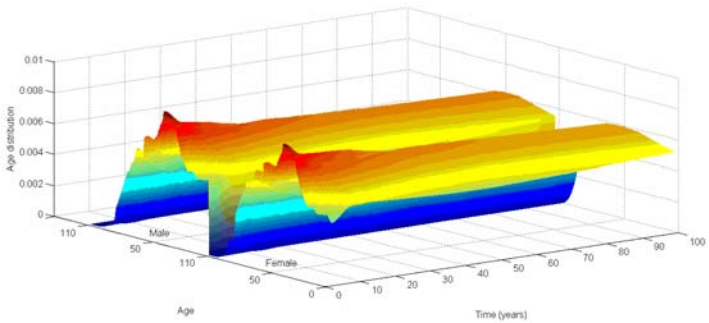


Figure 1. Dynamics of the age distribution for Scenario C

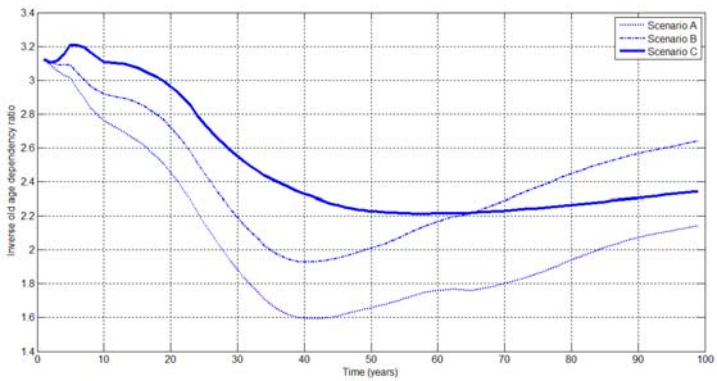


Figure 2. Inverse old-age dependency ratio for the three scenarios

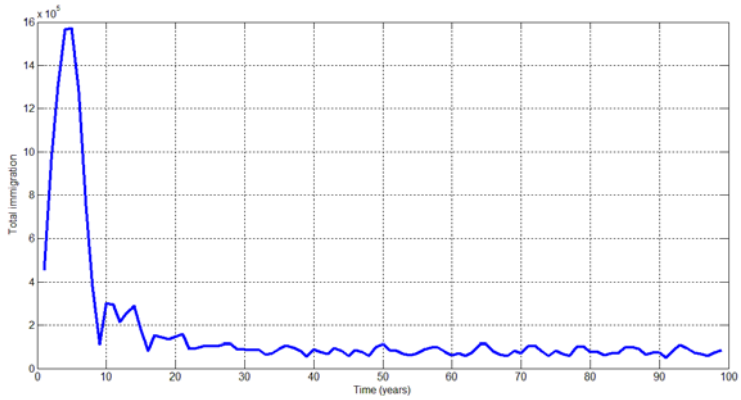


Figure 3. Total immigration per year for Scenario C

Of course, the analysis of the demography underlying the pension system is the very first step towards a complex solution also including the important socio-economic components.

3. Observer design for trophic chains

At the beginning of this research line concerning monitoring, the study was concentrated on basic qualitative properties of population ecological models. One of them is o*bservability* which, for population systems, means that from partial observation of the system (observing only one or several indicator species), in principle, the whole state process can be recovered. Recently, for different ecosystems, the so-called *observer system* (or state estimators) have been constructed that enables us to effectively estimate the whole state process from the observation. Below, from Varga et al. (2010) we shortly recall the observer design for certain trophic chains of the type *resource – producer – primary consumer*.

The considered model describes how a resource moves through a trophic chain. A typical terrestrial trophic chain consists of the following components:

resource, the *0th trophic level* (solar energy or inorganic nutrient),
which is incorporated by
a plant population, the *1st trophic level* (*producer*),
which transfers it to
a herbivorous animal population, the *2nd trophic level* (*primary consumer*).

According to the possible types of 0th level (energy or nutrient), two types of trophic chains can be considered: *open chains* (without recycling) and *closed chains* (with recycling). At the *0th trophic level, resource* is the common term for energy and nutrient.

Let x_0 denote the time-varying quantity of free resource, say nutrient present in the system, x_1 and x_2, in function of time, the biomass (or density) of the

producer (species 1) and the primary consumer (species 2), respectively. Let Q be the resource supply considered constant in the model. Let $\alpha_0 x_0$ be the velocity at which a unit of biomass of species 1 consumes the resource, and assume that this consumption increases the biomass of species 1 at rate k_1. A unit of biomass of species 2 consumes the biomass of species 1 at velocity $\alpha_1 x_1$, converting it into biomass at rate k_2. Both the plant and the animal populations are supposed to decrease exponentially in the absence of the resource and the other species, with respective rates of decrease (Malthus parameters) m_1 and m_2.

Finally, in a closed system the dead individuals of species 1 and 2 are recycled into nutrient at respective rates $0 < \beta_1 < 1$ and $0 < \beta_2 < 1$, while for an open system (where there is no natural recycling) $\beta_1 = 0,\ \beta_2 = 0$ holds. Then with model parameters

$$Q, \alpha_0, \alpha_1, m_1, m_2 > 0\ ;\ k_1, k_2 \in]0,1[;\ \beta_1, \beta_2 \in [0,1[,$$

for the trophic chain the following dynamic model can be set up:

$$\begin{aligned} \dot{x}_0 &= Q - \alpha_0 x_0 x_1 + \beta_1 m_1 x_1 + \beta_2 m_2 x_2 \\ \dot{x}_1 &= x_1(-m_1 + k_1 \alpha_0 x_0 - \alpha_1 x_2) \\ \dot{x}_2 &= x_2(-m_2 + k_2 \alpha_1 x_1) \end{aligned} \tag{3.1}$$

It can be shown that if the resource supply is high enough then this system has a unique equilibrium $x^* = (x_0^*, x_1^*, x_2^*) > 0$. Let us suppose that, for a monitoring of the system, the time-varying quantity of free resource (more precisely, its deviation from its equilibrium value) is observed:

$$y = x_0 - x_0^* \tag{3.2}$$

In Varga et al. (2010), it has been proved that observation system (3.1)-(3.2) is locally observable which means that near the equilibrium x^*, from observed function y, the state of system (3.1) can be uniquely recovered. For the actual state estimation an auxiliary system, the so-called observer system can be constructed, the solutions z of which exponentially approaches the unknown solution x of the original system.

Example 3.1. As a numerical example, we consider the following parameter values:

$$\begin{gathered} Q := 10;\ \alpha_0 := 0.3;\ \alpha_1 := 0.1;\ \beta_1 := 0.2; \\ \beta_2 := 0.3;\ m_1 := 0.1;\ m_2 := 0.4;\ k_1 := 0.5;\ k_2 := 0.5. \end{gathered}$$

Now the a positive equilibrium of system (3.1) is $x^* = (4.52,\ 8\ ,\ 5.78)$, and the observer system corresponding to (3.1)-(3.2) is

$$\begin{aligned}
\dot{z}_0 &= 10 - 0.3 z_0 z_1 + 0.2 \cdot 0.1 z_1 + 0.3 \cdot 0.4 z_2 + 10[y - (z_0 - x_0^*)] \\
\dot{z}_1 &= z_1(-0.1 + 0.5 \cdot 0.3 z_0 - 0.1 z_2) \\
\dot{z}_2 &= z_2(-0.4 + 0.5 \cdot 0.1 z_1) + 1[y - (z_0 - x_0^*)].
\end{aligned} \tag{3.3}$$

If we set initial condition $x(0) := (3, 7, 2)$ near the equilibrium $x^* = (4.52, 8, 5.78)$, and similarly, we consider another nearby initial condition, $z(0) := (2.9, 7.2, 1.8)$ for the observer system (3.3), Figure 2 shows that the corresponding solution *z* approaches the solution *x* of the original system (3.1).

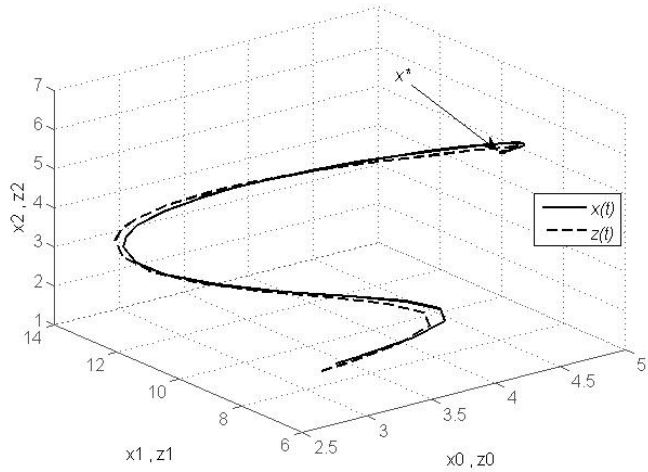


Figure 3.1. Solutions of system (3.1) and its observer (3.3)

4. Equilibrium control of a fishery model with reserve area

In addition to observation (or monitoring), control is also an important issue in conservation ecology. In particular, for a population system, in an appropriate setting, controllability implies that a disturbed ecosystem can be steered back to an equilibrium state, or to a new equilibrium by an abiotic human intervention. Recent research concerns the effective calculation of such control functions.

In Gámez et al. (2011) the following dynamic fishing effort model was considered. Let $x_1(t)$ and $x_2(t)$ be the biomass densities of the same fish population inside a free fishing area and a reserve area, respectively, at a time *t*. We suppose the population growth is logistic in each area, the fish subpopulation of the free area migrate into reserve area at rate m_{12}, and the inverse migration rate is m_{21}. Let *E* be the fishing effort applied in the free area. Then dynamics of the population system can be described by equations

$$\begin{aligned}\dot{x}_1 &= r_1 x_1\left(1-\frac{x_1}{K_1}\right) - m_{12}x_1 + m_{21}x_2 - qEx_1 \\ \dot{x}_2 &= r_2 x_2\left(1-\frac{x_2}{K_2}\right) + m_{12}x_1 - m_{21}x_2\end{aligned} \tag{4.1}$$

where r_1 and r_2 are the intrinsic growth rates, K_1 and K_2 the carrying capacities in the free and reserve areas, respectively, and q is the catchability coefficient. Setting the right-hand sides of system (4.1) equal to zero, an equation for x_1 is easily obtained and the latter has a unique positive solution x_1^* if the following conditions hold:

$$\frac{r_2(r_1 - m_{12} - qE)^2}{K_2 m_{21}} < \frac{(r_2 - m_{21})r_1}{K_1}, \tag{4.2}$$

$$(r_2 - m_{21})(r_1 - m_{12} - qE) < m_{12}m_{21}. \tag{4.3}$$

Now, inequality

$$\frac{r_1 x_1^*}{K_1} > r_1 - m_{12} - qE \tag{4.4}$$

implies the existence of a unique positive solution x_2^* and hence we have a unique positive equilibrium $x^* = (x_1^*, x_2^*)$. Conditions (4.2)-(4.4) will be supposed throughout this section.

Let us suppose first that the fish population is controlled by a time-dependent fishing effort of the form $E+u(t)$, where u is defined on a fixed time interval $[0,T]$. Then from model (1.1) we obtain the control system

$$\begin{aligned}\dot{x}_1 &= r_1 x_1\left(1-\frac{x_1}{K_1}\right) - m_{12}x_1 + m_{21}x_2 - q(E+u(t))x_1 \\ \dot{x}_2 &= r_2 x_2\left(1-\frac{x_2}{K_2}\right) + m_{12}x_1 - m_{21}x_2.\end{aligned} \tag{4.5}$$

Applying a linearization method, for control system (4.5) it can be proved that it is locally reachable from x^* in time T. The latter means that, using an appropriate control, the system can be steered from x^* to any point of a neighbourhood of x^*. In fact, for a given y^* from this neighbourhood, we can calculate a control function (a time-dependent fishing effort) such that for the corresponding solution x the distance between $x(T)$ and y^* is minimal (i.e. zero

by reachability). This optimal control problem can be numerically solved using an optimal control toolbox developed for MatLab in Banga et al. (2005) and Hirmajer et al. (2009). Once the system is in state y^*, with an appropriate constant fishing effort E1, this new equilibrium can be maintained, as illustrated in the following

Example 4.1. For a numerical illustration let us consider the following model:

$$\begin{aligned} \dot{x}_1 &= 0.7x_1\left(1-\frac{x_1}{10}\right)-0.2x_1+0.1x_2-0.25\cdot 0.9x_1 \\ \dot{x}_2 &= 0.5x_2\left(1-\frac{x_2}{2.2}\right)+0.2x_1-0.1x_2 \end{aligned} \tag{4.6}$$

For this system conditions (4.2)-(4.2) are satisfied, and it has a nonnegative equilibrium x^*=(4.85, 3.12), corresponding to a constant fishing effort E=0.9. Suppose that we want to steer the system in the new equilibrium y^*=(1.86, 2.43) by the time moment T=5. The obtained results are shown in Figure 4.1.

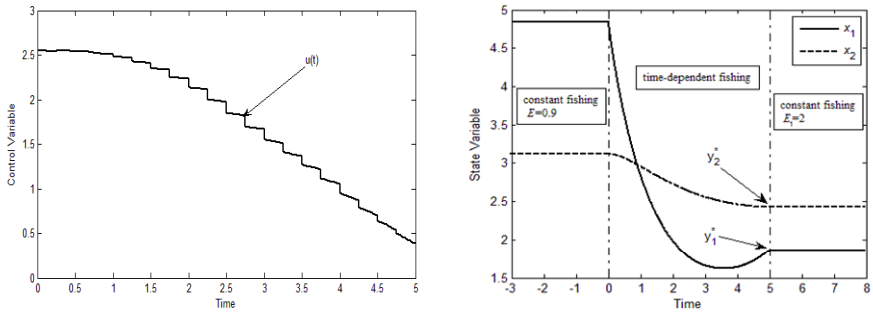
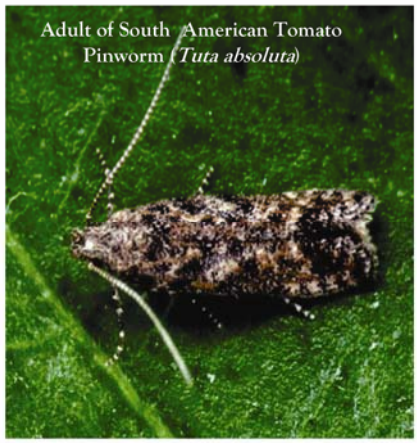


Figure 4.1. Control function and solution of system (4.5) for T=5, with initial value x(0)=(4.85,3.12)

5. Biological pest control

The South American tomato pinworm, Tuta absoluta (Meyrick), from its entrance in Spain in 2006, turned out to be a pest with hard economic consequences both in greenhouse and open field tomato crops (Cabello, 2009). From that time on, a work has been going on with autochthonous parasitoid natural enemies that could be adapted to this new prey/host and would admit the development of a biological pest control program (Cabello et al., 2009a,b). Principally, two species have been studied: the egg parasitoid Trichogramma achaeae (Nagaraja & Nagarkatti) and the predator Nesidiocoris tenuis (Reuter). This study was aimed at setting up a mathematical model that would enable us to program the use of the mentioned natural enemies (already commercially

available in Spain, France and Morocco), depending on the pest population dynamics in greenhouse tomato crops. The application of systems analysis techniques contributed to the better understanding of the biological control and its practical application (Gámez et al., 2000; Cabello et al., 2008; Tellez et al., 2009).



The present work has been motivated by the need to study the intra- and interguild relations between several species. These relations can play a very important role in the efficacy of the utilization of natural enemies. The objective of the present work has been to present mathematical phytophagous-predator (-parasitoid) models that would be applied in the utilization of T. absoluta in biological control.

The data used for the construction of the model have been collected in years 2008 and 2009 in experimental greenhouse, utilizing techniques of control by means of the application of Trichogramma achaeae and Nesidiocoris tenuis, as we can see in Cabello (2009). To these data the following mathematical models have been applied:

Phytophagous-parasitoid model

$$\dot{x}_1 = x_1(m_1 - \gamma_1 x_2)$$
$$\dot{x}_2 = x_2(-m_2 + k_{21}\gamma_1 x_1),$$

where x_1 and x_2 are the density (number/plant) of T. absoluta and T. acaeae, respectively, m_1 the growth rate of the pest, γ_1 the parasitation rate, m_2 the mortality rate and k_{21} the emergency rate of parasitoids.

Phytophagous-parasitoid-predator model

$$\dot{x}_1 = x_1\left(m_1 - \beta_1 x_1 - \gamma_1 x_2 - \alpha_1 x_3\right)$$
$$\dot{x}_2 = x_2\left(-m_2 + k_{21}\gamma_1 x_1 - \beta_2 x_2 - k_{23}\gamma_1 x_3\right)$$
$$\dot{x}_3 = x_3\left(m_3 + k_{31}\alpha_1 x_1 + k_{32}\gamma_1 x_2 - \beta_3 x_3\right).$$

Here x_3 is the density (number/plant) of N. tenuis, β_i (i = 1, 2, 3) the coefficients of intra-specific competition, α_i the predation rate, k_{23} the predation rate of parasitized prey; m_2 the mortality rate, k_{31} the conversion rate of prey, k_{32} the conversion rate of parasitized prey and the rest of coefficients have meanings analogous to those of model a). In both cases the adjustments have been done with the software SimFit 6.1 (Bardsley, 2007) for the data of the first and second generation of the pest (T. absoluta) collected in greenhouses.

The results obtained in Gámez et al. (2010a) are the following: To the data collected on the activity of predation by N. tenuis and parasitism by T. achaeae, on pest eggs (T. absoluta) in commercial tomato crops in Spanish greenhouses, a good fit was found for model a) of the phytophogous-parasitoid interaction. The model fitting is shown in Figure 1 for the first, and in Figure 2 for the second generation of pest. We can conclude, on the one hand, that the pest responds well to the presence of the parasitoid, with quickly reducing density, according to model a). On the other hand, we got good fit of for model b), therefore, the presence of the predator does not explain the dynamic behaviour of the pest population. Consequently, in the phytophagous-parasitoid-predator system, the main agent against T. absoluta is T. achaeae, and N. tenuis plays only a complementary role.

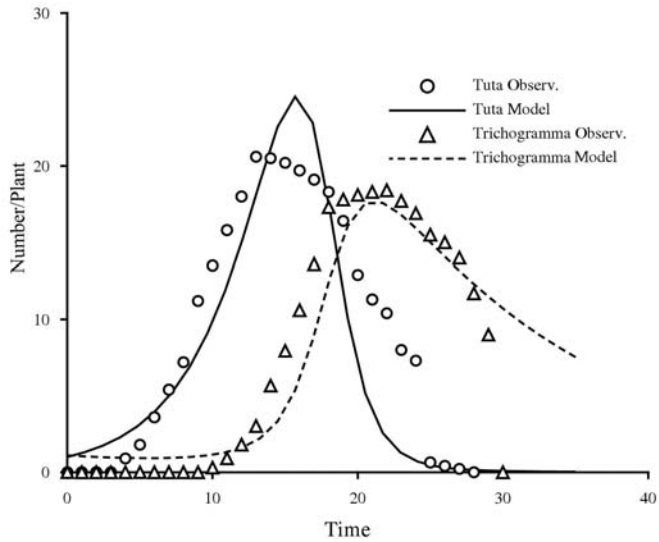


Figure 5.1. Dynamics of population of the first generation of T. absoluta, in comerciale greenhouse conditions of tomato, in the system phytophagous (T. absoluta)-parasitoid (T. achaeae)-predator (N. tenuis)

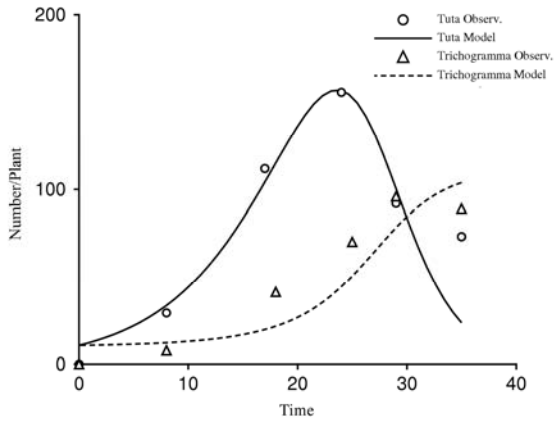


Figure 5.2. Dynamics of population of the second generation of T. absoluta, in comercial greenhouse conditions of tomato, in the system phytophagous (T. absoluta) --parasitoid (T. achaeae)-predator (N. tenuis)

6. Discussion and outlook

As an alternative to the optimal immigration control of the demographic model presented above, sustainability criteria of different forms can also be formulated, in order to find optimal immigration strategies. The presented dynamic analysis of the demography underlying the pension system is the very first step towards a complex model also including the socio-economic component. For a more comprehensive model, not only predictions on the demographic parameters should be taken into consideration, but a dynamic description of employment, economic growth (at least in terms of possible scenarios) and the social costs of the integration of immigrants have to be included.

Concerning the observer design for population systems, we note that and approach similar to the presented one can be applied for the monitoring of general food webs and many different multispecies dynamic population models, see e.g. Cressman and Garay (2003, 2006), Cressman et al. (2004), Garay (2002). The applied methodology can also be extended to more complex models of food webs, involving the observation of certain abiotic environmental components and/or certain indicator species. Furthermore, it can also be useful for the monitoring of population systems in changing environment, where the change of certain abiotic parameters of the ecosystem is governed by an "external" dynamic system (describing e.g. an industrial pollution or climatic changes), see Gámez et. al. (2008b). For an iterative method of observer design see Gámez et al. (2008a).

As for the dynamic fishing effort model with reserved area, the ongoing research is aimed at the monitoring and optimal harvesting problems. This fishery model can also be extended to general spatially structured (patchy) populations, with linear diffusion between patches. Recently, control problems

for trophic chains have been considered in López et al. (2010). Both monitoring and control problems of particular structured cell populations have been studied in the context of radiotherapy in Gámez et al. (2009).

For further recent research on biological control has been published in Gámez et al. (2010b), where with an appropriate functional response, a dynamic model was given for the application of the egg-larval parasitoid Chelonus oculator (Panzer) as an agent against the pest Beet Armyworm, Spodoptera exigua (Hubner) in greenhouse crops.

We also note that while in the study recalled in Section 5, the considered models are set up in terms of total densities of the species involved, in an ongoing research reported in Sebestyén et al. (2011), a stage-structured dynamic model has been built for the description of the temperature-dependent development of a host-parasitoid system of insect populations, in order to find the optimal timing of the release of parasitoid agents to control a pest population.

Finally, in the interface between applied population dynamics and population genetics, in Sebestyén et al. (2010) a dynamic harvesting model was set up, where in a size-structured population model, an additional genetic structure was also included, and the effect of harvesting on the genetic composition of the population was studied.

References

Angrisani, M., Attias, A., Bianchi S., Varga, Z. (2004), Demographic dynamics for the pay-as-you-go pension system. Pure Mathematics and Applications 15, 4, pp. 357-374.

Angrisani, M., Attias, A., Bianchi, S., Varga, Z. (2010), Dynamic analysis of the effect of immigration on the demographic background of the pay-as-you-go pension system. MAF 2010–Mathematical and Statistical Methods for Actuarial Sciences and Finance. 7-9 April, 2010, Ravello, Italy. Abstracts, pp. 6–7.

Banga, J. R., Balsa-Canto, E., Moles, C. G. and Alonso, A. A. (2005), Dynamic optimization of bioprocesses: Efficient and robust numerical strategies. Journal of Biothecnology 117, pp. 407-419.

Bardsley, W. G. (2007), SIMFIT: Simulation, fitting, statistics and ploting. Reference manual. University of Manchester, pp. 1-379.

Cabello, T. (2009), Cultivos hortícolas bajo abrigo: Control biológico de Tuta absoluta en tomate. XI Symposium Nacional de Sanidad Vegetal. Junta de Andalucía. Seville, pp. 199-217.

Cabello, T., Gámez, M., Varga, Z. (2008), An improvement of the Holling type III functional response in entomophagous species model. Journal of Biological Systems, 15, pp. 515-524.

Cabello, T., Gallego, J. R., Fernandez-Maldonado, F. J., Soler, A., Beltran, D., Parra, A., Vila, E. (2009a), The damsel bug Nabis pseudoferus as a new biological control agent of the South American Tomato Pinworm, Tuta absoluta, in tomato crops of Spain. Bulletin OIBC/wprs, 49, pp. 219-223.

Cabello, T., Gallego, J. R., Vila, E., Soler, A., Pino, M. del, Carnero, A., Hernández, E., Polaszek, A. (2009b). Biological control of the South American Tomato Pinworm, Tuta absoluta, with releases of Trichogramma achaeae in tomato greenhouses of Spain. Bulletin OIBC/wprs, 49, pp. 225-230.

Cressman, R., Garay, J.(2003), Stablility N-species coevolutionary systems. Theoretical Population Biology 64, pp. 519-533.

Cressman, R., Krivan, V., Garay, J. (2004), Ideal free distributions, evolutionary games, and population dynamics in multiple-species environments. American Naturalist 164, 4, pp. 473-489.

Cressman, R., Garay, J. (2006), A game–theoretical model for punctuated equilibrium: species invasion and stasis through coevolution. BioSystems 84, pp. 1-14.

Gámez, M., Carreño, R., Ándujar, A. S., Barranco, P., Cabello, T., (2000), Modelos matemáticos de depredador-presa en cultivos hortícolas en invernadero en el sudeste de la Península Ibérica. Boletín de Sanidad Vegetal-Plagas, 26, pp. 665-672.

Gámez, M., López, I. and Varga, Z. (2008a), Iterative scheme for the observation of a competitive Lotka–Volterra system. Applied Mathematics and Computation 201, pp. 811–818.

Gámez, M., López, I. and Molnár, S. (2008b), Monitoring environmental change in an ecosystem. Biosystems, 93, pp. 211-217.

Gámez, M., López, I., Garay, J., Varga. Z. (2009), Observation and control in a model of a cell population affected by radiation. Biosystems 96, pp. 172–177.

Gámez, M., Varga, Z., Garay, J., Vila, E., Cabello, T. (2010a), Mathematical Model for Biological Control of the South American Tomato Pinworm, Tuta absoluta (Lep.: Gelechiidae), with releases of Trichogramma achaeae (Hym.: Trichogrammatidae) and Nesidiocoris tenuis (Hem.: Miridae), in tomato greenhouses of Spain. IXth European Congress of Entomology, 22-27 August 2010, Budapest/Hungary Abstracts, p. 214.

Gámez, M., Torres, A., López, I., Garay, J., Varga, Z., Cabello, T. (2010b), Biological Control of Beet Armyworm, Spodoptera exigua (Lep.: Noctuidae), by Chelonus oculator (Hym.: Braconidae) in greenhouse crops. IXth European Congress of Entomology, 22-27 August 2010, Budapest/Hungary. Abstracts, p. 141.

Gámez, M., López, I., Garay, J. and Varga, Z. (2011), Monitoring and control in a spatially structured population model. International Workshop on

Biomathematics, Bioinformatics and Biostatistics, June 20-23, 2011, Sandander, Spain (submitted).

Garay, J. (2002), Many Species Partial Adaptive Dynamics. BioSystems 65, pp. 19-23.

Hirmajer, T., Balsa-Canto, E. and Banga, J. R., (2009). DOTcvpSB, a software toolbox for dynamic optimization in systems biology. BMC Bioinformatics 10, p. 199.

López, I., Gámez, M., Molnár S. and Varga, Z. (2010), Controlling a trophic chain to equilibrium. Third Conference on Computational and Mathematical Population Dynamics" (CMPD3). May 31-June 4, 2010, Bordeaux, France, Abstracts, p. 148.

Sebestyén Z., Varga Z., Garay J., and Cimmaruta R. (2010), Dynamic model and simulation analysis of the genetic impact of population harvesting. Applied Mathematics and Computation. 216, 2, pp. 565-575.

Sebestyén, Z., Varga, Z., Gámez, M., Carreño, R., Cabello, T. (2011), Dynamic model for a host-parasitoid interaction of insects. VII Congreso Nacional de Entomología Aplicada, 24-28 octubre, 2011, Baeza, España (submitted).

Tellez, M. M., Tapia, G., Gámez, M., Cabello, T., Emden, H.F. van (2009), Predation of Bradysia sp., Liriomyza trifolii and Bemisia tabaci by Coenosia attenuata in greenhouse crops. European Journal of Entomology, 106, pp. 199–204.

Varga, Z., Gámez, M., López, I. (2010), Observer design for open and closed trophic chains. Nonlinear Analysis: Real World Applications 11, pp. 1918-1924.

On the qualitative properties of hierarchical systems[3]

Sándor MOLNÁR[1], Alexandros SOUMELIDIS[2],
András EDELMAYER[2], Ferenc SZIGETI[3]
[1]Department of Informatics, Institute for Mathematics and Informatics,
[2] Systems And Control Laboratory, Computer and Automation Research Institute,
Hungarian Academy of Sciences
[3]Department of Control Engineering, University Los Andes, Merida, Venezuela

Abstract

Economic systems usually have a special structure, where subsystem i depends on only the states of the preceeding subsystems 1,2,.... i-1. These systems are called verticum-type systems. They have a very clear, systematic structure. Their special properties can be utilized both in theoretical and practical investigations. In addition, the computer implementation of such systems also has advantages. In this paper verticum-type linear system are analysed. Furthermore sufficient and necessary conditions are given for controllability and observability in these systems.

Keywords

Linear systems, observability, controllability, verticum-type systems

1. Introduction

Let $k \in N, m_i, n_i, r_i \in N\ (i = 1,\ldots,k)$. Assume that there are given matrices

$$A_{00} \in \mathbf{R}^{n_0 \times n_0}, B_0 \in \mathbf{R}^{n_0 \times m_0}, C_0 \in \mathbf{R}^{r_0 \times n_0}, \tag{1.1}$$

and for all $i = 1,2,\ldots,k$,

$$A_{i,i-1} \in \mathbf{R}^{n_i \times n_{i-1}}, A_{ii} \in \mathbf{R}^{n_i \times n_i}, B_i \in \mathbf{R}^{n_i \times m_i}, C_i \in \mathbf{R}^{r_i \times n_i}, \tag{1.2}$$

Let the 0^{th} subsystem be defined as

$$\begin{aligned} \dot{x}_0 &= A_{00}x_0 + B_0u_0 \\ y_0 &= C_0x_0 \end{aligned} \tag{V_0}$$

[3] This work was supported by the OTKA 68187 project.
This work was supported (in part) by the project TAMOP 421B, at the Széchenyi University of Győr

and for all $i = 1, 2, \ldots, k$, let the i^{th} subsystem be defined as

$$\begin{aligned} \dot{x}_i &= A_{ii}x_i + A_{ii-1}x_{i-1} + B_i u_i \\ y_i &= C_i x_i \end{aligned} \qquad (V_i)$$

Let $n = \sum_{i=0}^{k} n_i$, $m = \sum_{i=0}^{k} m_i$, $r = \sum_{i=0}^{k} r_i$ and define matrices $A \in \mathbf{R}^{n \times m}$, $B \in \mathbf{R}^{n \times m}$, $C \in \mathbf{R}^{r \times m}$ as follows

$$A = \begin{pmatrix} A_{00} & 0 & & 0 \\ A_{10} & A_{11} & & \\ 0 & A_{21} & A_{22} & \\ & & & 0 \\ 0 & & 0 & A_{kk-1}A_{kk} \end{pmatrix}$$

$$B = \begin{pmatrix} B_0 & & \\ & B_1 & 0 \\ 0 & & B_k \end{pmatrix} \quad C = \begin{pmatrix} C_1 & & \\ & C_2 & 0 \\ 0 & & C_k \end{pmatrix}$$

The system

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx \end{aligned} \qquad (V)$$

is called a verticum, and subsystem (V_i) is called the i^{th} subsystem of verticum (V) (Molnár, 1994).

We can define weaker types of verticums, as well. Assume that matrices (1.1) and (1.2) are given, and let

$$\widehat{B}_i \in \mathbf{R}^{n_i \times m}, \quad C_i \in \mathbf{R}^{r \times n_i}.$$

Define the 0^{th} subsystem as

$$\begin{aligned} \dot{x}_0 &= A_{00}x_0 + \widehat{B}_0 u \\ y_0 &= C_0 x_0 \end{aligned} \qquad (\hat{V}_0)$$

and for all $i = 1, 2, \ldots, k$ define the i^{th} subsystem as

$$\begin{aligned} \dot{x}_i &= A_{ii}x_i + A_{ii-1}x_{i-1} + \widehat{B}_i u_i \\ y_i &= C_i x_i. \end{aligned} \qquad (\hat{V}_1)$$

Furthermore a new 0^{th} subsystem

$$\begin{aligned} \dot{x}_0 &= A_{00}x_0 + B_0 u \\ y_0 &= C_0 x_0 \end{aligned} \qquad (\breve{V}_0)$$

is defined, and for all $i = 1, 2, \ldots, k$ a new i^{th} subsystem can be defined as

$$\begin{aligned} \dot{x}_i &= A_{ii}x_i + A_{ii-1}x_{i-1} + B_i u_i \\ y_i &= C_i x_i. \end{aligned} \qquad (\breve{V}_i)$$

As one may see neither system $\{(\hat{V}_1)\}$, nor $\{(\breve{V}_i)\}$ can be considered well defined systems, since besides the dynamic relations other conditions also hold between them. Define now matrices

$$A \in \mathbf{R}^{n\times m}, \quad B \in \mathbf{R}^{n\times m}, \quad C \in \mathbf{R}^{r\times n}$$

or

$$\widehat{B} \in \mathbf{R}^{n\times m} \text{ and } \widehat{C} \in \mathbf{R}^{r\times n}.$$

Construct the matrices

$$\widehat{B} = \begin{pmatrix} \widehat{B}_0 \\ \widehat{B}_1 \\ \vdots \\ \widehat{B}_k \end{pmatrix} \text{ and } \breve{C} = (\breve{C}_0 \breve{C}_1 \ldots \breve{C}_k).$$

By using their notations, subsystems $(\hat{V}_1)$ and $(\breve{V}_i)$ can not be considered as regular subsystems of systems

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx \end{aligned} \qquad (\hat{V})$$

and

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx = \sum_{i=0}^{k} \breve{C}_i x_i \end{aligned} \qquad (\breve{V})$$

respectively (Molnár, 1989).

Therefore system $\hat{V}$ is called an observability verticum consisting of subsystems $\hat{V}_i$. Similarly, system $\breve{V}$ is called a controllability verticum consisting of subsystems $\breve{V}_i$. It is obvious, that if a system is observability and also controllability verticum, then it is a verticum in the sense given above.

We may define an even weaker type of verticum as follows. Consider therefore the 0^{th} subsystem

$$\begin{aligned} \dot{x}_0 &= A_{00}x_0 + \widehat{B}_0 u \\ y_0 &= \breve{C}_0 x_0 \end{aligned} \qquad (V_0')$$

and for all $i = 1, 2, \ldots, k$, the i^{th} subsystem

$$\begin{aligned} \dot{x}_i &= A_{ii}x_i + A_{ii-1}x_{i-1} + \widehat{B}_i u_i \\ y_i &= \breve{C}_i x_i. \end{aligned} \qquad (V_i')$$

We can see that (V_0'), … , (V_i') may not be considered as subsystems of the system

$$\begin{aligned} \dot{x} &= Ax + \widehat{B}u \\ y &= Cx \end{aligned} \qquad (V')$$

in the usual sense. In the case of these weaker conditions (V') is called a system with verticum type dynamism consisting of subsystems (V_i').

2. Controllability and Observability of Verticum Type Systems

In this section conditions will be derived for observability and controllability of verticum type systems. In our investigations the results of Kalman will be applied.

It is well known that the basis solutions of any system can be obtained as the solution of the initial value problem

$$\dot{x} = Ax, \quad x(0) = I_{n\times n},$$

where $I_{n\times n}$ is the n dimensional identity matrix. It is also known that with constant matrix A, the solution has the form $\exp(At)$, but this exponential can not be expressed by using only the exponentials of the blocks A_{ii}, … A_{ii-1}, since these matrices usually do not commute. That is, relations

$$A_{ii-1}A_{i-1i-1} = A_{ii}A_{ii-1} \quad (1 = 1, 2, \ldots, k) \tag{2.1}$$

are not necessarily satisfied. Therefore the computation of the exponential $\Phi(t) = \exp(At)$ will be replaced by the recursive solution of the initial value problem

$$\begin{aligned} \dot{x}_0 &= A_{00}x_0 \quad x_0(0) = \xi_0 \\ \dot{x}_1 &= A_{11}x_1 + A_{10}x_0 \quad x_1(0) = \xi_1 \\ &\vdots \\ \dot{x}_k &= A_{kk}x_k + A_{kk-1}x_{k-1} \quad x_k(0) = \xi_k. \end{aligned}$$

The first equation implies that

$$x_0(t) = \exp A_{00}t_0 \tag{2.2}$$

and from the j^{th} equation we conclude that

$$x_j(t) = \exp A_{jj}t_j + \int_0^t \exp A_{jj}(t - t_1)A_{jj-1}x_{j-1}(t_1)dt_1. \tag{2.3}$$

These relations enable us to determine the blocks of $\exp(At)$ in a recursive manner:

$$t_0 \to x_1(t_0) = \exp A_{11}t_{01} \int_0^t \exp A_{11}(t_0 - t_1)A_{10}x_0(t_1)dt_1 =$$

$$= \exp A_{11}t_{01} + \int_0^t \exp A_{11}(t_0 - t_1)A_{10} \exp A_{00}t_1 dt_{10}.$$

By using induction we can easily verify that the solution of the j^{th} equation is as follows:

$$t_0 \to x_j(t_0) = \sum_{i=0}^{j} \int_0^{t_0} \dots \int_0^{t_{j-i-1}} \prod_{l=j}^{i+1} \exp A_{ll}(t_{j-l} - t_{j-l+1}) \cdot \\ \cdot A_{ll-1} \exp A_{ii} t_{j-i} dt_{j-i} \dots dt_{1i}, \tag{2.4}$$

which can be verified by substituting relation for $j := j+1$:

$$t_0 \to x_{j+1}(t_0) = \exp A_{j+1\,j+1} t_{0\,j+1} + \int_0^{t_0} \exp A_{j+1\,j+1}(t_0 - t_1) \cdot$$

$$\cdot A_{j+1\,j} x_j(t_1) dt_1 = \int_0^{t_0} \exp A_{j+1\,j+1}(t_0 - t_1) A_{j+1\,j} \left(\sum_{i=0}^{j} \int_0^{t_1} \dots \right.$$

$$\left. \dots \int_0^{j-1} \prod_{l=j}^{i+1} \exp A_{ll}(t_{j+1-l} - t_{j-l}) A_{ll-1} \exp A_{ii} t_{j+1-i} dt_{j+1-i} \dots dt_2 \right)$$

$$dt_{1i} + \exp A_{j+1+1} t_{0\,j+1} = \sum_{i=0}^{j+1} \int_0^{t_0} \dots \int_0^{t_{j+i-i-1}} \prod_{l=j+1}^{i+1}$$

$$\exp A_{ll}(t_{j+1-l} - t_{j+1-l+1}) A_{ll-1} \exp A_{ii} t_{j+1-i} dt_{j+1-i} \dots dt_{1i}.$$

Thus, x_j is really given by . Therefore the blocks Φ_{ij} of Φ can be given as

$$\Phi_{ji}(t_0) = \int_0^{t_0} \dots \int_0^{t_{j-i-1}} \prod_{l=j}^{i+1} \exp A_{ll}(t_{j-l} - t_{j-l+1}) \cdot \\ \cdot A_{ll-1} \exp A_{ii} t_{j-i} dt_{j-i} \dots dt_1 \tag{2.5}$$

for $i \le j$ and $\Phi_{ji}(t_0) = 0$ if $i > j$.

By applying the method of Kalman, we can verify the following result.

Theorem 2.1 The verticum type system is reachable from 0 *(*or controllable to 0*)* if and only if functions

$$t_0 \to B_i^* \sum_{j=i}^{k} \int_0^{t_0} \dots \int_0^{t_{j-i-1}} \exp A_{ii}^* t_{j-i} \prod_{l=i+1}^{j} A_{ll-1}^* \cdot \\ \cdot \exp A_{ll}^*(t_{j-l} - t_{j-i+1}) dt_{j-i} \dots dt_{1j} \quad (j = 0,1,\dots,k)$$

are all zeros, then all vectors $\xi_0, \xi_1, \ldots, \xi_k$ *are necessarily zero.*

Proof If function $t_0 \to B^* \exp(A^* t_0)$ is expressed by its blocks, then the functions given in the Theorem are obtained. By using the Theorem of Kalman we conclude that reachability from 0 is equivalent to the fact that if function $t \to B^* \exp A^*(T-t)$ is identically zero, then $\xi 0$ is necessarily zero. This condition is equivalent to the assertion. +

Similar results are true for the observability of verticum type systems as it is given next.

Theorem 2.2 The verticum type system (V) is observable if and only if functions

$$t_0 \to C_j \sum_{i=0}^{i} \int_0^{t_0} \ldots \int_0^{t_{j-i-1}} \prod_{l=j}^{i+1} \exp A_{ll}(t_{j-l} - t_{j-l+1}) A_{ll-1} \cdot$$
$$\cdot \exp A_{ii} t_{j-i} dt_{j-1} \ldots dt_{1i} \quad (j = 0,1,\ldots,k)$$

are all zeros, then all vectors $\xi_0, \xi_1, \ldots, \xi_k$ *are necessarily zero.*

Since the proof is analoguous to that of Theorem 2.1, the details are omitted. Let $i, j \in Z_0, i \le j$. Denote the set

$$(l \in Z_0, \quad i \le l \le j)$$

by $[i, j]$. The subsystem $(V_{[i,j]})$ of verticum (V) is defined as

$$\dot{x}_i = A_{ii} x_i + B_i u_i$$
$$x_l = A_{ll} x_l + A_{ll-1} + B_l u_l \quad (l \in i+1, j),$$

which is itself a verticum. One may easily verify that the controllability and observability of (V) do not imply the same for subsystems $(V_{[i,j]})$, but the following opposite-type statement holds.

Theorem 2.3 Let $0 \le i_1 \le j_1, i_2 \le j_2, \ldots, i_l \le j_l \le k$ integers such that

$$\bigcup_{p=1}^{t} [i_p, j_p] = [0, k] \tag{2.6}$$

holds. If all subverticums $(V_{[i_p, j_p]}\ (p = 1,2,\ldots,)$ are reachable from 0 *(*or controllable to 0*)*, then verticum (V) is also reachable from 0 *(*or controllable to 0*)*.

Proof. We shall use Theorem 2.1 in proving the assertion. Let $\xi_0 \in \mathbf{R}^{n_0}$, $\xi_1 \in \mathbf{R}^{n_1}, \ldots, \xi_k \in \mathbf{R}^{n_k}$ be vectors such that all functions defined in Theorem 2.1 are identically zero. Select a subverticum $(V_{[i_p, j_p]})$ such that $k \in [i_p, j_p]$. The controllability of $(V_{[i_p, j_p]})$ implies that all vectors $\xi_{i_p}, \xi_{i_p+1}, \ldots \xi_k$ are zero. Assume that it is already proved that all vectors $\xi_g, \xi_{g+1}, \ldots, \xi_k$ are zero. If $g = 0$ then the assertion is verified. If $g \geq 1$, then we can find a subverticum $(V_{[i_p, j_p]})$ such that $g - 1 \in [i_p, j_p]$.
Then

$$t_0 \to B_i^* \sum_{j=1}^{k} \int_0^{t_0} \ldots \int_0^{t_{j-i-1}} \exp A_{ii}^* t_{j-i} \prod_{l=i+1}^{j} A_{ll-1}^* \cdot$$
$$\cdot \exp A_{ll}^* (t_{j-l} - t_{j-l+1}) dt_{j-1} \ldots dt_{1j} =$$
$$= B_i^* \sum_{j=1}^{q-1} \int_0^{t_0} \ldots \int_0^{t_{j-i-1}} \exp A_{ii}^* t_{j-i} \prod_{l=i+1}^{j} A_{ll-1}^* \cdot$$
$$\cdot \exp A_{ll}^* (t_{j-l} - t_{j-l+1}) dt_{j-i} \ldots dt_{1j} = 0.$$

$(i \in [i_p, j_p])$, and the controllability of subsystem $(V_{[i_p, j_p]})$ implies that all vectors $\xi_i, (i \in [i_p, j_p])$ are zero. Therefore $\xi_{g-1} = 0$. Then repeat the above argument with g being replaced by $g - 1$ and so on, until it is verified that all vectors $\xi_0, \xi_1, \ldots, \xi_k$ equal zero. Then Theorem 2.1 implies the assertion.

Similar result holds for observability. This assertion is formulated next.

Theorem 2.4 Let $0 \leq i_1 \leq j_1$, $i_2 \leq j_2, \ldots, i \leq j_l \leq k$ integers such that condition holds. If all subverticums $(V_{[i_p, j_p]})$ are observable, then verticum is also observable.

The proof of this result can be performed analoguously to that of Theorem 2.3, and the assertion of Theorem 2.2 is now applied. The details are not given here.
Some special cases are analysed next.

Corollary 2.1 If all subsystems

$$\dot{x}_i = A_{ii} x_i + B_i u (i \in [0, k])$$

are reachable from 0 (or controllable to zero), then verticum is also reachable from 0 or controllable to zero. It is obvious that the assertion of the Theorem holds

in the special case, when verticum (V) is a simple decomposition of subverticums $(V_{ii})(i \in [0,k])$. We can also prove this special statement directly. Let $T > 0$, and let $W_i(t)$ denote the Kalman-matrix of the i^{th} subsystem, that is,

$$W_i(T) = \int_0^{t_0} \exp A_{ii}(T-t) B_i B_i^* \exp A_{ii}^*(T-t) dt.$$

Then the matrices are invertable. Let $\xi_0 \in \mathbf{R}^{n_0}, \xi_1 \in \mathbf{R}^{n_1}, \ldots, \xi_k \in \mathbf{R}^{n_k}$ be arbitrary vectors. Let controls $u_i : [0,T] \to \mathbf{R}^{m_i}$ be furthermore defined recursively in the following way. Set

$$u_0(t) := -B_0^* \exp A_{00}^*(T-t) W_0(T)^{-1} \exp A_{00} T \xi_0,$$

and let the corresponding trajectory be the solution of the initial value problem

$$\dot{x}_0 = A_{00} x_0 + B_0 u_0, \quad x_0(0) = \xi_0$$

If controls $u_i : [0,T] \to \mathbf{R}^{m_i}$ and trajectories $x_i : [0,T] \to \mathbf{R}^{n_i}$ are already defined for $i \le j < k$, then control $u_{j+1} : [0,T] \to \mathbf{R}^{m_{j+1}}$ is given as

$$u_{j+1}(t) = -B_{j+1}^* \exp A_{j+1\,j+1}^*(T-t) W_{j+1}(T)^{-1} \cdot$$
$$\cdot \left(\exp A_{j+1\,j+1} T_{j+1} + \int_0^T \exp A_{j+1\,j+1}(T-t) A_{j+1\,j} x_j(t) dt \right).$$

These inputs $u_j, (j \in [0,k])$ control vectors $\xi_0, \xi_1, \ldots, \xi_k$ to zero on the interval $[0,T]$. To prove this assertion observe that the Cauchy-formula for $i = 0$ implies that

$$x_0(T) = \exp A_{00} T_0 + \int_0^T \exp A_{00}(T-t) B_0 u_0(t) dt =$$
$$= \exp A_{00} T_0 + \int_0^T \exp A_{00}(T-t) B_0 \left(-B_0^* \exp A_{00}^*(T-t) W_0(T)^{-1} \cdot \right.$$
$$\left. \cdot \exp A_{00}\, T_0 \right) dt = \exp A_{00} T_0 - \int_0^T \exp A_{00}(T-t) B_0 B_0^* \cdot$$
$$\cdot \exp A_{00}^*(T-t) dt W_0(T)^{-1} \exp A_{00} T_0 = \exp A_{00} T_0 - W_0(T) W_0(T)^{-1} \cdot$$
$$\cdot \exp A_{00} T_0 = 0,$$

and for all $i>0$,

$$x_i(T)=\exp A_{ii}T_i+\int_0^T \exp A_{ii}(T-t)\left(A_{ii-1}x_{i-1}(t)+B_iu_i(t)\right)dt=$$

$$=\exp A_{ii}T_i+\int_0^T \exp A_{ii}(T-t)\Bigg(A_{ii-1}x_{i-1}(t)+B_i\left(-B_i^*\exp A_{ii}^*(T-t)\right)\cdot$$

$$\cdot W_i(T)^{-1}\Bigg(\exp A_{ii}T_i+\int_0^T \exp A_{ii}(T-\tau)A_{ii-1}x_i-1(\tau)d\tau\Bigg)\Bigg)dt=$$

$$=\exp A_{ii}T_i+\int_0^T \exp A_{ii}(T-t)A_{ii-1}x_{i-1}(t)dt-$$

$$-W_i(T)W_i(T)^{-1}\left(\exp A_{ii}T_i+\int_0^T \exp A_{ii}(T-t)A_{ii-1}x_{i-1}(t)dt\right)=0.$$

By aplying the rank-condition of Kalman we obtain the following

Corollary 2.2 If $\mathbf{rank}(B_i,A_{ii}B_i,\ldots,A_{ii}^{n_{i-1}}B_i)=n_i\ \ (i\in[0,k])$ then verticum (V) is reachable from 0 (or controllable to zero) on any interval $[0,T]$ that these rank-condition do not depend on matrice $A_{i,i-1}$. This fact also illustrate that the controllability of subverticum is much stronger property than the controllability of the entire verticum (V).

Consider next the same special case also for observability (Molnár, 2009).

Corollary 2.3 Consider verticum (V). If all subsystems

$$\dot{x}_i=A_{ii}x_i+B_iu\ \ (i\in[0,k])$$
$$y_i=C_ix_i$$

are observable, then verticum (V) is also observable. Obviously, this special case arises when verticum (V) is a simple composition of subverticums (V_{ii}). We could give again a direct proof for this simplified statement, which is analoguous to that of Corollary 2.1, so the details are omitted. One may also easily verify that the conclusion of Corollary 2.2 remains valid for observability, as well.

Corollary 2.4 Reachability (or controllability) is independent of the observability matrix, and also observability is independent of the controllability

matrix. Therefore Theorem 2.1 remains valid for controllability verticum $(\breve{V})$, and also, Theorem 2.2 holds for observability verticum $(\hat{V})$.

3. Conclusions

In this paper we have introduced verticum type systems and we have proved that if the subsystems are controllable (or observable) then the whole verticum is also controllable and observable. Further conditions for controllability and observability of verticum type systems can be a subject of a future paper.

References

KALMAN, R.E., FALB, P., and ARBIB, M. (1969): Topics in Mathematical System , McGraw-Hill, New York, 1969.

KAILATH THOMAS (1980): Linear Systems. Prentice-Hall., Englewood Cliffs, 1980.

MOLNÁR S. (1989): A Special Decomposition of Linear Systems, Belgian Journal of Operations Research, Statistics and Computer Science, Vol. 29. No. 4, pp. 1-19, 1989.

MOLNÁR S. (1994): "On "Verticum"-Type Linear Systems with Time-Dependent Linkage", Applied Mathematics and Computation, Vol. 60., pp. 89-102., 1994, (társszerző: Szigeti F.).

MOLNÁR S., SZIGETI F. (2009): Generalized Fuhrmann's rank condition for controllability and reachability of time-varying discrete-time linear systems, PU.M.A. Vol. 19 (2008), No. 1, pp. 55–66.

Institute for Process Engineering

Professor Dr. János BEKE
Director of the Institute

Dear Reader,

Following our tradition, on behalf of my colleagues in Institute for Process Engineering I would like to present our most significant scientific activities from the year of 2010.

The Institute for Process Engineering – as it is well known – consists of three professional working areas as follows:

- Department of Automotive Technology
- Department of Energetics
- Department of Measurement Technology

According to our traditions the main scientific goals of the institute in the scientific area are to research knowledge, needed to meet the expectations of qualified environmental-conscious activities in development as well as establishments of engineering systems, processes, technologies and equipment.

Department of Energetics carried on its research of caloric, energetic and food-industrial operation disciplines which give theoretical as well as practical knowledge to create, develop or derive sustainable environment-industrial technologies.

As its scientific mission *Department of Automotive Technology* kept on several institutional purposes by researching different questions related to vehicle constructions, off-road technologies automotive mechatronics.

Department of Measurement Technology continued its research activities in developing knowledge of electrotechnics, electronics as well as metrological technologies and equipment

Within the main research fieds mentioned above the selective papers represent which kind of concrete topics we focused on last year.

More details and other information can be found:

www.fomi.gek.szie.hu, www.beke.gek.szie.hu

Development of vegetable drying process by combining convective and microwave methods

János BEKE, Zoltán KURJÁK
Department of Energetics
Institute for Process Engineering

Abstract

In this study the empirical analysis of convective and microwave drying of vegetables and greens are presented. The final purposes of experiments were to reduce the energy consumption of drying process while keeping the appropriate quality of dried material.

During the experiment carrot, potato, parsley, tomato and onion were investigated as vegetable and as greens alfalfa, parsley leaves and dill were analyzed.

Experiments proved that tubers show different drying parameters, depending on their morphological compositions.

The drying features of stalk and leaf – as two typical parts of greens – separately were investigated. As tests proved in the case of microwave drying the stalk drying rate exceeds the similar parameter of the leaf and this phenomenon contradicts the convective experiences. It can be supposed that an intensive moisture movement is established in the vascular systems, which leads the majority of water content from the stalk to the leaf causing more intensive water loss in the stalk. Nevertheless, these phenomena give a good chance to apply a pertinent combination of convective and microwave energy that can result a well efficient drying process and end product of good quality.

Keywords

microwave drying, combined drying process, material quality

1. Introduction

The drying is one of the most energy-intensive unit operations in treating agricultural products. Various studies report energy consumption for drying of agricultural materials ranging from 20 to 35% of the total energy that is used producing agricultural products, depending on their harvesting moisture contents. At the same time the influence of drying on the final product quality and on the prime cost is dominant.

The drying regime of agricultural product is more complicated chemicophysical process as usual, because in this case drying materials are living organisms in

which with parallel the dewatering process morphologically determined vital functions take place. As a result of the thermal drying the significant parts of natural flavors are off, shrinkage caused by drying deforms the final product and the cell walls become rigid.

More than 95% of drying equipment used nowadays in the agriculture operates by convective method. As a possible non-convective drying technology – especially for drying of greens – the microwave drying can be taken into consideration.

After entering of drying regime the asymptotically falling drying rate period the microwave dewatering method becomes more economical and the moisture content of leaves is low enough to be able to separate the two components. In this way two types of foodstuff of different properties and an efficient drying technology can be obtained.

2. Theory and background

For analyzing the energetic aspects of convective drying process, in theory, the following balance equations of classical energy- and mass transfer can be used especially in that cases when the drying materials well approach some regular geometric shapes.

$$\frac{\partial X}{\partial \tau} = D\frac{\partial X}{\partial r^2} + \frac{D}{r}\frac{\partial X}{\partial r} + D\frac{\partial^2 X}{\partial z^2} \tag{1}$$

$$\rho c\frac{\partial t}{\partial \tau} = \lambda\frac{\partial^2 t}{\partial r^2} + \frac{\lambda}{r}\frac{\partial t}{\partial r} + \lambda\frac{\partial^2 t}{\partial z^2} + L\rho\frac{\partial X}{\partial \tau} \tag{2}$$

Certainly, as initial conditions (τ=0) should be used the $X=X_o$ and $t=t_o$ substitutions. If $\tau>0$ then the boundary conditions are:

$$D\left(\frac{\partial X}{\partial \tau} + \frac{\partial X}{\partial z}\right) + \beta_s\left(X - X_a\right) = 0 \tag{3}$$

$$\lambda\left(\frac{\partial t}{\partial \tau} + \frac{\partial t}{\partial z}\right) + \alpha_s\left(t - t_a\right) = 0\,. \tag{4}$$

Since agricultural products and foodstuffs are living organisms there is a close connection between the water loss process and the material morphology. During convective drying the vaporization starts on the material surface causing significant thermal as well as mechanical stresses between the different parts of drying material. Consequently, in the falling drying rate period local overheating, deformation and mechanical damage occur (1).

During the microwave treatment the inner moisture and temperature distribution according to the mass and heat transfer equations is the function of place and time. If we put the drying material in the microwave cavity, it causes space perturbation and changes the resonance frequency of the dissipation area as follows:

$$\frac{\Delta\omega}{\omega_o} = -\frac{1}{W}\int_V \left(\frac{d\varepsilon}{\varepsilon} w_e + \frac{d\mu}{\mu} w_m \right) \frac{dV}{V} \tag{5}$$

For biological materials $\mu=1$ and $d\mu=0$. Therefore the temperature variation – assumed homogeneous temperature field – can be calculated:

$$\frac{dt}{d\tau} = 9.3 \cdot 10^{-12} \cdot E^2 \cdot f \frac{\varepsilon''}{c\rho} \tag{6}$$

For absorbents with high moisture content the ratio is *(dε"/dt)<0* while for those that content of few moisture *(dε"/dt)>0,* thus overdrying can be occurred (2). For biological materials we usually cannot calculate with inner homogenous moisture and temperature field (3). When different kinds of greens are dried the dissimilar behaviors of leaves and stalks can be easily proved (4).
By taking the mentioned scientific antecedents into considerations there is a good reason for assuming that the well known theoretical functions of energy and mass transport for the cases of food stuff drying must be modified, thus to find the right energetic conclusions extensive experimental works are needed.

3. Equipment and concept of experiments

For analysis of the convective drying process special laboratory equipment was developed, which can be seen in Figure 1. The features of the apparatus gave possibility to measure the input and output temperature, the velocity and the relative humidity of drying rate, as well as the weight of the drying sample. The moisture content of material was determined according to the relevant ASAE standard.

During microwave drying – to serve a nearly homogeneous microwave field – a resonator of basic mode was used as the test chamber (Figure 2) in which waves are reflected onto the so-called Brewster-angle (5)

As the typical parameter of convective experiments the inlet air temperature was 80 ^oC and 100 ^oC explicitly, while the velocity of the incoming drying air was fixed on 0.4 and 0,8 m/s.

The incident microwave power was P_i=1000 W with a frequency of 2.45 GHz. Furthermore, the P_r reflected performance was measured by means of a $\lambda/4$ probe of square cross-section, built in the wave guide. The dissipated

performance was calculated as the difference of incident and reflected powers. The output sign of power meter was coupled to a PC's measure card.

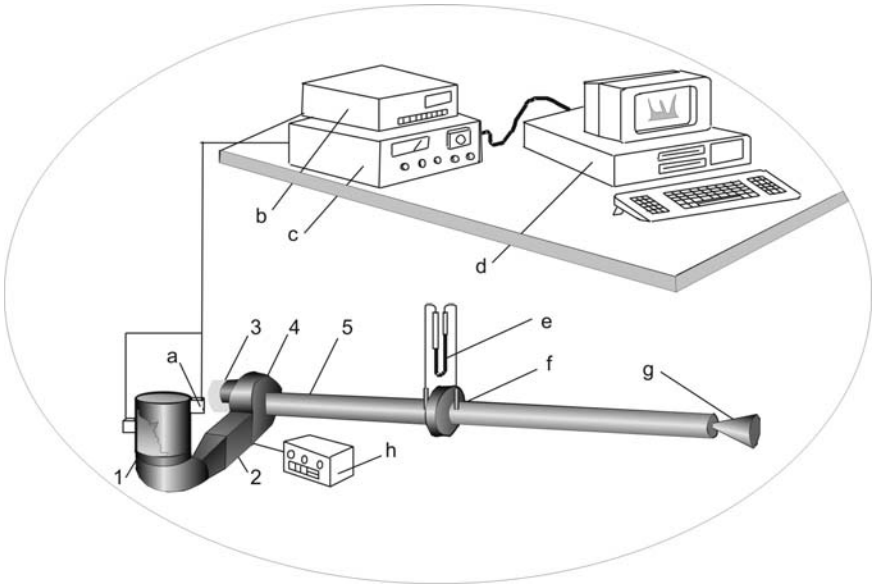


Figure 1. The scheme of the apparatus for analyzing convective drying
1: Sample holder, 2: Air heater, 3: Electric engine, 4: Fan, 5: Measuring tube.
a: thermocouples, b: dew point meter, c-interface,

For experiments of greens drying, as samples, alfalfa, dill and parsley leaves were prepared. One measuring procedure was repeated five times under the same conditions. The moisture content was determined with the help of a convective oven according to the relevant ASAE standards. To develop the kinetic curves samples were dried to equilibrium moisture content. During the drying experiments stalks and the leaves created one natural unit and before drying there were not any mechanical or chemical treatments. The samples with the same physical properties were dried different length of residence time, and then the stalks and the leaves were taken apart.

For vegetable drying, as samples, potato, green pepper, tomato, parsley root carrot and onion were used. In the case of root drying slices with thickness of 10 mm were prepared. The dried slices were chopped into small cube-shaped blocks with measure of 5x5x5 mm. In the lengthwise of the root 6 measuring cross-sections were formed. To develop the kinetic curves sample were dried to equilibrium moisture content. To plot the inner heat map the temperature ratio was introduced. The temperature ratio shows the temperature modification as a function of the place when the temperature in the origin of coordinate system (x=y=z=0) is a unit (ϑ=1). The preparation of tuber simples was similar to those of roots but cubes of 1 cm^3 were created to dry.

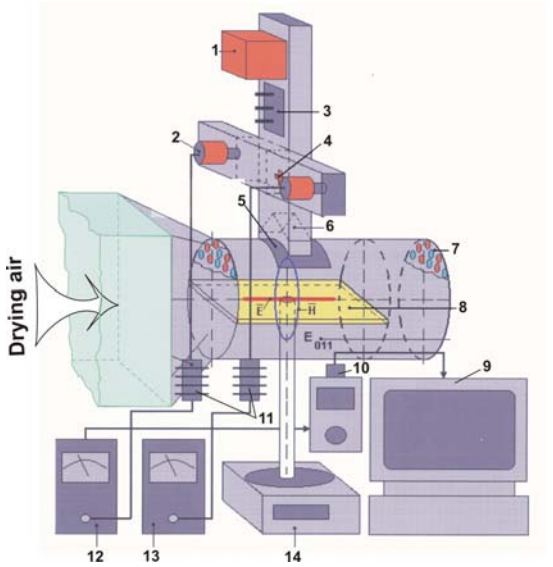


Figure 2. Microwave drying and measuring equipment with resonator of basic mode.
1: magnetron, 2: measuring probes, 3: tuner probe, 4: directional couplers,
5: incident clap, 6: adjusted link, 7: terminating plate, 8: sample holder,
9: data analyzer, 10: interface, 11: attanuators, 12: power meter for incident
performance, 13: power meter for reflected performance

For the analysis of the relationship between the moisture content and the energy consumption of the drying process – as a characteristic parameter – the relative energy consumption was created. By measuring the input and output temperatures and the airflow rate the utilized convective heat energy could be calculated. Similarly, by knowing the dissipated performance the temporary absorbed microwave energy could be determined. The relative energy consumption was given as the quotient of the temporary and the minimum values of the energy consumption measured during the same experiment.

4. Results and discussion

Since in the constant drying rate period the moisture of samples everywhere exceed the critical value, it is possible to dry with applying an appropriate high specific energy input without the drying rate would be limited by the capability in water loss of material. In aspects of the analysis the falling drying rate period is very insistent because this is the interval in which the interactions between the biological features of material and the parameters of drying technology have pertinent influences on the energy- and mass transport.

Drying experiments of tubers
When tubers are dried by convective method during the process of falling drying rate the peel thickness of samples more and more increases, the pores on the

surface gradually narrow that greatly hinder the dewatering process. This is the explanation, why the drying rate curves at the end of drying approach asymptotically the x-axis.

In the case of microwave dehydration similar drying rate curves to those of convective drying are obtained but they show significantly more rapidly increasing characteristics.

In this case the falling drying rate period is divided into a linear and an asymptotical interval. It means that the morphological structure has a dominant influence on the energy consumption, depending on the kind of energy transfer. The biological construction of the drying material does not support the surface energy- and mass transport. Specific drying rate curves of some tubers can be seen in Figure 3 for the cases of convective and microwave drying. It is evident that during microwave drying the incident energy is better utilized.

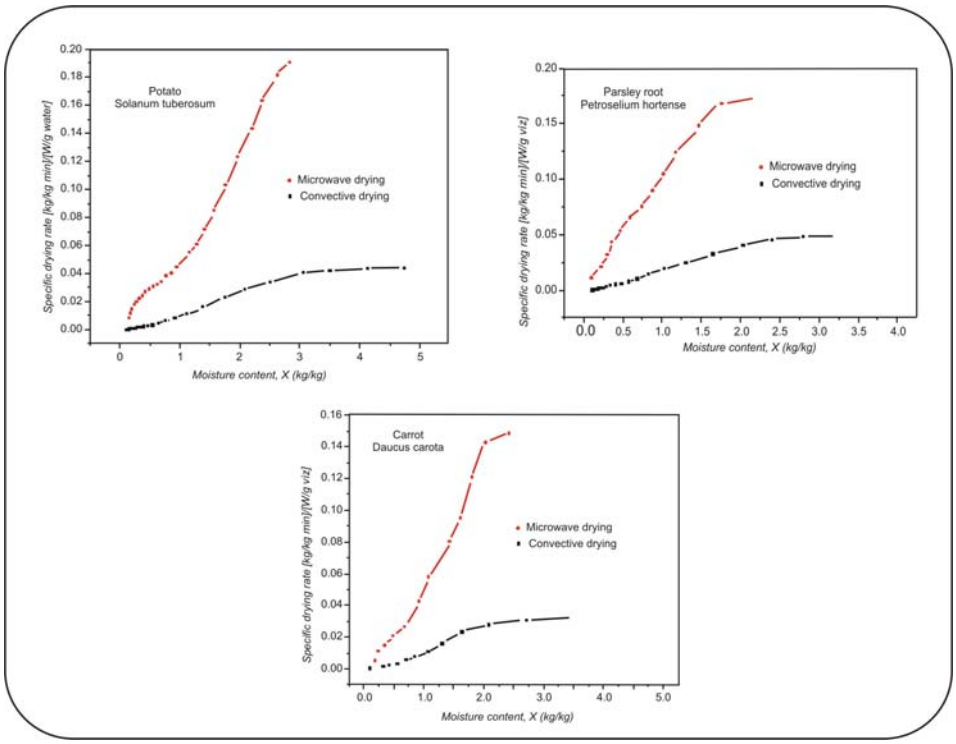


Figure 3. Drying rates of tubers, caused by 1 W/g water incident energy

Drying experiments of greens

Although the plotted kinetic curves of *convectively dried greens* show typical forms of biological products, stalks and leaves have significantly different water loss properties. The integrated moisture content comes from the weighted average of the topical water content of the two components thus; the kinetic

curves – strictly speaking – are not symmetrical to the mean moisture content curve.

The intensity of the water loss process is demonstrated by the drying rate curves which show that in convective conditions the stalk has lower drying rate (Figure 4). The probable reasons for that can be found in its smaller special surface and in the peculiarity of the inner moisture movement. Namely it can be assumed from the morphological structure of plants because the outer epitheliums create some water confining layer and the inner vascular system promotes the longitudinal moisture movement. There are similar reasons for the higher maximum drying rate of 20% in the case of leaf drying. However, a difference of a similar degree cannot be experienced in the average integrated moisture content for the whole drying rate because in the last period of the water loss process the difference is gradually equalized.

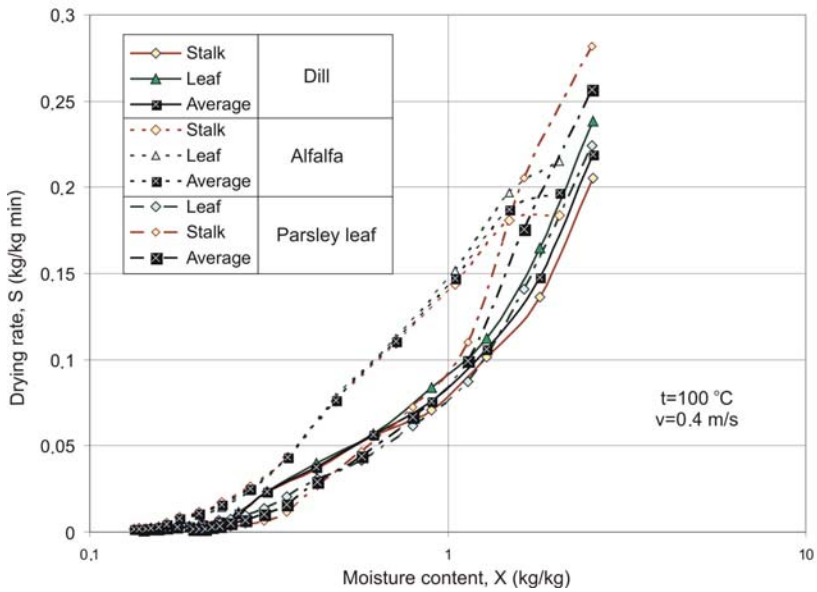


Figure 4. Typical drying rate curves of vegetative components in convective conditions

Carrying out microwave drying experiments the forms of the obtained kinetic curves show a significant similarity to that of convective energy transport, but the typical reversed S-shape becomes deformed in the high-frequency field. The beginning and the last flat intervals are shorter and the middle section has a larger angular coefficient. In these conditions the confining effect of surface epitheliums is nearly negligible. Because of the inner heat production, a significantly larger diffusion coefficient can be experienced, consequently the moisture movement rate increases (Figure 5). On all these bases it can be supposed that an intensive moisture movement is established in the vascular

systems, which leads the majority of water content from the stalk to the leaf causing more intensive water loss in the stalk.

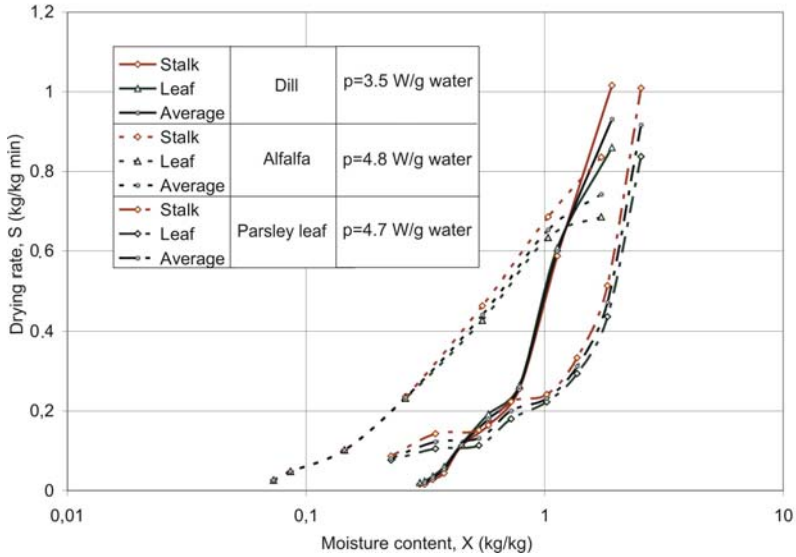


Figure 5. Typical drying rate curves of vegetative components in microwave field

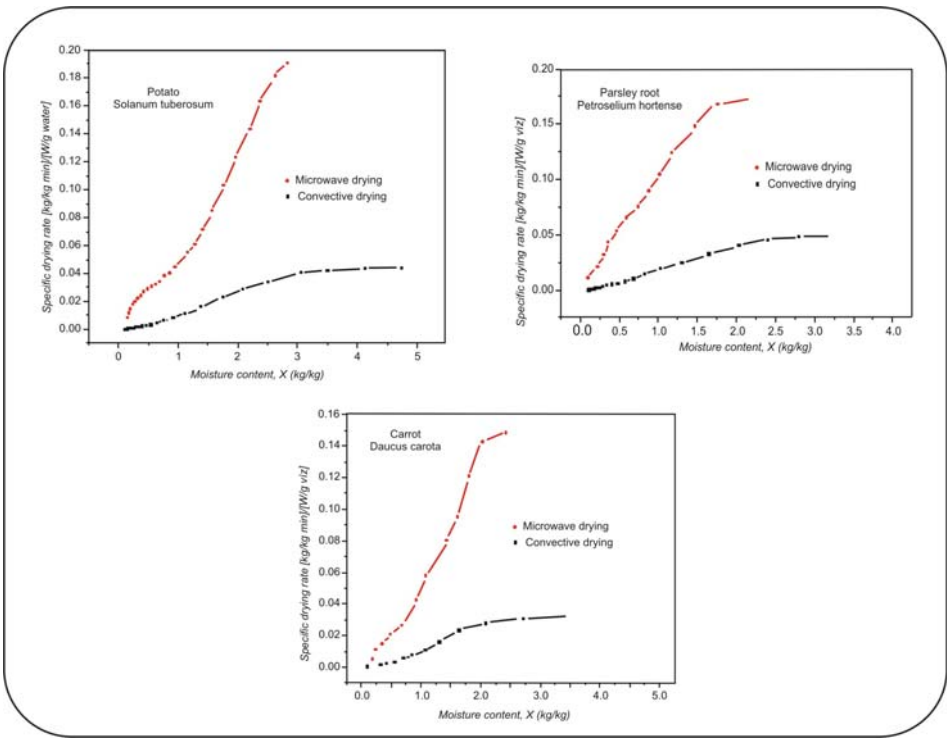


Figure 6. Drying rates of greens, caused by 1 W/g water incident energy

Since the energy absorption is in direct proportion with moisture content and the thermal effect of microwave energy takes place where the drying material consists of water the utilization of the incident microwave energy is more efficient (Figure 6).

Energetic analysis

On the basis of data that partly illustrated on Figures firstly the energy consumption of convective drying was analyzed and the following equation was derived to describe the relationship between the relative energy consumption (q_{rc}) and the initial moisture content of drying material (X):

$$q_{rc} = e^{a_K + b_K \ln X} \tag{7}$$

As a result of similar analysis to that of the convective method the energy consumption of microwave drying – as a function of the moisture content – can be described by the following equation:

$$q_{rm} = a_M + b_M \ln X \tag{8}$$

The "a" and "b" constants in Equations (7) and (8) were determined by regression analysis, henceforth its values can be seen in Table 1.

Table 1. Values of constants in Equation (7) and (8)

Drying material	a_K	$-b_K$	$-R_K$	a_M	$-b_M$	$-R_M$
Potato	-0.144	1.411	0.996	0.942	1.185	0.986
Green pepper	0.002	1.415	0.990	1.052	1.230	0.992
Tomato	-0.499	1.402	0.991	0.663	1.340	0.965
Parsley root	0.284	1.527	0.987	1.344	1.454	0.991
Carrot	0.272	1.535	0.965	1.378	1.499	0.940
Onion	0.142	1.485	0.994	1.181	1.343	0.989

By taking the mean energy consumption in the constant drying rate interval as unit, dimensionless (it means relative) energy consumption can be created for the falling rate period.

As the Equation 8 and Figure 7 prove under microwave conditions the energy consumption as a function of the moisture content shows a purely exponential trend.

For energetic comparison the ratios of maximum energy consumptions in convective and microwave conditions were created for different kinds of vegetables. On the one hand we can realize that the microwave drying needs less entering energy by at least one order of magnitude (in aspect of the pure energy

and mass transport). On the other hand significant differences are demonstrable in the energy consumption of types of vegetables that have morphological reasons. For instance the xylem-type built-up or the modified leaf-structure of the onion does not support the moisture flow in the cross direction. The longitudinal xylem system of the carrot as well as that of the parsley root support the longitudinal moisture flow during drying. In this way the waxes layer on the surface of the onion leafage hinders the peripheral mass transport.

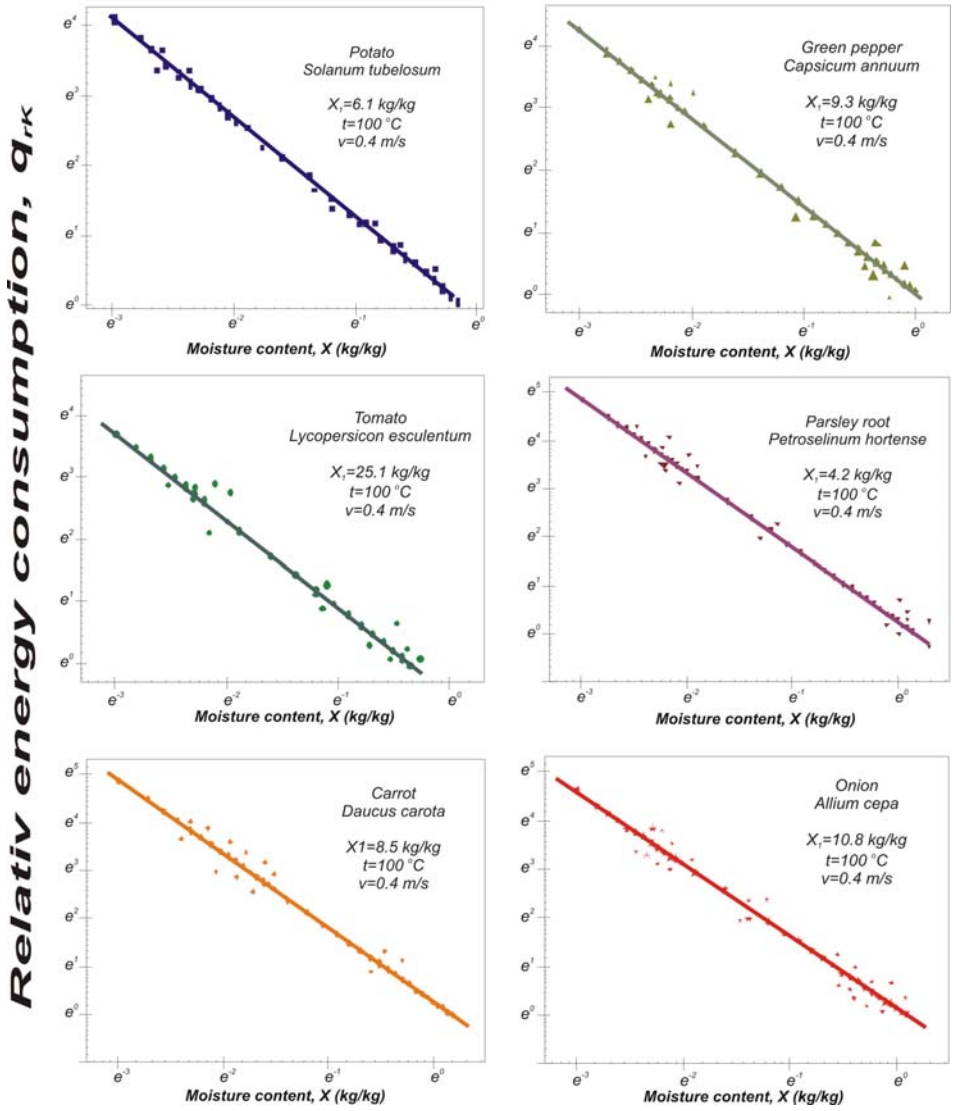


Figure 7. Energy consumption of drying vegetables as a function of moisture content in convective conditions

By following up the same procedure but in convective conditions it turns up that in this case the exponential character of the relative energy consumption is much powerful (Equation 7, Figure 8).

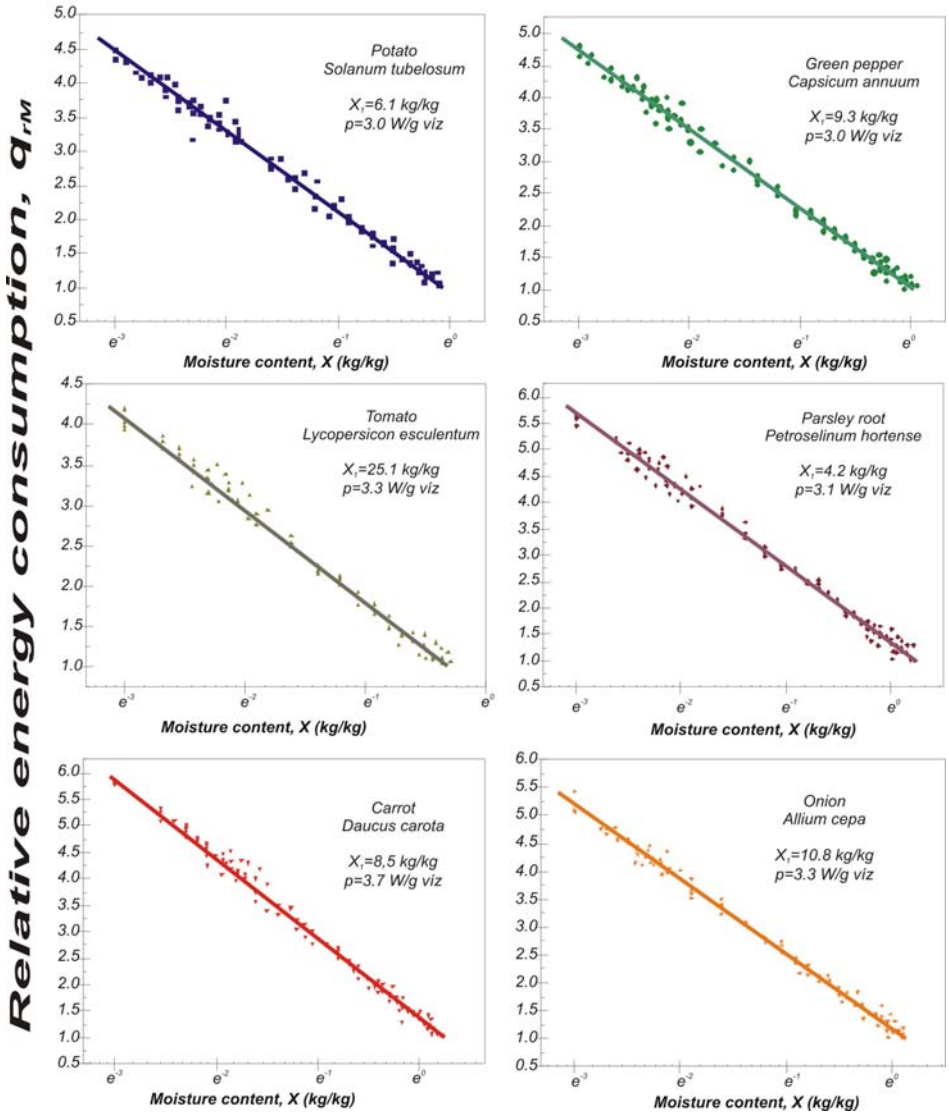


Figure 8. Energy consumption of drying vegetables as a function of moisture content in microwave field

In order to analyze the relationship between the moisture content (on dry basis) and the energy consumption of the greens drying process it was supposed an equalized distribution of inlet energy between the stalk and the leaf. On the

basis of the carried out experiments larger heat and mass transport rate is established in the case of leaf drying. In these conditions the leaf drying process needs less energy, nearly to the same extent. It should be mentioned that the above findings are valid on the basis of equivalent moisture content. It means it is hard to avoid overdrying the components because of the quicker leaf water get-off (Figure 9).

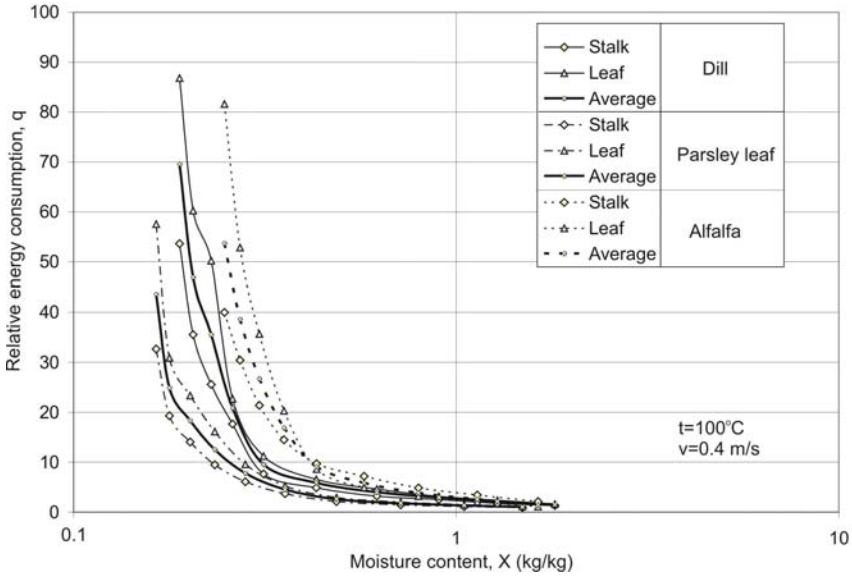


Figure 9. Relative energy consumption of greens during convective drying

By analyzing the drying process under microwave conditions it turns out that in this case the stalk drying rate exceeds the similar parameter of the leaf (Figure 10). This phenomenon contradicts the convective experiences. This phenomenon can be originated in at least two probable reasons. On one hand as a result of the microwave absorption the temperature gradient has an opposite direction, compared to the convective case.

On the other hand the degree of energy absorption is not in close connection with the surface heat and mass transfer coefficient or the specific water get-off surface

By comparing the convective and microwave drying process from energetic point of view it is evident that from the place where the curves enter the asymptotically falling drying rate period the energy curves of convective dehydration increase dramatically (Figure 11). It is a good reason for assuming that in this period the microwave drying could be favorable from both economical and quality point of view. At this point applying the convective process the moisture content of leaves is low enough to be able to separate the leaf and stalk components by using a mechanical method. In this way two types

of foodstuff of different properties can be obtained. Furthermore, combining the two drying methods and separating the leaves and the stalks an efficient drying technology and an end product of excellent quality can be produced.

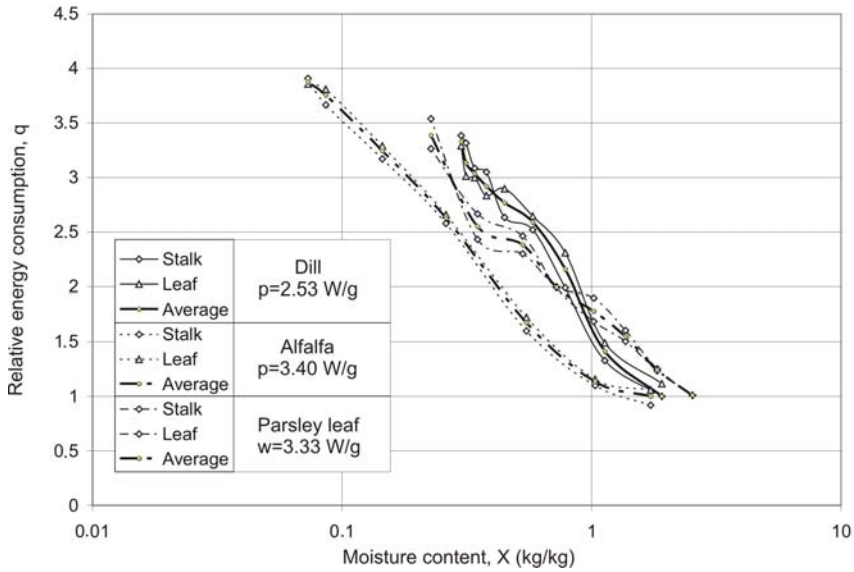


Figure 10. Relative energy consumption of greens during microwave drying

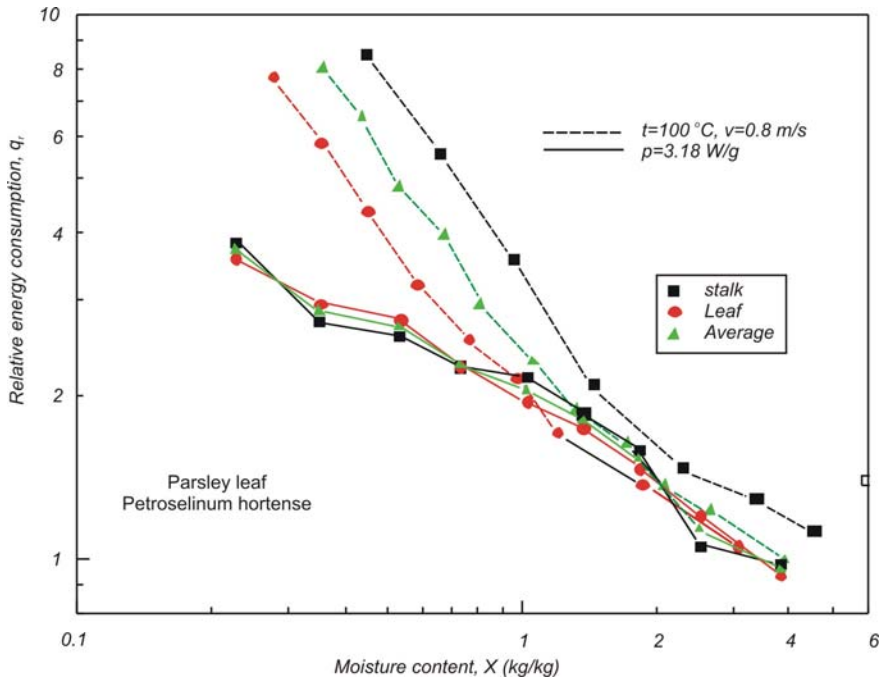


Figure 11. Comparison of energy consumption in convective and microwave conditions

5. Conclusions

In the falling drying rate period – when microwave energy is used for drying – the energy consumption as a function of the moisture content shows a purely exponential trend. At the same time in convective conditions this exponential character of the relative energy consumption is much powerful.
From the point when the drying rate curve enters the asymptotically decreasing water loss rate interval the energy consumption of convective dehydration increases dramatically and exceeds by an order of magnitude that of the microwave dewatering process.

Under these conditions using microwave energy become more favorable from energetic and also economical point of view. Combining the two drying methods and separating the leaves and the stalks an efficient drying technology and an end product of excellent quality can be obtained.

Symbols

D	moisture diffusivity
E	electric field strength vector
H	magnetic field strength vector
L	heat of evaporation
P_d	Dissipated performance
S	Drying rate
V	volume of dissipated field
W	energy stored by resonator
X	moisture content, dry basis (kg/kg)
X_a	ambient moisture content
X_o	initial moisture content
c	specific heat (KJ/kgK)
c	phase velocity
f	frequency
t	temperature, ^oC
t_a	ambient temparture
w	moisture content, wet basis (%)
w_e	electrical energy density
w_m	magnetic energy density
α_s	surface heat transfer coefficient
β_σ	surface mass transfer coefficient
$\omega\Delta$	frequency drift
ε	dielectric permittivity
ε'	the real part of complex dielectric constant
ε''	the imaginary part of complex dielectric constant
ϑ	temperature ratio
λ	thermal conductivity (W/mK)

μ	magnetic permeability
ρ	density
ρ	Density of absorbent
τ	time (min)
ω_o	Resonance frequency in the empty chamber

References

1. Beke, J.; Vas, A. Testing the Drying Process of Shelled Corn on Aspects of Nutritional Quality and Thermal Efficiency. Drying Technology. 1994, 12 (5), 1007-1027.

2. Bengston, N.; Risman, P. Dielectric properties of foods at 3 GHz as determined by a cavity perturbation technique. Journal of Microwave Power 1971. 6. 107-123.

3. Beke,J.; Kurjak, Z.; Bihercz, G. Microwave field test of inner moisture and temperature conditions of beetroot. *Proccedings of International Drying Conference*, Beijing, China August 27-30, 2002; 670-679.

4. Beke, J. Experimental study of dewatering process on vegetative parts of some agricultural products. *Proccedings of International Drying Conference*, San Paulo, Brazil, August 22-25, 2004;1553-1560. p.

5. Beke, J.; Ludányi, L. Microwave Drying of Biological Materials as Metrological Problem. *Proceedings of IV. Asia Drying Conference*, Hong Kong, China, August 13-15, 2007; 326-332.

Strain Test of Farm Machine Chassis

Károly PETRÓCZKI, Péter KORZENSZKY, Zoltán GERGELY
Department of Metrology
Institute for Process Engineering

Abstract

During the construction of self-propelled and tractor-pulled farm machines several tests take place in laboratory and in the field. The functional and the stress-strain tests are mostly used. The mechanical load on the chassis is derived from the load of the normal machine operation and from the mechanical motion on the terrain. The normal machine operation is a specific problem and depends on the operation of the specific machine. This article deals with the strain measurement on the critical parts of chassis and the accelerations on the machine during motion. The mobile measuring system consists of 3 pieces of HBM made Spider8 universal digital PC measurement electronics with 4.8 kHz carrier frequency all in all 24 measuring channels and a notebook. The measuring channels can process strain gages with ¼ bridge configuration together with external compensating resistors and inductive half bridge acceleration transducers. The purpose of the article is to give an example of a complete mobile measuring system and a couple of problems during the calibration and the interpretation of measuring data.

Keywords

strain measurement, strain gage, acceleration measurement, farm machine testing

1. Introduction

Several kinds of farm machines have more or less universal unit namely wheeled chassis. During the construction and the production of it several aspects are into considerations: the weight of the unit, the durability, reliability, the universality, the technology and the costs of the production, etc… The theoretical and experimental stress-strain analysis of the structure has an important role in this process. The correct analysis needs the determination of the real mechanical load excitation onto the chassis. These real values can be determined during field investigations.

The purpose of this paper is to describe a mobile measuring system which is capable of performing multichannel measurement of the strain and the acceleration at the proper points of the chassis during field tests in case of a tractor pulled baler. The article deals with some aspects of calibration and data

interpretation. The test conditions and the other relations of the investigations are described by Kiss, P., and Mezei, T. (2009).

2. Methods and materials

Strain measurement

After the preliminary finite element analysis and a discussion about the structure and the manufacturing process strain measurement was done with strain gages at the critical points of the machine e. g. at the crossbeam, cantilever of landing wheels, etc… The block diagram of measuring system is shown on the Figure.

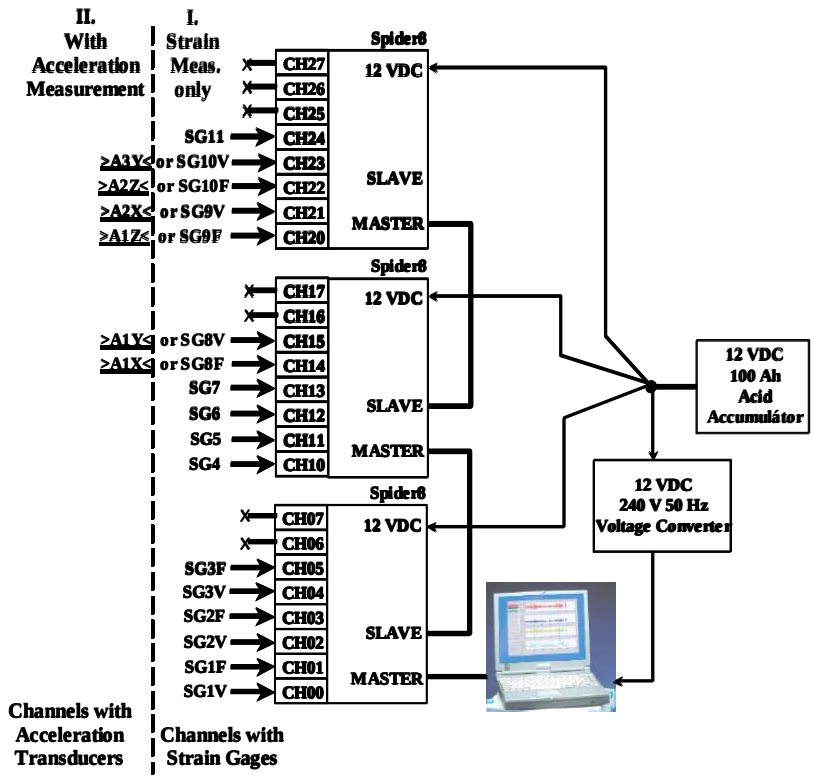


Figure 1. Block diagram of strain and acceleration measurement

The strain measurements was done with Hottinger Baldwin Messtechnik (HBM) made K-KY41-3/120 type strain gages. The preparation of the measuring surface was made according to the rules of bonding. The painting, dirt and rust was removed and after grinding and roughening the position of the strain gage (SG) was scribed. Having completed the degreasing with RMS1 cleaning agent, positioning and provisional fixing of strain gages the bonding

was done with HBM Z70 single component adhesive. LS5 solder terminal was bonded together with strain gages and lead wires of strain gages and the ends of the measuring cable was soldered to them. After checking the measuring points the SG together with the wires and connection points was covered with AMB75 protective coatings against external influences (mechanical influences, humidity, dust…). The ¼ bridge configuration was applied at every measuring point except 5 and 6. The Spider8 universal digital PC measurement electronics is applicable only with ½ SG bridge configuration, that's why SG ¼ bridge arms was completed with external compensating resistors. The compensating resistor was configured from the same type of strain gages bonded on a separate steel sheet in the neighborhood of the measuring point approx. at the same temperature, so this one together with the active SG the thermo compensation of the resulting ½ bridge configuration was excellent.

The measuring points except 5 and 6 have 1 active strain gage, so the so called bridge factor is 1. At the 5 and 6 points a bending beam are measured with 2 active strain gages, so the bridge factor is 2. The bridge factor is necessary in the system calibration. The gauge factor of strain gages were 2,01±1 % (Dally, J. W., Hoffmann,K.; see Ref. below: PC meas. Electr.Spider8…., Strain Gages…)

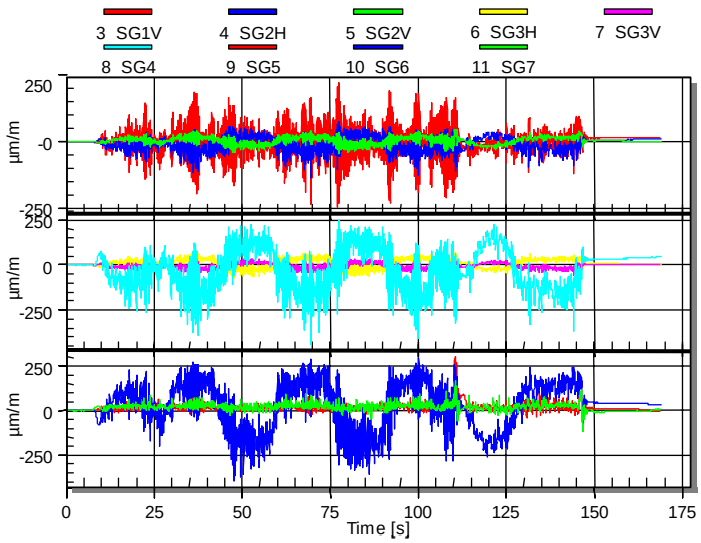


Figure 2. Strain vs. time diagram of "8" shape motion

The measurement and the evaluation was done with HBM CATMAN 4.5 Rel.2. software. The functions of the "Define device and IO channels" module are: "Device type definition" (Spider8), interface (USB), name of the measuring channels (e.g. SG1V, SG1H, SG2V….), hardware connection of the measuring channel and slot number. We can step to the "Device Setup" module from here. In the "Device Setup" module the real physical measuring channels can be set

up: configuration of the measuring channel (bridge, frequency, counter, voltage….), in our case bridge configuration (1/2 or full bridge) and the range (3mV/V). From the "Define device and IO channels" module one can step into the "Scaling" module, where the channel calibration can be carried out. The "Strain Gage" calibration was chosen and in this window the Gauge Factor (k-Factor or sensitivity, here: 2,01) as well as the bridge factor. If more than one strain gage in the bridge circuit is active in case of channel 4 and 5 are 2, at the others 1. After setting up channels together and the calibration the measurement and the data saving can be executed in the "Measure" module. The sampling frequency, the format of the measuring diagram and format of the exported data can be set in this module (See Ref.: CATMAN…).

After the preparation the strains resulting from the own weight was measured. The unloading process can not be executed exactly, correction can be made with the analysis of the structure. During unloading zero balance was done and after that the machine was replaced to the normal position to the concrete of the laboratory. As the strain gages are bonded in the directions of the principal strains the stresses can be calculated directly from the Hook's law. As an example Figure 3. shows strain vs. time diagram during "8" shape motion.

Acceleration measurement

Degree of freedom of the rigid body is six during free mechanical motion in space, so six measuring channels and acceleration transducers are necessary. The rigid and protected, but well mountable surfaces were taken into consideration. The allocation of the transducers is on the Figure 3. Further particulars can be found in paper Kiss, P., and Mezei, T. (2009).

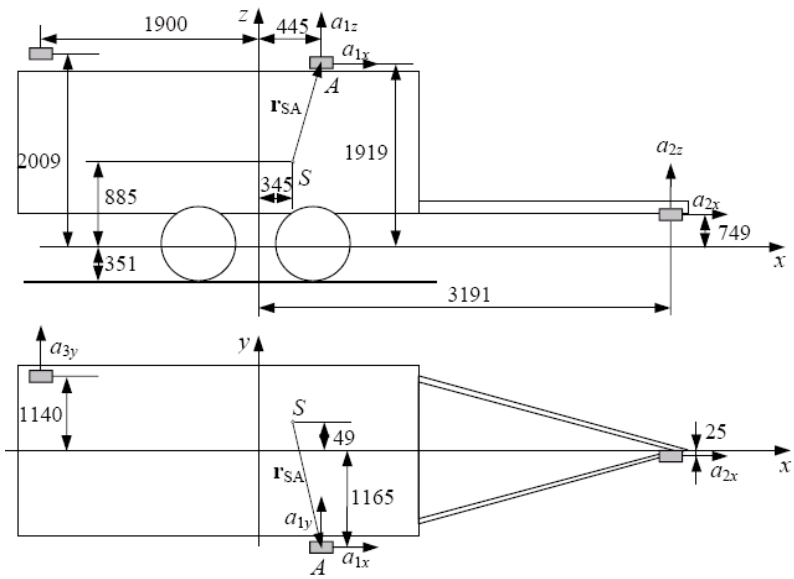


Figure 3. The allocation of the acceleration transducers

The measurement of the accelerations was carried out with Spider8 PC measurement electronics, too. As the instrument works with 4,8 kHz carrier frequency and 2,5 V bridge excitation voltage, so the B12 type HBM inductive acceleration transducer was chosen to measurements. The transducer is actually a one degree of freedom vibrating system: seismic mass is assembled to a spring and the displacement of the mass is proportional with the acceleration. The displacement is measured by a half bridge inductive transducer. The block diagram of the measuring system is shown on the Figure 1. The system is the same as at the strain measurement, only the CH14….CH23 SG channels are replaced with A1X…A3Y acceleration transducers.
As the B12 acceleration transducer works from the 0 Hz frequency, the calibration process is quite simple and can be performed in gravitation field. When the transducer is in the vertical position, the mass is in a defined, fixed position. The amplifier should be adjusted to zero while the transducer is in this position. Next the transducer is rotated through 180° to the opposite vertical direction, 2g acceleration acts on the mass. In "Scaling" menu 19,62 m/s2 value is written and the now the measuring channel has been calibrated (See Ref.: Acc. Transd. B12…).

The great advantage of the 0 Hz low frequency transducer is the simple calibration method and the simple checking method of operability, but it has a disadvantage: the transducer works as a position sensor and so the position signal and the acceleration signal are added. Furthermore the position sensitivity are different in horizontal and vertical directions.

An example of the acceleration vs. time diagram can be seen on Figure 4. The following statements can be submitted:

1. After magnification of the diagram and from the praxis we can establish, that the diagram includes significant relatively high frequency machine vibration which doesn't belong to the tested process.
2. A1X, A1Y, and A3Y acceleration signals in their tendency show relatively slow continuous changing. In case of A1X continuous increase and in case of A1Y and A3Y continuous decrease can be observed. This effect can not be interpreted with the real machine motion. In case of A1Z and A2Z similar effect can not be observed, the transducers in Z direction don't "creep". The reason of it is the higher position sensitivity of the transducers in horizontal direction. For example the Table 1. contains the apparent acceleration in case of two angle of slope.

Table 1. Apparent acceleration on slope in horizontal and vertical transducer's directions

Angle of slope [°]	Percentage inclination [%]	AZ [m/s^2]	AX, AY [m/s^2]
5,7	10	-0,05	0,97
8,5	15	-0,1	1,45

After a long considerations in order to suppress the disturbing signals curve smoothing was executed with running average method. To suppress the relatively high frequency machine vibration signal the window size 6 and to suppress the position signal the window size 100 was applied. On the Figure 5. the filtered signals are shown.

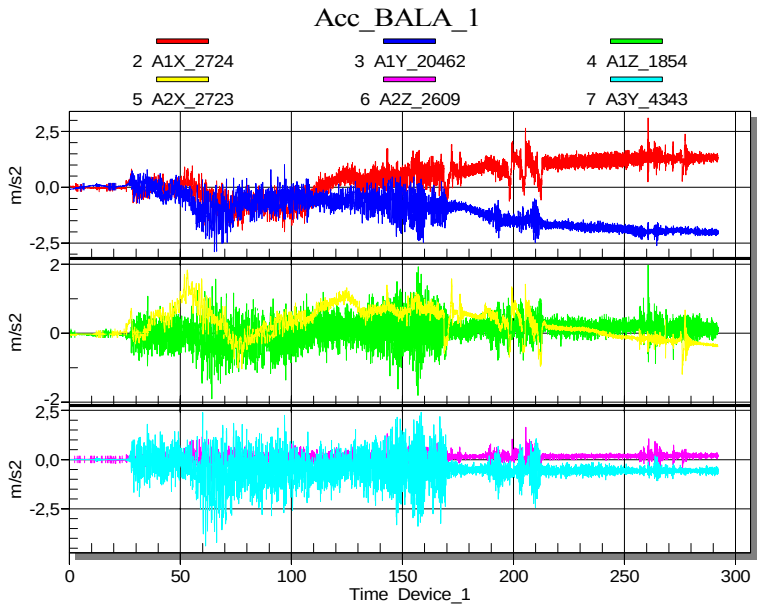


Figure 4. An example of the acceleration vs. time diagram

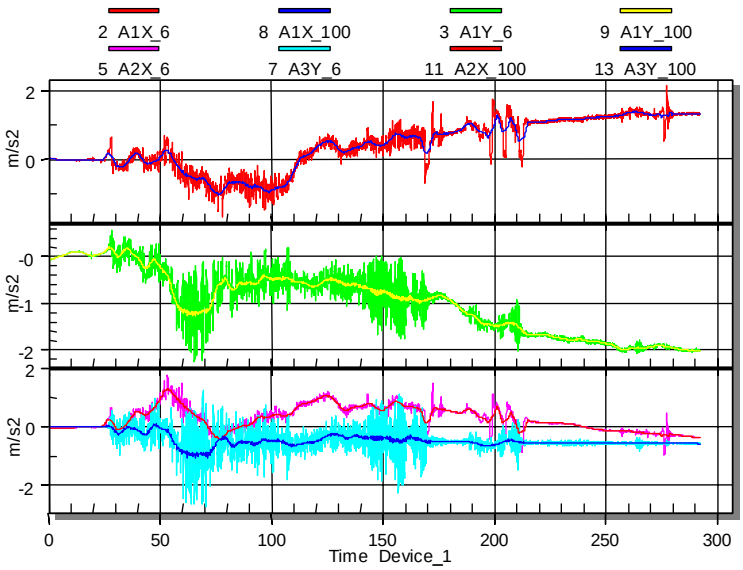


Figure 5. Running averaged signals with window size 6 and 100

The signals with window size 100 are the position signals of the acceleration transducers. They mean the X and Y direction transducers' position from the horizontal plane. The investigations were carried out on slops that's why the averaging considerations seems to be valid. On the Figure 6. the corrected acceleration signals are shown with suppression of high frequency machine vibration and the position signal.

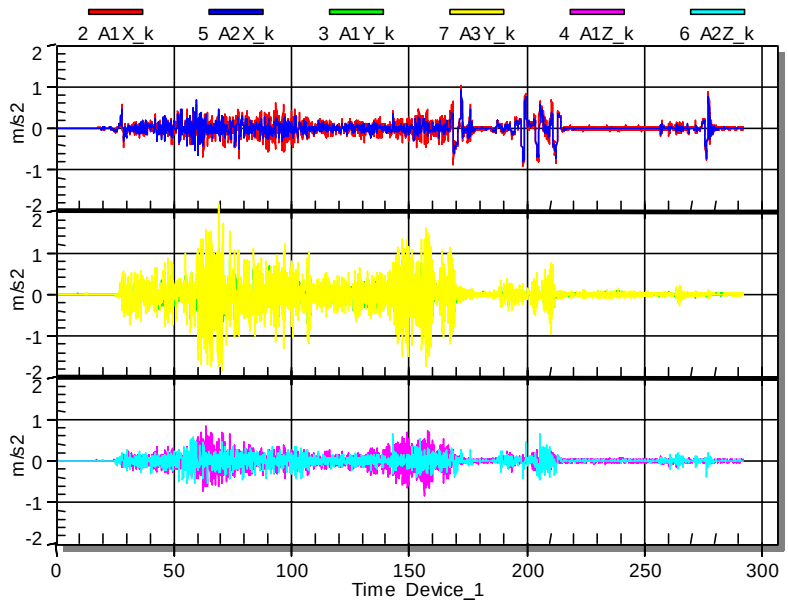


Figure 6. The corrected acceleration signals

3. Conclusion

The mobile measuring system based on HBM made universal Spider8 PC digital measurement electronics and CATMAN measuring and data processing software is a proper choice to carry out the strain and acceleration measurements not only in laboratory but in the field as well. The well-arranged simple and even so correct calibration in case of strain gages and acceleration transducers gives the possibility of increase of the measurements' reliability. The HBM strain gages together with bonding technology are constant and confidential parts of the experimental stress-strain analysis for a long time. The measurement with traditional B12 type acceleration transducers are correct, but nowadays new and cheap transducers so called iMEMS® (integrated Micro Electrical Mechanical System) have appeared in the market.

During the motion on the terrain main statistical parameters of the strain was determined for the complex stress-strain analysis. Many times, but one can say

every time, the interpretation of the acceleration signal is rather complicated, the filtering is necessary, it needs the knowledge of the signal theory at a high level, but sometimes useful information can be obtain with simple averaging as was presented in the article above. The cross-correlation between the rigid body's accelerations and strains/stresses in case of a given machine can be calculated from the data to forecast the stress at a recorded states of acceleration. The publication of concrete and comprehensive strain and acceleration values didn't belong to this paper, only typical examples were presented.

References

Dally, J. W. (1991): Experimental Stress Analysis, McGraw-Hill, Inc. 1991.

Hoffmann, K.: An Introduction to Measurements using Strain Gages, HBM Reference Book.

Kiss, P., Mezei, T. (2009): Determining static load distribution and geometrical parameters required for dynamic testing of a baler. Proc. Of Synergy and Technical Development International Conferences in Agricultural Engineering (Synergy2009). Gödöllő, Hungary, 30. August-02. September 2009.

Acceleration Transducer B12. Mounting Instructions. Hottinger Baldwin Messtechnik GmbH.
http://www.hbm.com/hbm_doc/navi/start-en_prod_archives.htm

CATMAN Professional and Express. HBM Software. Operating Manual.
http://www.hbm.com/hbm_doc/navi/start-en_prod_archives.htm

iMEMS: http://www.analog.com/en/mems/products/index.html

PC measurement electronics Spider8, Spider8-30 and Spider8-01. Operating Manual. Hottinger Baldwin Messtechnik GmbH.
http://www.hbm.com/hbm_doc/navi/start-en_prod_amplifiers_((overview)).htm

Strain Gages and Accessories. Hottinger Baldwin Messtechnik GmbH.
http://www.hbm.com/hbm_doc/navi/start-en_brochures_(strain-gages).htm

The Effects of the Off-Road Vehicle on the Soil Cohesion and Internal Friction

Lajos LAIB, László MÁTHÉ, György PILLINGER
Department of Automotive Technology
Institute for Process Engineering

Abstract

The movement of the off-road vehicle on a deformable track such as soil and grass modifies the soft soil profile and mechanical properties – cohesion and internal friction – of the soil. It is assumed there is a dynamic physical change in the tire-soil interface, the cohesion and internal friction soil properties, being also not constant.

In our paper we summarize the most important methods which can be applied to compute the towed force in tire-soil interface situations. We have carried out many in situ soil shearing tests with different soil moisture contents and in different soil depressions. We conducted the soil shearing test before and after the vehicle movement, which means there are different soil compactions.

Based on our soil shearing tests we can say that the soil cohesion increase depends on the soil depth. The value of the increase depends on the soil compaction and the soil depth. The soil internal friction hardly decreases and its changing value is less than that of soil cohesion. The soil cohesion and shearing values decrease as a function of soil moisture content. The higher values of both soil cohesion and soil shearing are characteristics of compacted soil.

The results of our tests were as follows: soil cohesion and soil shearing values are not constant with different types of soil. Both of them change as a function of the soil moisture content and soil depth. The equation for the towed force which occurs during the tire-soil interface is more precise if the soil cohesion would be a function of the soil moisture content and the soil depth in the equation.

Soils are multifarious, hence our results are accurate for the above soil conditions and test methods.

Keywords

terra-mechanic, soil- vehicle interaction, mobility of the off-road vehicle

1. Introduction

The development of the off-road vehicle was very intensive in the last twenty – five years. The all wheels drive improves the stability and the mobility of the

off-road vehicle significantly. The research results of the army off-road vehicles are used in civil vehicles as soon as appropriate.

The intensive activity on this topic began in 1950. Terra- mechanics is an interdisciplinary science, because it is based on physics and mechanics and it applies mathematics and computational methods and the soil mechanics, machine elements, automotive design and vehicle dynamics. [1., 2., 3.]

In the first decade of the XX century, Bernstein (1913) [4., 5.] was first of those who has investigated the relationship between soil bearing capacity and rolling resistance in the beginning of this century. In the nineteen forties the U.S. Army charged the Corps of Engineers Waterways Experiment Station (WES) with the research of the basics of cross-country mobility and with the development of the principles of modern cross-country vehicle design. (Earlier this institute had already been heavily involved in hydrological research concerning rivers and harbors.) The reason for this new interest was W.W.II. The military vehicles have had many problems with the mud on the Pacific Islands and became immobilized, creating severe problems.

In 1954 Dr. M.G. Bekker founded the Land Locomotion Laboratory which was located on the buildings of the U.S. Army Detroit Arsenal. The researchers of the LAB realized that soil has two stresses under the rigid wheel. One of them is caused by the weight of the vehicles and is called normal stress. The other is the shearing stress which is caused by the driving moment and created by the peripherical force. [3., 6., 7.]

The researchers concentrated on the mathematical description of the physical action at the soil-wheel interface. Bekker developed formulas for the sinkage, and rolling resistance.

Janosi worked out an equation for the shearing action of the soil. Both methods from Janosi and Bekker are based on soil shearing function and they used empirical factors for describing the actual vehicle mobility phenomenon. Bekker summarizes his research results in his book which was published in 1956, titled “Theory of Land Locomotion” [3.]

The lack of universal applicability of Bekker's method has long been proven by researchers.

Janosi's formula has been successfully used during the past thirty years. However, the formula, originally developed for tracked vehicles, has been improved significantly by several researchers. [7.]

Komandi applied the formula [1] for cross-country vehicles running on pneumatic tires. Komandi did not use soil shear curves to determine Kred, rather he used traction-slip curves which came from real drawbar pull tests. Professor Sitkei examined the soil wheel interaction in sand and he found that wheel slip has an important role here. [8., 9., 10., 11., 12]

$$F_t = A \cdot \tau_m \cdot \left\{ 1 - \frac{K_{red}}{s \cdot l} \cdot \left[1 - \exp\left(- \frac{s \cdot l}{K_{red}} \right) \right] \right\} \quad [N] \qquad (1)$$

$$\tau_{max} = \mu\sigma + c$$

where :

" Ft " Towed force which was bourn in the tire-soil interaction
" A " is the contact area of the tire – soil interaction
" K_{red} " is an empirical factor in the equation
" τ_{max} " is the maximum shearing stress of the soil
" μ " is the internal friction of the soil
" σ" is the normal stress of the soil
" c " is t he cohesion of the soil
" s " is the wheel slip
" l " the length of the contact area the in the tire – soil interaction

These methods are only valid for stationary conditions. This is a serious problem because there are non-stationary situations even when the vehicle moves slowly. On the other hand the soil profile generates vibrations and normal forces which exist are caused by the weight of the vehicle. These are not stationary either. This affects the shearing action in the soil-wheel interface. [13.]

Another serious problem is that cohesion and internal friction are changing in the soil. The deformation of pneumatic tires and the relative motion between the wheel and the soil - the slip - also influence the physical phenomena.

From the field tests of the Hungarian Army Mobility Model (HAMM) we conclude that the soil profile which is generated by the vibration of the vehicle - is significantly modified by the vehicle itself. The soil is compressed by the vehicle and part of the kinetic energy of the vibrating vehicle is absorbed by the soil, while the dynamic, normal and shear forces are affecting the interaction. [14., 15., 16., 17., 18., 19.]

We have studied the influences of vehicles on the soil cohesion and the internal friction during the last five years. The target of this research was to determinate the variation of the soil cohesion and internal friction as a function of the soil moisture content and the compression of the soil. Besides this, we investigated the influence of soil compression in different soil depths.

2. Materials and methods

For the above research target we used the following instruments and methods. First of all we chose the test fields. One test field was located in Godollo. It has sandy soil. The other test field was located in Veszprém in the middle of Hungary. It has compact loamy soil. We picked some samples from the soil with the instrument which you can see in figure 1.

Figure 1. Instrument for soil sample

We carried out the tests at different soil moisture contents. We used an in situ equipment, the type was 5910 –A which was produced by Soil-moisture Equipment Corp. U.S.A. Santa Barbara CA. The precision of the measured data was $\pm 10\%$. The figure 2 show the soil-moisture equipment. Besides this, we determinate the soil moisture content in the laboratory by exsiccator too so that we can control the data which came from the in situ field tests.

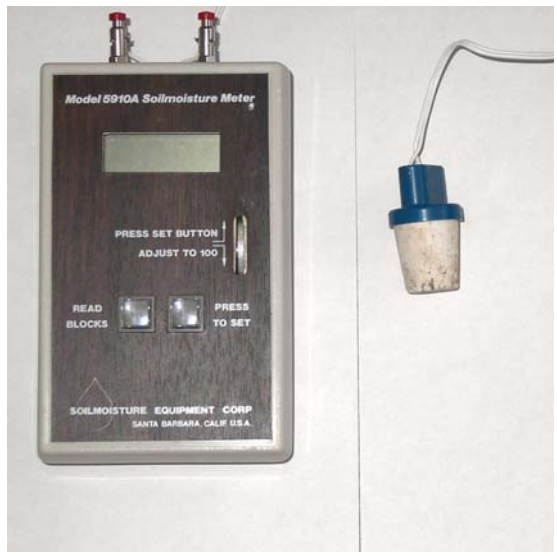


Figure 2. Soil-moisture equipment

We used torsional shearing equipment for the in situ soil shearing tests in the fields. The type of equipment was Eijkelkamp 14.5. This instrument measures only the maximum shearing stress in the field test. It has three rungs where the sizes of them are different. The following table shows the diameter and the length of the three rungs.

Diameter	Length
16 mm	32 mm
20 mm	40 mm
25.4 mm	50.8 mm

We simulate the normal stress which was necessary for shearing with different sizes of mass, on the torsional shearing instrument. We determinate the size of the mass, and hence the normal stress, of the soil in its original virgin condition was equal to the normal stress of the soil when an off-road vehicle moves on the off- road soil.

The measured maximum shearing 260 kPa was when the diameter of the rung was 20 mm, and 520 kPA was when the diameter of the rung was 16mm. The figure 3 shows the torisonal soil shearing instrument.

The normal stress was born by a 3, 5, 7, kg which means 93.7 kPa, (that is 1 bar soil compression) 156.2 kPa, (that is 1.5 bar soil compression) and 218.6 kPa (that is 2 bar soil compression). The figure 4 shows the soil shearing instrument with the mass.

We made the in situ soil shearing tests in three levels. First we made the test above the ground The depth of the second test was 25 cm under the ground, the third one was 50cm under the ground.

Figure 3. Torsional soil shearing instrument with three rungs

We made the soil shearing tests on the three different levels of soil compactness. The first test was made with natural soil. Before the second test we compacted soil with 2000 kg mass. The compacting mass was 4000 kg and 8000 kg before the third and fourth soil shearing tests.

The density of the sandy soil was 2.36 kg/dm^3 and the density of the compact loamy soil was 2.79 kg/dm^3. The soil density was determinated by the Soil Sciences department of the Szent Istvan University.

The moisture content of the soil was between 8 and 23 % in the field test. We determinated the exact value of the soil compaction with the Cone Penetrometer. The type of the cone penetrometer was Eijkelkamp 69.87, the maximum measured value was 5000kPa at 20.6 mm cone diameter. You can see the Cone Penetrometer in Figure 5.

Figure 4. Soil shearing test instrument with the mass

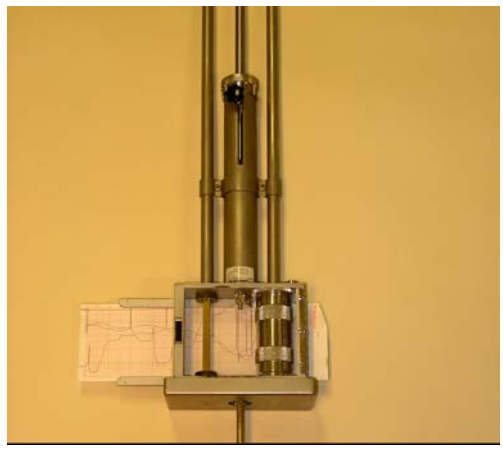


Figure 5. Eijkelkamp penetrograf

3. Test data

We summarized the field test data in the tables 1 – 20. See the Appendix 1. In the tables the data belongs to one soil moisture content and to three ground levels. In every ground level the soil shearing test was made at three different soil compactions. The soil compaction equals 1., 1.5, and 2., bar soil pressure.

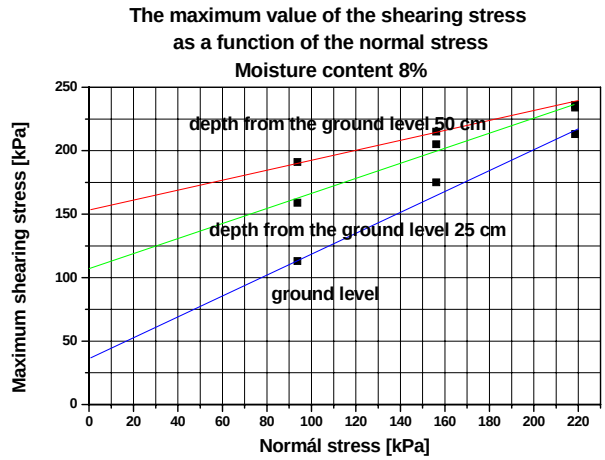


Figure 6.

After the computed of the field test data we fixed Coulomb lineal line to computed data and with it we determinate the soil cohesion and the internal friction of the soil.

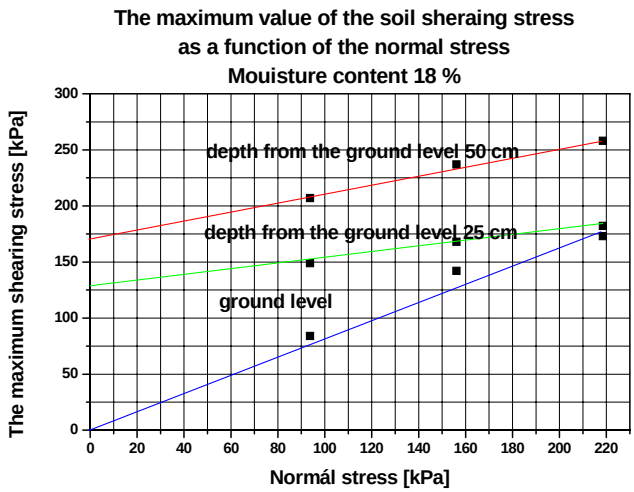


Figure 7.

You can see the soil data in the tables 1. – 4. at 23 % moisture content, in the tables 5. – 8. at 20 % soil moisture content, in the tables 9. – 12. at 18 % moisture content, in the tables 13. – 16 at 13 % moisture content, in the tables 17. – 20. at 8 % moisture content.

Every table has a computed Coulomb linear line which you can see in the figures 6., and 7. as samples. You can see the maximum soil shearing stress as a function of the normal soil stress in the figure 6. and 7. Every figure has one constant soil pressure value and soil moisture value. With help of these figures we can compute the soil cohesion –which one can read at the crossing of the vertical axis and the Coulomb linear line, and the internal friction of the soil, which is the tangent function of the Coulomb line angle.

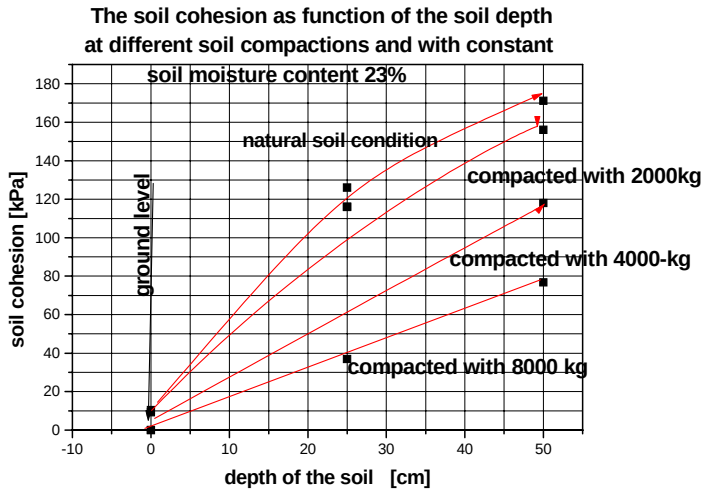


Figure 8.

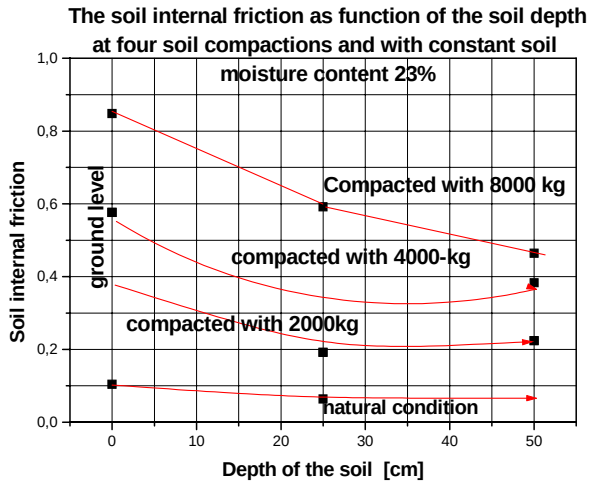


Figure 9.

We wanted to study the soil cohesion and the soil internal friction as a function of the moisture content, ground level, and the soil compaction. The figure 8 show the soil cohesion as a function of the soil depth at four soil compactions at 23 % constant moisture content. The figure 9 shows the soil internal friction as a function of the soil depth at four compactions at 23 % soil moisture content.

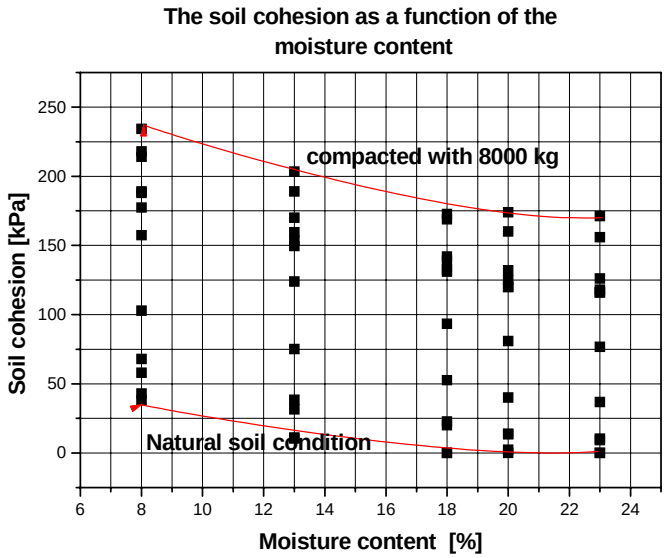


Figure 10.

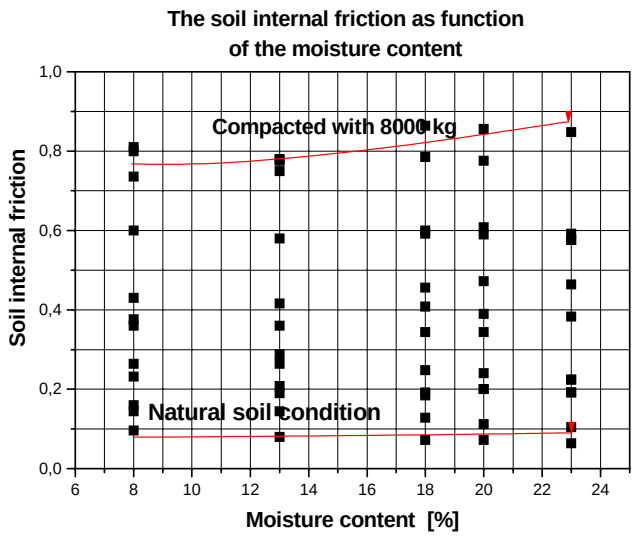


Figure 11.

Finally, the figure 10 shows the data of the soil cohesion as a function of the moisture content from the natural soil condition to the compacted soil condition, and figure 11 shows the data of the soil internal friction as a function of the soil moisture content from the natural soil condition to the compacted soil condition too.

In the in situ field test the soil cone index was between 75 and 120 CI which is an acceptable value.

4. Discussion

From the field tests and computed data we determinate the soil cohesion increase as a function of the depth of the soil. The size of the increase is reduced by the soil compaction. See figure 8.

This determination means that the off-road vehicle breaks and compacts the soil when it moves over the soil, therefore decreasing the soil cohesion.

The soil internal friction decreases as a function of the depth of the soil. See figure 9. The high value belongs to compacted soil. The difference of the soil internal friction value between the ground level and 50 cm depth was not equal. The difference in the natural condition was small, but in compacted soil was relative high. The changing of the soil internal friction was not significant, in spite of the soil cohesion.

The soil cohesion as a function of soil moisture content decreases. The high value belongs to the compacted soil condition. See figure 10. The changing of the internal friction as a function of the moisture content was not significant. See figure 11.

According to the above diagrams (figure 8. – 11.) we can determinate that the two important soil parameters, cohesion and the internal friction, are not constant in different soil types. Both of them change as a function of the soil depth and soil compaction and the cohesion as a function of the soil moisture content as well.

These determinations were established with the tested soil types and the soil conditions.

5. Conclusion

The target of my paper was to determinate how the soil cohesion and internal friction change as a function of the moisture content, the soil compaction and depth of the soil.

The well-known Janosi [7] equation of the towing force (which originates in the tire-soil interaction) is as follows:

Jánosi, Komandi equation:

$$F_t = (Ac + \mu Q)\{1 - \frac{K}{sL}[1 - \exp-(\frac{sL}{K})]\} \text{ [N]}$$

where:

A = the size of the caterpillar-soil or tire-soil* interaction area [m^2]
L = the length of the caterpillar-soil or tire –soil interaction area [m]
τ_{max} = The maximum value of the soil shearing stress [N/m^2]
$\tau_{max} = c + \mu\sigma$
c = soil cohesion (N/m^2)
μ = the internal friction of the soil
σ = the normal stress in the soil under the caterpillar or tire (N/m^2)
K = empirical factor [m]
s = the slip
* As professor Komandi use K_{red} empirical factor [m]

In the equation the two soil parameters, cohesion and internal friction, are constant.

From my research results this is not the case, as the two soil parameters change depending on the soil moisture content, soil compaction, and depth of the soil.

I can determinate that the soil cohesion increases as a function of the depth of the soil. The soil compaction reduces the cohesion. The soil cohesion and the soil shearing stress decrease as a function of the moisture content.

The soil internal friction decreases as a function of the depth of the soil.

If we want to calculate the towing force more accurately, we have to apply the functions of the soil cohesion and internal friction.

The soil type and soil condition are very variable, hence these findings were established specifically for the tested soil types and soil conditions.

The basis of the above conclusion is that we suggest to clarify the two mechanical soil parameters, cohesion and internal friction, as functions of the moisture content, compaction and depth of the soil. In addition more in-situ soil tests are suggested as well as more in-situ soil shearing tests.

References

1. BEKKER, M. G.: (1976) Theory of land locomotion. The mechanics of Vehicle mobility. Ann Arbor, The University of Michigan Press.
2. BEKKER, M., G.: Introduction of Terrain Vehicle Systems, University of Michigan Press, 1961
3. BEKKER, M.,G.: Theory of Land Locomotion, The Mechanics of Vehicle Mobility, University of Michigan, Press Ann-Arbor, 1956
4. BERNSTEIN, R.: (1913) Probleme zur experimentiellen Motorpflugmechanik. Der Motorwagen 1913. 16. sz.
5. BERNSTEIN.E.: Probleme zur experimentellen Motorplugmechanic., Der Motorwagen 16.heft. 1913

6. FREITAG, D. R.: (1965) A dimensional analysis of the performance of pneumatic tyres on sort soils. Technical Report No. 3-688. Us-Army Engineering Waterways Experimet Station, Vickburg. Ms. USA.

7. JÁNOSI, Z. - HANAMOTO, B.: (1961) The analytical determination of drawbar pull as a function of slip for tracked vehicles in deformable soil. Int. Conf. Mechanics of Soil-Vehicle System. Torino-Saint Vincent. 12/16.

8. KOMÁNDI, Gy.,: Bestimmung der physikalischen Bodenkennwerte auf Ground der Zusammenhange zwischen Scherdiagram und Zugkraftcharakteristik, Landmaschinen Forschung, Heft 4 1966.

9. KOMÁNDI, GY.: (1989): Kerék és a talaj adhéziós kapcsolatának elemzése traktor gumiabroncsokon. Akadémiai doktori értekezés. Budapest-Gödöllő.

10. SITKEI, GY.: (1978) Allgemeine Zusammenhänge zwischen der Leistung dem Gewicht und den optimalen Betriebsparamentern von Traktoren. Grundl. Landtechnik. 5/1978.

11. SITKEI, GY.: (1981) Mezőgazdasági anyagok mechanikája. Akadémiai Kiadó, Budapest

12. SITKEI, Gy.: Mezőgazdasági gépek talajmechanikai problémái, Budapest, Akadémia kiadó,1967

13. REECE, A.R.: (1964) Theory and practice of off-the-road locomotion. The Annual Conference, London

14. LAIB, L. - URBÁN, Z.: (1994) Gépjárművek terepjáró-képességének vizsgálata, geomorfológiai szempontok figyelembe vételével. Járművek, Építőipari és Mezőgazdasági Gépek, 41. Évf. 6. Sz. 210-218.

15. LAIB, L.: (1989) Terepjáró járművek mozgékonysága terepen. Kandidátusi értekezés, Gödöllő.

16. LAIB, L.: (1994) Terepjáró járművek mozgékonysága terepen I. Járművek, Építőipari és Mezőgazdasági Gépek, 41. Évf. 6. Sz. 203-209.

17. ROHANI, B. - BALADI, G.Y.: Correlation of mobility cone index with fundamental engineering properties of soil. 7th Int. Conf. ISTVS. Calgary

18. RULA, A. - NUTTALL, C.: (1971) An analysis of ground mobility models. Technicak Report M-71-4. US. Army Engineering Waterways Experiment Station. Vickburg, Ms

19. Kiss, P.: Effect of Soil Deformation on the Energy Balance of Tractors In: *Hungarian Agricultural Engineering* No. 12. p. 35-40. 1999. HU ISSN 0864-7410

Appendix 1

Table 1. The maximum soil shearing and normal stress, soil cohesion,internal friction

Number	Depth	Mass	Normal stress	Shearing stress	Cohesion	Internal friction	Moisture content	Compaction
	z	m		Tau	c		x	
	cm	kg	kPa	kPa	kPa		%	
1	0	3	93,7	19	9,3	0,104	23	natural
2	0	5	156,2	22	9,3	0,104	23	natural
3	0	7	218,6	32	9,3	0,104	23	natural
4	25	3	93,7	132	126	0,064	23	natural
5	25	5	156,2	134	126	0,064	23	natural
6	25	7	218,6	140	126	0,064	23	natural
7	50	6	93,7	160	118	0,383	23	natural
8	50	5	156,2	190	118	0,383	23	natural
9	50	7	218,6	202	118	0,383	23	natural

Table 2. The maximum soil she aring and normal stress, soil cohesion, internal friction

Number	Depth	Mass	Normal stress	Shearing stress	Cohesion	Internal friction	Moisture content	Compaction
	z	m	σ	τ	c	μ	x	
	cm	kg	kPa	kPa	kPa		%	kg
1	0	3	93,7	64	10,4	0,576	23	2000kg
2	0	5	156,2	96	10,4	0,576	23	2000kg
3	0	7	218,6	136	10,4	0,576	23	2000kg
4	25	3	93,7	134	116,1	0,192	23	2000kg
5	25	5	156,2	144	116,1	0,192	23	2000kg
6	25	7	218,6	158	116,1	0,192	23	2000kg
7	50	6	93,7	192	171,1	0,224	23	2000kg
8	50	5	156,2	205	171,1	0,224	23	2000kg
9	50	7	218,6	220	171,1	0,224	23	2000kg

Table 3. The maximum soil shearing and normal stress, soil cohesion, internal friction

Number	Depth	Mass	Normal stress	Shearing stress	Cohesion	Internal friction	Moisture content	Compaction
	z	m	σ	τ	c	μ	x	
	cm	kg	kPa	kPa	kPa		%	kg
1	0	3	93,7	70	0	0,576	23	4000kg
2	0	5	156,2	124	0	0,576	23	4000kg
3	0	7	218,6	168	0	0,576	23	4000kg
4	25	3	93,7	134	116	0,192	23	4000kg
5	25	5	156,2	150	116	0,192	23	4000kg
6	25	7	218,6	158	116	0,192	23	4000kg
7	50	6	93,7	192	156	0,224	23	4000kg
8	50	5	156,2	220	156	0,224	23	4000kg
9	50	7	218,6	240	156	0,224	23	4000kg

Table 4. The maximum soil shearing and normal stress, soil cohesion, internal friction

Nuber	Depth	Mass	Normal stress	Shearing stress	Cohesion	Internal friction	Moisture content	Compaction
	z	m	σ	τ	c	μ		
	cm	kg	kPa	kPa	kPa		%	kg
1	0	3	93,7	46	0	0,848	23	8000kg
2	0	5	156,2	88	0	0,848	23	8000kg
3	0	7	218,6	152	0	0,848	23	8000kg
4	25	3	93,7	92	36,9	0,592	23	8000kg
5	25	5	156,2	134	36,9	0,592	23	8000kg
6	25	7	218,6	166	36,9	0,592	23	8000kg
7	50	6	93,7	120	76,8	0,464	23	8000kg
8	50	5	156,2	144	76,8	0,464	23	8000kg
9	50	7	218,6	178	76,8	0,464	23	8000kg

Table 5. The maximum soil shearing and normal stress, soil cohesion, internal friction

Number	Depth	Mass	Normal stress	Shearing stress	Cohesion	Internal friction	Moisture content	Compaction
	z	m	σ	τ	c	μ	x	
	cm	kg	kPa	kPa	kPa		%	
1	0	3	93,7	24	13,5	0,112	20	natural
2	0	5	156,2	29	13,5	0,112	20	natural
3	0	7	218,6	38	13,5	0,112	20	natural
4	25	3	93,7	137	125	0,072	20	natural
5	25	5	156,2	141	125	0,072	20	natural
6	25	7	218,6	146	125	0,072	20	natural
7	50	6	93,7	165	132	0,344	20	natural
8	50	5	156,2	197	132	0,344	20	natural
9	50	7	218,6	208	132	0,344	20	natural

Table 6. The maximum soil shearing and normal stress, soil cohesion, internal friction

Number	Depth	Mass	Normal stress	Shearing stress	Cohesion	Internal friction	Moisture content	Compaction
	z	m	σ	τ	c	μ	x	
	cm	kg	kPa	kPa	kPa		%	kg
1	0	3	93,7	69	14	0,59	20	2000
2	0	5	156,2	101	14	0,59	20	2000
3	0	7	218,6	143	14	0,59	20	2000
4	25	3	93,7	139	120	0,2	20	2000
5	25	5	156,2	157	120	0,2	20	2000
6	25	7	218,6	164	120	0,2	20	2000
7	50	6	93,7	197	174	0,24	20	2000
8	50	5	156,2	211	174	0,24	20	2000
9	50	7	218,6	227	174	0,24	20	2000

Table 7. The maximum soil shearing and normal stress, soil cohesion, internal friction

Number	Depth	Mass	Normal stress	Shearing stress	Cohesion	Internal friction	Moisture content	Compaction
	z	m	σ	τ	c	μ	x	
	cm	kg	kPa	kPa	kPa		%	kg
1	0	3	93,7	75	2,36	0,776	20	4000kg
2	0	5	156,2	131	2,36	0,776	20	4000kg
3	0	7	218,6	172	2,36	0,776	20	4000kg
4	25	3	93,7	139	120	0,2	20	4000kg
5	25	5	156,2	157	120	0,2	20	4000kg
6	25	7	218,6	164	120	0,2	20	4000kg
7	50	6	93,7	197	160	0,39	20	4000kg
8	50	5	156,2	227	160	0,39	20	4000kg
9	50	7	218,6	246	160	0,39	20	4000kg

Table 8. The maximum soil shearing and normal stress, soil cohesion, internal friction

Number	Depth	Mass	Normal stress	Shearing stress	Cohesion	Internal friction	Moisture content	Compaction
	z	m	σ	τ	c	μ	x	
	cm	kg	kPa	kPa	kPa		%	kg
1	0	3	93,7	51	0	0,856	20	8000kg
2	0	5	156,2	95	0	0,856	20	8000kg
3	0	7	218,6	158	0	0,856	20	8000kg
4	25	3	93,7	97	40	0,608	20	8000kg
5	25	5	156,2	141	40	0,608	20	8000kg
6	25	7	218,6	173	40	0,608	20	8000kg
7	50	6	93,7	125	80,82	0,472	20	8000kg
8	50	5	156,2	151	80,82	0,472	20	8000kg
9	50	7	218,6	184	80,82	0,472	20	8000kg

Table 9. The maximum soil shearing and normal stress, soil cohesion, internal friction

Number	Depth	Mass	Normal stress	Shearing stress	Cohesion	Internal friction	Moisture content	Compaction
	z	m	σ	τ	c	μ	x	
	cm	kg	kPa	kPa	kPa		%	
1	0	3	93,7	32	20	0,128	18	natural
2	0	5	156,2	40	20	0,128	18	natural
3	0	7	218,6	48	20	0,128	18	natural
4	25	3	93,7	146	139	0,072	18	natural
5	25	5	156,2	152	139	0,072	18	natural
6	25	7	218,6	155	139	0,072	18	natural
7	50	6	93,7	175	142	0,344	18	natural
8	50	5	156,2	207	142	0,344	18	natural
9	50	7	218,6	218	142	0,344	18	natural

Table 10. The maximum soil shearing and normal stress, soil cohesion, internal friction

Number	Depth	Mass	Normal stress	Shearing stress	Cohesion	Internal friction	Moisture content	Compaction
	z	m	σ	τ	c	μ	x	
	cm	kg	kPa	kPa	kPa		%	kg
1	0	3	93,7	79	22,8	0,6	18	2000kg
2	0	5	156,2	111	22,8	0,6	18	2000kg
3	0	7	218,6	154	22,8	0,6	18	2000kg
4	25	3	93,7	150	132,7	0,184	18	2000kg
5	25	5	156,2	158	132,7	0,184	18	2000kg
6	25	7	218,6	173	132,7	0,184	18	2000kg
7	50	6	93,7	207	172,7	0,248	18	2000kg
8	50	5	156,2	223	172,7	0,248	18	2000kg
9	50	7	218,6	227	172,7	0,248	18	2000kg

Table 11. The maximum soil shearing and normal stress, soil cohesion, internal friction

Number	Depth	Mass	Normal stress	Shearing stress	Cohesion	Internal friction	Moisture content	Compaction
	z	m	σ	τ	c	μ	x	
	cm	kg	kPa	kPa	kPa		%	kg
1	0	3	93,7	84	0	0,785	18	4000kg
2	0	5	156,2	142	0	0,785	18	4000kg
3	0	7	218,6	182	0	0,785	18	4000kg
4	25	3	93,7	149	131	0,192	18	4000kg
5	25	5	156,2	168	131	0,192	18	4000kg
6	25	7	218,6	173	131	0,192	18	4000kg
7	50	6	93,7	207	168,8	0,408	18	4000kg
8	50	5	156,2	237	168,8	0,408	18	4000kg
9	50	7	218,6	258	168,8	0,408	18	4000kg

Table 12. The maximum soil shearing and normal stress, soil cohesion, internal friction

Number	Depth	Mass	Normal stress	Shearing stress	Cohesion	Internal friction	Moisture content	Compaction
	z	m	σ	τ	c	μ	x	
	cm	kg	kPa	kPa	kPa		%	kg
1	0	3	93,7	61	0	0,864	18	8000kg
2	0	5	156,2	106	0	0,864	18	8000kg
3	0	7	218,6	169	0	0,864	18	8000kg
4	25	3	93,7	108	52,588	0,592	18	8000kg
5	25	5	156,2	152	52,588	0,592	18	8000kg
6	25	7	218,6	182	52,588	0,592	18	8000kg
7	50	6	93,7	136	93,31	0,456	18	8000kg
8	50	5	156,2	162	93,31	0,456	18	8000kg
9	50	7	218,6	193	93,31	0,456	18	8000kg

Table 13. The maximum soil shearing and normal stress, soil cohesion, internal friction

Number	Depth	Mass	Normal stress	Shearing stress	Cohesion	Internal friction	Moisture content	Compaction
	z	m	σ	τ	c	μ	x	
	cm	kg	kPa	kPa	kPa		%	
1	0	3	93,7	52	38,52	0,144	13	natural
2	0	5	156,2	62	38,52	0,144	13	natural
3	0	7	218,6	70	38,52	0,144	13	natural
4	25	3	93,7	167	159,5	0,08	13	natural
5	25	5	156,2	174	159,5	0,08	13	natural
6	25	7	218,6	177	159,5	0,08	13	natural
7	50	6	93,7	197	170	0,288	13	natural
8	50	5	156,2	229	170	0,288	13	natural
9	50	7	218,6	233	170	0,288	13	natural

Table 14. The maximum soil shearing and normal stress, soil cohesion, internal friction

Number	Depth	Mass	Normal stress	Shearing stress	Cohesion	Internal friction	Moisture content	Compaction
	z	m	σ	τ	c	μ	x	
	cm	kg	kPa	kPa	kPa		%	kg
1	0	3	93,7	81	11,05	0,75	13	2000kg
2	0	5	156,2	133	11,05	0,75	13	2000kg
3	0	7	218,6	175	11,05	0,75	13	2000kg
4	25	3	93,7	171	153,5	0,19	13	2000kg
5	25	5	156,2	180	153,5	0,19	13	2000kg
6	25	7	218,6	195	153,5	0,19	13	2000kg
7	50	6	93,7	228	203,3	0,264	13	2000kg
8	50	5	156,2	245	203,3	0,264	13	2000kg
9	50	7	218,6	261	203,3	0,264	13	2000kg

Table 15. The maximum soil shearing and normal stress, soil cohesion, internal friction

Number	Depth	Mass	Normal stress	Shearing stress	Cohesion	Internal friction	Moisture content	Compaction
	z	m	σ	τ	c	μ	x	
	cm	kg	kPa	kPa	kPa		%	kg
1	0	3	93,7	104	31,5	0,78	13	4000kg
2	0	5	156,2	162	31,5	0,78	13	4000kg
3	0	7	218,6	202	31,5	0,78	13	4000kg
4	25	3	93,7	169	149,5	0,208	13	4000kg
5	25	5	156,2	190	149,5	0,208	13	4000kg
6	25	7	218,6	195	149,5	0,208	13	4000kg
7	50	6	93,7	228	189	0,416	13	4000kg
8	50	5	156,2	257	189	0,416	13	4000kg
9	50	7	218,6	280	189	0,416	13	4000kg

Table 16. The maximum soil shearing and normal stress, soil cohesion, internal friction

Number	Depth	Mass	Normal stress	Shearing stress	Cohesion	Internal friction	Moisture content	Compaction
	z	m	σ	τ	c	μ	x	
	cm	kg	kPa	kPa	kPa		%	kg
1	0	3	93,7	83	10,36	0,776	13	8000kg
2	0	5	156,2	135	10,36	0,776	13	8000kg
3	0	7	218,6	180	10,36	0,776	13	8000kg
4	25	3	93,7	129	75,2	0,58	13	8000kg
5	25	5	156,2	174	75,2	0,58	13	8000kg
6	25	7	218,6	202	75,2	0,58	13	8000kg
7	50	6	93,7	158	124	0,36	13	8000kg
8	50	5	156,2	184	124	0,36	13	8000kg
9	50	7	218,6	203	124	0,36	13	8000kg

Table 17. The maximum soil shearing and normal stress, soil cohesion, internal friction

Number	Depth	Mass	Normal stress	Shearing stress	Cohesion	Internal friction	Moisture content	Compaction
	z	m	σ	τ	c	μ	x	
	cm	kg	kPa	kPa	kPa		%	kg
1	0	3	93,7	82	68	0,144	8	2000kg
2	0	5	156,2	95	68	0,144	8	2000kg
3	0	7	218,6	100	68	0,144	8	2000kg
4	25	3	93,7	197	188	0,096	8	2000kg
5	25	5	156,2	206	188	0,096	8	2000kg
6	25	7	218,6	209	188	0,096	8	2000kg
7	50	6	93,7	250	214	0,376	8	2000kg
8	50	5	156,2	260	214	0,376	8	2000kg
9	50	7	218,6	297	214	0,376	8	2000kg

Table 18. The maximum soil shearing and normal stress, soil cohesion, internal friction

Number	Depth	Mass	Normal stress	Shearing stress	Cohesion	Internal friction	Moisture content	Compaction
	z	m	σ	τ	c	μ	x	
	cm	kg	kPa	kPa	kPa		%	kg
1	0	3	93,7	112	43	0,736	8	2000kg
2	0	5	156,2	166	43	0,736	8	2000kg
3	0	7	218,6	204	43	0,736	8	2000kg
4	25	3	93,7	204	189	0,16	8	2000kg
5	25	5	156,2	212	189	0,16	8	2000kg
6	25	7	218,6	224	189	0,16	8	2000kg
7	50	6	93,7	259	234,3	0,264	8	2000kg
8	50	5	156,2	270	234,3	0,264	8	2000kg
9	50	7	218,6	292	234,3	0,264	8	2000kg

Table 19. The maximum soil shearing and normal stress, soil cohesion, internal friction

Number	Depth	Mass	Normal stress	Shearing stress	Cohesion	Internal friction	Moisture content	Compaction
	z	m	σ	τ	c	μ	x	
	cm	kg	kPa	kPa	kPa		%	kg
1	0	3	93,7	134	57,93	0,81	8	4000kg
2	0	5	156,2	192	57,93	0,81	8	4000kg
3	0	7	218,6	235	57,93	0,81	8	4000kg
4	25	3	93,7	199	177,3	0,232	8	4000kg
5	25	5	156,2	223	177,3	0,232	8	4000kg
6	25	7	218,6	228	177,3	0,232	8	4000kg
7	50	6	93,7	258	218	0,43	8	4000kg
8	50	5	156,2	289	218	0,43	8	4000kg
9	50	7	218,6	312	218	0,43	8	4000kg

Table 20. The maximum soil shearing and normal stress, soil cohesion, internal friction

Number	Depth	Mass	Normal stress	Shearing stress	Cohesion	Internal friction	Moisture content	Compaction
	z	m	σ	τ	c	μ	x	
	cm	kg	kPa	kPa	kPa		%	kg
1	0	3	93,7	113	38,12	0,8	8	8000kg
2	0	5	156,2	135	38,12	0,8	8	8000kg
3	0	7	218,6	213	38,12	0,8	8	8000kg
4	25	3	93,7	159	102,8	0,6	8	8000kg
5	25	5	156,2	205	102,8	0,6	8	8000kg
6	25	7	218,6	234	102,8	0,6	8	8000kg
7	50	6	93,7	191	157,3	0,36	8	8000kg
8	50	5	156,2	215	157,3	0,36	8	8000kg
9	50	7	218,6	236	157,3	0,36	8	8000kg

Institute for Environmental Engineering Systems

Professor Dr. István FARKAS
Director of the Institute

Dear Reader,

The Institute for Environmental Engineering Systems consists of two departments as Department of Environmental and Building Engineering and Department of Physics and Process Control.

From the results of the research activity carried out in the framework of the Institute and its partner organizations in 2010 the following main topics are selected out to publish in the recent issues of the Journal of Mechanical Engineering Letters:

Solar energy for greenhouse heating:

- The solar wall surface temperature during the daylight increased rapidly as compared with the wall deep thickness and vise versa at night. Using the adobe and red brick storage walls led to increased the interior air and soil temperatures above the greenhouse that without walls.

Use of geothermal energy in buildings:

- Utilization of thermal water heat at heating and domestic hot water production is more effective than using conventional gas boilers. Regarding geothermal energy utilization it is favourable to apply a so called staged use or multipurpose use.

Green roofs:

- The calculation and measurements definitely show economical advantages of green roofs in contrast of gravel or bare flat roofs. Creating extensive green roofs appears to be a good solution considering the changing weather conditions.

Coupled transport process:

- The general formalism of the classical irreversible thermodynamics in linear approximation is connected to the concept of quasipolynomials. It is demonstrated, that this refinement of the existing descriptions emanates directly from the Curie principle and is based on the effective use of group representation theory.

Heat and mass transfer:

- An analytical model gives a tool to determine the mass transfer coefficient for evaporating moist from fruit surface. The model can be used for calculating the surface temperature which is in accordance to the measurements, using thermal imaging. This work was conducted by a honorary professor of the Institute.

Storage Wall Techniques for Greenhouse Heating by Solar Energy

Sameh KISHK[1] and István FARKAS[2]
[1]Department of Agricultural Engineering, Faculty of Agriculture,
Suez Canal University, Egypt
[2]Department of Physics and Process Control,
Institute for Environmental Engineering Systems

Abstract

In order to overcome greenhouse air temperature drop within the cold winter nights solar storage wall used to increase inside air temperature at night. Trials were carried out at the farm of the Faculty of Agriculture, Suez Canal University, Ismailia, Egypt (latitude of 30.62^o, longitude 32.27^o and 5m above sea level) between periods from 30th November 2004 till 22nd of February 2005. Adobe and red brick solar storage wall were used as an auxiliary solar heating and compared with a control trial greenhouse without storage walls. Soil covers were applied to study the effect of mulching on the soil heating. Straw mulch of 5cm thick and black plastic mulch of 0.1 mm thick were compared with the uncovered soil. The adobe and red brick storage walls increased the greenhouse interior air temperature by 11 and 11.8 % above that without storage wall, respectively. Average soil temperatures for the covered soil with straw and black plastic mulch were increased above the uncovered soil.

Keywords

Solar storage wall, North wall, soil mulch, soil heating

1. Introduction

A greenhouse is usually built to intensify the agricultural crops productions. Solar radiation is a limiting factor for greenhouse production, especially during winter months. The amount of solar radiation available to the plants cultivated in a greenhouse is affected by the greenhouse structural frame, covering material, surrounding topography and orientation (Abdel-Ghaffar and Helmy, 1988 and Abdel- Latif and Helmy, 1988). A greenhouse is essentially an enclosed structure, which traps short wavelength solar radiation and stores long wavelength thermal radiation to create a favourable micro-climate for higher productivity. Thus, greenhouses are known as controlled environment greenhouses (Jain and Tiwari, 2003).

A greenhouse can be managed to protect the plants by creating a favourable environment, allowing intensive use of soil, and helping sanitary plant control (Ozgener and Hepbasli, 2005). From an economic point of view, the main objective of horticultural greenhouses is to advance the normal season

production or to obtain a completely out-of-season production, which corresponds with higher crop prices. Many variables must be controlled in order to provide the good environmental conditions.

The most important parameters to be controlled inside a greenhouse are temperature, humidity and light. Especially, the temperature at night appears as an important critical variable to be controlled. Greenhouse makes it possible to increase crop productivity by maintaining a favourable environment for plants; and thus production by using greenhouse has become more popular than in the past (Cemek et al. 2006). About 96% of the greenhouse production area is concentrated in the Mediterranean and Aegean regions (Benli and Durmus, 2009).

Greenhouse is a structure, which provides the most suitable microclimate for the maximum plant growth during off-season. Air temperature is one of the most dominant parameters affecting the plant growth. (Sethi, 2009) study, five most commonly used single span shapes of greenhouses viz. even-span, uneven-span, vinery, modified arch and quonset type have been selected for comparison. The length, width and height (at the center) are kept same for all the selected shapes for both east-west and north-south orientation. Experimental validation of both the models is carried out for the measured total solar radiation and inside air temperature for an east-west orientation. Results show that uneven-span shape greenhouse receives the maximum solar radiation during each month of the year at all latitudes, whereas Quonset shape receives the minimum solar radiation during each month of the year at all latitudes. Air temperature remains the highest inside an uneven span shape and the lowest in a Quonset shape as compared to other shapes during different months of the year. East-west orientation in North semi spherical is best suited for year round greenhouse applications at all latitudes as this orientation receives greater radiation in winter.

North wall insulation of an east west oriented gothic arch greenhouse saved 30% in heating costs (Gupta and Chandra, 2002). (Khedari et al. 2003) showed that, Trombe wall is populated application composed of glass, an air gap and concrete wall is a widely used for heating in winter. (Hassanain and Hokam 2005) found that, the greenhouse with the passive Trombe wall gave higher greenhouse air temperature as compared with that without walls for the same greenhouse.

The plastic cover was affected the plant growth by decreasing days from planting to emergence and the maturity of many vegetable crops including tomato and cucumber. Also, it increased the total yield and the marketable yield over no plastic mulch (Kwabiah, 2004). (Perez-Diaz et al. 2004) found that, dark mulches (black, red and gray) had higher soil warming ability than light coloured mulches, as indicated by their higher root zone temperature value. Also, they found that, the root zone temperature under coloured mulches significantly influenced the growth of tomato plants greater than those of the un-mulched soil. Straw mulch might be useful against degeneration i.e. for virus control in potato seeds but straw mulching in potatoes was disappeared from the commercial practice due to its function to increase the soil moisture (Doring et al., 2005). Soil temperatures influence the cultivation biochemical reactions such

as seed germination, plant growth, nutrient availability and pesticide degradation (Salh and Medany, 2003) found that, average soil temperature was influenced by air temperature under the different types of plastics greenhouses.

This study aims to overcome the reduction in the greenhouse air temperature, especially in the cold winter nights. This achieved using solar storage walls. Beside solar storage wall soil covers were studied.

2. Greenhouse construction

Experiments were carried out at the Faculty of Agriculture Farm Suez Canal University, Ismailia, Egypt (latitude of 30.62^{o}, longitude 32.27^{o} and 5m above sea level) through the period from 30th November 2004 till 22nd February 2005. Three experimental gable uneven-span greenhouses with plane dimensions of 4 m long and 3 m wide (area of 12 m^{2}) were used. Figs 1-2 represent all dimensions and orientation for greenhouses. Galvanized pipes of 25.4 mm diameter were used to construct the frame of the greenhouse; while a transparent polyethylene sheet of 0.2 mm thick was used as greenhouse sheet. Each greenhouse was divided to three soil treatments, the first covered by black polyethylene mulch thickness of 0.1 mm, second one covered with straw (horse bean straw) of 5 cm average length and the final one uncovered soil. Two different materials were used to build the solar storage wall, adobe and red brick walls. One greenhouse was included adobe bricks while the second one included red bricks storage and the third control greenhouse without storage wall.

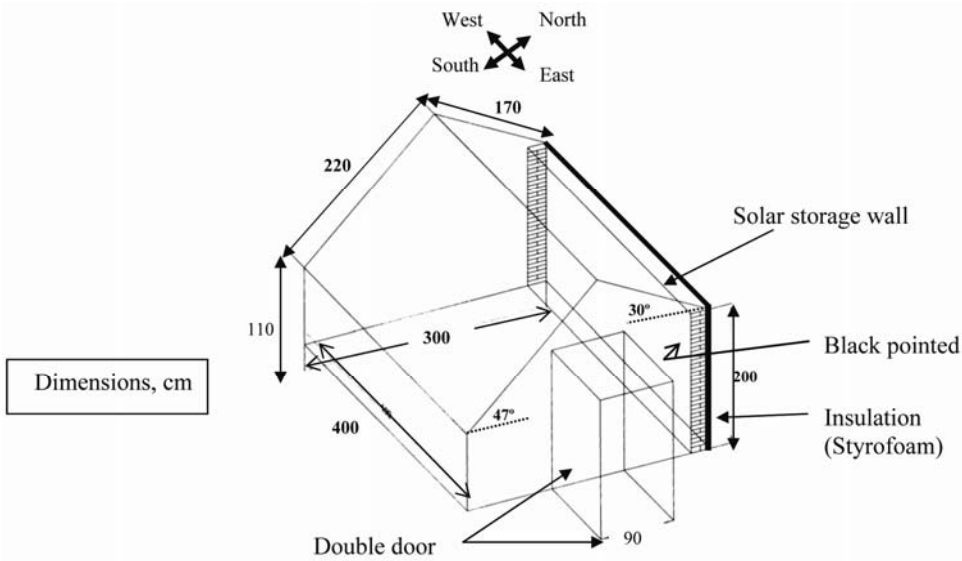


Figure 1. Schematic diagram of gable uneven-span greenhouse

Walls dimensions were 4.0 m long, 2.0 m high and 0.23 m thick. Internal wall surface was painted with matt black paint of 0.93 absorptivity. Meanwhile, the external surface was insulated by Styrofoam layer of 30 mm thick and 0.03 $Wm^{-1}K^{-1}$ thermal conductivity. The greenhouse soil was found sandy texture of 95.97% sand, 3.81% silt and 1.22% clay. Nile squash variety Seeds belongs to cucurbitaceae family was sown in the soil on 30th November 2004 at 0.5 m a part on the two sides (El-Shatoury, 2005). Soil was irrigated and kept at 0.5 or (50 %) from the field capacity, F.C. (Hassanain and Hokam, 2005).

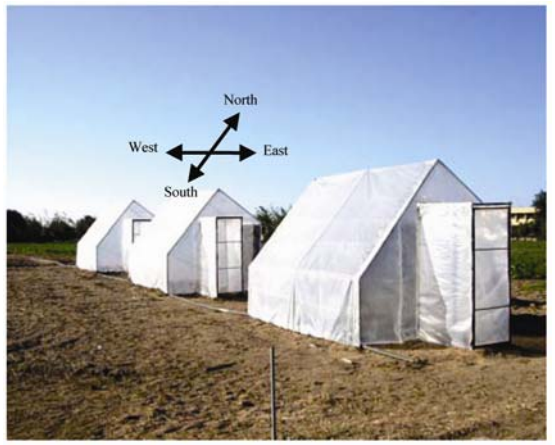


Figure 2. Greenhouses configuration

3. Measurements carried out

In this paragraph the methodology and the equipment used for measurements are discussed and evaluated.

3.1. Methodology of measurements

Determination of the Incident radiation and the greenhouse covering sheet Transmittivity

The transmitted solar radiation was measured to determine the amount of energy stored by the solar wall. The transmitivity of the used covering sheet was found to be 89% from the incident solar radiation.

Solar storage wall, soil and air temperatures determination

Five distributed locations were selected on the wall depth to measure the storage wall temperatures (around the three axes X, Y and Z) with five wall thick measuring depths at the surface of 0, 5, 10, 15 and 20 cm, respectively. Four measuring points were distributed on the interior wall side; these points were far 0.30 m from the four wall-corners, while the fifth measuring point was located exactly on the wall diagonal point i.e. 2 m from X (East and West sides) and 1m from the greenhouse ground level. Soil temperatures were measured at the soil

surface (0), 5 and 10 cm soil depths. These depths include the most of squash roots 20-30 cm for the surface soil. Interior air temperature was measured at different points inside the greenhouses using hangings thermometers.

Specific heat determination
The solar wall specific heat (C_p) also the greenhouse soils as well were determined in $Jkg^{-1}K^{-1}$ using a locally made calorimeter according to the method of (Klute, 1986). The average specific heat was found to be 882, 588 and 924 $Jkg^{-1}K^{-1}$ for the adobe, red bricks storage walls and the greenhouse soil, respectively.

Walls density
Storage wall density (ρ) was carried out using the paraffin wax method (Black, 1965). It was found to be 1900 kg m^{-3} for the adobe wall. Meanwhile, the red brick storage wall was found to be 1800 kg m^{-3}.

Squash (Cucurbita Pepo L) seeds vitality
Indoors squash seeds vitality under the indoor lab condition of 13.7 ^{o}C average ambient air temperature for 8 days was found in average of 90%.

3.2. Equipment used for measurements

Data were taken each two hours around 24 hours within the experiments. Weather conditions were measured, i.e. hourly global solar radiation, ambient air temperature, relative humidity (inside and outside greenhouse), inside greenhouse air temperature, wall thickness temperature and temperature of the different soil depths. The following instrumentations were used:

Incident solar radiation
Due to the cost a solar cell connected to a digital multimeter was used to determine the incident solar radiation (G) according to (Mujahid and Alamoud, 1988). A previously calibration was carried out against an American made Apply Pyranometer before the experimental work. The short circuit reading obtained from the cell was converted into W/m^2 (Duffie and Beckman, 2006).

Temperature measurements
Ambient and interior greenhouse air temperatures also the soil depths and wall thickness temperature were measured by digital thermocouples (BTC type range -50 - 120 ºC).

Relative humidity
Relative humidity was determined by the psychrometric chart using dry and wet bulb temperatures.

Wind speed
A TESCO 405-V1 Hot Wire Anemometer was used to measure the prevailing wind speeds outside the greenhouse.

4. Results and discussions

In this chapter first the weather conditions of the measurement are shown. After that the different measurement results are discussed as evaluated as Trombe solar storage wall temperature, soil temperature, air temperature and the fresh and dry weight for the squash plant

4.1. Weather conditions

Weather conditions throughout the experiments period were averaged, summarized and presented in Table 1.

Table 1. Average weather conditions throughout the field experimental work

Stage	Stage length, day	Average day length, hr	Ambient air temperature, ^{o}C	Wind speed, ms^{-1}	Solar radiation intensity, Wm^{-2}	Relative humidity, %
Germination	12	10.32	14.8	0.9	496.7	74.0
Vegetation	22	10.30	13.8	1.1	407.5	71.0
Flowering	13	10.30	12.5	0.7	327.6	74.9
Fruit	32	10.43	14.6	1.2	393.3	59.2

4.2. Trombe solar storage wall temperature

Trombe solar storage wall was acting as heat storage; during the daylight it stored heat energy, while at night time it dissipated the stored heat to the greenhouse environment. Average solar storing walls temperatures are given in Fig. 3 for the different plant stages. For a day in the seeds germination stage (30^{th} November) maximum temperature within the day from sunrise to sunset were recorded at 2.00 afternoons for adobe and red brick solar storage wall. Surface temperature of the adobe wall reached its maximum (within the storing period) of 57 ºC, while other thicknesses of the same wall were 50, 47, 37, 27 ºC for the wall thick of 5, 10, 15 and 20 cm, respectively. This affected the inside greenhouse air temperature which was found as 38.8 ºC, while the outside air temperature was 22.0 ºC. Meanwhile for the red brick solar wall the beak temperature was 66 ºC for wall surface, while other wall thickness were 54, 45, 33 and 31 ºC for the wall thick of 5, 10, 15 and 20 cm, respectively when the interior air temperature was 39 ºC.

Taken into consideration the greenhouse soil has no green cover when these measurements were carried out (germination stage). Within the night time (discharging period) which occurred from sunset to next day sunrise the stored heat dissipated from the previous mentioned stage. From Fig. 3A the solar wall surface dissipated the heat faster that the other wall thickness. As within this stage wall temperature is cooling down in opposite way of the heating stage. Wall thickness temperature at night can be arranged as 20, 15, 10, 5 cm and wall surface

respectively. The adobe wall temperatures reached 21, 19, 17, 17 and 15 ºC for the 20, 15, 10, 5cm and wall surface thickness, respectively when the interior air temperature was 9.3 ºC. Meanwhile, it reached to 21, 19, 17, 17 and 15 ºC for the 20, 15, 10, 5cm and wall surface thickness, respectively when the interior air temperature was 9.0 ºC.

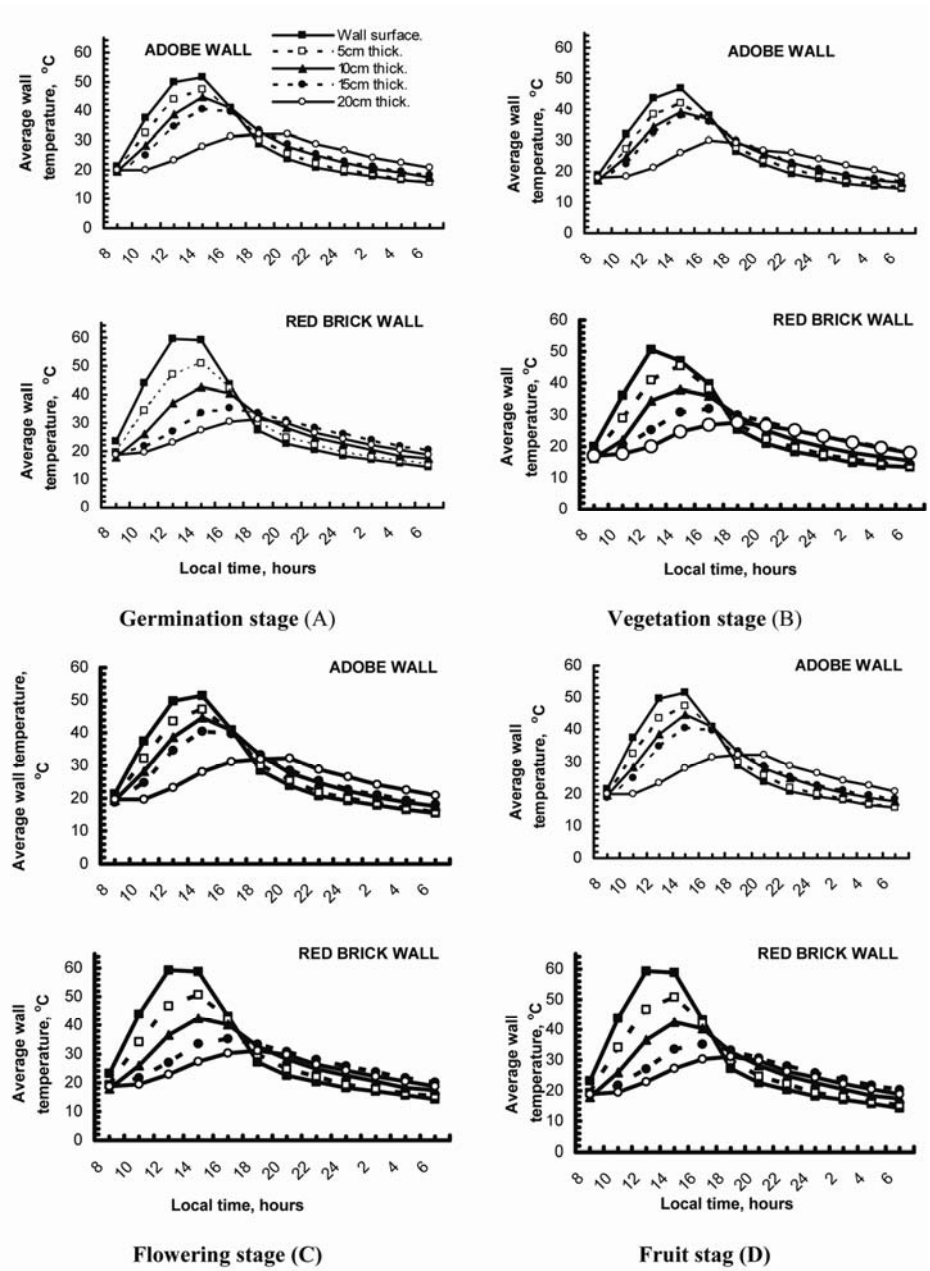


Figure 3. Average temperatures of solar storage wall at the different plant stages

Fig. 3B represented solar storage wall temperatures within vegetation stage for adobe and red brick walls within the period 13th December 2004 till the 4th January 2005 under the average weather conditions in Table 1. Solar storage wall within flouring and fruit stages represents in Figs 3C and 3D respectively. The temperatures distribution for the adobe and red brick walls for all thickness same in the first stage (germination).

4.3. Soil temperature

Average soil temperatures distributions within the all plant stages are shown in Fig. 4. The Figure is based on the average of soil temperature under different treatments (uncovered soil, covered soil with black plastic and straw mulch) under the average prevailing weather conditions given in Table 1. Average soil temperature inside the greenhouses includes adobe and red brick solar wall are increased above that inside control greenhouse for all stages.

In Fig. 5 as a result of the presence of solar storage wall for a day in the germination stage on the 30th November, 2004 it is shown that, the uncovered average soil temperature under the greenhouse that comprised adobe and red brick solar storage wall increased by 1.0 and 1.3 ºC, respectively above that inside the greenhouse without storage wall. Covered soil with a plastic mulch increased soil temperature by 2.0 and 1.5 ºC, respectively and covered soil by straw increased by 0.4 and 1.0 ºC, respectively above that inside the greenhouse without storage wall under the same weather conditions. One the other hand as a result of covered soil under the same any greenhouse soil temperature for soil covered with plastic or straw mulch were increased above that uncovered soil.

4.4. Air temperature

The effect of the solar storage wall presence on the greenhouse interior air temperature for all plant stages are shown in Fig. 6 for all plant stages, around 24 hours with 2 hours apart. It is noticed from the shown Figure the average interior air temperature for 24 hours were found to be 17.8, 19.8, and 19.9 ºC for the greenhouse without storage walls, the greenhouse that comprised adobe wall and the greenhouse that comprised red brick storage wall, respectively. When average ambient outside air temperature of 13.9 ºC and average weather conditions were mentioned before in Table 1.

In another word it can be said that, the Trombe wall increased the inside air temperature by 11 and 11.8% above that without storage wall if the Trombe wall of adobe and red brick were attached to the greenhouse, respectively. This due to the solar storage wall collects and stores solar energy during the day time and it emits this energy during the night. From the previous results we can see the solar storage wall were significant for increase the interior air temperature beside increase soil temperature but the solar wall material not significant.

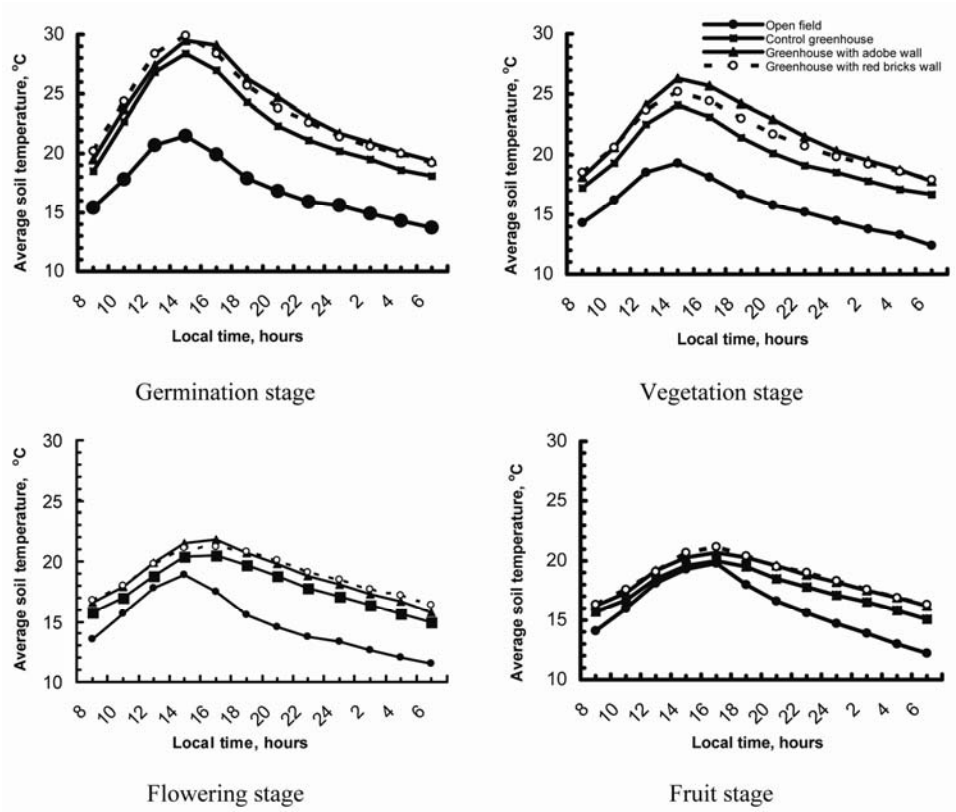


Figure 4. Effect of solar storage wall on the average soil temperature for different plant stages

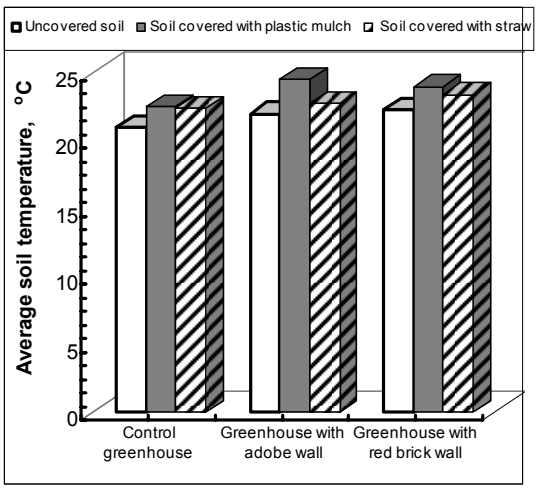


Figure 5. Average soil temperatures as affected by solar wall beside soil covers

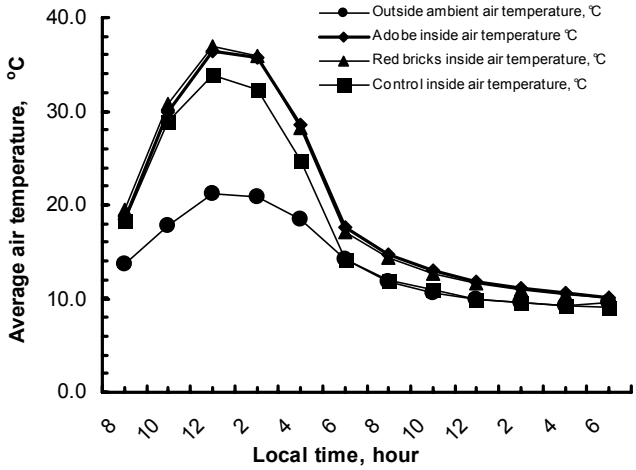


Figure 6. Average interior air temperature as affected by solar wall for all plant stages

4.5. Fresh and dry weight for the squash plant

Data represented in Table 2 indicates the average plant fresh and dry weight at the end of the season, which cultivated in different soil covering type. The average weight was highly affected by the solar storage wall type. The plant sowing under the greenhouses with solar storage wall had a greater dry weight as compared with that from the unheated greenhouse (control trials).

Table 2. Average fresh and dry weight of one squash plant at the end of the cultivation season, kg

Greenhouses	Greenhouse with adobe wall	Greenhouse with red brick wall	Greenhouse without solar wall
Fresh weight	1.785	1.425	1.365
Dry weight	0.139	0.127	0.115

5. Conclusions

The study conducted to the following conclusions:

- The solar wall surface temperature during the daylight increased rapidly as compared with the wall deep thickness and vise versa at night (dissipating period). As its heat lost faster than the other wall thickness temperatures. It lost its heat rapidly as compared with the other wall depths.
- Using the adobe and red brick storage walls led to increased the interior air and soil temperatures above the greenhouse that without walls.

- Using the plastic and straw mulch increased the soil temperatures above the uncovered soil for any greenhouse.
- Average fresh and dry weight of squash plants at the end of the cultivation season was weighty for the plants that were cultivated under the greenhouses with adobe solar storage wall above that was cultivated in the greenhouse with red brick storage wall above that cultivated inside the greenhouse without storage walls.

Acknowledgements

The authors express his acknowledge for the support of Hungarian Scholarship Committee and the Ministry of Higher Education in Egypt.

References

Abdel-Ghaffar, E.A. and Helmy, M.A., 1988. Studies of the effect of external climatological variables in the air temperature inside a plastic greenhouse located at Kafr El-Sheikh. Misr Journal of Agricultural Engineering, 5(2): 179-190.

Abdel-Latif, S.M.; and Helmy, M.A., 1988. Some parameters affecting solar energy available inside the greenhouse under Kafr El-Sheikh conditions. Misr Journal of Agricultural Engineering, 5(2): 167-178.

Benli, H. and Durmus, A., 2009. Performance analysis of a latent heat storage system with phase change material for new designed solar collectors in greenhouse heating. Solar Energy, (83): 2109–2119.

Cemek, B., Demir, Y., Uzun, S. and Ceyhan, V., 2006. The effects of different greenhouse covering materials on energy requirement, growth and yield of aubergine. Energy, (31): 1780–1788.

Doring, T.F., Brandt, M., Heb, J., Finckh, M.R. and Saucke, H. 2005. Effects of straw mulch on soil nitrate dynamics, weeds, yield and soil erosion in organically grown potatoes. Field Crops Research, Germany, 94, 238-249.

Duffie, J.A. and Beckman, W.A., 2006. Solar engineering of the thermal process, Wiley Interscine New York, USA, 3-44.

El-Shatoury, R.S, 2005. Effects of the plants density and nitrogenous fertilizer on growth characters of summer squash, Master of Science, Horticulture department, Faculty of Agriculture, Suez Canal University, Egypt.

Gupta, M.J. and Chandra, P., 2002. Effect of greenhouse design parameters on conservation of energy for greenhouse environmental control. Energy, (27): 777-794.

Hassanain. A.A. and Hokam, E.M., 2005. Trombe wall storage effect on the greenhouse passive solar heating. Egypt Journal of Applied Science, Egypt, 20(8B): 710 - 740.

Jain, D. and Tiwari, G.N. (2003): Modeling and optimal design of ground air collector for heating in controlled environment greenhouse. Energy Conversion and Management, 44 (8): 357–1372.

Khedari, J.; Rachapradit, N. and Hirunlabh J., 2003. Field study of performance of solar chimney with air-conditioned building. Energy, (28): 1099-1114.

Klute, A, 1986. Methods of soil analysis. Physical and Minarological Methods, 2nd Edition, Madison Wisconsin, USA, 9, 1.

Kwabiah, A.B, 2004. Growth and yield of sweet corn (Zea mays L.) cultivars in response to planting date and plastic mulch in a short- season environment. Scientia Horticulturae, (102): 147-166.

Mujahid, A.M. and Alamoud, A.R.M., 1988. An easily designed and constructed photovoltaic pyrheliometer. Solar &Wind Technology, 5(2): 127-130.

Ozgener, O. and Hepbasli, A., 2005. Performance analysis of a solar assisted ground source heat pump system for greenhouse heating: an experimental study. Building and Environment, (40): 1040–1050.

Perez-Diaz, J.C., Granberry, D., Bertrand, D. and Giddings, D., 2004. Tomato plant growth during establishment as affected by root zone temperature under colored mulches. Acta Horticulturae, (631): 119-124.

Saleh, S.M. and Medany, M.A., 2003. Effect of polyethylene colour on the growth and production of the cucumber (*Cucumis Sativum*) under greenhouses during Autumn season. Acta Horticulturae, 608, 259-265.

Sethi, V.P., 2009. On the selection of shape and orientation of a greenhouse: Thermal modeling and experimental validation. Solar Energy (83): 21-38.

Utility of Geothermal Energy in Building Energy Supply with Exergy Theory

György né HALÁSZ
Department of Environmental and Building Engineering,
Institute for Environmental Engineering Systems

Abstract

The aim of the present article – knowing the main characteristics of the Hungarian geothermal possibilities – is the utilization of the geothermal energy and thermal water in energy supply, primarily in heat supply. Energetic and thermodynamic analization based on the measurement results of an existing district heating system supplied with thermal water.

Keywords

Geothermal energy, Utilization of thermal water in buildings, Exergy theory.

1. Introduction

In the interest of the future generation one of the most important challenges of mankind beside sustainable development is environment protection. Towards this to create energy systems with the minimized utilization of fossil primer energy source. This is not only in favour of environment protection yet because the unrevealed fossil primer energy source fields are limited.

In 2007 the primer energy consumption of Hungary according to Hungarian Energy Office database was 1.125,5 PJ, of which 62 % was imported. 79,3 % of the consumed primer energy sources was fossil fuel, while 5,2 % was the renewable energy sources (57,8 PJ) and 15,5 % was the primer electric energy. About 70 % (794,3 PJ) of the primer energy was delivered to the final consumers of which 29 % (232,5 PJ) is the domestic use – as the main consumer. 4 % (34,1 PJ) is the renewable energy sources of the final energy consumed.

In 2007 only 6,4 % of the total renewable primer energy sources (58,7 PJ) was geothermic energy. These values have not changed significantly in the past years.

Although Hungary has significant geothermic energy known worldwide - the heat flow density value can reach up to 100 mW/m^2, and the maximum geothermic gradient value approaches 60 $^oC/km^2$ – there are no large enthalpy geothermic fields in the country. According to deep drillings in most areas of the country 200 ˚C rock temperature can be reached in 3000-5000 m depth, which means the crystal or carbonate rock aquifers in most cases. From sandstone in

1200-2500 m depth maximum 100 ˚C, from limestone in 3200 m depth max. 150 ˚C, from 1000 m deep water aquifer 55-75 ˚C is the available water temperature. The exploited water amount of a well is 200-3000 m^3/day, in energetic 30-80 m^3/hour well output can be calculated (Tóth A. 2010.).

2. Geothermal Energy in Heat supply

The possible usage of the geothermal energy is based on the temperature and enthalpy of the exploited fluid. The utilization of geothermal energy has two main fields in energy supply: electric power production and heat utilization.

Low temperature thermal water utilization for direct power energy is not effective even considering optimal conditions. According to presently known technologies the real efficiency is about 10 %.

Low enthalpy geothermic energy, lower than 100 ˚C temperature can be exceedingly used in central heat supply. The utilization can possibly work with or without thermal water exploitation. Without exploitation – some literatures call it closed cycle heat production – is a technology when a fluid in a closed system transports the energy to the surface. In practice known and applied systems without thermal water exploitation are geothermal heat pump systems. The 50-150 m deep 10-15 ˚C temperature heat transfer fluid (water) is exploited. This water with heat pump equipment supports building heating and hot water supply. In the closed cycle the 10-15 ˚C temperature flowing fluid is suitable even without heat pump to decrease the winter and summer energy consumption of the building by tempering the boundary structures.

3. Heat Utilization with Exploitation of Thermal Water

The centralized heat supply with thermal water can be as follows:

- *Directly*, when thermal water has favourable physical and chemical qualities, has no corrosion effect to pipes and equipment and the thermal water is not inclined to precipitate formation. In this case thermal water is directly transported to final consumer. Buildings can be heated directly and hot water equipment can be supplied directly by thermal water.
- *Indirectly*, when thermal water has unfavourable physical and chemical qualities. In this case thermal water occurs as primer fluid in heat utilization and a secunder fluid transports the heat to the consumer by an installed heat exchanger. In this case consumers are basically: heating systems, domestic hot water producing systems, absorption and adsorption chillers.

In Hungary the salt and chemical content of the exploited thermal water does not allow direct heat utilization. Thermal water exploited exclusively for energy utilization purposes must be reinjected according to 219/2004. (VII.21.) Government Edict.

When utilizing thermal water it is important to aim for rational, complex utilization in order to utilize the heat energy of the fluid at a possible maximum rate. The possible results are:

- less environmental load/impact
- saving the treasures of the Earth
- economic increase of high investment demanded systems

In Hungary the higher temperature thermal water sources can be found in the Southern Great Plains, along river Tisza, North and South part of Transdunabia. The evolved townships and their needs not always match the geothermal possibilities.

Before establishing geothermal energy used systems extensive information of the needs and possible solutions are important.

The possible utilization of geothermal energy, the development of the utility systems, the type, size, extent of system equipment, the dependability of geothermal energy used systems, the economic factor of the systems are affected by the following main factors:

- temperature of thermal water as heat transmitting fluid
- output of the thermal well
- fluctuation of volume flow according to production pressure change
- thermal water physical and chemical attributes (gas content of the fluid, consistency of the gas content, salt content, metal content, sand content, incrustation amount according to production pressure and temperature…etc.)
- types of consumers and their needs
- the fluid temperature to fulfill the requirements
- temporal change of consumer needs
- spatial position of the consumers (low temperature thermal water transported to long distance as well as the heat loss during transport and the pump work can question the economic factor of the complete system)

Utilization of the exploited thermal water is possible:

- directly – through simply water-water heat exchanger
- indirectly – through heat pump

This can be constructed with a system that contains less and less heating need elements connected in a row. (Figure 1.) The heat exchangers with one possible calculated primer and secunder side temperatures can be seen in following figure. There is a two-way regulation valve installed into the circuit side of the linearly connected heat exchangers.

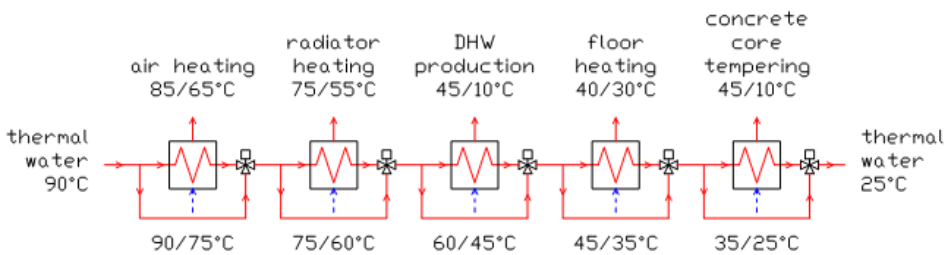


Figure 1. Heat exchangers connected linearly

If the thermal water temperature is considered to be constant, then the result of this connection is that the thermal water temperature exiting from the heat exchanger increase by the decline of the consumer needs – within extreme circumstances, if there is no heat loss on the heat exchanger the entering water temperature is parallel to the exiting water temperature, namely the water quitting without cooling, while the mass flow rate of the primer fluid/thermal water is constant.

If the thermal water temperature considered being constant, but the consumer heat exchangers connected parallel, decline of consumer needs results drop of temperature of exiting thermal water and drop of its mass flow rate.

The following figure shows an example of a connection of a primer side of thermal water based district heat supply system – not including the connecting scheme of the consumer heat centrals. This can be identified as a typical system, mostly used in constructing thermal water heat supply systems in Hungary. The consumer heat centrals may differ according to the type of the system and the typical parameters of the consumer needs.

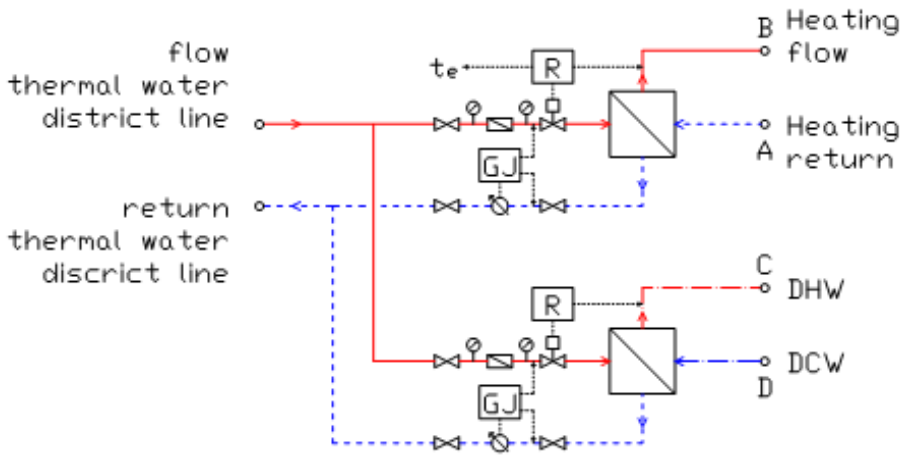


Figure 2. Indirect type heat central with altering primer mass flow rate – connected to a thermal water district line

Figure 2. Shows indirect type, altering primer mass flow rate based heat centrals connected to a thermal water supplied district line system built in South-Hungary in 2007. All the heat exchangers are connected parallel. Before the heat exchangers there have been a one-way regulation valve installed into the primer circuit. These one-way regulation valves change the power of the heat exchangers by changing the amount of the primer fluid and hence regulating the temperature of the secunder fluid. According to the outdoor air temperature, the secunder flow heating water is being regulated by the one-way regulation valve and during producing domestic hot water; the hot water temperature is set to a constant value. The consumer systems are primarily the heating and domestic hot water supply systems of local government schools, kindergartens, offices

and hotels. In one building also a thermal water bath has been established. (Figure 4.) In the buildings, thermal water as primer fluid supplied heat centrals were constructed during the reconstruction of the existing gas boiler system. This leaves us the opportunity to be able to operate the system by gas supplied heating system.

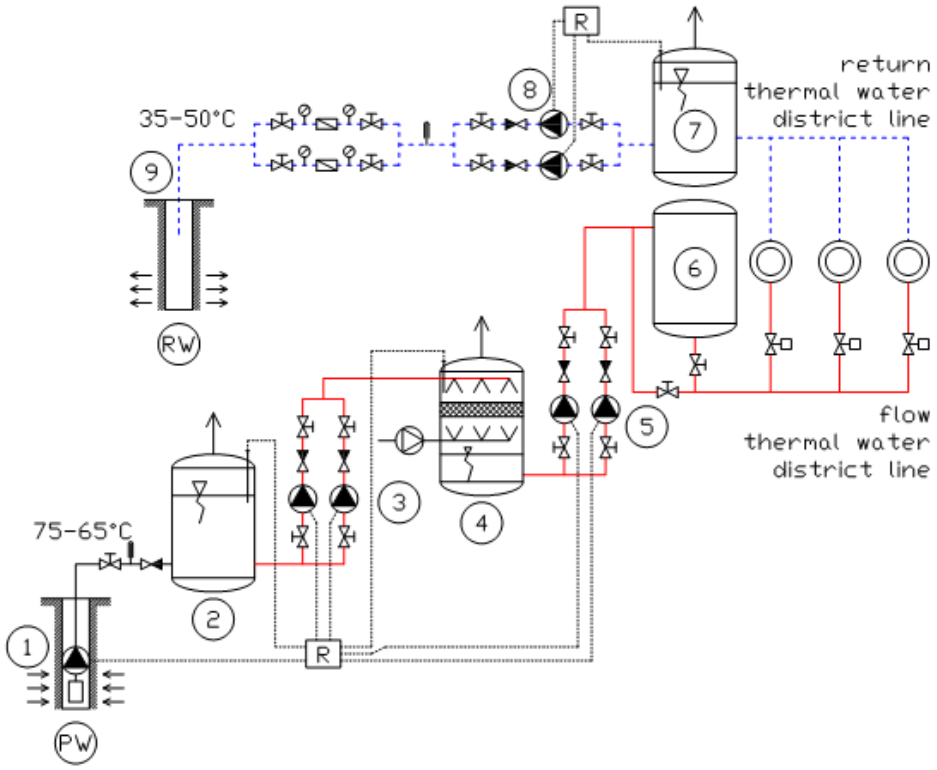


Figure 3. District heating system supplied with thermal water (exploiting and reinjection system elements)

Using the markings of Figure 3., the introduction of the system in brief is:

From the production well a submersible pump (1) transports the thermal water to the prior degassing tank (2) which is opened to the atmosphere. From here pressure rising pumps (3) transport the fluid to the degassing tank (4). Then from the degassing tank the fluid is transported through the demetalling (6) to all the heat centrals by pressure rising pumps (5). After the fluid returns to the opened resting tank (7) of the „heat central" established at the exploiting well. Finally the fluid used only for heating purpose is reinjected to the reinjection well by pressure rising pumps (8).

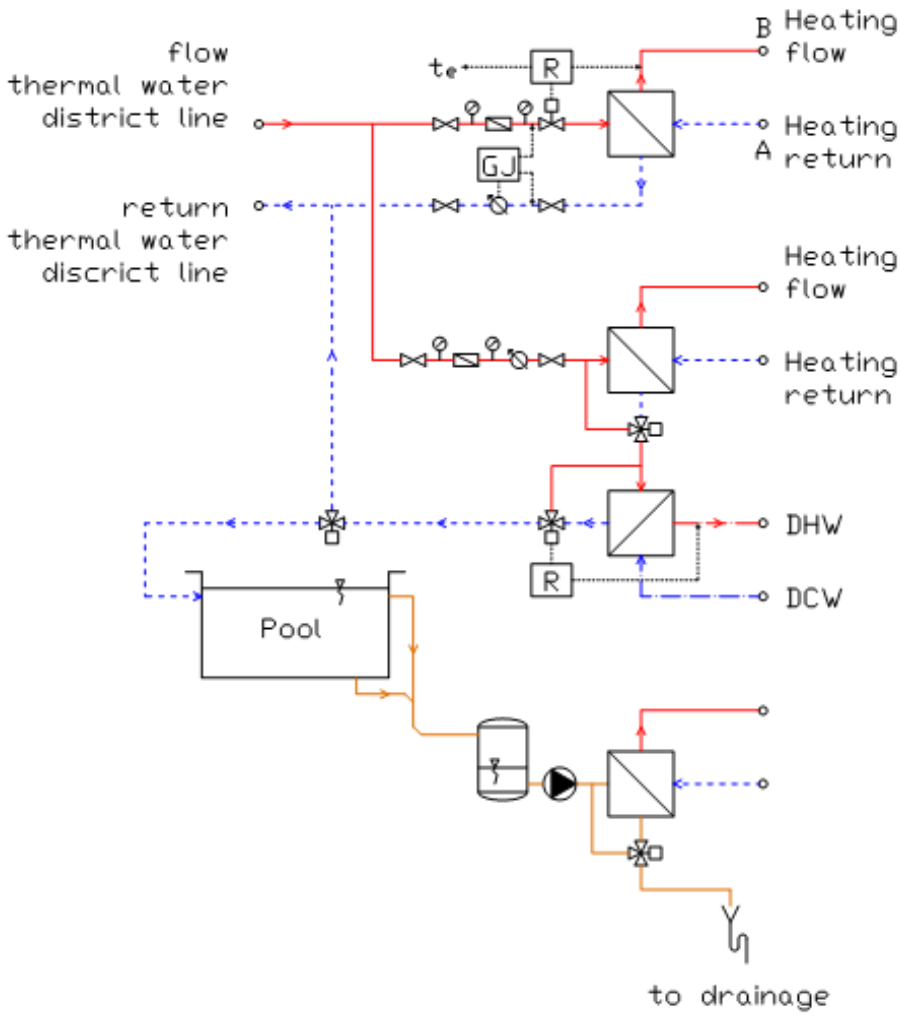


Figure 4. A possible solution of a thermal water supplied bath, when the bath supplying thermal water is pre- and after-utilized

To analyze the sample system by exergy aspect there are several years' measurement data available that contains the followings:

- daily well-mass flow rate at the production well
- daily average temperature at the well head
- daily water volume at the reinjection well
- daily average water temperature of the reinjected water
- summarized electric power use of all the pumps operating the complete system by month
- heat consumption measured by heat amount meters installed to the consumer side heat centrals by month

In our calculations we used data of the year 2009. Knowing the volume of the exploited thermal water and the temperature drops the gained heat amount can be calculated. The results are shown in Figure 5. The electric power data used to operate the system are shown in Figure 6.

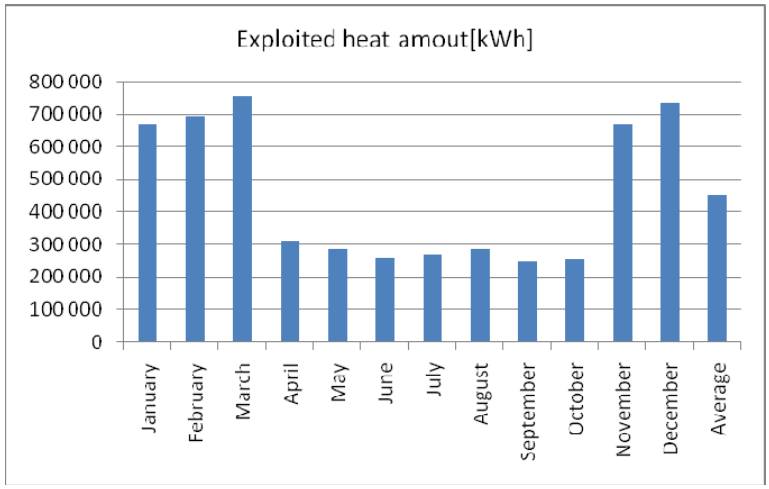


Figure 5. The exploited heat amount from thermal well by month in 2009.

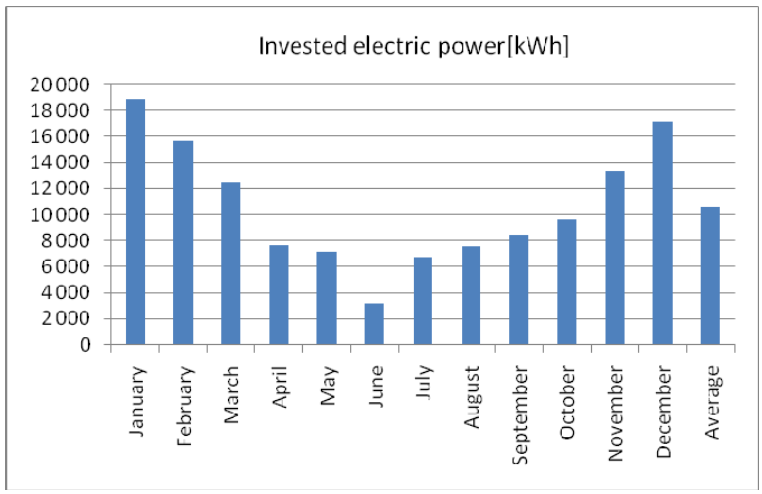


Figure 6. The electric power data used to operate the thermal system by month in 2009

It is clearly visible in the figures that the exploited water and heat amount followed the heat demand drop of warmer periods/season and summertime. In connection to this the electric power use decreased remarkably. At small mass flow rate the submersible valve does not operate as the well becomes „positive" - this fact also takes a roll in the decrease of the electric power use.

If the entire exploited heat amount is utilized then the quotient of the annual exploited heat and the annual electric power use give us the annual efficiency factor of the system. In deed the exploited heat is not completely utilized in heat supply. This is caused partly by the heat loss through the tanks, system, equipments, on the other hand the thermal bath supplying water does not take a role in heat supply or if it does, its effect is not being measured. Heat amount meters have been installed into all consumer heat centrals. The measured results are shown in Figure 7. Supposing that the meters measure the actual consumption – between tolerance range – the proportion of the exploited and the utilized heat can be calculated, as well as the proportion of the invested and utilized energy.

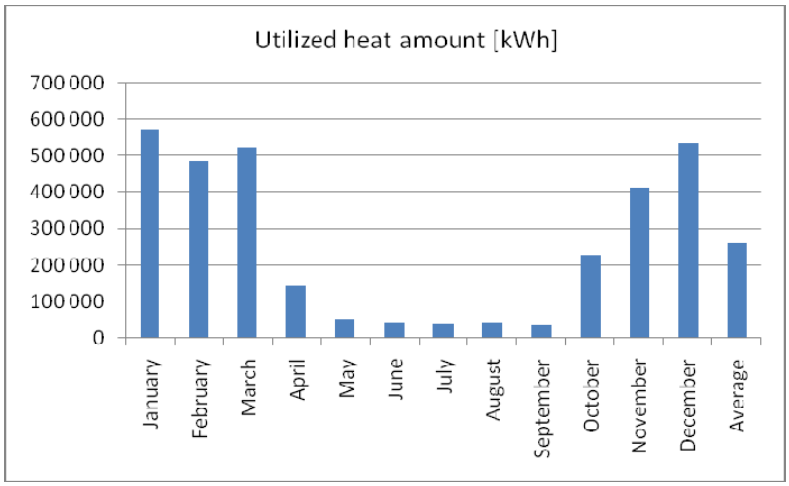


Figure 7. The utilized heat amount based recorded data on the consumer side by month in 2009

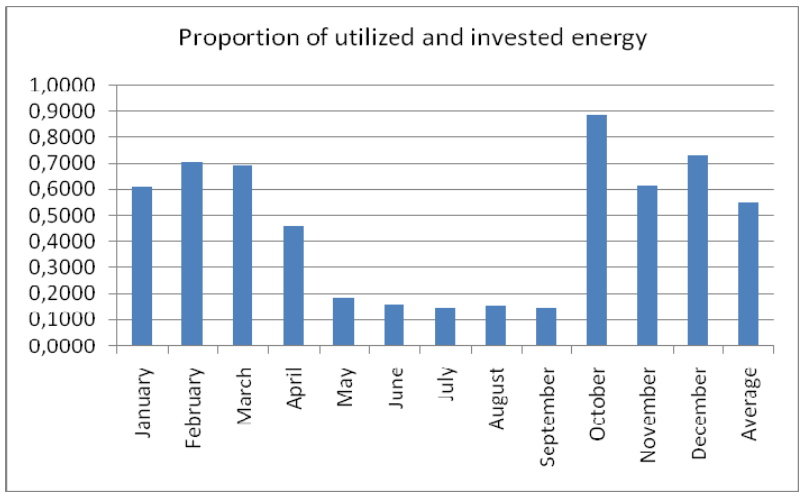


Figure 8. The proportion of the utilized and exploited heat amount by month in 2009

In wintertime and sometime of the transitional months in average 60-70 % of the exploited heat amount used for heat production. In summer the heat demand decreases significantly, thus the proportion of the utilized and exploited energy decreases to 15 %. Figure 9 shows the actual monthly energy factors.

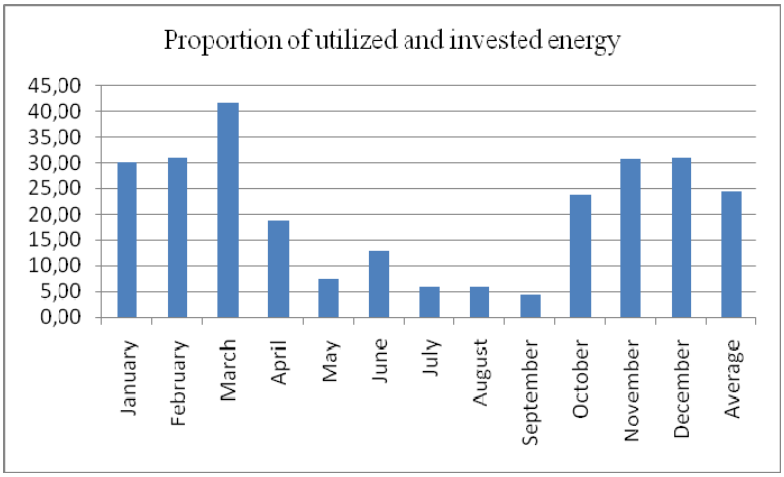


Figure 9. The proportion of the utilized and the invested energy by month in 2009

For practicing specialists one of the main technical parameter is ζ value that shows the relation between utilized heat amount (Q_H) and necessary utilized primer energy (Q_{PE}) – between boundaries.

$$\zeta = \frac{Q_H}{Q_{PE}} \tag{1}$$

In the case when utilizes heat amount is for building heating or domestic hot water supply and this heat is produced by a gas boiler installed within the building, then the ζ value equals to the gas boiler's efficiency (η_k), provided that the energy from the heat producer is transported to the rooms and domestic hot water taps without losses. Energy specialists usually use reciprocal value instead of (1), this is called the specific primer energy consumption (Büki G. 2007.):

$$g = \frac{Q_{PE}}{Q_H} = \frac{1}{\zeta} \tag{2}$$

To evaluate the utilization of the exploited thermal water we have to consider the complete energy supplying system. In Hungary, natural gas as primer energy has a large role in energy supply - according to 2007 statistical data 41 % of the

total primer energy utilization was natural gas based. Therefore it is expedient to determine the amount of spared natural gas primer energy source with utilization of thermal water. The specific natural gas consumption as specific primer energy consumption is a value that makes us able to compare different systems that use gas as primer energy source.

At the calculations it is assumed, that electric power use to exploit thermal water is produced by gas. The specific natural gas consumption was calculated with two different efficiency values: 0,35 (η_{E1}) considering the former power plants and 0,52 (η_{E2}), a more favourable value considering the modern power plants (Table 1) (Büki G. 2007). The monthly stats regarding the thermal water system also points that the decrease of the summer heat demand, the decrease of the utilization of the exploited heat affect badly the specific natural gas consumption efficiency. On possible way to increase summer heat demand is to cool the building with heat. *The geothermal energy used in heating makes it possible to transform low exergy energy into valuable cool energy and thus to increase summer utilization of the production wells (Halász Gyné-Halász Gy.2010).*

Table 1. Specific natural gas consumption at different heat production cases

Heat production method	**Equipment (ζ)**		**Specific natural gas utilization (g)**			
	efficiency; coefficient of performance (COP) (ηk;ε)		$1/\eta_k$ ill. $1/\varepsilon\eta_E$			
Natural gas boiler: conventional	0,7		1,43			
new	0,9		1,11			
condensing	1,0		0,95			
Thermal water based district heating system (utilization with heat exchenger)	**exploited/invested ε_1**	**utilized/invested ε_2**	**$1/\varepsilon_2\eta_{E1}$**	**$1/\varepsilon_{2\eta E2}$**	**$1/\varepsilon_1\eta_{E1}$**	**$1/\varepsilon_1\eta_{E2}$**
January	49,96	30,295	0,094	0,063	0,057	0,038
Febuary	44,36	29,385	0,097	0,065	0,064	0,043
March	60,52	41,83	0,068	0,046	0,047	0,032
April	40,99	18,78	0,152	0,102	0,070	0,047
May	39,92	7,298	0,391	0,264	0,072	0,048
June	81,87	12,946	0,221	0,149	0,035	0,023
July	40,17	5,84	0,489	0,329	0,071	0,048
August	38,10	5,826	0,490	0,330	0,075	0,050
September	29,24	4,264	0,670	0,451	0,098	0,066
October	26,86	23,74	0,120	0,081	0,106	0,072
November	50,37	30,87	0,093	0,062	0,057	0,038
December	43,02	31,20	0,092	0,062	0,066	0,045
Average	45,448	20,190				

The ξ (ε η) and g values cannot give complete picture about how effectively the natural possibilities have been exploited. Thermodynamic efficiency is what describes the procedure.

The thermodynamic efficiency of the procedure can be expressed with exergy efficiency. To determine exergy efficiency in complex systems we have to split it into system parts and sessions and determine the exergy efficiency for these processes:

$$\nu^{x}{}_{i} = \left(\frac{E^{x}{}_{out}}{E^{x}{}_{in}} \right)_{i} \tag{2}$$

Where $E^{x}{}_{out}$ is the exergy of exiting energy of a session and $E^{x}{}_{in}$ is the exergy of entering energy of a session:

$$\nu^{x} = \Pi_{i=1}^{n} \nu^{x}{}_{i} \tag{3}$$

In real processes there are exergy losses. For example: frictional force or similar processes and temperature loss during heat transfer. The exergy efficiency can be separated into two parts: $\nu^{x}{}_{mech/el}$ that means mechanical and electrical losses and $\nu^{x}{}_{th}$ that involves temperature losses and irreversibility caused losses:

$$\nu^{x} = \nu^{x}{}_{mech/el} . \nu^{x}{}_{th} \tag{4}$$

In our calculations values of $\nu^{x}{}_{th}$ were calculated knowing temperature data, while in case of $\nu^{x}{}_{mech/el}$ we calculate with empirical efficiency values (3).

At the analyzed sample system before the reconstruction hot water heating fluid was produced by old gas boiler installed in each building, and heating and hot water consumers were supplied by this way, therefore we give below the necessary correlation to calculate resultant exergy efficiency with the two different heat producing systems.

The heat producing during burning transmits to the heat transfer fluid (water) in the boiler. The exergy efficiency (H.Müller.2007):

$$\nu^{x}{}_{burn} = \frac{H_{u0}}{H_0} \cdot \eta_k \cdot (1 - \frac{T_0}{T_{fm}}) \tag{5}$$

Where T_{fm} is the average temperature of the transfer fluid, η_K is the boiler's efficiency, H_0 is the combustion heat of the fuel, H_{u0} is the heating value, T_0 is the temperature of the environment, so called dead-state temperature.

Producing domestic hot water, the heat transfer fluid exiting from the boiler reaches an indirect heated reservoir or heat exchanger, where transmits its heat to the domestic cold water. In this case the final element of the system is a reservoir or heat exchanger. At the heating system the heating fluid transmits its

heating value through the radiators, then the radiators to the room, that has to be equal to the heat loss of the room. The final element of the system shall be the radiator which is a heat exchanger itself.
The exergy efficiency of the heat transfer in the heat exchanger:

$$\nu^{x}{}_{he} = \frac{E^{x}{}_{Qout}}{E^{x}{}_{Qin}} = \frac{Q_{out}(1-\frac{T_0}{T_{exit}})}{Q_{in}(1-\frac{T_o}{T_{enter}})} = \eta \frac{1-\frac{T_0}{T_{exit}}}{1-\frac{T_0}{T_{enter}}} \tag{6}$$

where $E^{x}{}_{Qout}$ is the exiting, $E^{x}{}_{Qin}$ is the entering exergy of the heat energy, T_0 is the outdoor temperature, T_{exit} is the average temperature of the exiting, warming up fluid, T_{enter} is the average temperature of entering heating fluid, η is the efficiency of the heat exchanger. If the efficiency of the heat exchanger: $\eta = \frac{Q_{out}}{Q_{in}} = 1$, namely if there is no quantity loss, the exergy efficiency:

$$\nu^{x}{}_{he} = \frac{(1-\frac{T_0}{T_{exit}})}{(1-\frac{T_0}{T_{enter}})} \tag{7}$$

The resultant exergy efficiency:

$$\nu^{x}{}_{burn}\nu^{x}{}_{he} = \frac{H_{u0}}{H_0} \cdot \eta_k \cdot (1-\frac{T_0}{T_{exit}}) \tag{8}$$

The exergy efficiency of the radiator as a heat exchanger based on Formula 7. is (H.Müller.2007) :

$$\nu^{x}{}_{rad} = \frac{(1-\frac{T_0}{T_i})}{(1-\frac{T_0}{T_{fm}})} = \frac{\eta_{C0i}}{\eta_{C0fm}} \tag{9}$$

Where T_i is the room temperature, T_{fm} is the average temperature of the radiator heating fluid, η_{C0i} is the Carnot efficiency for T_0/T_i, η_{C0fm} is the Carnot efficiency for T_0/T_{fm} .
Resultant exergy efficiency:

$$\nu^{x}{}_{burn}\nu^{x}{}_{rad} = \frac{H_{u0}}{H_0} \cdot \eta_k \cdot (1 - \frac{T_0}{T_i}) \tag{10}$$

Necessarily assumed, that the heat from the heat producer reaches heat rejection without any losses. In our calculations regarding to gas fuel based direct heat production, the reasonable efficiency of the old boilers was calculated with: $\eta_k=0,8$ and the gas was calculated with: $H_u/H_0=0,9$ value.

At the two different systems, the environmental temperature or so called dead-state temperature (T_0) was calculated with an assumed value: 0 ^{o}C. However it is important to point out, that the end-state is not optional. It means the environment within the system exists. A system, such as a heating system can have different end-states. These different end states of one system can result in different exergy energy values. Therefore there are no exergy tables, unlike steam or gas tables. The exergy is not a thermodynamic parameter.

At the exergy efficiency it was taken into consideration that how much electric power was needed for exploiting and transporting thermal water. During this procedure electric power was produced at a 0, 52 efficiency by gas fuel. At one value it was assumed that the exploited heat was completely utilized, while at the other, lower value, the utilized heat amount was calculated. There was no different pump work calculated at linearly and parallel connected heat exchangers. In both cases the electric power was calculated and used from parallel state data.

Table 2. Sample system with gas boiler and thermal water - exergy and energetics efficiency at a certain operation condition

Heat consumer primer side fluid temperature °C	Heat consumer secunder side fluid temperature °C	Power station energy efficiency	Heat producing: boiler, heat exchanger exergy efficiency	η mech./elektr ν^{x} mech./elektr.	Heat consumer: radiator, heat exchanger exergy efficiency	Resultant exergy efficiency ν^{x} %	Energy factor η;ε	
Heat producing: conventional gas boiler η_k=0,8 assumed constant, H_u/H_o=0,9								
75/40 °C	45/10 °C		0,125	1	0,52	6,52	0,8	
46/34 °C	19/21 °C		0,092	1	0,535	4,92	0,8	
Thermal water primer fluid heat exchangers connected paralell								
75/40 °C	45/10 °C	0,52	0,522	1		52,2	23,6*	10,5*
75/40 °C	46/34 °C	0,52	0,737	1	0,535	39,43		
Thermal water primer fluid heat exchangers connected linearly								
55/40 °C	45/10 °C	0,52	0,611	1		61,1	23,6*	10,5*
75/55 °C	46/34 °C	0,52	0,665	1	0,535	35,5		

*energy efficiency calculated with the annual average coefficient of performance (COP) assuming gas primer energy.

Room temperature: 20 °C – *Outdoor temperature:* 0 °C

The exergy efficiency shown in the tables were calculated for the end-state temperatures in both cases, while same values were used at heating and domestic hot water production. Table 2 shows the calculation results.

4. Summary

According to the results we can state, that utilization of thermal water heat at heating and domestic hot water production considering energy aspect and thermodynamic efficiency is more effective than using conventional gas boilers. The reason of this is simple: using fuel for low exergy tasks, such as room heating or domestic hot water production low is a „waste". It is better to use termically less valued heat for these purposes.

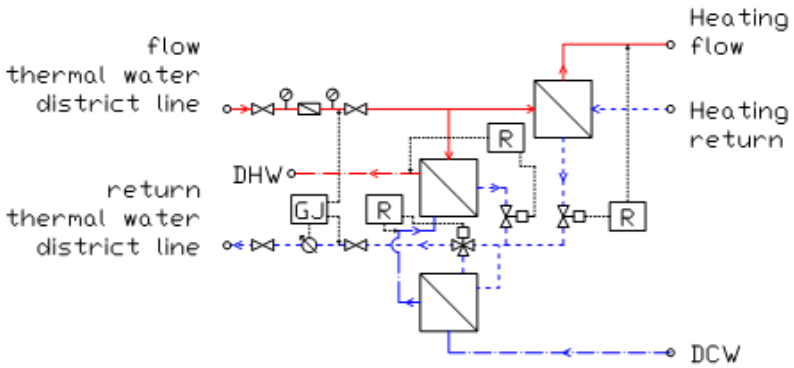


Figure 10. Indirect type, altering primer current domestic hot water production with preheater and afterheater heat exchangers connected to thermal water district line, combined connection

Regarding geothermal energy utilization it is favourable to apply a so called staged use or multipurpose use - the consumers are connected in a row, or we use combined connection as Figure 10 shows.

It is important to aim for utilizing larger rate of exploited heat. It is for considering that during geothermal energy utilization that has been produced by pumps, heat pumps supplied with fossil primer energy source - renewable energy source is only partly used.

References

[1] Müller, H.: Technische Thermodynamik, Wismar, 2000.

[2] Büki, G.: Kapcsolt energiatermelés, Műegyetemi kiadó, 2007.

[3] Tóth A.: Hungary Country Update 2005-2009. WGC 2010 World Geothermal Congress 25-30 april 2010 Bali International Convention Centre Nusa Dua Bali Indonesia

[4] Halász, Gyné. Halász Gy. (2009): Geothermic Energy in Enery Supply in Hungary WGC 2010 World Geothermal Congress 25-30 april 2010 Bali International Convention Centre Nusa Dua Bali Indonesia

This research was supported by New Hungary Development Plan project no. TÁMOP-4.2.2-08/1/2008-0017 „Integrated modelling of sustainability of geothermal systems". The project is supported by European Union and co-financed by European Regional Development Fund.

Symbolic solutions of ordinary differential equation systems used for coupled transport processes

Csaba MÉSZÁROS[1], István KIRSCHNER[2], Klaus GOTTSCHALK[3], László SZÉKELY[4], Ágnes BÁLINT[5]

[1]Department of Physics and Process Control,
Institute for Environmental Engineering Systems
[2]Eötvös Lóránd University, Institute of Physics, Budapest, Hungary
[3]Leibniz-Institut für Agrartechnik Bornim e.V. (ATB), Potsdam, Germany
[4]Department of Mathematics, Institute for Mathematics and Informatics
[5]Szent István University, Faculty of Agricultural and Environmental Sciences,
Department of Chemistry and Biochemistry

Abstract

The system of partial differential equations describing the coupled heat and mass transfer through capillary-porous bodies is based on the application of the zone picture. A general description of the basic system of ordinary differential equations of coupled transport processes is given within the framework of a linear approximation and treated by tools of matrix analysis and group representation theory. It is shown, that the technique of hyperdyads directly generalizes the method of simple dyadic decomposition of operators used earlier in the classical linear irreversible thermodynamics and leads to possible new applications of the concept of quasi-polynomials at the descriptions of coupled transport processes.

PACS Numbers: 05.60. Cd; 05.70.Ln

Keywords

coupled transport processes, matrix analysis, group representation theory

1. Introduction

Since the non-equilibrium thermodynamics plays a decisive role [de Groot and Mazur, 1962], [Gyarmati, 1970], [Jou et al., 2001] in accurate description of the genuine (and usually: coupled) transport processes, the refined mathematical modelling of such phenomena represents a permanent objective of the contemporary mathematical physics and numerical mathematics, too. Although its powerful methods are able to provide effective tools for modelling the essential features of different complex systems, the well-known open problems of thermo-hydrodynamics also call for elaboration of more refined mathematical models.

Namely, many classical transport processes also show features of the anomalous diffusion [Jou et al., 2001] of non-local character and with memory effects on macroscopic level (corresponding usually to the percolative-fractal (P-F) character on the mesoscopic level), and require e.g. use of the Riemann-Liouville operators, leading to a widespread use of the fractional partial derivatives with respect to time and spatial coordinates, as well [Zaslavsky, 2002], [Uchaikin, 2003]. The relevant mathematical questions have also been analyzed in detail in very deep studies related to general problems of fractional diffusion e.g. [Tarasov and Zaslavsky, 2007], [Atanacković et al., 2009], [Sun et al., 2009]. The Extended Irreversible Thermodynamics (EIT), has also many open fundamental questions in some earlier developed areas of its, too. In this sense it is probably enough to mention the archetypal problem of infinitely large propagation velocity connected with the application of separate parabolic type partial differential equations (PDEs), to which the classical equations of Fourier (describing heat propagation) and Fick (describing diffusion) belong, too. Since existence of the infinitely large propagation velocities contradicts to the basic principles of physics, this problem induced introduction of novel-type formalisms based on existence of non-zero relaxation time constants, which gave us [Jou et al., 2001] possibility to work with hyperbolic type PDEs instead of parabolic ones. One of the most general variants of these descriptions is the formalism elaborated by Gyarmati [Gyarmati, 1977] resulting in systems of PDEs composed from coupled hyperbolic type PDEs.

Therefore, within frame of the Classical Irreversible Thermodynamics (CIT), the mathematical modelling of the transport processes is performed mainly in the linear approximation (i.e. when the conductivity and coupling coefficients are assumed to be constants – and without necessity for applying of the fractional partial derivatives mentioned above) and even this well-elaborated branch of thermodynamics represents an active research domain, which also may lead to further new and valuable informations about the topic. In order to justify the validity of this statement, we mention here the symmetry of the thermodynamic cross-effects formally explained by the crucial Onsager symmetry relations (see e.g. [de Groot and Mazur, 1962], [Landau and Lifshitz, 2000]) relevant for the elements of the conductivity tensor $L_{ik} = L_{ki}, 1 \le i,k \le F_f$ with F_f being the number of thermodynamic degrees of freedom. The incorporation of the cross-effects into the general formalism of thermodynamics significantly modifies the modelling results emanating from simple, separated PDEs e.g. [Luikov, 1965] led us to a mathematical modelling tool for deep studies of the relations between systems of parabolic-, and hyperbolic type PDEs [Mészáros et al., 2004]. Besides, some very general and simple symmetry assumptions about these generalized systems of the coupled heat and mass transfer processes through porous media has also given us (via application of the symbolic computer algebra software package [MAPLE]) [Mészáros et al., 2007] a possibility for the introduction of special functions previously not used in transport theories, and a direct, natural way for the elimination of the infinitely large propagation velocity problem.

Firstly, we recall here the most frequently used system of PDEs for mathematical modelling of drying processes (which represent a typical coupled transport process in natural manner) [Luikov, 1965], [Pandey et al., 1999]:

$$\begin{aligned}
\frac{\partial M}{\partial t} &= \nabla^2 a_{11} M + \nabla^2 a_{12} T + \nabla^2 a_{13} P, \\
\frac{\partial T}{\partial t} &= \nabla^2 a_{21} M + \nabla^2 a_{22} T + \nabla^2 a_{23} P, \\
\frac{\partial P}{\partial t} &= \nabla^2 a_{31} M + \nabla^2 a_{32} T + \nabla^2 a_{33} P,
\end{aligned} \tag{1}$$

where M denotes the moisture content, T the temperature and P the pressure. The system of equations (1), as well as its simplified variants, is usually written in „criterial" form, i.e. the time t is to be understood as the Fourier-criterion ($t \to \frac{a_T t}{l^2}$, where a_T denotes the heat conductivity coefficient of the mixture and l is the characteristic length of the body being dried) [Luikov, 1965] and except situations, when the temperature has too high values and the pressure change is negligible, usually the next simplified variant of this system is applied:

$$\frac{\partial}{\partial t}\begin{pmatrix} M(x,t) \\ T(x,t) \end{pmatrix} = \begin{pmatrix} 1 + \varepsilon KoLu_m Pn (\equiv a_{11}) & -\varepsilon KoLu_m (\equiv a_{12}) \\ -Lu_m Pn (\equiv a_{12}) & Lu_m (\equiv a_{22}) \end{pmatrix} \cdot \begin{pmatrix} \nabla^2 M(x,t) \\ \nabla^2 T(x,t) \end{pmatrix}, \tag{2}$$

The physical meaning of the factors of coefficients a_{11}, a_{12}, a_{21}, a_{22} is explained at the end of the paper and we work here in the zone-picture mentioned above, according to which the state-dependent, and therefore non-linear matrix elements in (1) can be replaced by their average values in thin enough layers (zones) of the porous matter being dried [Luikov, 1965]. The relevant system of coupled parabolic type PDEs with constant coefficients has been studied extensively in the literature of transport processes and represents a very active research object from the point of view of modelling such phenomena [Coles, 2006], [Pandey et al., 1999], [Mészáros et al., 2001], [Kirschner et al., 2004].

Finally, it must also be mentioned that forms of the temperature and moisture-level functions are well-known from the classical literature on thermodynamics and hydrodynamics e.g. [Landau and Lifshitz, 2000] and in the one-dimensional case for e.g. the temperature function can be written as

$$T(\vec{r},t) \equiv T(x,y,z,t) \to T(x,t) = \sum_n c_n T_n(x) \cdot e^{-\lambda_n t}. \tag{3}$$

(the series expansion coefficients c_n are constants, and the quantities λ_n denote the reciprocal values of relaxation time constants of the relevant n-th harmonics).

2. Some earlier applications of matrix analysis in the CIT

In order to summarize the most general features of the mathematical formalism of coupled transport processes taking place in dissipative macroscopic continua (but still within a linear approximation), we recall here briefly the results of the theory based on classical results of the linear approximation of the CIT and later refined by Kirschner and co-workers [Kirschner et al., 1991-94]. According to this theory, the basic thermodynamic equations of motion are given by

$$\dot{\vec{\alpha}} = \hat{L}\vec{X}, \tag{4}$$

and

$$\vec{X} = -\hat{g}\vec{\alpha}, \tag{5}$$

where $\vec{X} = \vec{X}(t)$ denotes the abstract vector-column of thermodynamic forces (i.e. the complete ordered set of thermodynamic gradient vector components is denoted here by capital $\vec{X}$ vector, in order to distinguish it from the simple spatial *x*-coordinate), $\vec{\alpha} = \vec{\alpha}(t)$ is similarly the relevant vector-column composed from α – parameters (non-equilibrium deviations of extensive thermodynamic parameters from their equilibrium values), while $\hat{g}$ is the entropy matrix and $\hat{L}$ is the conductivity matrix (both matrices are quadratic of *n*-th order, positive-definite and symmetric ones). Throughout this paper the symbols of type $\hat{A}$ will be used for general notation of matrices, while without operator notation (*A*) they will be assigned to their concrete representation corresponding to a given basis of the actual abstract vector space. More explicit examples for elements of the conductivity matrix are given in the Appendix (relations (A,2)-(A,7)). Since we are studying changes in the vicinity of equilibrium states (where the entropy matrix is a matrix composed from entirely constant elements $g_{ik} = -\left(\frac{\partial^2 \alpha_s}{\partial \alpha_i \partial \alpha_k}\right)_0 = -\left(\frac{\partial^2 \alpha_s}{\partial \alpha_k \partial \alpha_i}\right)_0 = g_{ki}, i,k = 1,2,...,F_f$ e.g. [de Groot and Mazur, 1962]), the total derivative (with respect to the time variable) of system (5) can be written in the form $\dot{\vec{X}} = -\hat{g}\dot{\vec{\alpha}}$. Then, in the most general case of the linear approximation of the non-equilibrium thermodynamics we arrive at the following systems of ordinary differential equations written again in matrix form

$$\dot{\vec{\alpha}} = -\hat{L}\hat{g}\vec{\alpha} - \vec{G} \equiv -\hat{A}^{+}\vec{\alpha} - \vec{G}, \tag{6}$$

$$\dot{\vec{X}} = -\hat{g}\hat{L}\vec{X} + \vec{G} \equiv -\hat{A}\vec{X} + \vec{G}, \tag{7}$$

where by symbol "+" the operation of the hermitian conjugation is denoted and $\vec{G} = \vec{G}(t)$ represents the abstract vector-column of thermodynamic forces of the external generation. Therefore, the usual common application of constitutive-, and balance equations e.g. [de Groot and Mazur, 1962], [Gyarmati, 1970] is omitted in the present formalism, and we obtained directly separated equations in the systems for both thermodynamic forces and extensive thermodynamic parameters. The solutions of these systems can immediately be given in the form

$$\vec{\alpha}(t) = e^{-\hat{A}^+ t} \cdot \vec{\alpha}_0 - e^{-\hat{A}^+ t} \int_0^t e^{\hat{A}^+ \tau} \vec{G}(\tau) d\tau, \tag{8}$$

$$\vec{X}(t) = e^{-\hat{A} t} \vec{X}_0 - e^{-\hat{A} t} \int_0^t e^{\hat{A} \tau} \vec{G}(\tau) d\tau, \tag{9}$$

where $\vec{X}_0 = \vec{X}(0)$ and $\vec{\alpha}_0 = \vec{\alpha}(0)$ denote the initial values of the vectors of thermodynamic forces and extensive thermodynamic parameters, respectively. Similar analysis has been performed in the famous work by de Groot and Mazur [de Groot and Mazur, 1962] – particularly in the field of Gaussian-type Markoff - processes. It was repeatedly demonstrated [Kirschner et al., 1991-94], that even in the cases, when only one of the thermodynamic forces differs relevantly from its equilibrium value for a given system, it may induce non-equilibrium values for all the other forces, too. Then, according to the basic properties of the systems of ODEs, each from the thermodynamic forces may change its sign (F_f-1) times and approach its equilibrium value asymptotically. Therefore, each perturbation propagates through the system via all possible types of interactions present in it (we think here of interaction types represented by systems (4-5) and (6-7)). Furthermore, since the matrix $\hat{A} = \hat{g}\hat{L}$ has only positive eigenvalues: $\lambda_i > 0, \left(1 \le i \le F_f\right)$, [Kirschner et al., 1991-94], it is very suitable to represent it in the below given form of the sum of dyadic products

$$\hat{A} = \sum_{i=1}^{F_f} \lambda_i \vec{\psi}_i \circ \vec{\varphi}_i^+, \tag{10}$$

where the abstract vectors $\left\{\vec{\psi}_1, ..., \vec{\psi}_{F_f}\right\}$ and $\left\{\vec{\varphi}_1, ..., \vec{\varphi}_{F_f}\right\}$ are elements of a biorthogonal system of basis vectors, i.e. $\vec{\psi}_i^+ \cdot \vec{\varphi}_k = \vec{\varphi}_k^+ \cdot \vec{\psi}_i = \delta_{ik}, \left(1 \le i, k \le F_f\right)$. The dyadic product of basis vectors is of idempotent character, i.e.

$$\left(\vec{\psi}_i \circ \vec{\varphi}_i^+\right)^2 = \vec{\psi}_i \left(\vec{\varphi}_i^+ \vec{\psi}_i\right) \vec{\varphi}_i^+ \equiv \vec{\psi}_i \hat{I} \vec{\varphi}_i^+ = \vec{\psi}_i \circ \vec{\varphi}_i^+, i = 1, 2, ..., F_f \tag{11}$$

($\hat{I}$ is the identity operator). Then, it is obvious, that all matrix polynomials $F\left(\hat{A}\right)=\sum_{l} c_l \hat{A}^l$, can (with all c_l-s as constants) can also be represented as

$$F\left(\hat{A}\right)=\sum_{i=1}^{F_f} F\left(\lambda_i\right)\vec{\psi}_i \circ \vec{\varphi}_i^{+}, \tag{12}$$

implying immediately, that $e^{-\hat{A}t}=\sum_{i=1}^{F_f} e^{-\lambda_i t}\vec{\psi}_i \circ \vec{\varphi}_i^{+}$. In the present case, when external generating forces are sequal to zero, i.e. $\vec{G}\left(t\right)=\vec{0}$, we will obtain the following particular forms for the homogeneous variants of the solutions (6-7)

$$\vec{\alpha}\left(t\right)=\sum_{i=1}^{F_f} c_i e^{-\lambda_i t}\vec{\varphi}_i, \left(c_i=\vec{\psi}_i^{+}\cdot\vec{\alpha}_0\right), \tag{13}$$

and

$$\vec{X}\left(t\right)=\sum_{i=1}^{F_f} c_i^{*} e^{-\lambda_i t}\vec{\psi}_i, \left(c_i^{*}=\vec{\varphi}_i^{+}\vec{\alpha}_0\right). \tag{14}$$

Therefore, the coefficients c_i and c_i^{*} can be calculated on the base of the general solution functions of type (8-9) and expressions in the parentheses in (13) and (14).

3. Application of the theory of quasipolynomials

- Novel solutions of the systems of ODEs of CIT (Comparison with results based on applications of the Jordan blocks)
  In this section, we give a novel proposal for further generalization of the above discussed abstract formalism of the coupled transport processes. Although this analysis will be again performed within frame of the strictly linear approximation, it can trace out pathway for applications of some well-known and effective methods of the theory of the systems of ODEs. Therefore, we start from the formalism used by Gyarmati [Gyarmati, 1977] at funding of the wave approach of thermodynamics (WAT) in its original form. Accordingly, in the case of convection-free coupled transport processes in dissipative systems, the systems of differential equations related to intensive (Γ) and specific extensive (a) thermodynamic quantities can be given concisely as

$$\rho\frac{d\vec{\Gamma}}{dt}+\left(\hat{s}\nabla\cdot\hat{L}\nabla\right)\vec{\Gamma}=\hat{s}\vec{\sigma},$$
$$\rho\frac{d\vec{a}}{dt}+\left(\nabla\cdot\hat{L}\hat{s}\nabla\right)\vec{a}=\vec{\sigma}, \qquad (15)$$

i.e. we arrange them into vectors, as well as the different contributions of the general entropy production function into the vector $\vec{\sigma}$. For the sake of completeness and further forthcoming applications we mention here immediately, that $\sigma=\overset{(s)}{\sigma}+\overset{(v)}{\sigma}+\overset{(a)}{\sigma}+\overset{(t)}{\sigma}\geq 0$ [Gyarmati, 1970], and that this inequality emanating from the second law of thermodynamics must be fulfilled separately for each of the scalar (s)-, polar-vector (v)-, axial-vector-(a), and tensorial-(t) type contributions of the entropy function. Since within the frame of CIT all these components of σ are separated one from each other on the base of the Curie principle, it is obvious, that the group representation theory e.g. [Wigner, 1959], may play a more important role in the refinements of the general solutions emanating from the systems of type (15). The elements of the quadratic and symmetric matrix $\hat{s}$ are composed from elements defined by $s_{ik}\equiv -g_{ik}$ and in the strictly linear case we have

$$\frac{d\vec{\Gamma}}{dt}-\hat{\kappa}_0\Delta\vec{\Gamma}=\frac{1}{\rho_0}\hat{s}_0\vec{\sigma},\left(\hat{\kappa}_0\equiv-\frac{1}{\rho_0}\hat{s}_0\hat{L}_0\right),$$
$$\frac{d\vec{a}}{dt}-\hat{D}_0\Delta\vec{a}=\frac{1}{\rho_0}\vec{\sigma},(\hat{D}_0\equiv-\frac{1}{\rho_0}\hat{L}_0\hat{s}_0), \qquad (16)$$

where $\Delta\equiv\frac{\partial^2}{\partial\vec{r}^{\,2}}$ denotes the Laplace-operator and the lower indices zero correspond to constant quantities. Besides, the matrices $\hat{\kappa}_0$ and $\hat{D}_0$ are related to each other [Gyarmati, 1977] via similarity transformation $\hat{\kappa}_0=\hat{s}_0\hat{D}_0\hat{s}_0^{-1}$. It is a characteristic feature of the systems (15) and (16), that they contain total time derivatives, but the spatial derivatives are of partial character. For further purposes, we will firstly perform here a Fourier-transformation procedure $f\left(\vec{r},t\right)\rightarrow\tilde{f}\left(\vec{k},t\right)\equiv\iiint\limits_{\left(\square^3\right)}f\left(\vec{r},t\right)e^{i\vec{k}\cdot\vec{r}}d^3\vec{r}$ for both systems (with respect to the spatial coordinate), and obtain the following vectors

$$\vec{\Gamma} = \vec{\Gamma}\left(\vec{r},t\right) \equiv \begin{pmatrix} \Gamma_1\left(\vec{r},t\right) \\ . \\ \Gamma_i\left(\vec{r},t\right) \\ . \\ \Gamma_{F_f}\left(\vec{r},t\right) \end{pmatrix} \rightarrow \tilde{\vec{\Gamma}}\left(\vec{k},t\right) \equiv \begin{pmatrix} \tilde{\Gamma}_1\left(\vec{k},t\right) \\ . \\ \tilde{\Gamma}_i\left(\vec{k},t\right) \\ . \\ \tilde{\Gamma}_{F_f}\left(\vec{k},t\right) \end{pmatrix},$$

$$\vec{a} = \vec{a}\left(\vec{r},t\right) \equiv \begin{pmatrix} a_1\left(\vec{r},t\right) \\ . \\ a_i\left(\vec{r},t\right) \\ . \\ a_{F_f}\left(\vec{r},t\right) \end{pmatrix} \rightarrow \tilde{\vec{a}}\left(\vec{k},t\right) \equiv \begin{pmatrix} \tilde{a}_1\left(\vec{k},t\right) \\ . \\ \tilde{a}_i\left(\vec{k},t\right) \\ . \\ a_{F_f}\left(\vec{k},t\right) \end{pmatrix}. \qquad (17)$$

This formalism developed by Gyarmati in 1977 is valid strictly speaking in convection-free cases

($\vec{v} = 0$) [Gyarmati, 1977] and therefore, on the base of general balance equations of the non-equilibrium thermodynamics [Gyarmati, 1970], we obtain $\rho \dot{a} = \frac{\partial \rho a}{\partial t}$, which, together with the continuity equation results in $\rho_0 \frac{da}{dt} = \rho_0 \frac{\partial a}{\partial t}$ where $\rho \rightarrow \rho_0 = const.$

As a novel result within CIT, we write down the solutions of these systems of ODEs by taking into account ***the possibility of multiple eigenvalues of their characteristic equations*** [Rózsa and Elsner, 1981], [Rózsa, 2009]. Then, for the inhomogeneous systems of ODEs $\dot{\vec{x}}\left(t\right) = \hat{A}\left(t\right)\vec{x}\left(t\right) + \vec{f}\left(t\right)$ obeying initial condition $\vec{x}\left(t_0\right) = \vec{x}_0$ and omitting the intermediate steps at its solving (see e.g. [Gantmakher, 1959], [Rózsa, 2009]) we will have

$$\vec{x}\left(t;t_0,\vec{x}_0\right) = \sum_{k=1}^{s} e^{\lambda_k (t-t_0)} L_k\left(\hat{A}\right)\vec{x}_0 + \sum_{k=1}^{s}\int_{t_0}^{t} e^{\lambda_k (t-\tau)} L_k\left(\hat{A}\right)\cdot \vec{f}\left(\tau\right) d\tau, \qquad (18)$$

in the case, when the minimal equation of the matrix $\hat{A}$ has only ("s" number of) single roots ($L_k\left(\hat{A}\right)$ denotes here the Lagrangian-type matrix-polynomial). In this case, we have a direct prescription of the formula (12) in the form of

$$F\left(\hat{A}\right)=\sum_{k=1}^{s}F\left(\lambda_k\right)L_k\left(\hat{A}\right)\equiv\sum_{k=1}^{s}\sum_{\nu=1}^{\alpha_k}F\left(\lambda_k\right)\vec{u}_{k\nu}\vec{v}_{k\nu}^{T}.$$

In the case, when the „minimal equation" of the matrix $\hat{A}$ has also multiple solutions, we have

$$\begin{aligned}\vec{x}\left(t;t_0,\vec{x}_0\right)&=\sum_{k=1}^{s}e^{\lambda_k\left(t-t_0\right)}\sum_{\upsilon=0}^{\gamma_k-1}\left(t-t_0\right)^{\upsilon}H_{k\upsilon}\left(\hat{A}\right)\vec{x}_0+\\&+\sum_{k=1}^{s}\int_{t_0}^{t}e^{\lambda_k\left(t-\tau\right)}\sum_{\upsilon=0}^{\gamma_k-1}\left(t-\tau\right)^{\upsilon}H_{k\upsilon}\left(\hat{A}\right)\cdot\vec{f}\left(\tau\right)d\tau,\end{aligned}\tag{19}$$

where $H_{k\upsilon}\left(\hat{A}\right)$ denotes the Hermitian-type interpolation matrix-polynomial and γ_k is the multiplicity of the root λ_k in the minimal equation. In this case, we have $F\left(\hat{A}\right)=\sum_{k=1}^{s}\sum_{\nu=0}^{\gamma_k-1}F^{(\nu)}\left(\lambda_k\right)H_{k\nu}\left(\hat{A}\right)$, which is a direct generalization of (12). In our paper this general case is of interest and the dyadic decomposition of the interpolation Hermite-polynomials will be of crucial interest.

Firstly, in order to justify a posteriori the above-discussed results explained by (18 – 19) at solving of systems of the type (6-7) in the case of absence of the forces of external generation, we recall here the fact, that (by a suitable similarity transformation via adequate non-singular matrices) all quadratic matrices can be transformed into matrices of Jordan normal form. Generally speaking, these Jordan – type matrices are given by direct sum of Jordan-blocks, i.e. $\hat{J}=\hat{J}_1\oplus\ldots\oplus\hat{J}_q\oplus\ldots\oplus\hat{J}_p$, where each of these blocks is given in the form of e.g.:

$$\hat{J}_q=\begin{pmatrix}\lambda_q & 1 & 0 & . & 0\\ 0 & \lambda_q & 1 & . & 0\\ 0 & 0 & \lambda_q & . & 0\\ . & . & . & . & 1\\ 0 & 0 & 0 & . & \lambda_q\end{pmatrix}\equiv\lambda_q\hat{I}+\hat{N}_q,1\le q\le p,\tag{20}$$

where λ_q denotes one of the eigenvalues of the matrix $\hat{A}$, $\hat{I}$ is the unit matrix of adequate order corresponding to the identity operator, while $\hat{N}_q=\begin{pmatrix}0 & 1 & 0 & . & 0\\ 0 & 0 & 1 & . & 0\\ . & . & . & . & .\\ 0 & 0 & 0 & . & 1\\ 0 & 0 & 0 & . & 0\end{pmatrix}$ is a nilpotent matrix of the order of u_q, i.e.

$\left(\left(\hat{N}_q\right)^{u_q} = \hat{0}, u_q \in \square\right)$. Then, for a given matrix function $f = f\left(\hat{A}\right)$ (f is defined on the spectrum of the matrix $\hat{A}$) we will have

$$f\left(\hat{J}_q\right) = \begin{pmatrix} f\left(\lambda_q\right) & \frac{f'\left(\lambda_q\right)}{1!} & \frac{f''\left(\lambda_q\right)}{2!} & . & \frac{f^{\left(u_q-1\right)}\left(\lambda_q\right)}{\left(u_q-1\right)!} \\ 0 & f\left(\lambda_q\right) & \frac{f'\left(\lambda_q\right)}{1!} & . & \frac{f^{\left(u_q-2\right)}\left(\lambda_q\right)}{\left(u_q-2\right)!} \\ 0 & 0 & f\left(\lambda_q\right) & . & \frac{f^{\left(u_q-3\right)}\left(\lambda_q\right)}{\left(u_q-3\right)!} \\ . & . & . & . & . \\ 0 & 0 & 0 & . & f\left(\lambda_q\right) \end{pmatrix}. \tag{21}$$

It is known, if the similarity transformation between matrices $\hat{A}$ and $\hat{J}$ is realized by use of the non-singular matrix $\hat{T}$, i.e. $\hat{A} = \hat{T}^{-1}\hat{J}\hat{T}$, and we will have $f\left(\hat{A}\right) = \hat{T}^{-1}\left[f\left(\hat{J}_1\right) \oplus ... \oplus f\left(\hat{J}_q\right) \oplus ... \oplus f\left(\hat{J}_p\right)\right]\hat{T}$). Particularly, it can be shown [Arnold, 2001], that for operator functions given in the general form of $e^{\hat{A}t}$, according to (21), the following representation appears

$$e^{At} = \begin{pmatrix} e^{\lambda t} & \cdots & e^{\lambda t}\cdot\frac{t^{n-1}}{(n-1)!} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & e^{\lambda t} \end{pmatrix}, \tag{22}$$

i.e. all matrix elements on the main diagonal are equal to $e^{\lambda t}$ (λ is an eigenvalue emanating from the basic equation $\hat{A}\vec{X}(t) = \lambda\vec{X}(t)$), all matrix elements below it are zeros, while all the other matrix elements are given as products of $e^{\lambda t}$ with polynomial terms $\frac{t^k}{k!}, \left(0 \le k \le n-1\right)$ of the relevant order.

According to our assumptions on the eigenvalues of $\hat{A}$, the basis functions of the abstract vector spaces of functions leading to representation in the form of Jordan-blocks are realized by quasi-polynomials, which in general case have the form [Arnold, 2001]

$$\varphi(t) = \sum_{l=1}^{k} e^{\lambda_l t} \sum_{m=0}^{\nu_l - 1} c_{l,m} t^m, \; c_{l,m} = const., \tag{23}$$

where ν_l denote the algebraic multiplicities of different eigenvalues $\lambda_1, \lambda_2, ..., \lambda_l, ..., \lambda_k$ in the eigenvalue-problem being investigated.

– Symmetry consequences
  The time-dependent part of the solution form (3) widely used in the literature may be taken as a special case of (23), where ν_l denotes the multiplicity of the eigenvalue λ_l. For a function $t \mapsto \varphi(t)$ the expression (23) gives possibility to connect the linear thermodynamic formalism demonstrated in this paper by formulae (6-7) to other well-established descriptions of the classical irreversible thermodynamics, e.g. [de Groot and Mazur, 1962]. Namely, according to this formalism, the coupling between thermodynamic forces and fluxes of different character and tensorial order at point group symmetry transformations of their Cartesian components is restricted: from all couplings only those with identical tensorial order are allowed according to the Curie principle. In order to simplify the forthcoming analyses, we also give here the abstract group-theoretical formulation of the principle; accordingly, we consider complex physical systems, whose subsystems are characterized by particular symmetry groups. If the subsystems of the given complete system do not affect each other significantly, the symmetry of the whole system is given as a direct product (relation (A,8) in the Appendix) of symmetry groups of the subsystems. If, however, there exists a significant interaction between subsystems, the symmetry group of the whole system is intersection between symmetry groups of the interaction and those of the subsystems. Often, this intersection is restricted to symmetry groups of the subsystems only, because the whole system remains invariant in the case of coincident action of common symmetries.

Thus, the Curie principle makes a direct generalization of mathematical formalism of the classical irreversible thermodynamics possible. Since the actual thermodynamic systems being examined are characterized by invariance property with respect to some symmetry transformations (e.g. in the case of electrical conductivity, the relevant system is simultaneously invariant [Cracknell, 1975] to the actual crystallographic point group-, as well as to the Onsager-Casimir symmetry transformations, whose crucial role in the formalism being developed here will also be discussed below), the relevant results of the group representation theory may be directly applied here. The concrete general formula for describing transformation properties of tensorial quantities according to a given irreducible representation is given by (A,9) in the Appendix. Due to the product structure $\hat{A} = \hat{g}\hat{L}$ of the matrix $\hat{A}$ we can formulate a result of the present investigation in the following general form: since the tensor $\hat{L}$ can be decomposed into irreducible blocks (according to the

actual symmetry group of the given dissipative system), and the $\hat{g}$ matrix can be transformed into direct sum of matrices of the Jordan-type normal form, the matrix $\hat{A}$ can be represented as a simple product of two matrices having block-diagonal forms, i.e.

$$\hat{A} = \hat{g}\hat{L} = \left[\hat{J}_1^{\hat{g}} \oplus ... \oplus \hat{J}_q^{\hat{g}} \oplus ... \oplus \hat{J}_p^{\hat{g}}\right]\left[\hat{L}^{(1)} \oplus ... \oplus \hat{L}^{(\mu)} \oplus ... \oplus \hat{L}^{(\mu_f)}\right] \equiv$$

$$\equiv \begin{pmatrix} \hat{J}_1^{\hat{g}} & \hat{0} & \hat{0} & . & \hat{0} \\ \hat{0} & . & . & . & \hat{0} \\ \hat{0} & . & \hat{J}_q^{\hat{g}} & . & \hat{0} \\ . & . & . & . & . \\ \hat{0} & \hat{0} & \hat{0} & . & \hat{J}_p^{\hat{g}} \end{pmatrix} \begin{pmatrix} \hat{L}^{(1)} & . & \hat{0} & . & \hat{0} \\ \hat{0} & . & . & . & \hat{0} \\ \hat{0} & . & \hat{L}^{(\mu)} & . & \hat{0} \\ . & . & . & . & . \\ \hat{0} & \hat{0} & \hat{0} & . & \hat{L}^{(\mu_f)} \end{pmatrix}. \tag{24}$$

In the expression (24), the zero-(sub-) matrices denoted by general symbols $\hat{0}$ may be of different order and of non-quadratic type, too. Moreover, the transformation procedures leading to quasi-diagonal forms may result multiple appearance of the same blocks according to relation (A,10) in the Appendix, relevant for the conductivity tensor. Analogous statement is valid for the block-diagonal decomposition of the entropy matrix realized via Jordan-blocks, too. The symbol μ_f indicates the last non-equivalent irreducible component of the tensor representation of symmetry group G_S of the conductivity tensor). Taking into account the general form (24) of the separate Jordan-blocks (together with nilpotent character of their constituent matrices $\hat{N}_q^{\hat{g}}$), we may incorporate them into $\hat{A} = \hat{g}\hat{L}$ using

$$\hat{g} = \hat{J}_1^{\hat{g}} \oplus ... \oplus \hat{J}_q^{\hat{g}} \oplus ... \oplus \hat{J}_p^{\hat{g}} \equiv \left(\lambda_1 \hat{I}_1 + \hat{N}_1^{\hat{g}}\right) \oplus ... \oplus \left(\lambda_q \hat{I}_q + \hat{N}_q^{\hat{g}}\right) \oplus ... \oplus \left(\lambda_p \hat{I}_p + \hat{N}_p^{\hat{g}}\right)$$

supplemented by relations $\left(\hat{N}_q^{\hat{g}}\right)^{u_q^{\hat{g}}} = \hat{0}, 1 \le q \le p, \forall u_q^{\hat{g}} \in \square$. Having collected all of these basic formulae, we are now in position to generalize the functions (3) and (23).

At first, since the symmetry theory plays a fundamental role in the above proposed generalization procedure, we base the introduction of the spatial dependence of the coefficients from (23) on it. Namely, via extension $c_{l,m} = const. \to c_{l,m} \cdot S_{l,m}(x)$, the functions $S_{l,m}(x)$ must obey certain spatial (≡ point group symmetry transformations, characterizing the given macroscopic dissipative system) invariance properties. This extension means also, that the replacement $t \mapsto \varphi(t) \to (x,t) \mapsto \varphi(x,t)$ must be used in (23).

Secondly, if we take into account the particular general forms of the basis functions allowing transformation of matrices to their Jordan normal forms, the formula (24) leads directly [Arnold, 2001] to the proposed generalization

$$\varphi(t)=\sum_{l=1}^{k} e^{\lambda_l t}\sum_{m=0}^{\nu_l-1} c_{l,m}S_{l,m}(x)t^m,\ c_{l,m}=const. \tag{25}$$

of the function system (23) (i.e. the choice $\nu_l=1$ corresponding to the case of non-multiple eigenvalues will lead back the form of (3)). Namely, taking into account the general method of the decomposition of the product of matrices by using the so-called hyperdyads [Elsner and Rózsa, 1981], [Rózsa, 2009], on the base of (24) we may write

$$\left[\hat{J}_1^{\hat{g}}\oplus\ldots\oplus\hat{J}_q^{\hat{g}}\oplus\ldots\oplus\hat{J}_p^{\hat{g}}\right]\left[\gamma_1\hat{I}_1\oplus\ldots\oplus\gamma_\mu\hat{I}_\mu\oplus\ldots\oplus\gamma_{\mu_f}\hat{I}_{\mu_f}\right]\left[\hat{L}^{(1)}\oplus\ldots\oplus\hat{L}^{(\mu)}\oplus\ldots\oplus\hat{L}^{(\mu_f)}\right]\equiv$$
$$\equiv\begin{pmatrix}\hat{J}_1^{\hat{g}} & \hat{0} & \hat{0} & . & \hat{0}\\ \hat{0} & . & . & . & \hat{0}\\ \hat{0} & . & \hat{J}_q^{\hat{g}} & . & \hat{0}\\ . & . & . & . & .\\ \hat{0} & \hat{0} & \hat{0} & . & \hat{J}_p^{\hat{g}}\end{pmatrix}\begin{pmatrix}\gamma_1\hat{I}_1 & . & \hat{0} & . & \hat{0}\\ \hat{0} & . & . & . & \hat{0}\\ \hat{0} & . & \gamma_\mu\hat{I}_\mu & . & \hat{0}\\ . & . & . & . & .\\ \hat{0} & \hat{0} & \hat{0} & . & \gamma_{\mu_f}\hat{I}_{\mu_f}\end{pmatrix}\begin{pmatrix}\hat{L}^{(1)} & . & \hat{0} & . & \hat{0}\\ \hat{0} & . & . & . & \hat{0}\\ \hat{0} & . & \hat{L}^{(\mu)} & . & \hat{0}\\ . & . & . & . & .\\ \hat{0} & \hat{0} & \hat{0} & . & \hat{L}^{(\mu_f)}\end{pmatrix}, \tag{26}$$

where the scalar matrices $\gamma_\mu\hat{I}_\mu$, $\mu\in\{1,\ldots,\mu_f\}$ are chosen according to Schur's lemma e.g. [Wigner, 1959], [Vilenkin, 1968] to commute with all matrices of the adequate irreducible representation $D^{(\mu)}(G_S)$ of the group G_S. Then, according to the relations (A,11) and (A,12) of the Appendix we may realize the following decomposition of (26) into hyperdyads with coefficients γ_μ of the scalar matrices as linear combination coefficients

$$\gamma_1\begin{pmatrix}\hat{J}_1^{\hat{g}}\\ \hat{0}\\ \hat{0}\\ .\\ \hat{0}\end{pmatrix}\begin{pmatrix}\hat{L}^{(1)} & \hat{0} & \hat{0} & . & \hat{0}\end{pmatrix}+\ldots+\gamma_\mu\begin{pmatrix}\hat{0}\\ .\\ \hat{J}_q^{\hat{g}}\\ .\\ \hat{0}\end{pmatrix}\begin{pmatrix}\hat{0} & . & \hat{L}^{(\mu)} & . & \hat{0}\end{pmatrix}+\ldots+\gamma_{\mu_f}\begin{pmatrix}\hat{0}\\ \hat{0}\\ \hat{0}\\ .\\ \hat{J}_p^{\hat{g}}\end{pmatrix}\begin{pmatrix}\hat{0} & \hat{0} & \hat{0} & . & \hat{L}^{(\mu_f)}\end{pmatrix}. \tag{27}$$

This decomposition also justifies introduction of functions (25) in a direct manner, where the relevant basis functions are obtained as products of those belonging to abstract subspaces carrying Jordan-block decompositions of the entropy matrix (related explicitly to the time variable) and decomposition into irreducible blocks of the conductivity tensor (and related therefore explicitly to spatial coordinates). It shows also, that the dyadic decomposition (19) has a specific character, whose generalization is realized here by novel type

partitioning of the matrix $\hat{A} = \hat{g}\hat{L}$ represented by dyadic products of rows and columns of matrix blocks instead of the earlier applications of simple rows and columns of scalars. At writing down of the basis functions of type (35), which correspond to decomposition (36), the scalars $\gamma_\mu, \mu \in \{1,...,\mu_f\}$ may be incorporated into coefficients $c_{l,m} = const.$ It is also assumed here implicitly, that dimensions of the submatrices building up hyper-columns ($\hat{J}_q^{\hat{g}}$) and hyper-columns ($\hat{L}^{(\mu)}$) have values, which allow realization of the products of type (27).

As a direct continuation of this work, an additional treatment of the time-reversal symmetries via Onsager-Casimir relations $L_{ik} = \varepsilon_i \varepsilon_k L_{ki} \left(1 \le i,k \le F_f\right)$ are leading to partitioning of the entropy matrix according to the presence of α- and possible β-type parameters. Here $\varepsilon_i \varepsilon_k = +1$ in the case of cross-effects $(i) \leftrightarrow (k), 1 \le i,k \le F_f$ when either purely α or purely β parameters are involved, while $\varepsilon_i \varepsilon_k = -1$ when in the $(i) \leftrightarrow (k), 1 \le i,k \le F_f$ cross effects α and β parameters are simultaneously involved and the β-parameters are quantities which change their signs under the influence of the time-reversal symmetry operation $t \to (-t)$ [de Groot and Mazur, 1962]. This fact would make possible further refinement of description of the hyperdyads in (26) and (27) by taking into account the explicit "α-β" block structure of the matrix $\hat{A}$.

4. Conclusions

The general formalism of the classical irreversible thermodynamics in linear approximation is connected to the concept of quasipolynomials used in the theory of differential equations and matrix analysis. It is demonstrated, that this refinement of the existing descriptions emanates directly from the Curie principle and is based on the effective use of group representation theory. The hyperdyadic decomposition of the product of the entropy matrix and conductivity matrix leads directly to the introduction of new and very general basis functions, which naturally explain the meaning of degenerate eigenvalues of the matrix $\hat{A} = \hat{g}\hat{L}$ by use of the product of abstract vector spaces, with suitably chosen bases corresponding to the block-diagonal forms of them. The newly introduced concept of symmetry properties of the series expansion coefficients $c_{l,m} \cdot S_{l,m}(x)$ may further refine the mathematical formalism of the classical, linear irreversible thermodynamics, together with more detailed studies of some special functions on the base of the general group representation theory [Vilenkin, 1968]. Particularly, further direct continuation of this work may be realized by effective use of the Onsager-Casimir relations describing the influence of time-reversal symmetries on the general thermodynamic relations,

which may lead to the revealing of additional details of the block structure of the hyperdyads applied here.

Acknowledgements

The authors acknowledge the support of the project P-MÖB/843.

Appendix 1.

– Elements of the conductivity tensor

The central concept in all formalisms of the non-equilibrium thermodynamics is the entropy production function $\sigma = \sum_{i=1}^{F_f} J_i X_i$, where F_f again denotes the number of thermodynamic degrees of freedom in the given dissipative system and whose explicit form is (with scalar components of fluxes J_i, which obey constitutive relationships of type [Gyarmati, 1970])

$$\sigma = \sum_{j=1}^{R} J_j A_j + p^v X_v + \vec{J}_q \cdot \vec{X}_q + \sum_{k=1}^{K-1} \vec{J}_k \cdot \vec{X}'_k + \vec{P}^{va} \cdot \vec{X}^a_v + \overset{0}{\hat{P}}_{vs} : \overset{0}{\hat{X}}_{vs} \geq 0, \qquad \text{(A,1)}$$

where by double dot ": ", the so-called interior product of tensors is denoted.
In the case of isotropic systems, the relevant constitutive equations form the following system of coupled equations

$$J_j = \sum_{r=1}^{R} L_{jr} A_r + L^{cv}_j X_v, \left(j = 1,2,...,R\right) \qquad \text{(A,2)}$$

$$p^v = \sum_{r=1}^{R} L^{vc}_r A_r + L X_v, \qquad \text{(A,3)}$$

$$\vec{J}_q = L_{qq} \vec{X}_q + \sum_{k=1}^{K-1} L_{qk} \vec{X}'_k, \qquad \text{(A,4)}$$

$$\vec{J}_i = L_{iq} \vec{X}_q + \sum_{k=1}^{K-1} L_{ik} \vec{X}'_k, \left(i = 1,2,...,K-1\right) \qquad \text{(A,5)}$$

$$\vec{P}^{va} = \overset{(aa)}{L}\, \vec{X}^a_v, \qquad \text{(A,6)}$$

$$\overset{0}{\hat{P}}^{vs} = \overset{(tt)}{L}\, \overset{0}{\hat{X}}^s_v, \qquad \text{(A,7)}$$

where R denotes the number of chemical reactions in a given dissipative system, K is those of its constituents, A_r characterizes the chemical affinities, while $\overset{(aa)}{L}$ and $\overset{(tt)}{L}$ are tensors, representing the couplings between physical quantities of axial vector character and between zero trace symmetric tensors, respectively. $\vec{P}^{va}$ is an axial vector obtained from elements of the antisymmetric part of the pressure tensor. The explicit forms of the thermodynamic driving forces from (A,6) and (A,7) are given by $\vec{X}_v^a = -\frac{1}{T}\left(\nabla \times \vec{v} - 2\vec{\omega}\right)$ and $\overset{0}{\hat{X}}_v^s = -\frac{1}{T}\left(\nabla \circ \overset{0}{\vec{v}}\right)_s$, respectively. In these expressions $\vec{v}$ denotes the convection flow velocity, $\vec{\omega}$ the angular velocity, while $\left(\nabla \circ \overset{0}{\vec{v}}\right)_s$ is a symmetric tensor with zero trace.

– Product groups

The direct outer product of the symmetry groups marked by G_1 and G_2 is usually denoted by $G = G_1 \times G_2$ with multiplication rule of the relevant new product group elements given by

$$\left(g_1, g_2\right) \cdot \left(g_1', g_2'\right) = \left(g_1 g_1', g_2 g_2'\right), \forall \left(g_1, g_2\right) \wedge \left(g_1', g_2'\right) \in G_1 \times G_2. \qquad \text{(A,8)}$$

The order of these product groups (i.e. number of their elements) is given as a product of the orders of its constituent groups, i.e. $\left|G_1 \times G_2\right| = \left|G_1\right| \cdot \left|G_2\right|$. The concept of the product of groups has many applications mainly in crystallography, where the subgroup structure whether G_1 and G_2 are (or only some of them) simple $(<)$ or invariant $(\lhd)$ subgroups of G, i.e. if we speak about direct- ($G_1 \lhd G, G_2 \lhd G \Rightarrow G = G_1 \otimes G_2$), semi-direct- ($G_1 \lhd G, G_2 < G \Rightarrow G = G_1 \wedge G_2$), or weak-direct ($G_1 < G, G_2 < G \Rightarrow G = G_1 \cdot G_2$) products of the constituent groups) of a given group could also be of importance in the basic crystallographic symmetry theories e.g. [Cracknell, 1975], [Damnjanović and Vujičić, 1982].

– Symmetry properties of irreducible tensor operators

Within frame of the group representation theory e.g. [Cracknell, 1975] it is possible to give a concise description of the transformation properties of (e.g. macroscopic) physical properties with tensorial character according to an irreducible representation of an adequate symmetry group. Therefore, a set of operators $\hat{A}_1, \hat{A}_2, ..., \hat{A}_j, ..., \hat{A}_n$ obey the following general relationship

$$\hat{D}\left(h\right)\hat{A}_j\hat{D}^{-1}\left(h\right) = \sum_{p=1}^{n} D_{pj}^{(\mu)}\left(h\right)\hat{A}_p, h \in G, \qquad \text{(A,9)}$$

where the upper index (μ) denotes an irreducible representation of the group *G*. Such relations are defined for both discrete and continuous groups *G*, and the symbols $\hat{D}(h)$ denote symmetry operators corresponding to the group elements $h \in G$. In more general case and in the case of a given abstract vector space, these symmetry operators are represented by matrices given in the following block-diagonal form of direct sum of the irreducible representations

$$D(h) = \sum_{\kappa}^{\oplus} a_{\mu} D^{(\mu)}(h), (h \in G) \tag{A,10}$$

where the coefficients a_μ count the non-equivalent irreducible representations of the group *G*.

– Products of hyperdyads

Considering the product of matrices represented by blocks obtained on the basis of their suitable equal partitioning i.e. the numbers of blocks fit in the relevant rows and columns, we can form immediately the relevant products of them, which in the case of e.g. three matrices reads

$$\left(\hat{A}_{ik}\right)\left(\hat{B}_{kl}\right)\left(\hat{C}_{lj}\right) = \left(\sum_{l=1}^{n}\sum_{k=1}^{n} \hat{A}_{ik}\hat{B}_{kl}\hat{C}_{lj}\right), 1 \le i,k,l,j \le n, \tag{A,11}$$

which in the case of e.g. quasi-diagonal form of the block-matrix $\hat{B} = \left(\hat{B}_{kl}\right), (1 \le k,l \le n) \rightarrow \hat{B} = \left(\hat{B}_{k}\right), (1 \le k \le n)$ can be conveniently given by

$$\begin{pmatrix} \hat{A}_{11} & \hat{A}_{12} & . & . & \hat{A}_{1n} \\ \hat{A}_{21} & \hat{A}_{22} & . & . & \hat{A}_{2n} \\ . & . & . & . & . \\ . & . & . & . & . \\ \hat{A}_{n1} & \hat{A}_{n2} & . & . & \hat{A}_{nn} \end{pmatrix} \begin{pmatrix} \hat{B}_{1} & \hat{0} & . & . & \hat{0} \\ \hat{0} & \hat{B}_{2} & . & . & \hat{0} \\ . & . & . & . & . \\ . & . & . & . & . \\ \hat{0} & \hat{0} & . & . & \hat{B}_{n} \end{pmatrix} \begin{pmatrix} \hat{C}_{11} & \hat{C}_{12} & . & . & \hat{C}_{1n} \\ \hat{C}_{21} & \hat{C}_{22} & . & . & \hat{C}_{2n} \\ . & . & . & . & . \\ . & . & . & . & . \\ \hat{C}_{n1} & \hat{C}_{n2} & . & . & \hat{C}_{nn} \end{pmatrix} =$$

$$= \begin{pmatrix} \hat{A}_{11} \\ \hat{A}_{21} \\ . \\ . \\ \hat{A}_{n1} \end{pmatrix} \left(\hat{B}_{1}\right)\left(\hat{C}_{11} \quad . \quad . \quad . \quad \hat{C}_{1n}\right) + \ldots + \begin{pmatrix} \hat{A}_{n1} \\ \hat{A}_{n2} \\ . \\ . \\ \hat{A}_{nn} \end{pmatrix} \left(\hat{B}_{n}\right)\left(\hat{C}_{n1} \quad . \quad . \quad . \quad \hat{C}_{nn}\right). \tag{A,12}$$

The procedure may be directly applied to the case of block-matrices with non-quadratic block-structures, too.

List of symbols:

a_T – coefficient of temperature conductivity

a_m – coefficient of conductivity of the total transfer of vapour and liquid

ε – the liquid – vapour phase transition criterion (quantity of the moving vapour in the moist body in relation to the total current of the vapour and liquid)

F_f – number of thermodynamic degrees of freedom

K – number of constituents in a dissipative continuum

Ko – Kossovich criterion (relation between the heat expended in evaporation and the heat necessary for the heating of a moist body)

Lu_m – Luikov`s criterion of inertia ($Lu_m = \frac{a_m}{a_T}$: criterion of the lag of the temperature field relative to the transfer potential field of matter)

Pn – Posnov criterion (relative drop of mass-content and of the transfer potential matter in a body due to the drop of temperature)

p^v – 1/3 of trace of the viscosity pressure tensor

$\vec{P}^{va}$ – contribution corresponding to the asymmetric part of the viscous pressure tensor

$\overset{0}{\vec{P}}^{vs}$ – contribution corresponding to the symmetric part of the viscous pressure tensor

R – number of simultaneous chemical reactions taking place in a dissipative continuum

α_S – specific entropy production function

References

[1.] V.I. Arnold, Gewöhnliche Differentialgleichungen 2. Auflage (Übersetzung aus dem Russischen) Springer-Verlag: Berlin – Heidelberg – New York, 2001

[2.] T.M. Atanacković, S. Pilipović, D. Zorica, Time-distributed order diffusion-wave equation, I. Volterra-type equation, Proceedings of the Royal Society A – Mathematical, Physical and Engineering Sciences 465 (2009) 1869 – 1891

[3.] T.M. Atanacković, S. Pilipović, D. Zorica, Time-distributed order diffusion-wave equation, II. Application of Laplace and Fourier transformations, Proceedings of the Royal Society A – Mathematical, Physical and Engineering Sciences 465 (2009) 1893 – 1917;

[4.] C. Coles, D. Murio, Parameter estimation for a drying system in a porous medium, Computers & Mathematics with Applications, 51 (2006) 1519 – 1528

[5.] A.P. Cracknell, Magnetism in Crystalline Materials, Pergamon Press: Oxford – New York – Toronto – Sydney – Braunschweig, 1975

[6.] M. Damnjanović, M. Vujičić, Magnetic Line Groups, Physical Review B, 25 (1982) 6987 – 6994

[7.] L. Elsner, P. Rózsa, On Eigenvectors and Adjoints of Modified Matrices, Linear and Multilinear Algebra, 10 (1981) 235 – 247

[8.] F. R. Gantmakher, The theory of matrices, I., II., Chelsea: New York, 1959

[9.] S.R. de Groot, P. Mazur, Non-Equilibrium Thermodynamics, North-Holland Publ. Co., Amsterdam, 1962

[10.] I. Gyarmati, Non-Equilibrium Thermodynamics (Field Theory and Variational Principles), Springer – Verlag: Berlin-Heidelberg-New York, 1970

[11.] I. Gyarmati, On the wave approach of thermodynamics and some problems of non-linear theories, Journal of Non-Equilibrium Thermodynamics 2 (1977) 233 – 260

[12.] D. Jou, J. Casas-Vazquez, G. Lebon, Extended Irreversible Thermodynamics 3rd Ed., Springer – Verlag: Berlin – Heidelberg – New York, 2001

[13.] I. Kirschner, S. Leppävuori, A. Haasz, An irreversible thermodynamic theory of measuring sensors, Sensors and Actuators A 27 (1991) 677 – 682

[14.] I. Kirschner, S. Leppävuori, Non-Equilibrium, Irreversibility, Nonlinearity and Instability in the Operation of Sensors, Sensors and Actuators A 31 (1992) 275 – 282

[15.] I. Kirschner, Thermodynamic Stability of the Operation of Sensors, Sensors and Actuators A 42 (1994) 622 – 629

[16.] I. Kirschner, Cs. Mészáros, Á. Bálint, K. Gottschalk, I. Farkas, Surface changes of temperature and matter due to coupled transport processes through porous media, Journal of Physics A: Mathematical and General 37 (2004) 1193 – 1202

[17.] L.D. Landau, E.M. Lifshitz, Fluid Mechanics, Heinemann: Oxford, 2000

[18.] A.V. Luikov, Yu.A. Mikhailov, Theory of Energy and Mass Transfer, Pergamon Press: London, 1965

[19.] MAPLE 10 2005 *Symbolic Computation System* Maplesoft 615 Kumpf Drive Waterloo Ontario Canada N2V1K8

[20.] Cs. Mészáros, Á. Bálint, I. Farkas, J. Buzás, 2004 Modeling of the coupled heat and mass transfer through porous media on the base of the wave approach, Drying Technology 22 (2004) 71 – 80

[21.] Cs. Mészáros, Á. Bálint, I. Kirschner, K. Gottschalk, I. Farkas, Mathematical modeling of drying processes using methods of the non-equilibrium thermodynamics and percolation theory, Drying Technology 25 (2007) 1297 – 1304

[22.] Cs. Mészáros, I. Farkas, Á. Bálint, A new application of percolation theory for coupled transport phenomena through porous media, Mathematics and Computers in Simulation, 56 (2001) 395 – 404

[23.] R.N. Pandey, S.K. Srivastava, M.D. Mikhailov, Solutions of Luikov equations of heat and mass transfer in capillary porous bodies through matrix calculus: a new approach International Journal of Heat and Mass Transfer 42 (1999) 2649-2660

[24.] P. Rózsa: Introduction to Matrix Theory (in Hungarian), Typotex: Budapest, 2009

[25.] Sun HG, Chen W, Chen YQ, Variable-order fractional differential operators in anomalous diffusion modelling, Physica A – Statistical Mechanics and its Applications 388 (2009) 4586 – 4592

[26.] V.E. Tarasov, G.M. Zaslavsky, Fractional dynamics of systems with long-range space interaction and temporal memory, Physica A – Statistical Mechanics and its Applications 383 (2007) 291 – 308;

[27.] V.V. Uchaikin, Self-similar anomalous diffusion and Levy-stable laws, Physics – Uspekhi 46 (2003) 821 – 849

[28.] N.Ya. Vilenkin, Special functions and the theory of group representations *Translations of Mathematical Monographs*, Providence, RI: American Mathematical Society, 1968

[29.] E.P. Wigner: Group Theory and its Application to the Quantum Mechanics of Atomic Spectra, Academic Press: New York and London, 1959

[30.] G.M. Zaslavsky, Chaos, fractional kinetics, and anomalous transport, Phys. Rep., 371 (2002) 461-580

Institute for Mechanical Engineering Technology

Professor Dr. Gábor KALÁCSKA
Director of the Institute

Dear Reader,

The Institute for Mechanical Engineering Technology (consists of three departments: Department of Material and Engineering Technology, Department of Maintenance of Machinery, Department of Mechatronics) had the following main research area during 2010.

Concerning the materials science:

- Development of nano- and micro composite materials on the bases of magnesium catalyzed cast Polyamide 6 matrix. The composites are used in many areas of engineering. Antistatic, Esd, improved tribological and fire-resistant versions of cast PA 6 were developed.

Manufacturing technologies and Maintenance:

- In place of brittle ceramics used so far have appeared up-to-date so called tough ceramic materials resisting better against mechanical effects. Such material is the zirconium-dioxide, too. Extensive experiments were carried out to clarify the machining features of zirconium-dioxide comparing to other conventional iron materials.
- The formerly started research on the field of machining of semi-finished engineering polymers went on. The main- and feeding-directional cutting forces were measured and analyzed in the function of feed and cutting depth, while the chip removal was also investigated and analyzed with the tensile behaviour of the materials. General machining maps were created.
- Abrasive wear resistance of hot-dip galvanized multi-layers was investigated. Natural Technigalva and heat treated version of Technigalva surface layers were studied with different methods: microVickers hardness, EDS spectra analyses were taken into account during the evaluation of abrasive wear, which was determined based on the decrease of layer thickness during wear process.

Mechatronics:

- ER fluids composed with different additives were investigated possibly used in control enhancement of hydraulic systems.
- To establish further research, significant development in the field of PLC laboratory facilities went on.
- New research area has been established regarding mechatronics of vehicles: handling characteristics of off-road vehicles by applying active control of steering wheel torque, and development of haptic feedback device on mobile agricultural equipment.

Machining of ZrO_2 ceramics with PCD and CBN cutting tools

Gellért FLEDRICH, István PÁLINKÁS,
Róbert KERESZTES, László JÁNOSI
Department of Material and Engineering Technology,
Institute for Mechanical Engineering Technology

Abstract

In place of brittle ceramics used so far have appeared up-to-date so called tough ceramic materials resisting better against mechanical effects [6]. Such material is the zirconium-dioxide, too. The important advantage of hard-turning is the applicability of universal tool. Various outlines can be formed by a tool given. Machining ceramics in case of using traditional machining (turning, milling, drilling) requires special technological conditions (tools, machine-tools, technological parameters, etc.) which are developing presently [2]. We would like extending our research work in this course, too. To clarify the machinability – turning – of ZrO_2 ceramics we developed a cutting force measurements for the applied CBN and PCD cutting tools. The forces were studied in the function of cutting speed and feed, the surfaces were analized by SEM and the cutting process was controlled by thermo-camera. The failure of cutting edges were also studied. The summarized results suggest the possible turning possibilities of ZrO_2 ceramics.

Keywords

polycrystal diamond, cubic boron nitride, 3D-topography, heat ring, surface roughness, friction

1. Introduction

By industry development the demand is increasing for such materials to be applied at higher temperature beside at heavy physical and chemical load. The structural ceramics can have an important role exactly in this segment. The zirconium –dioxide is also such material [3]. The role of ceramics hard-machining is increasing presently [4]. The zirconium-dioxide deriving from its lower hardness and from other characteristics [1] is suitable to machine by tool having regular edge can become a potential material at piece or small- and medium series production. To ensure this it has to be known its cutting characteristics [5].

Our research work focusing a part of this in keeping with the recommendation of the company producing and developing zirconium-dioxide semi-finished products.

2. Materials tested and their forms.

The common properties of engineering ceramics are that they have outstanding physical and chemical characteristics in very high temperature range [7]. The ceramics tested by us have got high hardness (1250-1800 HV), because of this it can be cut by polycrystal diamond and cubic boron nitride tools. The material tested is zirconium-dioxide ceramics. The specimens used at turning tests were cylindrical, their diameters were 16 and 20 mm (Figure 1.)

Figure 1. The zirconium-dioxide (Zn40) ceramics used at tests.

The ceramic properties tested [6].

	Dimension	
Material		ZN 40
Main component		ZrO_2-MgO
Density	g/cm^3	5,74
Bending strength	MPa	500
Compression strength	MPa	1600
Young modulus	GPa	210
Poisson - ratio		0,3
Vickers hardness HV0,5		1240
Thermal convectivity	W/mK	3
Linear coefficient of thermal expansion	$10^{-6}K^{-1}$	10,2

Cutting tests

During turning the work-piece rotating movement is the main movement, the auxiliary movements are the turning tool movement in feeding and depth of cut directions. We have set the cutting speed by the work-piece revolution number.

To measure the axial and tangential components of the cutting force [8] we have used a measuring tool-head with strain gauge. We have used the Spider 8

measuring amplifier for the tests. We have connected four channels. We have coupled to the O-channel the revolution marker, we have measured the tangential and feeding forces on the 1 and 2 channels. The 3. channel served to measure the tool displacement, here we have connected an inductive displacement tele-transmitter restored by spring.

Heat affected zone tests.
The heat arising during cutting influences considerably the removal of stock process as well as influences very much the tool durability. We have also made shots with thermo-camera during cutting to study the heat affected zone formed.

3. Discussion

During cutting we have measured the main and feeding forces affecting the tool. With the measuring-system capable to measure active forces-developed by us in the institute – we show some diagrams (Figure 2 and 3) in the followings. We present the main cutting force with thin line the feeding force with crossed line. We show the change of feed-rate with sections having circle end point.

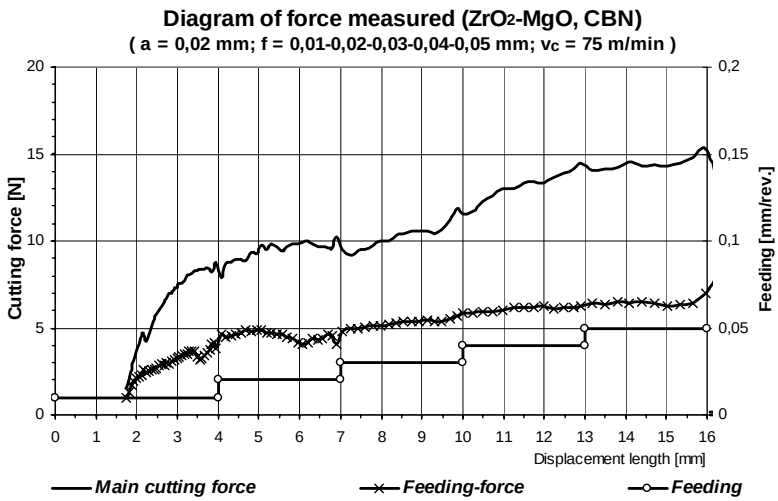


Figure 2. Diagrams of main cutting force and feeding force (v_c= 75 m/min., a= 0,02 mm; f= 0,01-0,02-0,03-0,04-0,05 mm, ceramic: zirconiumdioxide, tool: CBN)

The zirconium-dioxide was cut with cubic boron nitride tool in the above. Figure 2.

It can be seen that main cutting force and feeding force show increasing tendency with the increasing feed-rate.

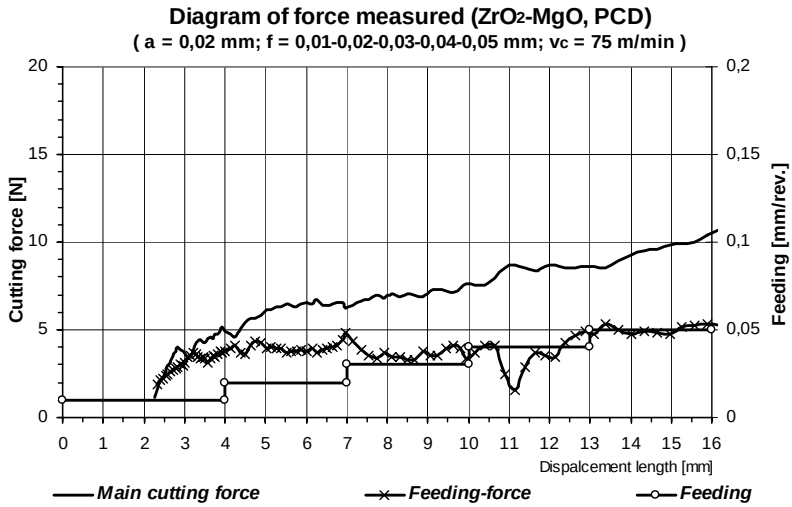


Figure 3. Diagrams of main cutting force and feeding force (v_c= 75 m/min., a= 0,02 mm; f= 0,01-0,02-0,03-0,04-0,05 mm, ceramic: zirconiumdioxide, tool: PCD)

In Figure 3 at the same conditions but cutting with polycrystal diamond resulted some different curves. The main cutting force increase is less rising in this case but it has got similar tendency, the feeding force following a short rising shows rather smaller increasing angular curve.

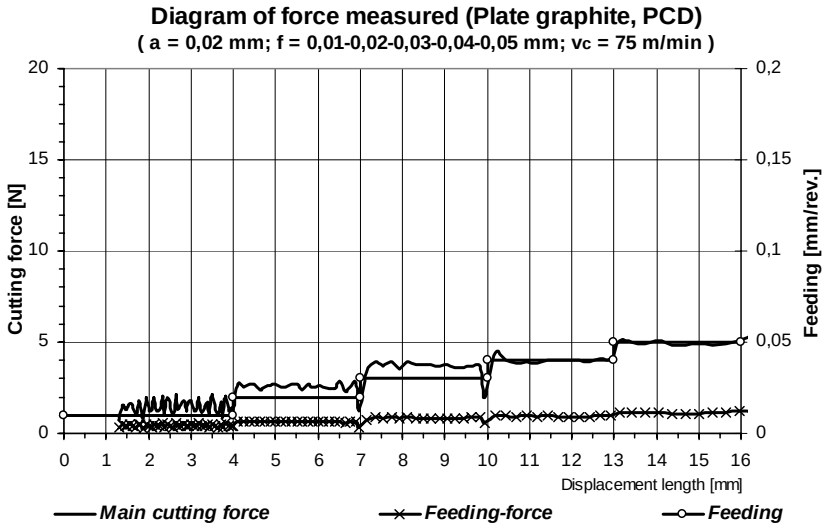


Figure 4. Diagrams of main cutting force and feeding force (v_c= 75 m/min., a= 0,02 mm; f= 0,01-0,02-0,03-0,04-0,05 mm, cast iron: plate graphite, tool: PCD)

We have cut plate graphite cast iron with PCD-tool similar with previous cutting data in case of Figure 4. We compared the known cast iron with zirconium-oxide unknown from turning standpoint is such way. The diagrams got are similar though with lower values. Difference can be experienced in the initial increasing section. The change of forces are directly proportional. In case of ceramic the initial increase is steeper at both tool materials.

SEM photos of the cut surfaces can be seen in figure 5-7 in 2000 x magnification. The different surface effects of the CBN and PCD tools can be realized.

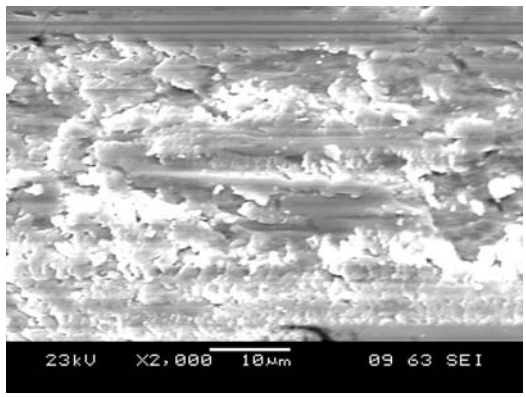


Figure 5. Ceramic surface, machined with CBN.
v_c=75 m/min., a = 0.02 mm, f = 0.04 mm/rev. magnification x2000

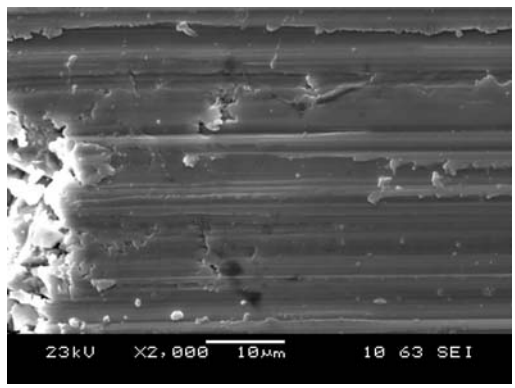


Figure 6. Ceramic surface, machined with PCD.
v_c=75 m/min., a = 0.02 mm, f = 0.04 mm/rev. magnification x2000

The ZrO_2 phase transformation tracks caused by mechanical effect can be seen in 2000x-magnification of the original surface grinded (Figure 7.). In case of PCD-tool the rate of greater shell-like tears is similar to grinding. The amount of micro-cracks can be reduced based on the Figures. Less micro-cracks can be seen on surfaces turned compared to grinded in the Figures.

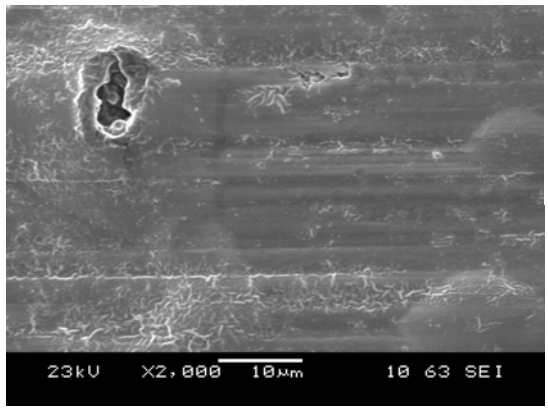


Figure 7. Grinded surface

3D-topography pictures can be seen in Figure 8-10.

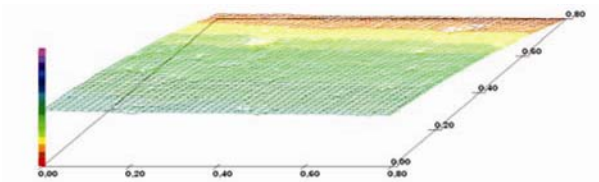


Figure 8. PCD tool turned
v_c =75 m/min., a = 0.02 mm, f = 0.05 mm/rev.

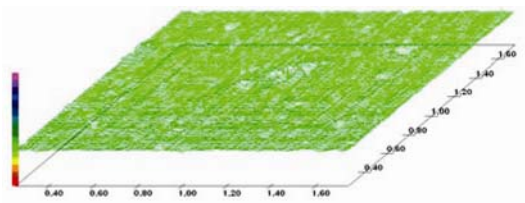


Figure 9. PCD tool turned
v_c =75 m/min., a = 0.02 mm, f = 0.02 mm/rev.

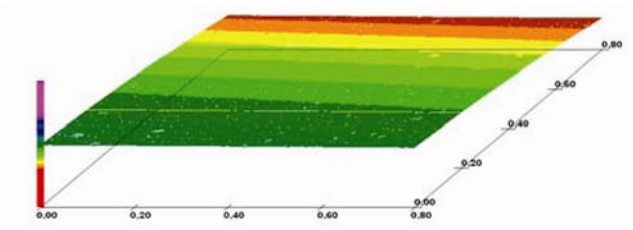


Figure 10. Grinded surface

In Figure 8-9. can be seen that the higher cutting speed (v_c =75 m/min.) resulted more favourable surface roughness than at lower cutting speed (v_c =25 m/min.).

From the point of applicability of cutting technology for ceramics, it is important to examine the friction behaviour of the different machined ceramic surfaces (grinded and machined surfaces with different cutting tools and technological settings) because the most common applications of such materials can be found among the slide bearing solutions.

In the dry friction tribological model system – steel pin on the machined ceramic ring surfaces - we have analyzed the friction force (coefficient) (Figure.11), and wear. Regarding the machining phenomena of the two tool materials, the PCD tool was selected to produce machined ceramic surfaces with different technological settings. In the tribological test system we have increased the normal load in three steps: starting with 50N ($\sigma_{Hertz\text{-}max}$ = 391 N, without wear) load the measurements were running for 5 minutes, then it was increased upto 100 N for 5 minutes and further 150N was applied to the seizure of samples.

Having S355 steel pin with grinded surface Ra 0,3 on the different machined ceramic ring surfaces, we have found that the applied higher cutting speed (v_c=75 m/min) and lower feed (f=0,01-0,02 mm/ford) resulted lower friction between the steel and ceramic surface than it was with the original grinded ceramic surface. The original grinded ceramic surface has more but smaller sized "pits", which can stick the steel worn particles, thus the friction process turns quickly to a steel/steel like contact having relatively high adhesion.

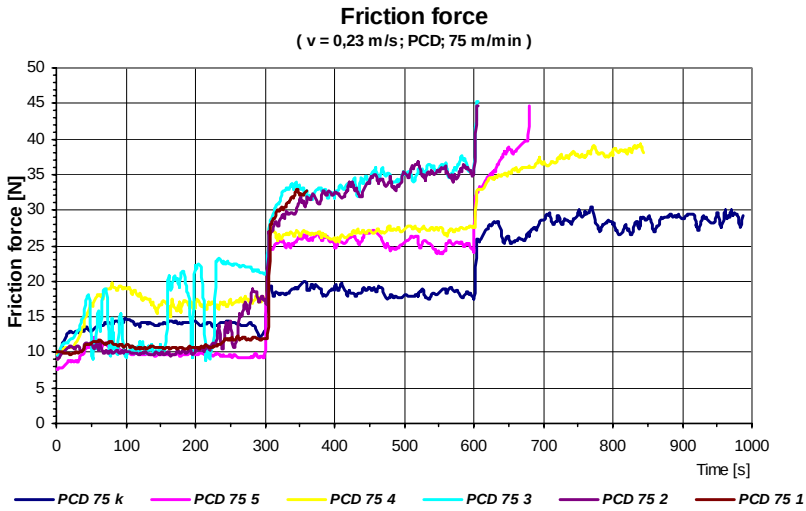


Figure 11. Friction force diagram
(steel/grinded and machined ceramic surfaces)

4. Conclusions

Trend characterizing steel turning appears at using CBN turning tool, local cutting force maximum can be identified.

In case of PCD-tool at higher cutting speed more favourable surface roughness can be got, comparing to CBN-tool.

It is possible turning zirconium-dioxide semi-finished product. The condition of the starting surface has got definite importance how is possible to turn the surface point of view. The raw-product surface grinded has damaged the PCD-tools in case of all cutting parameters tested.

The hardness recommendation accepted for tools in technical literature (3-4 times higher hardness difference in favour for the tool) can not be used at raw-products grinded in case of zirconium-dioxide ceramics, Softer, 2 times higher hardness can be used.

Based on the tests can be established that given cutting speeds in case of increasing feedings at different tool-materials and workpiece materials show similar tendency, only in steepness is change. The phenomenon validity can be extended from the steel/carbide connection to the zirconium-dioxide/CBN friction, chip removal connection, too. The CBN and PCD-tools have resulted significantly different cutting forces. It can be stated that the cutting speed increase has resulted well measured increase in cutting force in case of diamond turning tool which has called forth significant heat evolution.

This is significantly unfavourable tribological connection.

The standard PCD –tool used did not result different surfaces regarding the surface roughness.

It can be stated on the basis of SEM photos that on the surfaces grinded phase transformations take plate at microcracks. This results the higher cutting forces at turning in case of grinded surfaces. At surfaces turned phase transformation can not be seen. The surface cut by PCD –tool results favourable surface roughness.

We have found that it is possible to produce machined ceramic surface, which can offer better sliding properties against steel than the original grinded ceramic surface.

References

1. CURTIS, C.E.: Development of zirconia resistant to thermal shock. *J. Am. Ceram. Soc. 30* 180-196., 1947.

2. A. H. Fritz, G. Schulze (Hrsg.):Fertigungstechnik. 8. Aufl *Springer Verlag, Berlin* 279. o., 2007

3. H. SALMANG, H. SCHOLZE: Keramik, *Springer Verlag, Berlin,* 2007

4. SWAIN, M.V.; ROSE, L.R.F.: Strength Limitations of Transformation-Toughened Zirconia Alloys. J. *Am. Ceram. Soc.* 69 511-518., 1986.

5. MÉSZÁROS I., SZEPESI D.: Edzett acélok nagypontosságú megmunkálása IV. *Gépgyártás* XLVI. évf. 4. sz. 29-33 old., 2006.

6. SZÉPVÖLGYI J.: Korszerű műszaki kerámiák, *Magyar Tudomány*, 4. sz. 7-13 old., 1994.

7. CeramTec AG Innovative Ceramic Engineering Medical Products Division: *Broschüre* 2008

8. FARKAS G., KALÁCSKA G.:The effect of the different cutting tools on the micro-geometrical surface of engineering plastics. *Sustainable Construction & Design. 2010. Vol.1. p. 102-107.*

Novel cylinder positioning system with solenoid valves

László FÖLDI, Zoltán BÉRES, Eszter SÁRKÖZI, László JÁNOSI
Department of Mechatronics,
Institute for Mechanical Engineering Technology

Abstract

This paper presents a novel control design, developed to realise fast and accurate position control of a pneumatic actuator using inexpensive on/off solenoid valves. In contrast to conventional control methods, the proposed control method operates chatter free, based on air compression. The control principle was developed by investigating the dynamics of a pneumatic actuator with an identified mathematical model. This new approach is applied to a pneumatic double acting cylinder, controlled by a pair of 5/3 way directional single solenoid valves. The described closed-loop circuit copes with the discontinuities associated with the valve's switching dynamics, and relatively long response time. The experimental apparatus uses an analogue displacement encoder for metering the piston's position and velocity, and doesn't incorporate pressure sensors thus ensuring a low cost system design. The results of experiments with various step responses show that the proposed control method performs well. The measured steady-state position errors are equal to the used potentiometer's travel resolution, which is 0,01 mm. Therefore this novel control and the related pneumatic system design could be a cost effective alternative to the servo-pneumatic positioning systems.

Keywords

Position Control, Solenoid Valve, Pneumatic Cylinder, Control Method, Actuator

1. Introduction

In order to achieve linear motion, pneumatic, electromagnetic and hydraulic actuators are typically used. Due to their advantageous characteristics in position control applications the latter two are more widespread. Though using hydraulic actuators high velocity and great force is achievable, and their position control can also be relatively easily solved, a handicap is that the leaking of hydraulic fluids might contaminate the workpiece. Electromagnetic actuators on the other hand are clean and reliable in their operation but often require a mechanical transmission, both to convert high speed and low torque to a more useful combination and to convert rotary motion to linear motion. While linear motors overcome the need for transmission, they can be expensive [9].

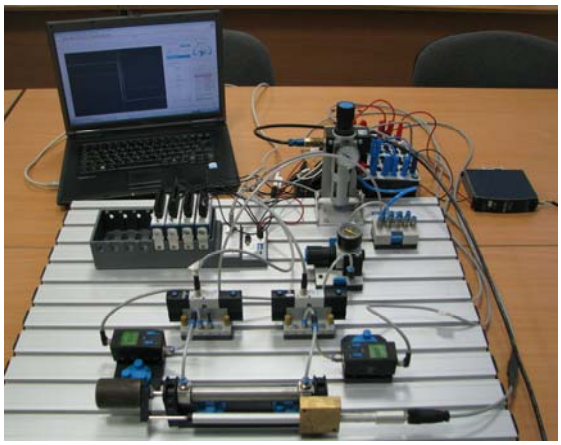


Figure 1. The experiment apparatus

Pneumatic actuators have several advantages: they are fast, cheap, have an outstanding power-to-weight ratio, are easily maintainable and they don't contaminate the work piece. The challenge to the use of pneumatic drives is that due to piston friction and the characteristics of compressed gas flow their behaviour is non-linear. As a result their industrial use is only widespread in applications which require linear motions between end positions.

Table 1. A review of the papers on the topic

Number in References	Authors	Control method	Number of required valves	Positioning accuracy
[9]	Thomas, M.B.; Maul, G.P.; Jayawiyanto, E.	Modified PD + PWM control	3	±0,1 mm
[1]	Ahn, K.; Yokota, S.	MPWM + neural network (LVQNN)	8	±0,2 mm
[7]	Parnichkun, M.; Ngaecharoenkul, C.	hybrid of fuzzy and PID + PWM control	1	±3,5 mm
[6]	Nguyen, T.; Leavitt, J.; Jabbari, F.; Bobrow, J.E.	Sliding mode control	4	±0,1 mm
[8]	Shih, M.-C.; Ma, M.-A.	Fuzzy + modified differential PWM control	2 + Proportional pressure valve	±0,075 mm
[5]	Messina, A.; Giannoccaro, N.I.; Gentile, A.	Individual control + PWM	2	±0,1 mm
[3]	Barth, E.J.; Zhang, J.; Goldfarb M.	Individual linear continuous + PWM	2	N/A
[2]	Akdağ, F.N.; Kuzucu, A.	Sliding mode control	2	±0,05 mm
[10]	van Varseveld, R.B.; Bone, G.M.	PID with friction compensation + PWM	2	±0,21 mm

In the last decade such industrial controllers became available which have adequate computing capacity for real-time usage. Thus there is now opportunity to develop pneumatic systems which don't require costly proportional valves for positioning and hence the usage of the more cost-efficient solenoid valves became possible (Figure 1.). These simple on/off valves are cheap and easily maintainable. Their drawback however is that as yet the required control method is not appropriately elaborate: the solutions offered in the scarce publications on the topic consist of applying conventional control methods (PID, sliding mode, fuzzy logic) or their hybrid versions, with varying success.

A review of the papers on the topic is given in Table 1, where the applied control methods, the number of required valves and the highest positioning accuracy is shown. Even though it is not indicated in the table, in some cases certain speed or force decreasing solutions (eg. throttle valve, reduced pressure) were used, which naturally may have an effect on the operating range of the positioning system and the steady-state error. The employed valve's switching time is also omitted from the table.

2. Mathematical and simulation model

We carried out the design work of our proposed control strategy based on an identified simulation model [4] of the actual apparatus. In order to develop this system we applied equations number (1) and (2); using these we can express the mass flow filling the chambers of the cylinder as follows:

$$\dot{m}_1 = A_1 \cdot p_{\text{sup}} \cdot \sqrt{\frac{2}{R \cdot T}} \cdot \psi_{(p_1/p_{\text{sup}})} \tag{1}$$

$$\dot{m}_2 = A_2 \cdot p_2 \cdot \sqrt{\frac{2}{R \cdot T}} \cdot \psi_{(p_{atm}/p_2)} \tag{2}$$

The outflow functions (ψ) in the equations are one of the major causes of the nonlinear behaviour of the pistons. Knowing the mass flows, the pressures in the cylinder's chambers can be expressed through the gas law's derivative with respect to time:

$$\dot{p}_1 = \frac{1}{V_1} \cdot \left(\dot{m}_1 \cdot R \cdot T - p_1 \cdot \dot{V}_1 \right) \tag{3}$$

$$\dot{p}_2 = \frac{1}{V_2} \cdot \left(p_2 \cdot \dot{V}_2 - \dot{m}_2 \cdot R \cdot T \right) \tag{4}$$

Using the pressures and knowing the physical parameters of the piston the mechanical equation complemented with the mass loads is as follows. Again, a term causing non-linearity appears: friction (Ff).

$$m \cdot \ddot{x} = p_1 \cdot A_1 - p_2 \cdot A_2 - |F_f| \cdot sign(\dot{x}) - m \cdot g \cdot \sin \alpha - j \cdot (\dot{x})^2 \cdot sign(\dot{x}) \qquad (5)$$

The last term in the equation is the pipe friction loss which, according to our observations is a significant factor in practice but is ignored in the specialised literature we made use of. This friction loss was taken into consideration with a friction factor (j) multiplied by the square of velocity.

Using the presented equations we were able to construct the proposed control system's simulation model in Matlab/Simulink software. After completing the identification process of the model's parameters, we were able to effectively analyse some of the conventional control methods. As a result we decided to elaborate a novel control strategy, which copes with the discontinuities associated with the valve's switching dynamics, and relatively long response time.

3. Control method

Designing the control method we determined to apply two solenoid valves (one for each chamber), the biggest benefit of which is that the number of control signals is thus raised to the second power.

Table 2. Available variations using two solenoid valves

Number of variation	Signal of solenoid valve 1	Signal of solenoid valve 2	Movement of the cylinder piston
1	fill	empty	positive direction, fast
2	fill	close	positive direction, slow
3	fill	fill	immobile
4	close	empty	positive direction, uncontrolled
5	close	close	first immobile, then uncontrolled
6	close	fill	negative direction, slow
7	empty	empty	accidental direction, uncontrolled
8	empty	close	negative direction, uncontrolled
9	empty	fill	negative direction, fast

Out of the variations presented in Table 2, we ignored the uncontrolled ones (where pressure in both chambers changes in an uncontrolled way) which thus leaves us with five useful solenoid valve variations. However, in order to stop the movement of the cylinder piston in variation number 3, a complementary element had to be introduced (due to the asymmetric construction of the cylinder

we used). Using a pressure regulator, the positive chamber was filled with supply pressure decreased in proportion to the surfaces of the piston. As a result a balance of force was reached based on equation number (5).

As a first step we started out from a graph representation, where time and position were placed on the axes. This way the most important control parameter, the error (difference between a measured process variable and a desired set-point) can be visualized easily if we represent both the ideal and the actual position values (reference signal and measured output respectively) plotted against the elapsed time.

Based on the scale of the error we might form predictions as to which valve variation has to be realized by the control method in order to decrease the absolute value of the error.

Table 3. Realized valve variations

Number of control signal	**Relation between reference signal and measured output**	**Desired piston movement**	**Number of variation in Table 2**
1	$r >> y_m$	positive direction, fast	1
2	$r > y_m$	positive direction, slow	2
3	$r \approx y_m$	immobile	3
4	$r < y_m$	negative direction, slow	6
5	$r << y_m$	negative direction, fast	9

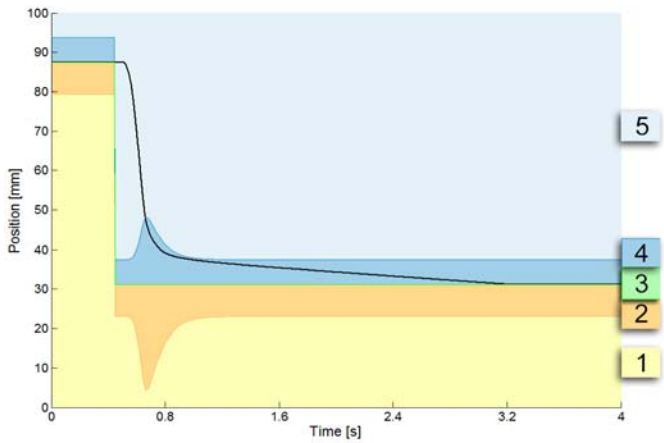


Figure 2. Control Strategy

The variations in Table 3 can be easily indicated in the graph representation we have already created (Figure 2.), since control signal number 3 has to take effect

within the close vicinity of the ideal position, namely in a predefined tolerance band where further changing of position is unnecessary. Nearby, above and below this band we can assign the zones where the piston is relatively close to the ideal position, thus slow movement is required (control signals number 2 and 4). Outside these bands those parts of the stroke length are situated which are relatively far from the ideal position, hence a fast piston movement is needed in order for a fast decrease in the absolute value of the error to take place (control signals number 1 and 5). Since the width of the tolerance band is a parameter given during utilization, the control method only has to determine the width of bands number 2 and 4 in a given moment. By rendering the respective valve variations to the given bands, the solenoid valves can be controlled harmoniously and in an essentially discrete way, with few valve switches, i.e. chatter free.

In order to determine the width of the bands we can set up the following equation based on the work-energy theorem:

$$0 - \frac{1}{2} \cdot m \cdot v^2 = -\left(\frac{p_b \cdot A \cdot (q+z)}{R \cdot T} \cdot R \cdot T \cdot \ln \frac{p_a}{p_b}\right) \tag{6}$$

Of the terms of the equation pa and pb are the initial and final pressures of the closed chamber, q is the distance needed to stop the piston while z is the remaining distance. Setting up the equation we regarded the process to be isothermal, thus we overestimated the necessary distance to stop the piston, this way compensating for the losses which appear in the positioning system.

After reducing the values regarded as constants and expressing q from the equation, we come to the following relation:

$$q = c_2 \cdot v^2 - z \tag{7}$$

The equation shows that the ideal deceleration distance of the piston (q) changes according to a second-degree function of piston speed. It follows that the widths of the deceleration bands of the piston have to be adjusted in a given moment proportional to the square of the measured output's first derivative; at this point a proportional coefficient (c2) has to be introduced. Furthermore, it is also necessary to set a constant c1 bandwidth which guarantees the slow movement near the ideal value.

As the next step, we have to assign the adequate control signs to the bands. This is done in a similar way to the operation of a relay, by sign functions, which decide whether the piston's measured output is below or above the band limit set. Summing up the values of these sign functions we arrive at 4 discrete numeric values which unequivocally determine which band the measured output falls into (e.g. we get -3 if it is below all band limits, i.e. in band number 1). By assigning the adequate control signs to these numeric values and to the tolerance band around the ideal position we achieve a control method which satisfies the criteria set.

In accordance with notations of control engineering, the mathematical description of the control method is as follows: r stands for the reference signal, while ym is the measured output, t means tolerance and F;C;E are respectively the filling, closing and emptying switch states of the first and second solenoid valves:

$$sign\left\{y_m - \left[r - (c_1 + c_2 \cdot (\dot{y}_m)^2)\right]\right\} + sign\,(y_m - r) + sign\left\{y_m - \left[r + (c_3 + c_4 \cdot (\dot{y}_m)^2)\right]\right\} = k$$

$$\begin{aligned} &if\ |y_m - r| < t, then \quad \Rightarrow \quad cont = [F;F] \\ &if\ |y_m - r| > t, then \begin{cases} if\ k = -3 \Rightarrow cont = [F;E] \\ if\ k = -1 \Rightarrow cont = [F;C] \\ if\ k = 1 \Rightarrow cont = [C;F] \\ if\ k = 3 \Rightarrow cont = [E;F] \end{cases} \end{aligned} \tag{8}$$

The constants c3 and c4 figuring in the equation have the same function as the already known c1 and c2 constants; their introduction into the equation is necessary because of the asymmetric setting options which reflect the asymmetry of the cylinder.

The behaviour of the system is calculable and compared to the examples in the specialized literature the command signal operates at a lower frequency; thus, due to the smaller amount of gas let into the environment the efficiency of the overall system is increased. It is worth noting that the system is capable of adjusting itself to the reference signal even if per chance we have chosen too high parameters at the control settings. In this case the balance of force required to stop the cylinder piston sets in earlier than necessary, but since we only fill one of the chambers (see bands number 2 and 4) in the other chamber the pressure will slowly decrease, which is a consequence of the ever-present cylinder loss. The force arising from this pressure difference will always slowly move the piston to its ideal position. This is advantageous because this way, as far as the positioning systems most significant quality factor, the steady-state error is concerned, we can say that our system operates without predefined working-positions. Another great advantage is that in spite of the low frequency and delays of the solenoid valves we are able to make the piston stop with a high accuracy thanks to the slow movement around the reference signal.

4. Apparatus

The circuit diagram of the pneumatic positioning system is presented on figure number 5. As an actuator we applied a Festo DSNU-20-100-PPV-A P606 cylinder of 100mm stroke length, to which we attached a Festo MLO-POT-225-LWG analogue displacement encoder, which has a 0,01 mm travel resolution. The applied encoder is a potentiometer which provides a voltage signal in

proportion to the displacement. In order to move the cylinder we applied two Festo VSVA-B-P53C-H-A2-1R2L 5/3 way solenoid valves, but we only used one output connection each and the remaining output ports were plugged. We measured the mentioned solenoid valve's ON and OFF switching time at 6 bar supply pressure; in case of switching on it was 14 ms (Figure 3.) while at switching off 36 ms (Figure 4.).

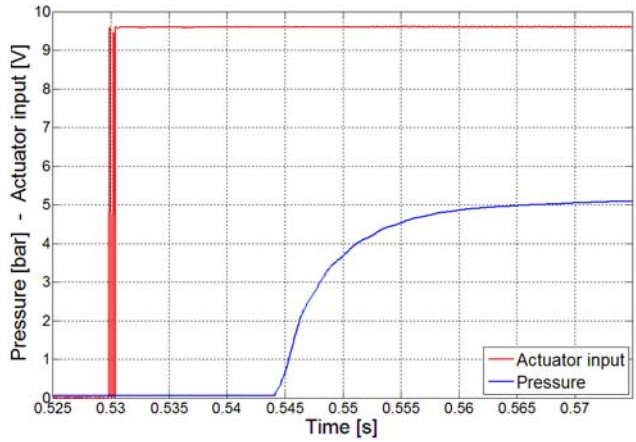


Figure 3. Valve switching – On

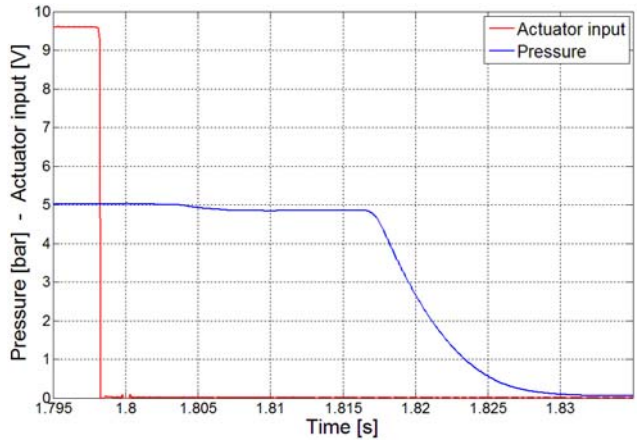


Figure 4. Valve switching - Off

A further constituent is a Festo D:LR-1/8-0-MINI pressure regulator, and we also connected a Festo SDE1-D10-G2-H18-C-PU-M8 pressure sensor to both chambers to serve as feedback, which however we did not use in the control process in order to minimize the number of sensors necessary for the operation of the system. This system was constructed to test the planned control method

but by changing the different elements of the scheme, it can be freely scaled to achieve a faster operation or the movement of heavier loads.

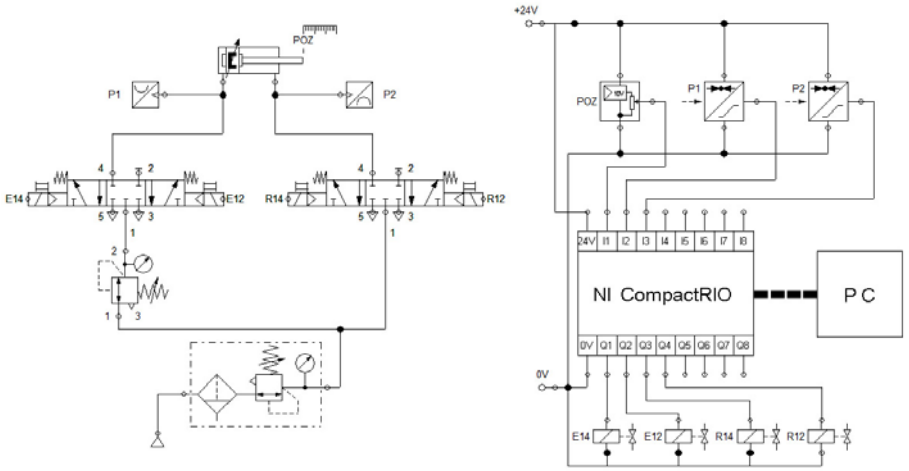


Figure 5. The circuit diagram of the pneumatic positioning system

The major elements of the electronic system are a 0-24 V direct current power supply (NI PS-15), an electronic instrument board (Festo), a NI CompactRIO™ (cRIO 9073) programmable automation controller and the already mentioned electro-pneumatic elements (displacement encoder, pressure sensors and solenoid valves). The applied NI CompactRIO™ programmable automation controller is a modular system; out of its modules we used the analogue-to-digital converter (NI 9201), for a dual purpose. On the one hand we applied it in the controlling process to measure the voltage signal (which is in proportion to the displacement) provided by the displacement encoder. On the other hand we used it in collecting data about the voltage signals corresponding to pressure values (expressed in bars) provided by the analogue pressure sensors. We controlled the two solenoid valves with the help of the digital output module (NI 9472). The communication between the CompactRIO™ and the computer was ensured by an Ethernet connection. We realised the real-time control based on equation number (8) by applying the FPGA module of CompactRIO™ programming it in the LabVIEW 2009 software.

5. Measurement results

The testing of the compiled system was done by determining the quality factors of the control method. During this process we have determined the settling time for step responses from end positions, overshoots and steady-state error graphically based on measurement results. Under settling time we conventionally understand

the time required for the measured output to finally reach the $\pm$ 5% vicinity of the reference signal.

The moved load was m=0.542 kg, the value of supply pressure in the case of the negative chamber was p2=6·105 Pa, and accordingly the decreased supply pressure of the positive chamber was $p_1 = \frac{A_2}{A_1} \cdot p_2 = 5{,}04 \cdot 10^5\ Pa$. The measurements were carried out at room temperature.

The control setting parameters were c1=18 [mm], c2=950 [-], c3=18 [mm], c4=250 [-]; based on previous experiences with the system the width of the tolerance band was set to be $\pm$ 0.025 mm, we regarded the position as adequate within this range.

First we tested the dislocations of the cylinder piston in the positive direction, i.e. when the piston was moving outwards of the cylinder.

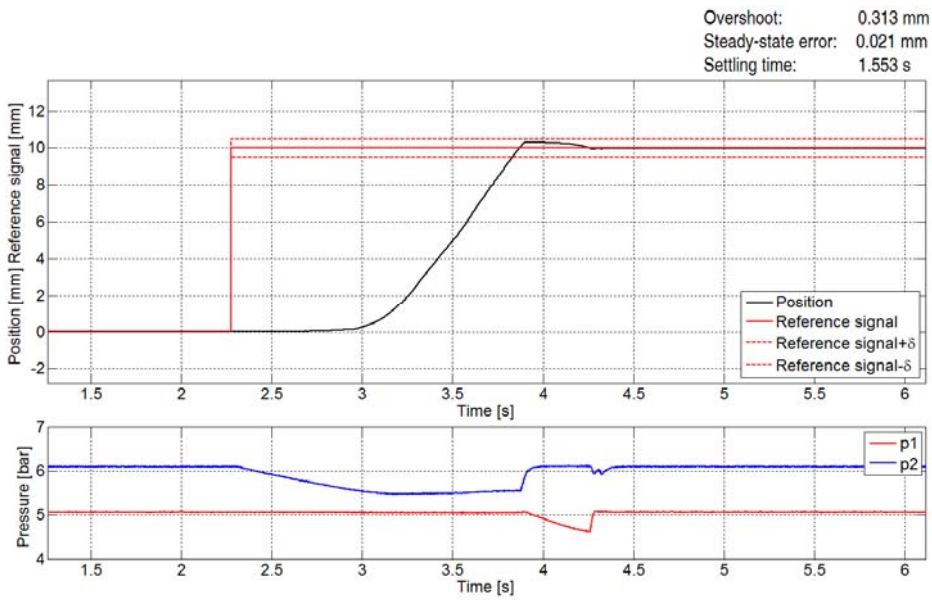


Figure 6. Step response - 0-10 mm

In Figure 6. we examined a settling onto the 10 mm reference signal from initial position, which is 10% of the stroke length of the cylinder. At such a small change of position the control system works in the deceleration zone (2) all the way through, i.e. the speed of the displacement of the piston is only dependent upon the cylinder loss. It is visible that the settling time is under 2 seconds, the overshoot can be considered minor. The steady-state error is greater than the resolution of the displacement encoder, but it nevertheless remained within the tolerance band and is still better than most of the best values to be found in the specialised literature.

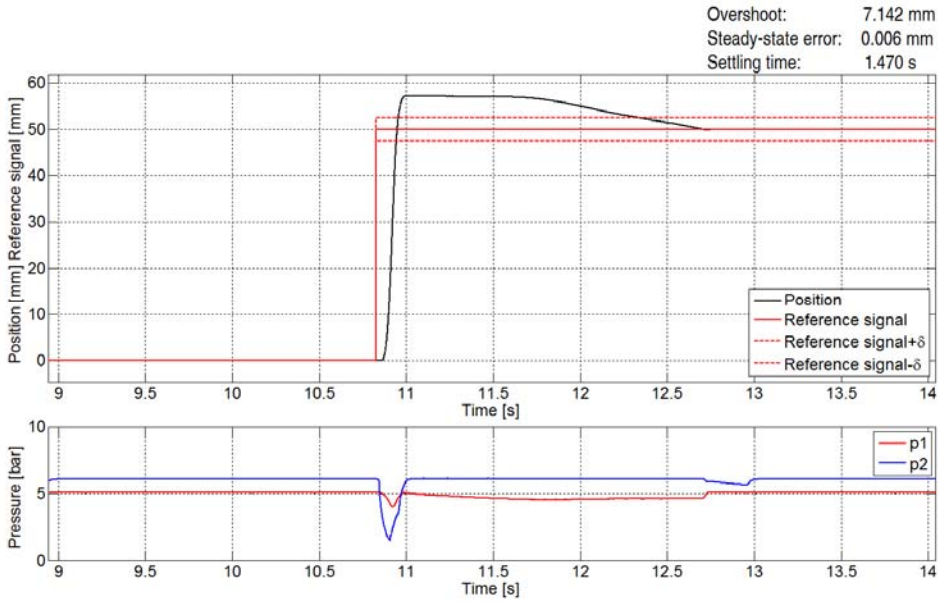


Figure 7. Step response - 0-50 mm

The evaluation of the second test (Figure 7.) sheds light on the limits of our positioning system, given that in the case of the settling to the 50 mm reference signal a significant overshoot was observed. This is the consequence of the fact that the chamber filling at the adjusted supply pressure is fast and because of this the acceleration of the low-friction cylinder piston is too great to be followed by the control system. This is the consequence of the operating limits of valves with long switching time. The deceleration of the piston movement, i.e. the closing of the ventilated chamber begins later than would be ideal, which is why the significant overshoot and the same settling time appear. It is important to be noted though, that all these do not influence the positioning accuracy, i.e. the steady-state error is equivalent to the travel resolution of the displacement encoder.

At the third measurement (Figure 8.) we have examined a displacement which is long compared to the stroke length of the cylinder by setting the reference signal at 90 mm. It is visible that in the case of large step size the control method is able to follow the dynamics of the cylinder, the overshoot is minimal while the settling time is 1 second. The steady-state error is again equivalent to the travel resolution of the displacement encoder.

After this we examined the movement of the piston in the negative direction, namely when it moves backwards into the cylinder.

In the course of the fourth test at Figure 9. we can see a displacement similarly great to the previous experiment, but in the negative direction. The overshoot is minimal, and the settling time is still under 2 seconds. The steady-state error is once again equivalent to the resolution of the displacement encoder.

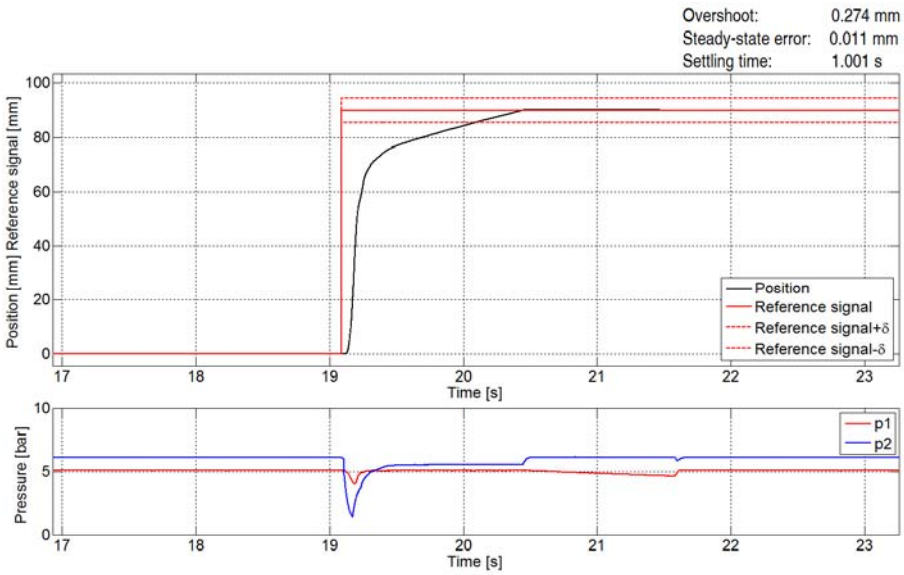


Figure 8. Step response - 0-90 mm

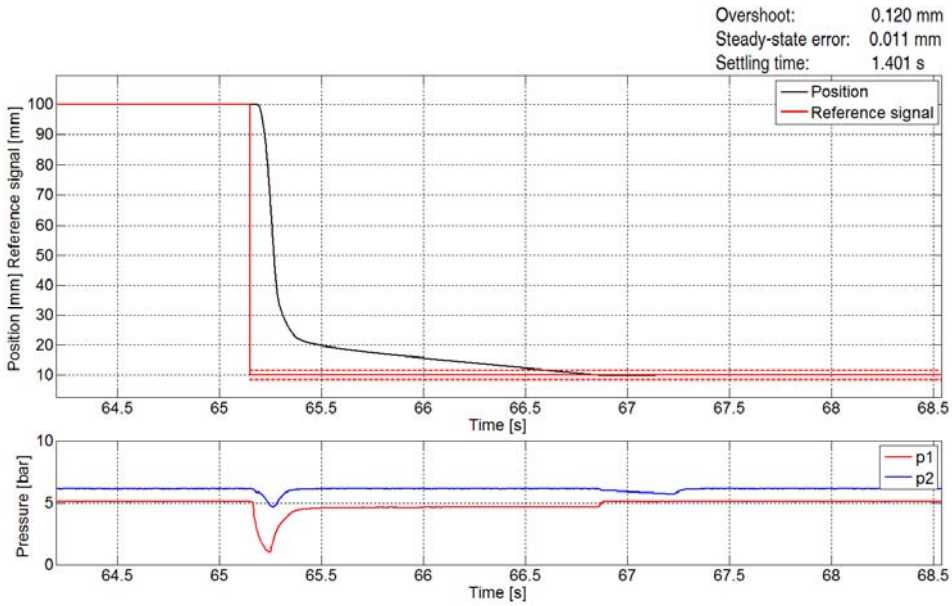


Figure 9. Step response - 100-10 mm

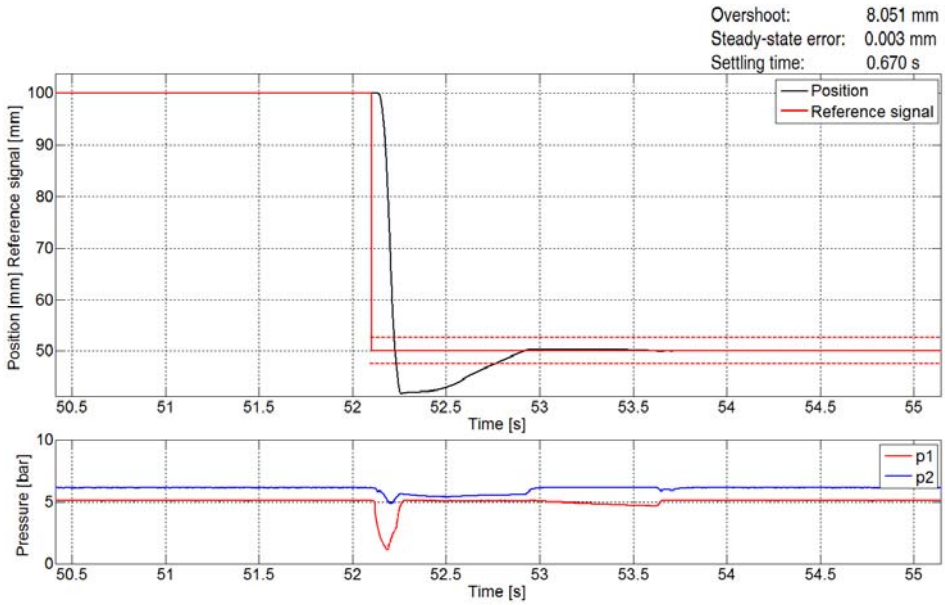


Figure 10. Step response - 100-50 mm

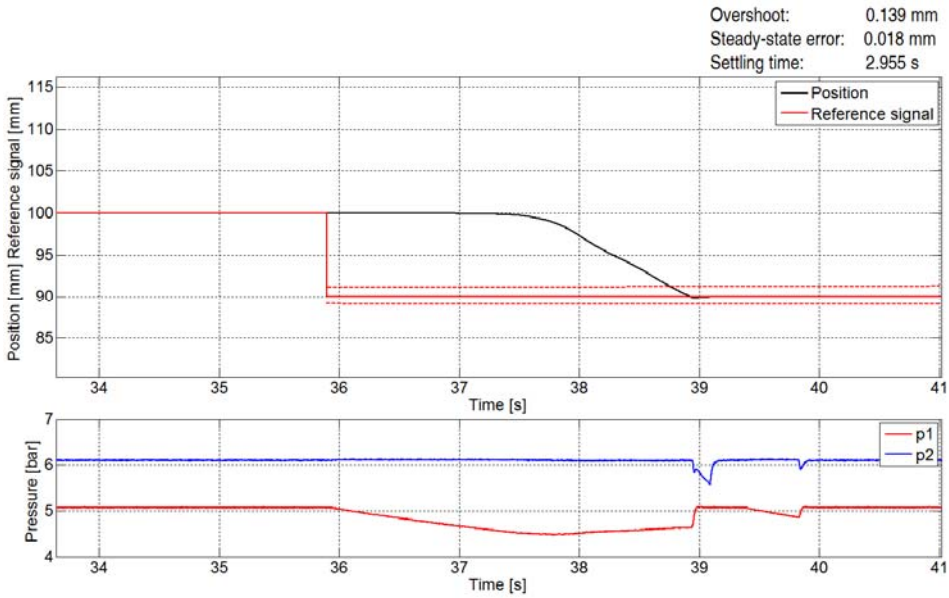


Figure 11. Step response - 100-90 mm

The fifth measurement (Figure 10.) examined a movement in the negative direction the scale of which is 50% of the stroke length of the piston. It is visible that an overshoot similar to that of the second testing appears. The steady-state error is minor, and in this case the settling time is exceptionally good, which is partly due to the fact that the piston moves faster in the negative direction because of the asymmetric construction of the cylinder.

In the last, sixth testing (Figure 11.) we observed the small-scale negative movement with a displacement from 100 mm to 90 mm. The movement of the piston remains in the deceleration band (4) all the way through, thus the speed of displacement is only dependent upon the cylinder loss. As a result, the settling time is significant, but the overshoot is minimal and the steady-state error is almost equivalent to the travel resolution of the displacement encoder.

6. Summary

A novel control strategy and the according experimental apparatus to achieve accurate positioning of a pneumatic cylinder using solenoid valves is presented. The most significant features of this pneumatic positioning system are the following:

- it substitutes the costly proportional valve with the conventional solenoid valve
- thanks to a novel control design, the system operates in a chatter free way
- the maximal operating velocity and force of the applied pneumatic actuator is not decreased
- it contains the least amount of sensors and the least expensive electro-pneumatic elements possible

Thanks to all these, the system can achieve an adequately high positioning accuracy and reach a favourable price/value ratio at the same time. The paper also sheds light on the fact that the system's steady-sate error is highly dependent on the displacement encoder's travel resolution. This holds out the promise that the application of more advanced technology in the area (e.g. using digital sensor with higher travel resolution, or a fast-switching solenoid valve to reduce the overshoot) will further improve the system's positioning accuracy.

References

[1] Ahn, K., Yokota, S., Intelligent switching control of pneumatic actuator using on/off solenoid valves, Mechatronics,15, 683–702, 2005.

[2] Akdağ, F.N., Kuzucu, A., Highly accurate pneumatic position control, Istanbul Technical University Mechanical Engineering Department, http://digital.ni.com/

[3] Barth, E.J., Zhang, J., Goldfarb, M., Control Design for Relative Stability in a PWM-Controlled Pneumatic System, Journal of Dynamic Systems, Measurement, and Control, 125, 504-508, 2003.

[4] Isermann, R., Mechatronic systems : Fundamentals, Springer, London, 2005.

[5] Messina, A., Giannoccaro, N.I., Gentile, A., Experimenting and modelling the dynamics of pneumatic actuators controlled by the pulse width modulation (PWM) technique, Mechatronics, 15, 859–881, 2005.

[6] Nguyen, T., Leavitt, J., Jabbari, F., Bobrow, J.E., Accurate Sliding-Mode Control of Pneumatic Systems Using Low-Cost Solenoid Valves, IEEE/ASME Transactions on mechatronics, 12(2), 216-219, 2007.

[7] Parnichkun, M., Ngaecharoenkul, C., Kinematics control of a pneumatic system by hybrid fuzzy PID, Mechatronics, 11, 1001-1023, 2001.

[8] Shih, M.-C., Ma, M.-A., Position control of a pneumatic cylinder using fuzzy PWM control method, Mechatronics, 8, 241-253, 1998.

[9] Thomas, M.B., Maul, G.P., Jayawiyanto, E., A Novel, Low-Cost Pneumatic Positioning System, Journal of Manufacturing Systems, 24(4), 377-387, 2005.

[10] van Varseveld, R.B., Bone, G.M., Accurate position control of a pneumatic actuator using on/off solenoid valves, IEEE/ASME Transactions on Mechatronics, 2(3), 195-204, 1997.

Abrasive wear tests of different hot-dip galvanized surfaces

László SZABADI, Gábor KALÁCSKA, Lajos PÉK
Department of Maintenance of Machinery,
Institute for Mechanical Engineering Technology

Abstract

More and more steel constructions are provided with zinc coatings as durable protection against corrosion. Hot-dip galvanizing process is widely used in Europe considering its favourable characteristics. Lately beside the corrosion resistance demand of hot-dip galvanized coatings as a new requirement is the abrasive wear resistance. The industrial floor boards, agricultural walking grids get heavy abrasion effect. The abrasive wear resistance of zinc coatings with multilayer structure is not tested yet, less known domain.

Keywords

hot-dip galvanized, diffusion, abrasive wear, gradient material structure.

1. Introduction, application fields of coating hot-dip galvanized, their producing and the coating characteristics

The hot-dip galvanizing technology is used for durable resistance against of surfaces of iron- and steel constructions and of goods sold by the piece, its protection effect depends decisively on the thickness of a layer [1]. Its main application fields suggested is the surface protection against atmospheric load in all corrosivity grades as well as to protect the metal structures of indoor ventilated spaces. Its application expands continuously because of its reliable protective effect, the process industrial-scale applicability, its comparatively high productivity and low need for human labor demand. Continually increasing proportion of steel constructions manufactured in Europe-yearly more than 6 million ton-are covered by coatings hot-dip galvanized. Nowadays the products get also mechanical loads beside corrosion effects at industrial filters, industrial, agricultural, public square pavement grids hot-dip galvanized meaning new and at the same time expanding application fields. At surfaces hot-dip galvanized exposed to abrasive wear, sand and breakstone spreading there is an application demand for today wear- and friction resistant coatings. Comparing test data regulated relating to wear resistance there are not available neither in technical literature nor in the data-base of companies producing and developing coatings. These data are indispensable to develop technology improving wear resistance.

Two basic groups developed hot-dip galvanizing [2]. The most characteristic phase of individual processes are the dip into the metal bath exists at each group, however there are significant differences between individual solutions considering the preceding surface preparation and subsequent treatments. There is no difference between the two process groups considering the zinc-layer forming which takes place by the same physical-chemical processes. They so called continuous technologies belong to the first group, during which the product to be galvanized „as spliced" continuously with suitable speed passes through on the technological system.

The continuous broad strip-, narrow strip- and wire galvanizing processes have similar methods essentially. The strip surfaces are oxide relieved in a closed technological system, their materials are heated up then are guided into zinc-bath where the coat develops.

The so called periodic technologies belong to the second group. At there the products get treatment by individually or in groups. The surface is here also pre-treated then this is followed with hot-dip galvanizing. Covering steel constructions belong to this group, too.

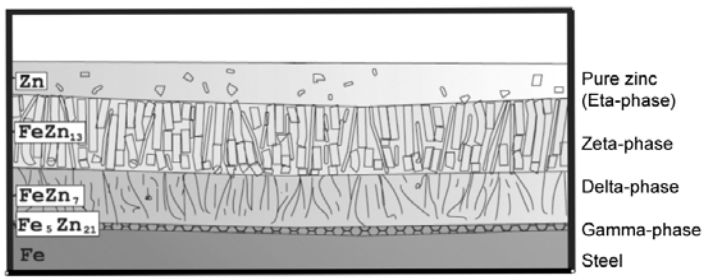


Figure 1. The zinc layer structure and the individual phases.

The pure zinc is readily-formed and soft metal. The Zn hardness- 55-70HBS [3] is far below of the mild steel hardness -120HBS-, thus its wear resistance is also significantly smaller. In case of alloyed steels this difference can be still significantly greater. The structure of zinc-layer is determined decisively by the chemical composition and texture of steel base-metal, the zinc-bath temperature, the dip duration, the zinc-bath alloying elements, the surface condition and thickness of the workpiece as well drawing out speed and the lifting method of the workpiece to be coated [4]. During dip-hot galvanizing an intermetallic, multi phase zinc layer develops on the products surfaces. The last step of technological treatment is the galvanizing when at about 450°C temperature the zinc atoms diffuse into the clean metal surface and produce atomic (cohesion) bond infiltrating into the metal-crystalline, namely they form an alloy with iron, on the zinc-iron boundary surface zinc rich Fe-crystallines develop. This forms essentially the „adhesive-bridge" between the zinc coating and the iron plate. The

further alloy layers are based on this which iron content decreases progressively to the outer layer and in case of optimal steel quality also cease to exist. The coating thickness is about 80-100 μm. Figure 1. shows the zinc layer structure.

The gradient material structure developed during diffusion is not homogeneous, its composition changes in the function of thickness.

2. Method and results

2.1 Specimens tested

We have chosen S235JRG2 steel as specimen material used for abrasive tests, which is the most definite base metal for example of pavement grid. As this material is desoxidized with aluminium- and not with silicium – and the mechanism of layer developing is determined first of all by the amount of silicium to be in steel, therefore it can be hot-dip galvanizing outstandingly. We have taken into account at deciding the specimens dimension the tool form of the abrasion tester, as well as those positions and the geometrical dimension of the container containing the abrasive medium ensuring abrasion. Figure 2 shows the dimension of specimen galvanized.

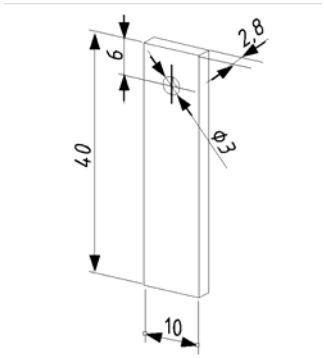


Figure 2. Specimen dimension.

Two types of coating were tested in the function of friction length, speed and pressure of abrasive medium.

- Technigalva:
  The determining components are the zinc (Zn), aluminium (Al), lead (Pb) and nickel (Ni) of the zinc-bath. The coating was made by dry periodic technology.
- Technigalva heat treated:
  The specimens were heat treated in order to the coating should have zinc-iron alloy phases in the total cross-section.

The abrasion resistance of coatings depends on its hardness, therefore it is needed to measure the micro-hardness. The micro-hardness tests showed that the

Techigalva coatings had 48HVM in average, while the heat treated coating had 106 HVM in average.The testing points on the specimen's surface can be seen in figure 3. we have selected 10 – 10 test speciemens and after measuring each ones the registered hardness values were averaged that are summarized in tabe 1.

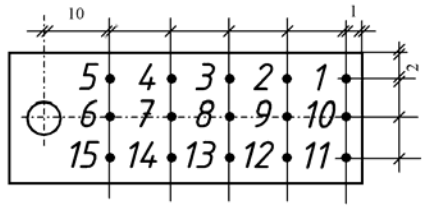


Figure 3. The hardness measuring points on the specimen's surface

Table 1. the measured micro Vickers hardness values

Measuring point	**„Technigalva" sample [HV M]**	**„Heat treated Technigalva" sample [HV M]**
1	48,7	98,2
2	50,7	97,3
3	45,9	114,2
4	46,8	92,5
5	49,7	96,3
6	45,9	89,5
7	51,8	101,4
8	49,7	105,4
9	45,9	108,2
10	54,0	112,2
11	45,9	112,2
12	42,5	103,5
13	41,7	96,3
14	43,3	91,2
15	44,7	91,2

Figure 4. and 5. show the SEM-pictures made from coatings, introducing the distinguished layers and EDS sampling areas. Certain layers, phases can be separated very good in the thousand-fold magnification. We have also carried out EDS (Electron Detector System) tests.

EDS spectra result are summarized in table 2., the measured graphs can be seen in Figure 6 –

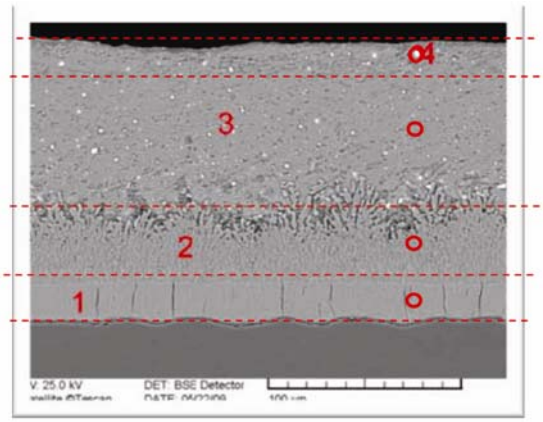


Figure 4. 1000X-magnification of Technigalva specimen zinc coating, (numbered layers with indicated EDS zone)

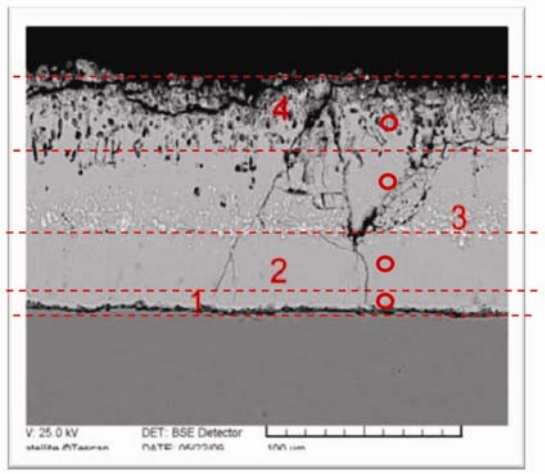


Figure 5. 1000X magnification of specimen heat treated zinc coating, (numbered layers with indicated EDS zone)

Table 2. EDS spectra result of Fe, Zn and Pb values (%)

Technigalva „T”				**Heat treated Technigalva „H”**		
Layer	**Fe (%)**	**Zn (%)**	**Pb (%)**	**Fe (%)**	**Zn (%)**	**Pb (%)**
1	16,4	83,6	-	47,5	52,5	-
2	10,1	89,9	-	16,3	87,7	-
3	-	100	-	19,6	79,6	0,8
4	-	21,2	78,8	17,9	82,1	-

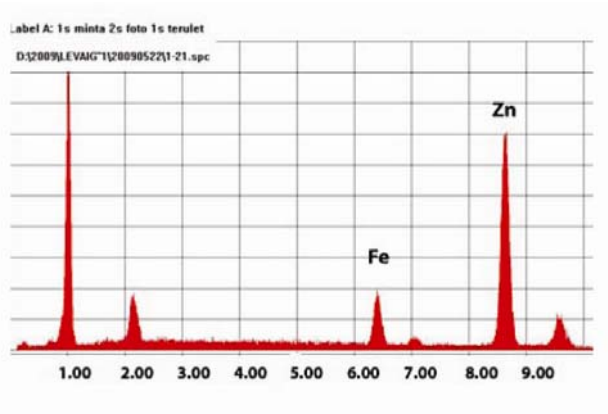


Figure 6. EDS spectra of Technivalva layer, phase Gamma (1)

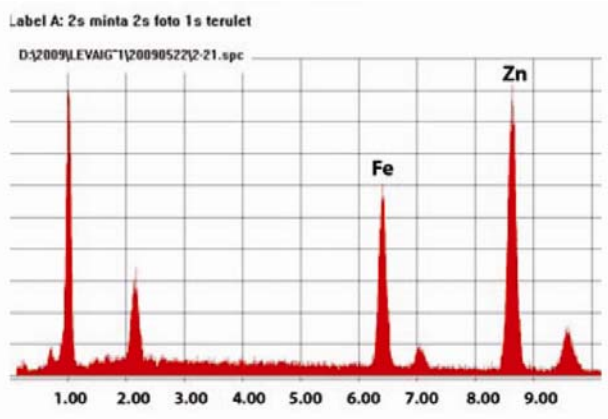


Figure 7. EDS spectra of heat treated Technivalva layer, phase Gamma (1)

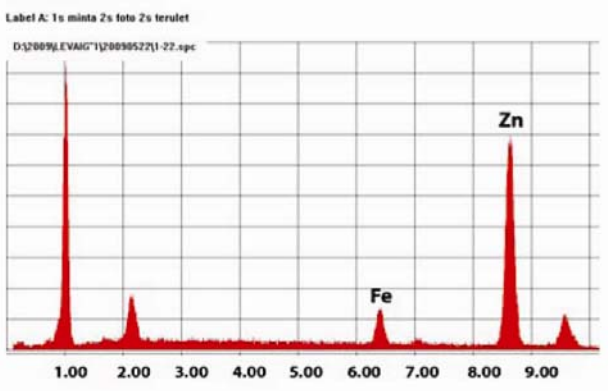


Figure 8. EDS spectra of Technivalva layer, phase Delta (2)

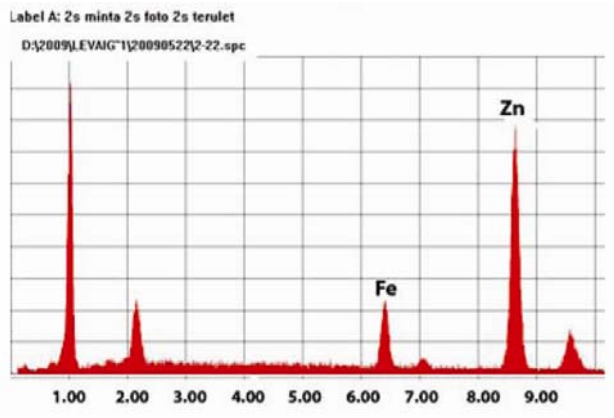


Figure 9. EDS spectra of heat treated Technivalva layer, phase Delta (2)

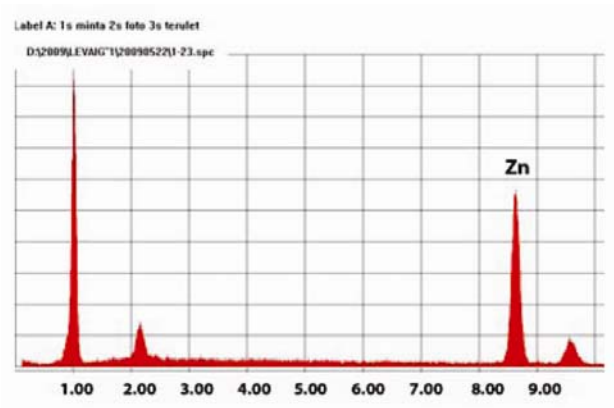


Figure 10. EDS spectra of Technivalva layer, phase Zeta (3)

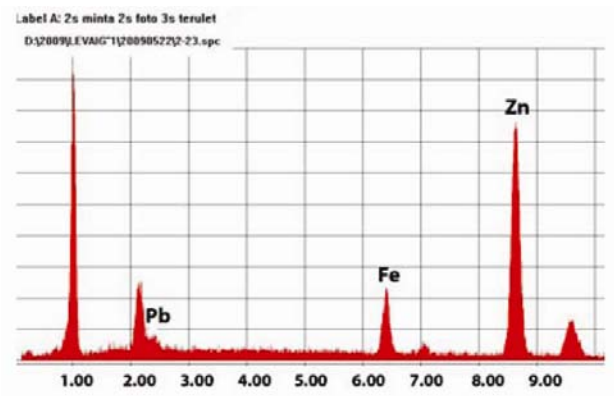


Figure 11. EDS spectra of heat treated Technivalva layer, phase Zeta (3)

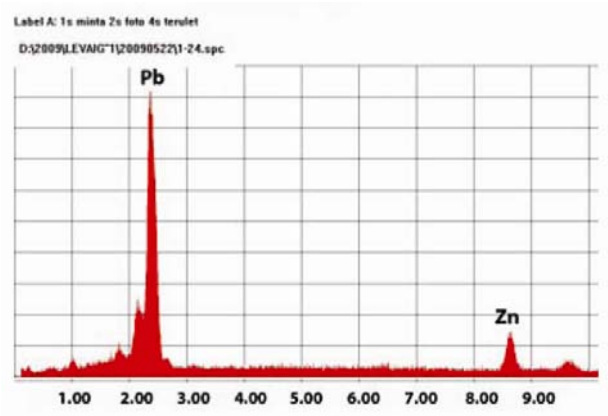


Figure 12. EDS spectra of Technivalva layer, phase Eta (4)

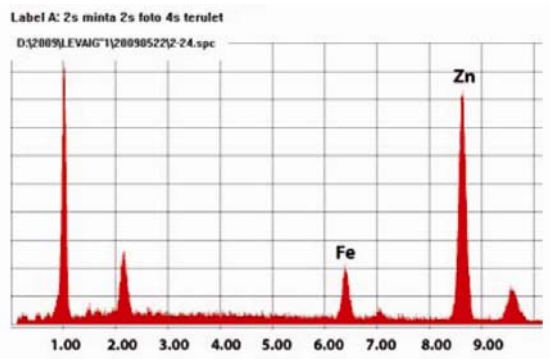


Figure 13. EDS spectra of heat treated Technivalva layer, phase Eta (4)

2.2 Abrasion test with „sand-slurry" equipment

The „sand-slurry" principle is well known in the VI. test category of tribological modeling. Several versions are wide-spread but they agree that for example in a sand with grain composition given - as in abrasive medium – a specimen moves in circular orbit generally with speed given. Great number of specimens can be measured at the same time in abrasive medium given as well as it can be well define but beside in different conditions with the abrasion tester developed in the Institute for Mechanical Engineering Technology. The results got made possible the evaluation according to various standpoints, too. Figure 14. shows the abrasion tester developed.

The electric motor shaft connects to a worm-gear which gear transmission in 22,58. The working shaft can be found at the exit side of the worm-gear on which 3 pcs. arm cross-clamps can be found – in different heights related to the base plate (Figure 15.) The specimens to be abraded can be fixed in suitable position on this. To one tool, to each arm 3 pcs. altogether 6-6 pcs. specimens can be fixed. Important characteristic of the cross-clamps is that the specimens

can be fixed with each other in 90^0 included position, at their sides in pairs altogether 6 various positions related to the centre of gyration. The tools are set turned away to one another on the working shaft in top view the circle is divided to 30^0 sectors. The container containing the abrasive medium can be put into an outer container in case of demand, which can be filled with cooling-heating water to the thermal dynamics of measuring procedure can be regulated.

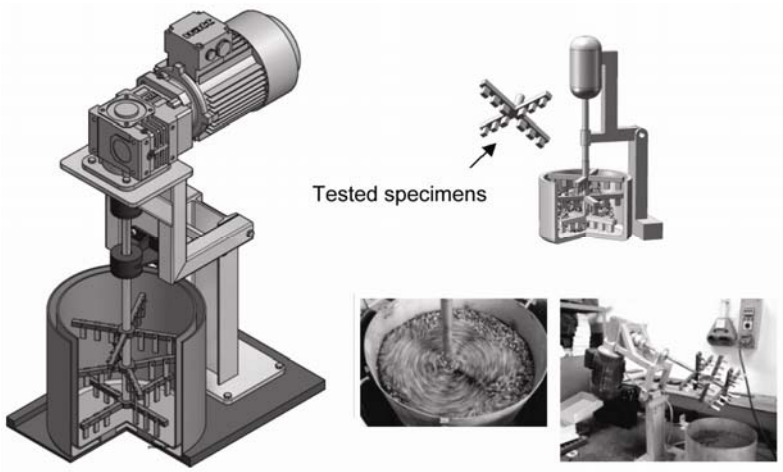


Figure 14. The abrasion with specimens mounted.

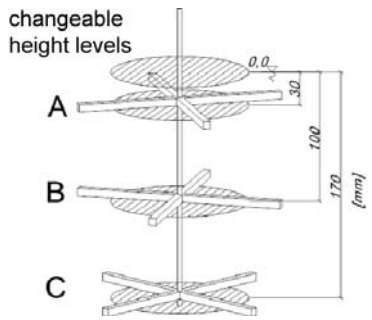


Figure 15. Specimens placed on three levels.

The abrasive medium was 0/8 OK – type ballast stone. Its average aggregation density in dry condition is 1,7 ton/m^3. There is no practically clay-sludge content as it is produced from washed, granulated gravels by knapping. The grain fraction is between 2 and 8 mm.

The specimens to be in different radiuses move with various different peripheral speed in the abrasive medium and the height position results different surface pressure rations (Figure 16). The abrasion tester makes possible exceptional complex evaluation in the function of these variables.

2.3 The abrasion and speed connection

We have measured the abrasion of the 6 pcs. heat treated and 6 pcs. not heat treated specimens placed in all three levels (A,B,C) after seven various abrasion time. The surface pressures in standing (not rotating) condition:

p_1 (A level) = 153,83 Pa
p_2 (B level) = 521,82 Pa
p_3 (C level) = 902,41 Pa
Testing speed range: 14 – 40 m/min

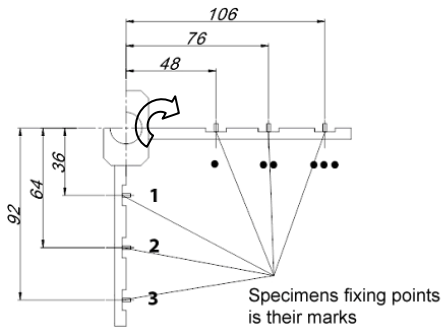


Figure 16. Position of specimens on different radiuses.

The specimens covered different length with different peripheral speed moving on various radiuses. After averaging the abrasion results measured at each specimen, repeating three times the measuring series the same incline could be seen at the lines to be adapteable to the plot at heat treated and not heat treated specimens, too. Based on these the speed independence supposed were proved by mathematical – statistical methods, by covariance analyses at all three levels at heat treated and not heat treated specimens, too. The abrasion values of all specimens to be in the different levels can be presented with a single regression straight line.

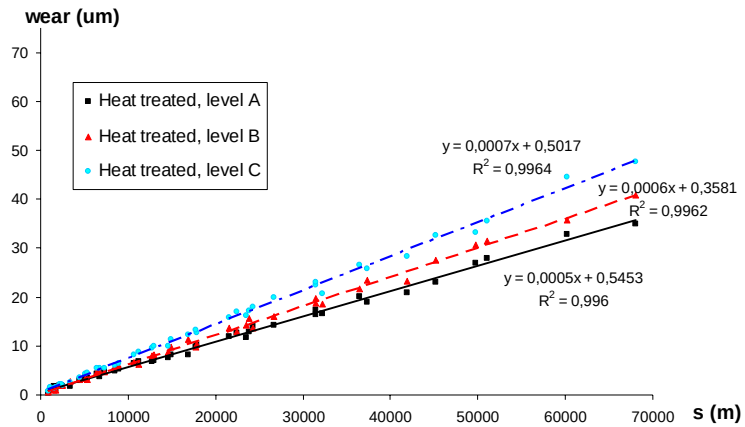


Figure 17. Abrasion values of heat treated specimens on "A", "B" and "C" level (h_1, h_2 and h_3 depth)

This means that the abrasion values do not depend on the abrasive speed in the speed domain tested (Figure 17. and 18). There is no significant difference between the specimens moved with various speeds but placed at the same level.

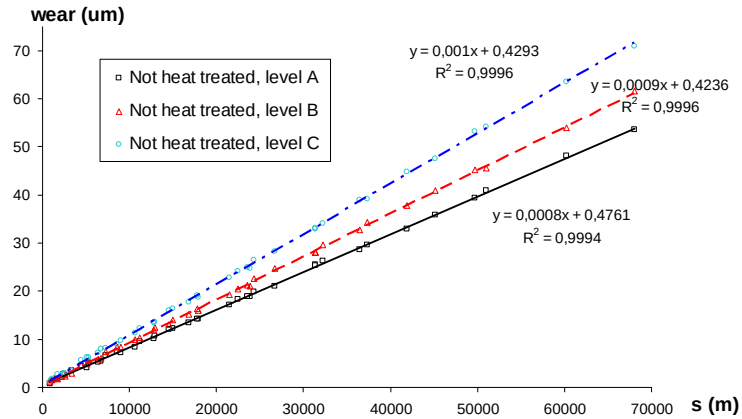


Figure 18. Not heat treated specimens abrasion values on "A", "B" and "C" levels.

2.4 Connection between surface load and abrasion

The specimens placed on the "A", "B" and "C" level get different surface loads because of this the abrasion values measured on "A", "B" and "C" levels has to be compared. According to the hypothesis the abrasion values depend on the load. It can be approximated with linear trend-line the appropriate data to different levels on the diagrams, the matching is close in each case. The data appropriate to different levels can be separated visibly however if there is significant difference between them it has to be examined by regression analysis. The worst case is where there is the smallest difference: this is the "A" and "B" level data of the specimens heat treated. The calculations carried out proved that there is significant difference between the abrasion values of specimens fixed on "A" and "B" level. This means that the amount of surface load has significant effect on the abrasion values in the system tested. In case of higher load the specimens have got higher abrasion.

2.5 Comparing the layer structure of coating and abrasion values

The abrasion measured as a resulting effect on the surface of specimens means the continuous decreasing of coatings with layer structure. In Figure 19. and 20. can be seen how changes the percentage rate of decisive chemical elements in certain layers of heat treated specimens. The chemical composition's changing does not influence the abrasion intensity. According to data tested by EDS spectroscopy the percentage rate of chemical elements in the layers of not heat treated specimens is formed otherwise, but this compound does not influence the abrasion intensity of certain layers.

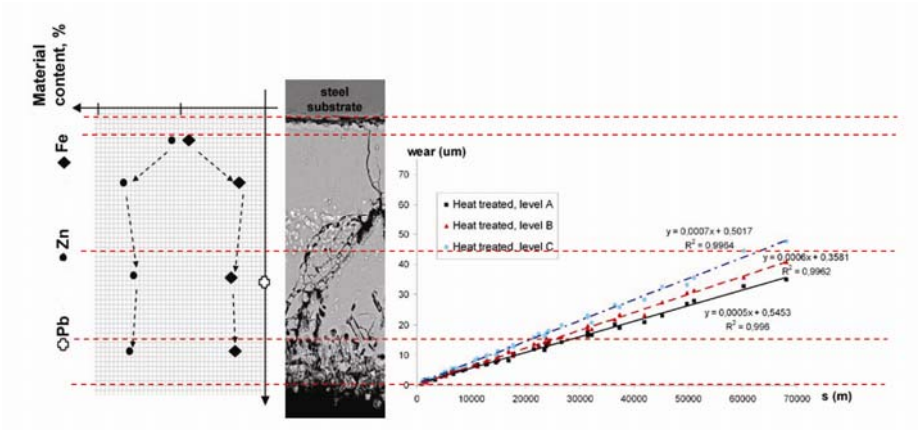


Figure 19. Heat treated specimens abrasion and the connection of layer structure.

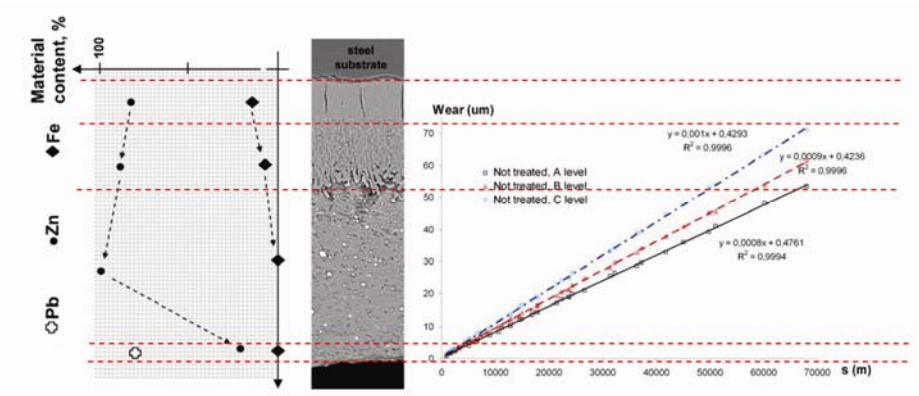


Figure 20. Not treated specimens abrasion and layer structure

3. Conclusion

- The multilayer coatings of hot-dip galvanized abrasion intensity in the test-system does not depend on the fiction speed neither at heat treated nor at not heat treated specimens, but it depends on the medium pressure and on the resistance of medium deriving from that.
- The heat treatment improves the abrasive resistance. The higher hardness results higher abrasive resistance improvement.
- SEM-pictures and EDS spectroscopy proved that heat treatment results different gradient layer structure and composition.
- The linear abrasion dynamics of hot-dip galvanized layers tested does not depend on that, which layer comes to friction connection with abrasive medium. The inner gradient friction character did not effect the abrasion intensity measured, but the resultant of different gradient structure has different abrasive resistance.

References

[1] Arenasa M. A., Damboreneaa J. Medranob J., A., J., García B. A., Rodríguez R. (2002): Corrosion behaviour of rare earth ion-implanted hot-dip galvanised steel. Surface and Coatings Technology, Volume 158-159, Page 615-619

[2] Magyar Tűzihorganyzók szövetsége: Szakmai Ismeretek, Dunaújváros, Dunatáj Kiadó Kft., Dunaújváros, 1997 (Association of Hungarian Hot Dip galvanizing Companies. Professional Knowledge. In Hungarian)

[3] Dr. Zorkóczi Béla: Metallográfia és anyagvizsgálat, Tankönyvkiadó Budapest, 1975. (Metallurgy and Material Testing. In Hungarian)

[4] Magyar Tűzihorganyzók szövetsége, Tűzihorganyzás, CD-kiadvány, 2008. (Association of Hungarian Hot Dip galvanizing Companies. Galvanizing Technologies. In Hungarian)

Institute for Engineering Management

Professor Dr. István HUSTI
Director of the Institute

Dear Reader,

There was a reform process in our Faculty last year, which has a result that the structure and the name of our Institute has changed to the Institute for Engineering Management from the previously called Institute for System's Engineering and Management. This process brought some changes not only among the research fields, and colleagues, but in the structure of the organization too.

The new Institute consists of three professional working areas called departments:

- Department of Applied Management
- Department of Engineering Economics
- Department of Material Handling and Logistics

The activity and the research fields of the Institute can be separated by these departments, and had the following main research areas during 2010:

Department of Applied Management:

- Quality management in the manufacturing processes.
- Adoption of product innovation and technology transfer. We analysed the diffusion of the innovation in the farm machinery manufacturing sector. We researched the influencing factors of the innovation.

Department of Engineering Economics

- Developing the system of agricultural machinery management. We analysed how to facilitate the operative management tasks based an integrated enterprise management system.

Department of Material Handling and Logistics

- Developing logistical processes by modelling and simulation.
- Research of material handling machines with special respect to conveyors.

The results of our activities were published not only in Hungary but worldwide in different papers.

The educational work is very important in our Institute, therefore we reviewed our books and notes for students and our list of the taught subjects was expanded by Engineering Economics and Industrial Marketing. Both courses are taught in English. The e-learning is spreading quickly and every year there is more demand for this way of learning. We developed several e-learning materials last year. In 2009 we announced the accreditation of our Engineering Manager degree on MSc level. This program started successfully last year.

More details about the Institute for Engineering Management are available at:
http://mumi.gek.szie.hu/

The Examination of the Innovation Activity of the National Agricultural Machinery Manufacturers

Árpád BAK, István HUSTI
Department of Applied Management
Institute for Engineering Management

Abstract

The current production of the Hungarian agricultural machinery manufacturing sector, which used to see better days, lags behind the production of the previous years to a great extent. Our study reveals the initial results of longer research work. We examine how the innovation activity of organisations changed during the period between 2007 and 2009.

The conclusions of our paper are based on the examination results of questionnaires and in-depth interviews that were carried out at 40 Hungarian agricultural machinery manufacturing companies. The characteristic features of the companies that were involved in the examination reflect the Hungarian conditions properly. Besides the brand-new or higly developed products and technological (procedure) innovations, novelties in organisation and marketing are also paid attention. Furthermore, some of the indicators of the innovation performance of the companies are also presented. The question of cooperation between the organisations taking part in innovation is referred to, as well.

The other results of the research are detailed in our further publications.

Key words

Innovation, agri-technical development

1. Introduction

It is an empirical fact that besides the financial constraint, structural troubles also prevent the national technical innovations from being successfully developed.

In Hungary the total expenditure on research and development reached 299.5 billion HUF in 2009, which was 1.14 percent of GDP (= GERD indicator).It means a 12.3 percent growth relative to the previous year at current prices (Statisztikai Tükör, 2010).

The share of the state of the GDP-related R&D expenditure is 0.42 percent of the total sum (state-owned research institutes and those in higher education altogether) while the expenditure of the business sector amounts to 0.58 percent of GDP. This proportion has been improving relative to the previous ones or, rather, approaches the international practice. However, approximately 60 percent

of the R&D expenditure of the national business sector is realised by exclusively foreign-owned enterprises or those which are mostly in foreign hands (KSH (Central Statistical Office), 2007). In most of the developed countries the national companies spend more of their revenues on R&D expenditure than the foreign-owned ones.

The high-level concentration of corporate R&D can also be observed: almost half of the expenditure derives from 17 big companies. The share of those employing fewer than 20 employees is only 12.6 percent.

Unfortunately, in the sector of the national small-and medium-sized enterprises not only research and development but also the number of licenses and know-how purchase is slight, so the demand (pull) side of innovation is weak under the present system of conditions.

- 72-74 percent of the segment is inactive regarding innovation or simply struggle to survive.
- 22-23 percent belongs to those catching-up, i.e. they show some initiatives in innovation that could prevail in the standard of their products and services by means of technological transfer, information and advisory institutions.
- Only 3-6 percent of companies make up the group of promising innovative companies (Losoncz, 2008).

According to our experience the above-mentioned facts can also hold true for agri-technical innovations more or less.

Before the change of the regime only 27 agricultural machinery plants operated mostly "embedded" in the system of the national "agri-business". Due to this fact (among others), 60 percent of the requirements for agricultural machinery in the country were covered by these plants at a more advanced standard than the average of the former Comecon countries. During the past 15-20 years the organisational structure of the Hungarian agricultural machinery production has totally been transformed. Generally, the machine manufacturers operating as small-or medium-sized enterprises appear on the market with „separate" products usually not developed by themselves. Consequently, they are not price-setters, rather price takers. The product line of the companies that are successful in the international competition primarily consists of mass-produced and highly automated products. The national agricultural machinery manufacturers-partly due to their size- are not able to mass-produce in such an extent that they could compete with the West-European, American and Asian companies of huge capital power either in productivity, price or product supply.

Regarding innovation, the national agricultural machinery manufacturers also significantly lag behind as they can only spend slight amounts on development relative to foreign-owned concerns. As a consequence, loss of market is not surprising as a bit more than one-quarter (26-27 percent) of the current total domestic market turnover comes from domestic manufacturers. The extent of market loss and the general situation of the national agricultural machinery manufacturers justify that the present of this sector must be dealt with by searching the ways-out of the crisis and make steps to develop.

In our paper first of all the method of the empirical research is presented where the structure of the questionnaire used in the research and the process of data record are shown in details. Due to the constraints of length, the results of the research based on monovariate descriptive statistics are published afterwards. The presentation of the results can be divided into four logical units: the characteristics of the sample subject to the research are shown with the most important indicators of innovation performance and finally the directions and sources of agri-technical innovations are analysed. Our study finishes with the conclusions drawn and marking the further directions of the examination.

2. Methods

Our examinations are mainly based on primary research. When formulating the research objectives, we relied on the theoretical conclusions drawn from the related specialist literature as well as the earlier publications of experts and empiric research results.

The basic objective of the research is to explore and analyse the innovation activity of the national agricultural machinery manufacturers, its results and influencing factors.

The more detailed questions and points of view of the examination that can be derived from the basic question of the research are the following:

- What are the main directions, assisting and hindering factors of the technical development activity of the companies?
- What are the dominant directions and bottlenecks of the agri-technical innovations?
- How can technical development be described compared to the research results of other sectors?

Finally our objective is to have a picture of the innovation activity of the organisations involved, the special features of innovations, the partners taking part in the processes and the impact of innovation on the general situation of the companies through our examinations. Besides the brand-new or significantly developed products and technological procedure innovations, organisational features, marketing activity and the environment of the innovation are also considered. The questionnaire serving as the basis of primary research embraces three years, from 2007 to 2009.

Question types used:

- According to the method of questioning:
  - close
  - open
  - filtering
- According to purpose:
  - shifting
  - checking
  - reminding

According to the estimations of experts the number of agricultural machinery manufacturing companies is between160-170 in Hungary. A great part of the enterprises are involved in more than one activity: a lot of predominantly small enterprises are also engaged in other activities besides machinery production so that is why it is difficult to define the actual number of 'agricultural machinery manufacturers' exactly. Most of the organisations that are subject to our analysis are small enterprises whose annual revenue does not reach one billion HUF.

As there was not an available list on all the companies on the basis of which a pattern of probability could have been compiled, the companies that could be drawn into the research had to be defined in another way. To find the companies necessary for carrying out the questionnaire, the address list of MEGOSZ (National Association of Agricultural Machinery Manufacturers) served as a basis and the heads of this professional organisation were also consulted. Of the altogether 60 members of MEGOSZ, 40 machinery manufacturers became part of the research.

In the data recording phase of the research a multi-channel approach was applied whose main points are:

- 15 machine manufacturers (hopefully the most significant companies of the sector) were questioned at personal in-depth interviews;
- 25 organisations were provided with a questionnaire sent by post requiring them to send back the material filled in;
- 9 of the 25 organisations were part of on-the spot discussion either because of the difficulties they faced while filling in the questionnaire (3 cases) or because the interpretation of the responses required further information (6 organisations).

Sample-taking cannot be regarded representative. However, during the research it was not our objective to draw conclusions that can be generalised for the basic population. Our basic objective was to give a thorough examination of innovation activity and to achieve it, we tried to select the organisations regarded to be suitable on the basis of preliminary professional considerations. As such a thorough examination dealing with the innovation activity of agricultural machinery manufacturers was not carried out in the past 25 years on a national level, we consider our research is to resolve discrepancies in the professional field.

In compliance with the general methodological requirements first of all some pilot questions were asked on the basis of which the questionnaire was finalised. Data recording took place between March 2010 and August 2010.

The duration of in-depth interviews was various, typically 90-100 minutes per interview. A positive feature of them was that data providers mainly come from the senior management (chief executive officer, head of production or technical manager). In this way first-hand information on the general situation, actual projects and strategic plans of the organisation involved was gained besides the reliability of data. The atmosphere of the interviews was typically of honesty and intimacy. Before starting the interviews we assured our interviewees of the confident nature of the research, i.e. their names and opinions were treated as

highly confidential and were not made recognisable for others. Some of our interviewees have already expressed their enquiry in our results.

The questionnaires compiled on the basis of the interviews and sent out by post were also accompanied by a guide to filling in. Seven-ten days after the questionnaires had been filled in, the enterprises were also contacted by phone asking them to cooperate. A kind of evaluation of our preliminary work is that all the responding organisations gave answers that could be assessed.

The statistical processing of data recorded by the questionnaires was carried out by using SPSS 13.0 programme.

3. Results

3.1 The presentation of the sample included in the examination

The breakdown of the concerned enterprises by size is presented by Figure 1. The types of enterprises based on the number of their employees were considered in accordance with Act XXXIV of 2004. The breakdown reflects that the SMEs, typical of the sector, are overrepresented. Of the responding enterprises, micro enterprises represent significantly less weight in comparison with the data of the sector. The reason for it is that it was not the objective of our research to analyse them in more details.

In our questionnaires we separately dealt with the analysis of the role of foreign-owned innovation. Twenty-one percent of the companies involved indicated that they had foreign owners or part of its shares in foreign hands. The responses reflected that at companies owned by foreigners in 100 percent the desire for innovation is rather pushed in the background and the tasks of the national subsidiaries are limited to the launch of the products developed by the parent company in their plans and to ensure the maximum cost efficiency in production.

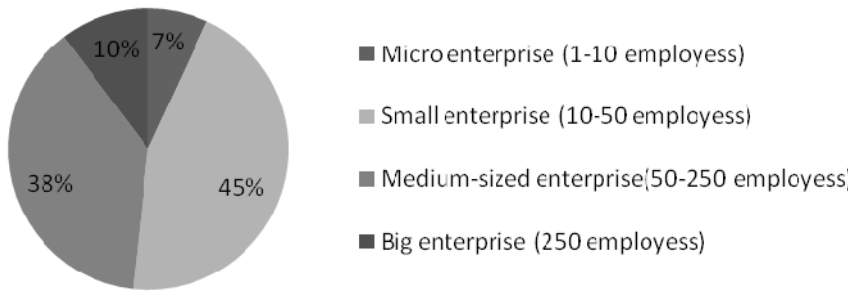


Figure 1. The breakdown of the enterprises concerned by size
(Source: own research results)

We also examined which markets the enterprises realise their revenue (Figure 2.) It was an outstanding fact that the enterprises concerned sold their products on the Hungarian market in 72 percent and during the past three years there has not been a considerable shift in this area. The exact numbers were completed by the in-depth interviews with the fact that one of the main intentions of the enterprises is to break successfully in foreign markets. However, most of them reported failures they faced. The reason for it derives from a lot of factors altogether: although their improved products have excellent value for money, they cannot compete with the local competitors mainly due to lack of a proper channel of distribution. Another problem is the resistance of the West-European farmers to less-known brands. The resources of the enterprises are depleted by demanding developments so they can hardly spend on marketing and PR activities, which would be essential in this sector.

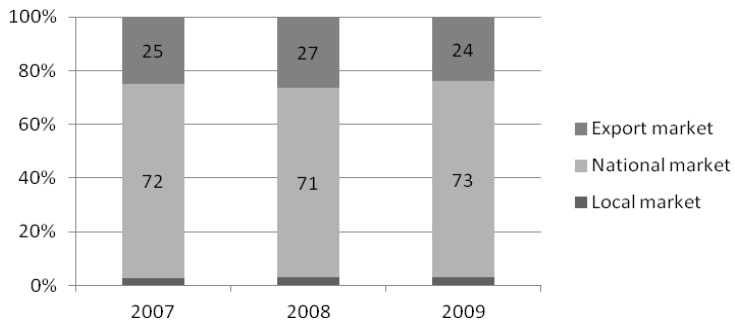


Figure 2. The breakdown of revenue by markets
(Source: own research results)

On the examined timeline 34 percent of the enterprises had a separate R&D department. The picture is made even more complicated by the fact that at the departments of such small-and medium-sized enterprises there are only one or two full-time employees to deal with the questions of innovation on the average compared to the big companies that can afford to operate R&D department with much bigger staff.

In case of the marketing departments, the result was the same as 30 percent of the examined organisations had such a section. In this department typically one person in full time, or rather, part-time is employed to deal with marketing issues. An interesting finding is that 48 percent of the enterprises operate in a linear while 38 percent of them in a simple organisational form, i.e. they do not have functionally separate organisational units.

3.2 The most important indicators of innovation activity

We examined how much of their revenues the enterprises spend on innovation, research-development and marketing (Figure 3). It is important to note that basic

and applied research as well as product-and production technological development were considered in the R&D activities while license-and know-how purchase, engineering activity or the introduction of a new marketing method or a new organisational solution into business practice, labour organisation or external relationships were regarded as innovation activities.

In our experience the examined enterprises have realised the necessity of innovation in the marked areas. In this respect revenues show an increasing tendency despite the fact that the global economic crisis had several drawbacks in the growth of most companies. The companies spend hardly more than one percent of their revenue on **marketing**. This can also be explained by the fact that this sector is a typically neglected one as marketing-and advertising costs are the part of expenditures that enterprises can save most of in time of crisis.

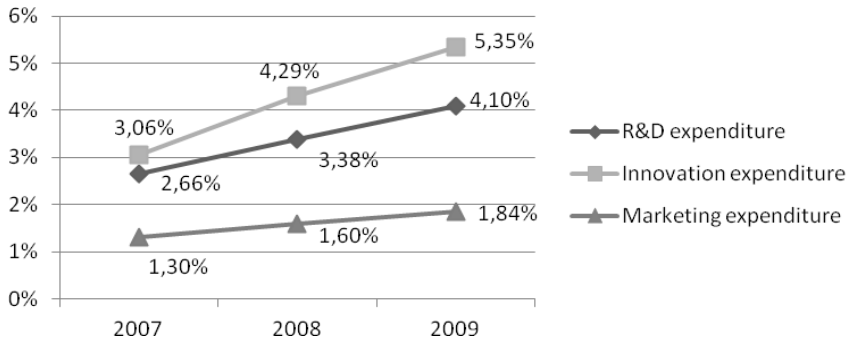


Figure 3. Expenditure on innovation, R&D and marketing
(Source: own research results)

3.3 Directions of agri-technical innovations

Based on our research 76 percent of he companies launched a new or a significantly modified product on the market during the examined period (2007-2009) and 69 percent carried out technological innovation.

The main directions of the R&D activities of companies are presented by Figure 4. The most determining direction of the R&D activity of agricultural machinery manufacturing companies is product development, which reflects the willingness to adapt to the continuously changing market needs. The common feature of the enterprises investing in innovation is their market proximity and user/customer-orientation. The sales opportunities of the manufacturers are determined by the fact to what extent they contribute to increasing the competitiveness of the farmers (enterprises) that make use of their new products. The improvement of existing products and constructions has a greater emphasis (71%) than the development of new products (66%). They are typically modifying-developing innovations that include the diversification of existing product lines with new products and the improvement of products, as well. The innovative features of the products are of wide range starting from the simplest solutions to the innovative products of great technical characteristic.

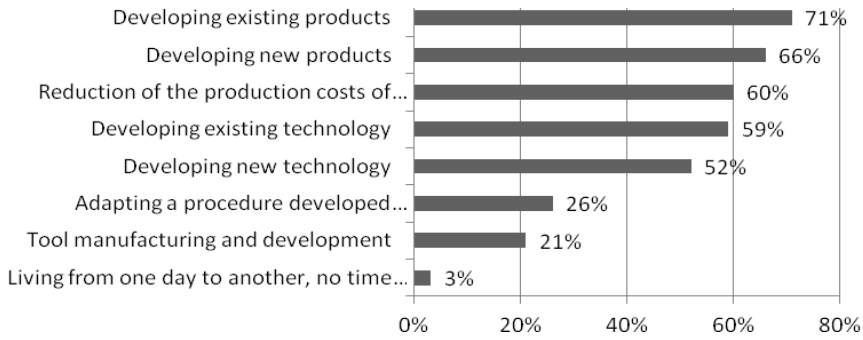


Figure 4. The main directions of the R&D activities of the companies
(Relative frequency, more than one answer could have been given)
(Source: own research results)

As far as technological innovation is regarded, the main objective of most of the national manufacturers is the reduction of the production costs of their existing products and catching up in terms of efficiency indicators (60%). Of the forms of R&D activity the development of new technologies is of slighter importance. The refinement of existing producing processes can also play an important role in the competition as it is less expensive and not so risky than the development of a radically new technology. At the same time, its positive effect could rapidly be traced back in increasing efficiency, decreasing energy consumption, refinement of processing s well as the better quality features of the products.

The adaptation of procedures developed somewhere else was marked by 26 percent of the companies as part of R&D activity. The fact echoed so many times proved to be true again in this respect, i.e. one of the bottlenecks of agri-innovation is adaptation. As a possible form of innovation it is rather underestimated in the sector.

3.4 The sources of agri-technical innovations

In research-development cooperation almost 80 percent of the companies concerned have already taken part in a form, which can be regarded a fairly good proportion. The influencing factors of R&D activity (Figure 5.) reflect the most decisive directions of knowledge flow. These results illustrate the demand-driven nature of innovations in the agricultural machinery sector. Of the different sources of technical knowledge, the companies prioritise getting to know the needs of the customers. The university research institutes were marked at the second place (64%) as the most important external partner in development. It can be regarded favourable as the gap between the national „science" and „industry" is traditionally deep (the so-called European paradox). It is in vain that Hungary (and principally Europe) produces balanced scientific performance (the number of PhD students, publication activity etc.) if the implementation of

innovation and scientific results in practice is weak and only of low standard. The sector seems to be willing to "buy" the results of the university sector as in their development practices the forms of partnership through which companies can join the scientific life of the profession have already been established.

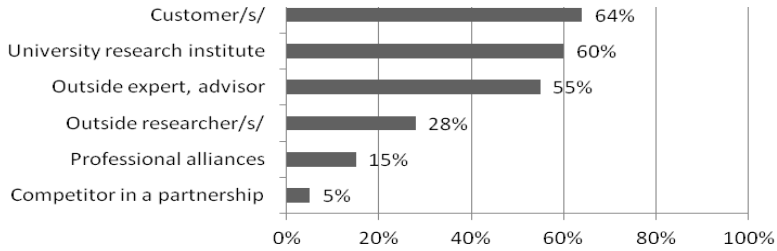


Figure 5. The proportion of R&D cooperation agreements of agricultural machinery manufacturing companies (Relative frequency, more than one answer could have been given) (Source: own research results)

It turns out from the results that the role of the different advisory organisations is low in development processes. The further inclusion of the intellectual capacity of experts from outside is made even more difficult by several factors. In the case of small-and medium-sized enterprises a significant hindering factor is the insufficiency of development resources so the rather expensive advisory companies seem to be beyond their reach (it can also explain why they turn to the university sector instead). Big companies employ professional staff and have a significant development capacity so they can cope with the problems arising on their own.

Only 15 percent of the responding companies stated that a kind of professional alliance had a role in their developments. Also, there are only few examples that competitors in the sector join their forces to achieve developments.

4. Conclusion

The characteristics of the Hungarian agricultural machinery manufacturers drawn in the examination illustrate the situation of the sector in the country. A decisive part of the organisations (83%) are small-and medium-sized enterprises. Their attitude in development is reflected by the fact that more than 70 percent of their products are sold on the domestic market. As a result of their intention to increase export, one-quarter of their products are launched on foreign markets. Regarding the factors that influence their sales results we concluded that almost one-third of the examined organisations had a separate R&D department and the proportion of organisations that have a separate marketing department is similar.

Among the examined indicators of innovation performance the following ones must be highlighted:

- Organisations were spending more and more on R&D in the examined years, in 2009 it comprised 4.1 % of their average revenue. This value means a nominal increase of 154% relative to the one in 2007.
- Marketing-related expenses show a similar tendency: they increased from 1.3% to 1.84% in the examined timeline, which nominally corresponds to 142%.
- Most of the examined organisations pay attention to R&D activity, only a slight proportion of them reported that they had no time for these activities.
- Regarding the main directions of R&D activities it can be stated that most attention is paid to the development of the existing products (71%) and technologies (59%). Fewer companies are willing to deal with novelties, 66% are striving to develop new products and 52% to develop new technologies.
- The proportion of those carrying out adaptation activity is rather low. This proportion of 26% is slight as most organisations do not have a separate R&D department.
- Moe than 80% of the companies concerned have already been taking part in R&D cooperation. In these forms of cooperation, the proportion of the relationship with the customers is the highest, which indicates the priority of demand-driven innovations. It is also important to note that 60% is the proportion of partnerships with university research institutes.

The further processing and publication as well as feedback of the research results to the organisations taking part in the examination are the tasks for the future.

References

Edquist, Ch. (1997): Systems of Innovation. Technologies, Institutions and Organisations. Pinter, London & Washington.

Halpen László-Muraközy Balázs (2010): Innováció és vállalati teljesítmény Magyarországon, Közgazdasági szemle, LVII évf.,4. pp. 293-317

Husti István: Az innovációmenedzsment elemei. Szent István Egyetemi Kiadó, Gödöllő, 2010.

KSH (2010): Kutatás fejlesztés 2009, Statisztikai tükör, 2010 IV. évfolyam 89. szám

Losoncz Miklós (2008): Az EU-csatlakozás és a magyar kutatás-fejlesztési és technológiai stratégia, Közgazdasági Szemle LV: évf, 2. pp. 169-182.

Modeling Stochastic Inventory Policy with Simulation

János BENKŐ
Department of Materials Handling and Logistics,
Institute for Engineering Management

Abstract

Keeping an inventory (stock of goods) for the future sale or use is very common in business. Retail firms, wholesalers, manufacturing companies have a stock of goods on hand. Management faces a lot of challenging issues involving difficult decisions which can be company-oriented and customer-oriented issues. For example how does such a firm decide upon its inventory policy which is an important part of supply chains. In studying this problem, modelers typically focus on the following key performance metrics, which eventually can be translated to monetary measures: customer service levels (the rate of totally satisfied customer demands), average inventory levels and backorder levels, rate and quantity of lost sales and inventory cost.

Keywords

simulation, inventory, inventory policies, supply chains

1. Introduction

The importance of inventory as a part of supply chains comes from the fact that they constitute a large segment of economic activities. To improve their activities, to gain competitive advantage companies are interested in effective and efficient inventory operations that meet customer expectations. Therefore the supply chain management (SCM) or logistics tasks to produce and distribute products in the right quantities to the right locations at the right time, while keeping costs down and customer service levels up. In essence, SCM aims to solve these tasks by searching for good trade-offs between system costs and customer satisfaction.

SCM faces a lot of challenging issues involving difficult decisions which can be company-oriented and customer-oriented issues. In this study we deal with customer-oriented issues. For example, a customer's demand may not be fully satisfied from stock on hand and the shortage may be backordered or the sale may be lost. Either way, it is desirable to balance the holding cost which depends on the magnitude of inventory against the cost of not fully satisfying customer demand.

In studying such issues, modelers typically focus on the following key performance metrics, which eventually can be translated to monetary measures: (1) customer service levels (the rate of totally satisfied customer demands), (2) average inventory levels and backorder levels, (3) rate and quantity of lost sales, (4) inventory cost.

To achieve good performance, supply chains employ inventory control policies that regulate the orders to replenish stocks. Inventories can be reviewed continuously or periodically. The well-known inventory policies used in industry: (t,Q), (t,S), (s,Q), (s,S), where t is the cycle time, Q is the order quantity, s is the reorder level and S is target level. In the first two cases the reorder point is determined by the cycle time. In the other cases replenishment of the inventory starts when the inventory level reaches the reorder level.

Typically, when an order is performed by a producer or retailer, there is a lag in time (called lead time) after which the order is received. The lead-time demand is the magnitude of demand that arises during lead time, and is typically random and therefore can result in stock-outs. Consequently, to decrease the uncertainty in lead-time demand, companies constitute safety stock, which is extra inventory stocked to maintain good customer service levels. Companies may also use to place new orders before previous ones are received. Therefore ordering decisions are typically made based on inventory positions rather than inventory levels, where the

$$\textit{Inventory position} = \textit{Inventory level} + \textit{Inventory on order} - \textit{Backorders}.$$

Companies to get feasible inventory management use scientific inventory policies which based on mathematical models. These models describe the behavior of the inventory system. Inventory models are classified according to whether the demand is known (deterministic) or whether it is random variable having a known probability distribution (stochastic). In the last case the chance of the exact solution is very limited therefore we use simulation.

2. Problem statement

Application of simulation will be demonstrated with a very simple production-inventory system consisting of a production facility which supplies a warehouse with one type of product. This generic model illustrates how an inventory control policy regulates the flow of product between production and inventory facilities.

Figure 1 depicts a schematic diagram of the system. The raw material storage feeds the production process, and finished product units are stored in the warehouse. Customers arrive at the warehouse with product requests (demands), and if a request cannot be fully satisfied by on-hand inventory, the unsatisfied portion represents lost sale.

The following assumptions are made:

There is always sufficient raw material in storage, so the process never starves.

Product processing is carried out in lots of 5 units, and finished lots are placed in the warehouse. Lot processing time is uniformly distributed between 10 and 20 minutes.

The production process experiences random failures that may occur at any point in time. Times between failures are exponentially distributed with a mean of 200 minutes, while repair times are normally distributed, with a mean of 70 minutes and a standard deviation of 30 minutes.

The warehouse operations implement the (*s*, *S*) inventory control policy with target level *S*=500 units and reorder point *s*=150 units.

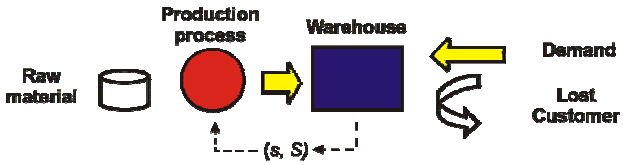


Figure 1. A generic production-inventory system

Customers arrive with interarrival times distributed as uniform with 3 to 7 hours, and individual demand quantities are also distributed uniformly between 50 and 100 units. On customer arrival, the inventory is immediately checked. If there is sufficient stock on hand, that demand is promptly satisfied. Otherwise, the unsatisfied portion of the demand is lost.

The initial inventory is 250, so the production process is initially idle.

Ordering cost has two components: *K* constant setup cost (regardless of the order quantity) is 10000 Ft/setup and *c* unit price 100 Ft/unit. The specific holding cost *h* is 2 Ft/unit/hour and the specific shortage cost *p* is 8 Ft/unit/hour.

We are interested in to determine the following performance parameters:

1. Production process utilization.
2. Downtime probability of the production facility.
3. Average inventory level at the warehouse.
4. Average demand level.
5. Percentage of customers whose demand is not completely satisfied.
6. Average lost demand quantity, given that it is not completely satisfied.
7. Average total operating (inventory) cost per hour.

The production-inventory problem described previously, although fairly elaborate, is still a simplification of real-life supply-chain systems. However, the applied method can be extended to more realistic problems, including multiple types of products, multiple production stages, production setups, and so on.

3. Arena model

Having studied the problem statement, we now construct the Arena model of the production-inventory system. Figure 2 shows our Arena model which is composed of two segments:

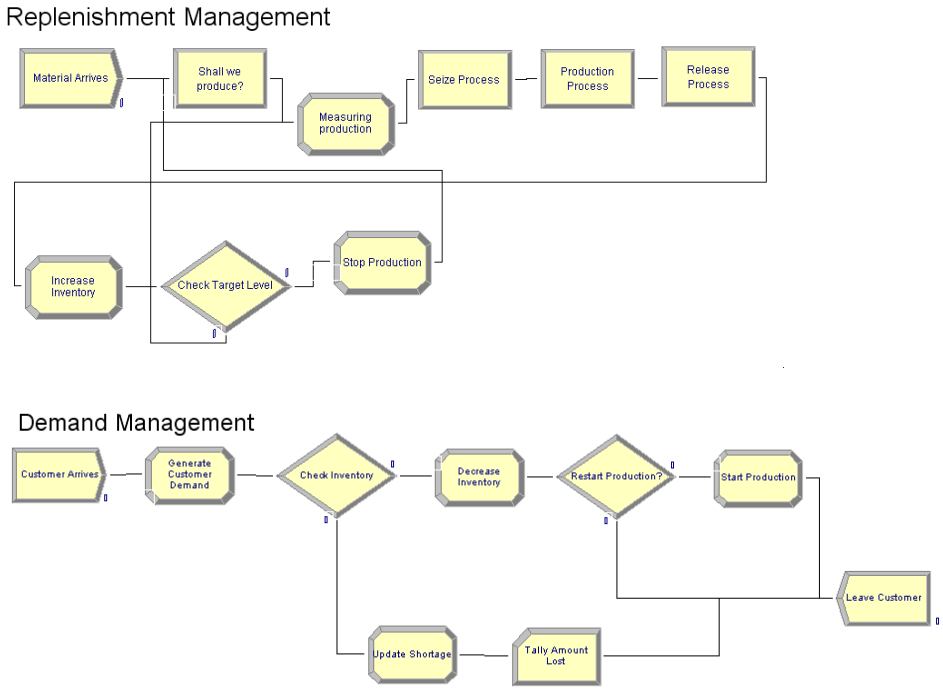


Figure 2. Arena model of the production-inventory system

Replenishment management segment. This segment keeps track of product unit entities. Entity definitions can be inspected and edited in the spreadsheet view of the Entity module from the Basic Process template panel. In this part of the model, the production process takes a unit of raw material from its queue, processes it as a batch of 5, and adds the finished lot to the warehouse inventory, represented by the variable *Inventory*. Thus, when a lot entity completes processing, *Inventory* is incremented by 5, and the simulation logic branches to one of two outcomes as follows: (1) If *Inventory* up-crosses the target level (variable *Target Stock*), then production stops until the reorder point (variable *Reorder Point*) is down-crossed again. (2) Processing of a new batch starts immediately when the reorder point is down-crossed. (Note that we always have sufficient raw material, so the production process never starves).

Demand management segment. This segment generates customers and their demands and adjusts variable Inventory upon customer arrival. It monitors the value of Inventory, and triggers resumption of suspended production when the reorder point is down-crossed. It also keeps track of lost demand (customers whose demand is not fully satisfied).

In addition, input and output data logic is embed in the two segments above. This logic consists of input/output modules (variables, resources, statistics, etc.) that set input variables, compute statistics, and generate summary reports. In the

following sections, we proceed to examine the Arena model logic of Figure 2. in some detail.

Replenishment management segment

We start with the replenishment management portion of the logic. The Create module, called *Raw Material*, generates product units for the production operation. A Hold module, called *Shall We Produce?*, serves to control the start and stop of the operation by feeding product units into a sequence of Seize, Delay, and Release modules (called *Seize Process*, *Production Process*, and *Release Process*), all drawn from the Advanced Process template panel. The actual processing takes place at the Delay module, called *Production Process*, where the process time of a batch is specified as *Unif*(10, 20) minutes.

The Create module *Raw Material* was initialized by populating it with a single product entity (at time 0), and therefore the interarrival time is irrelevant. Furthermore, by setting the *Max Arrivals* field to 1, the Create module will deactivate itself thereafter. The product entity circulates in the model repeatedly, with each cycle representing a production cycle.

The circulating product entity then proceeds to the Hold module, called *Shall We Produce?*, to test if production is turned on. The state of production (*Off*=0 or *On*=1) is maintained in the variable *Production*, which is initially set to 0. Recall that the Hold module performs a gating function on an entity by scanning for the truth or false of a logical condition. If the condition (in our case, *Production* = 1) is *true*, then the product entity proceeds to the next module. Otherwise, it waits in queue *Shall We Produce?.Queue* until the condition becomes *true* before being allowed to proceed. The circulating product entity then proceeds to queue *Production Queue* in the *Seize Process* module and seizes the resource called *Production facilities* immediately.

In the queue of the Seize module called *Production Queue* product entities await their turn for one unit of resource (*Production facilities*) to perform the process operation at module *Production Process*. Once the resource becomes available, it is seized by the highest-ranking product (in this case, rank 1). While the current process is in progress, the Seize module blocks any additional product entities from entering it. On service completion, the Release module (which never denies entry) releases one unit of resource *Production Process*.

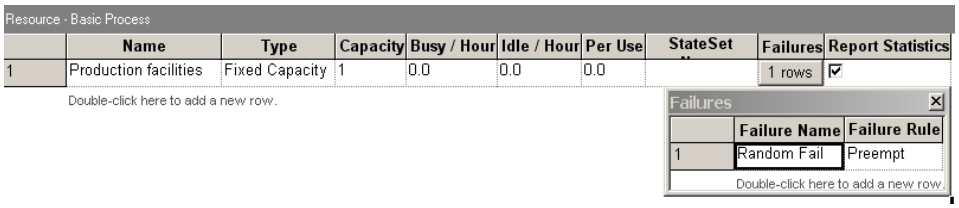


Figure 3. Dialog spreadsheet of the Resource module with its Failures dialog spreadsheet

A spreadsheet view of the Resource module with an entry for *Production facilities* is shown at the bottom of Figure 3. with its *Failures field* (associating resources to failures). All failure/repair data (uptimes and downtimes) are specified in the Failure spreadsheet module, as illustrated in Figure 4.

Failure - Advanced Process

	Name	Type	Up Time	Up Time Units	Down Time	Down Time Units	Uptime in this State only
1	Random Fail	Time	EXPO(200)	Minutes	NORM(70, 30)	Minutes	

Double-click here to add a new row.

Figure 4. Dialog spreadsheet of the Failure module

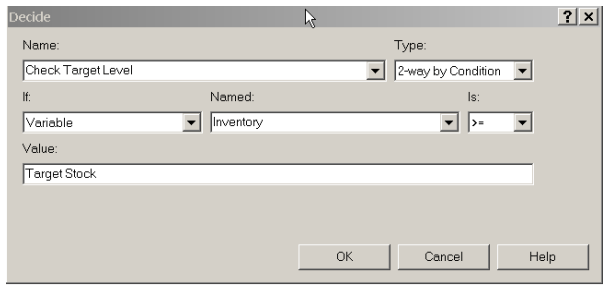


Figure 5. Dialog box of the Decide module *Check Target Level*

The inventory level at the warehouse is maintained by the variable *Inventory*, initially set to 250. Whenever the circulating product entity (batch of five units) enters the Assign module, called *Increase Inventory*, it adds a batch of 5 finished units (variable *Batch Size*) to the warehouse inventory by just incrementing *Inventory* by the batch size.

The circulating product entity then goes to the Decide module, called *Check Target Level*, whose dialog box is shown in Figure 5. Here, the circulating product entity tests whether the inventory target level has been up-crossed. The test can result in two outcomes:

If the inventory target level has been up-crossed, then the circulating product entity moves on to the following Assign module, called *Stop Production*, and sets *Production*=0 to signal that production is suspended.

Otherwise, the circulating product entity repeats productions.

Either way, the circulating product entity has completed its task through the system and would normally be disposed of (at a Dispose module). However, since the *Production Process* module is never starved and no delay is incurred since its departure from the *Production Process* module, we can "recycle" the circulating product entity by always sending it back to the *Production Process* module to play the role of a new arrival. This modeling device is logically equivalent to disposing of a product entity and creating a new one. However, it is computationally more efficient, since it saves us this extra computational effort, so that the simulation will run faster.

Demand management segment

The source of customer demand at the warehouse is the Create module, called *Customer Arrives*. The arrival pattern of customers is specified to be random with interarrival time distribution *Unif*(3, 7). Each arrival is an entity of type *Customer* (customer entity) with its private set of attributes (e.g. a demand quantity attribute, called *Demand*). On arrival, a customer entity first enters the Assign module, called *Generate Customer Demand*, where its *Demand* attribute is assigned a random value from the *Unif*(50, 100) distribution. The customer entity then proceeds to the Decide module called *Check Inventory* to test whether the warehouse has sufficient inventory on hand to satisfy its demand. The test can result in two outcomes:

If the value of variable *Inventory* is greater or equal to the value of attribute *Demand*, then the current demand can be satisfied and the customer entity takes the *True* exit to the Assign module, called *Decrease Inventory*, where it decrements the inventory by the demand amount. It next proceeds to the Decide module, called *Restart Production?*, to test whether the *Reorder Point* variable has just been down-crossed. If it has, the customer entity proceeds to the Assign module, called *Start Production*, to set *Production*=1, which would promptly release the circulating product entity currently detained in the Hold module *Shall We Produce?*, effectively resuming the production process. Either way, the customer entity proceeds to be disposed of at the Dispose module, called *Leave Customer*.

Variable - Basic Process

	Name	Rows	Columns	Clear Option	Initial Values	Report Statistics
1	Inventory			System	1 rows	☐
2	Production			System	1 rows	☐
3	Total Customer			System	0 rows	☐
4	Target Stock			System	1 rows	☐
5	Reorder Point			System	1 rows	☐
6	Batch Size			System	1 rows	☐
7	Lost Customer			System	0 rows	☐
8	Specific production cost Ft per Unit			System	1 rows	☐
9	Specific holding cost Ft per Unit per Unit time			System	1 rows	☐
10	Specific shortage cost Ft per Unit per Unit time			System	1 rows	☐
11	Number of Setup			System	0 rows	☐
12	Amount Production			System	0 rows	☐
13	Beginning of Cycle			System	1 rows	☐
14	Length of Cycle			System	0 rows	☐
15	Total Amount Lost			System	0 rows	☐
16	Total Demand			System	0 rows	☐
17	Setup cost Ft per number of setup			System	1 rows	☐
18	Demand			System	0 rows	☐

Double-click here to add a new row.

Figure 6. Dialog spreadsheet of the Variable module

If the value of variable *Inventory* is strictly smaller than the value of attribute *Demand*, then the current demand is either partially satisfied or not at all. Either way, the customer entity proceeds to the Assign module, called *Update Shortage*, where it sets the *Inventory* variable to 0. It also updates the variable *Lost Customer*, which keeps track of the number of customers to-date whose demand could not be fully satisfied (*Lost Customer*=*Lost Customer*+1), and the attribute called *Amount Lost*, which keeps track of the demand lost by the current customer (*Amount Lost* =*Demand*–*Inventory*). The customer entity next enters the Record module, called *Tally Amount Lost*, to tally the lost quantity per customer whose demand was not fully satisfied. Finally, the customer entity proceeds to be disposed of at module *Leave Customer*.

The user-defined variables involved in the model can be set or inspected by clicking on the Variable module of the Basic Process template panel. This yields the spreadsheet view of the module, as exemplified in Figure 6.

Statistics collection

The Statistic module is used to define nonstandard additional statistics (it is called *User Specified Statistic*) that are to be collected during the simulation and also to specify output data files.

Figure 7. displays the dialog spreadsheet of the Statistic module for the production-inventory model. It includes 4 Time-Persistent (DSTAT) type statistic, called *Average Inventory Level in Unit*, for the *Inventory* variable, *Average Demand Level in Unit*, for the *Demand* variable, *Average Cycle Length in Unit time*, for the *Length of Cycle* variable, and *Production On*, for the expression *Production*=1. The Time-Persistent type statistical outputs consist of the average value, 95% confidence interval, and minimal and maximal values of the variables ever observed during the replication. For example, the statistical output for *Production*=1 is the rate of time when this expression is true, or that is the probability that the production process is set to active. (Keep in mind, however, that the production process may experience downtimes.) Recall that this method of time averaging expressions can be used to estimate any probability of interest, including joint probabilities.

Statistic - Advanced Process

	Name	Type	Expression	Report Label
1	Average Inventory Level in Unit	Time-Persistent	Inventory	Average Inventory Level in Unit
2	Process State	Frequency		Process State
3	Production On	Time-Persistent	Production==1	Production On
4	Lost Rate in Customer per Total Customer	Output	Lost Customer/Total Customer	Lost Rate in Customer per Total
5	Average Demand in Unit per Customer	Output	Total Demand/Total Customer	Average Demand in Unit per Customer
6	Production Cost in Ft per Unit time	Output	Amount Production * Specific production cost Ft per Unit/TFIN	Production Cost in Ft per Unit time
7	Holding Cost in Ft per Unit time	Time-Persistent	Specific holding cost Ft per Unit per Unit time * Inventory	Holding Cost in Ft per Unit time
8	Shortage Cost in Ft per Unit time	Output	Total Amount Lost * Specific shortage cost Ft per Unit per Unit	Shortage Cost in Ft per Unit time
9	Setup Cost in Ft per Unit time	Output	Setup cost Ft per number of setup*Number of Setup/TFIN	Setup Cost in Ft per Unit time
10	Total Inventory Cost in Ft per Unit time	Output	OVALUE(Setup Cost in Ft per Unit time)+OVALUE(Production	Total Inventory Cost in Ft per Unit time
11	Average Cycle Length in Unit time	Time-Persistent	Length of Cycle	Average Cycle Length in Unit time
12	Average Demand Level in Unit	Time-Persistent	Demand	Average Demand Level in Unit

Double-click here to add a new row.

Figure 7. Dialog spreadsheet of the Statistic module

Figure 7 also contains a *Frequency* type statistic, called *Process States*, which estimates the state probabilities of the production process, namely, the probabilities that the production process is busy or down.

The first from the six Output type statistics in Figure 7 used to calculate the lost rate of customers that suffered some lost demand. The corresponding Expression field there indicates that when the replication terminates, the variable *Lost Customer* is divided by the variable *Total Customer* to yield the requisite rate. The other Output type statistics calculate the elements of inventory cost and the total inventory cost.

4. Result, discussion

Figure 8 displays reports of the results of a simulation run of length 20,000 hours (more than 2 years).

14:43:43 **User Specified** december 1, 2010

Production Inventory Replications: 1

Replication 1 Start Time: 0,00 Stop Time: 16 666,67 Time Units: Hours

Tally

Expression	Average	Half Width	Minimum	Maximum
Amount Lost Per Custumer	22.0868	1,51424	0.04322332	75.3473

Time Persistent

Time Persistent	Average	Half Width	Minimum	Maximum
Average Cycle Length in Unit time	1,037.18	(Insufficient)	0	3,267.90
Average Demand Level in Pc	75.1384		0	99.98
Average Inventoty Level in Pc	159.20	20,08239	0	504.58
Holding Cost in Ft per Unit time	318.40	40,16477	0	1,009.17
Production On	0.9854	(Insufficient)	0	1.0000

Output

Output	Value
Average Demand in Unit per Customer	75.2366
Lost Rate Customer per Total Customer	0.1214
Production Cost in Ft per Unit time	1,455.78
Setup Cost in Ft per Unit time	6.6000
Shortage Cost in Ft per Unit time	4.3043
Total Inventory Cost in Ft per Unit time	1,785.09

Figure 8. Simulation results for the production-inventory model

The results in Figure 8 give quite instructive information about performance of the production-inventory model.

The expression *Amount Lost Per Customer* in the Tally section indicates that the average lost demand per customer who experienced some loss was about 22 units. The Half Width column estimates the half-length of respective confidence intervals at the 95% confidence level.

The *Average Inventory Level in Unit* statistic in the Time Persistent section shows that occasionally the inventory up-crossed its target level. However, this happened only rarely, because the *Production* variable in the Time Persistent section assumed the value *On*=1 about 98.54% of the time. The average inventory level is 160.54 units. It means that often the inventory level was lower, than the reorder level, *s*=150. It seems in Figure 9. which shows the changing of the inventory and demand level in time.

The Output statistic *Lost Rate in Customer per Total Customer* in the Output section reveals that 12.02% of the customers had their demand either partially satisfied or not satisfied at all.

The Total Inventory Cost in Ft per Unit time also in the Output statistic is 1790.84 Ft/hour. Its components: the Setup Cost in Ft per Unit time is 6 Ft/hour, the Production Cost in Ft per Unit time is 1455.78 Ft/hour, the Holding Cost in Ft per Unit time is 321.09 Ft/hour, and the Shortage Cost in Ft per Unit time is 4.27 Ft/hour.

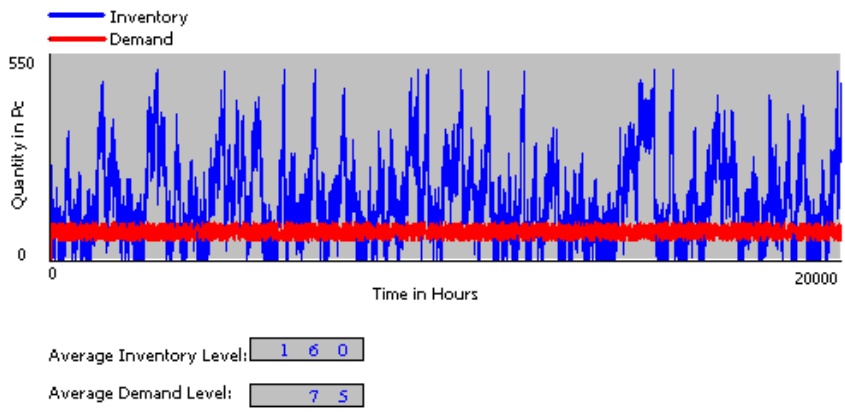


Figure 9. Inventory and *Demand* level as function of the simulation time

As we mentioned before, goal of SMC is to achieve economic balance between system costs and customer satisfaction. It known that simulation itself does not solve this problem directly, however, we can vary the input parameters of the system and we can observe the changes of the outputs. For example, we vary the value of Reorder point between 50 and 150 units, because we would like to decrease the inventory cost and to improve the service level (probability

that the demand of an arriving customer is fully satisfied) and the utilization of the production facility. Table 1 consists of the simulation results. Increasing Reorder point until 100 units results decreasing costs, improving facility utilization and the lost rate does not change significantly. Reorder point=100 units can be considered as an optimum, because above 100 units the cost increases and the facility utilization marginally falls down. This simple analysis leads us to conclude that a significant cost decrease can be achieved by choosing the appropriate reorder point.

Table 1. Cost analysis of the production-inventory system

Reorder Point in Unit	Holding cost in Ft/hour	Shortage cost in Ft/hour	Setup Cost in Ft/hour	Total Inventory Cost in Ft/hour	Lost Rate in %	Avg. Cycle Time in hour	Production on in %
50	349.87	3.21	6.0	1819,30	9.89	1314.49	98.25
75	288.09	4.20	4.0	1753,72	12.80	1532.36	98.91
100	296.68	3.75	3.5	1763,49	11.63	2477.92	99.02
125	303.74	4.35	4.0	1769,84	12.64	2261.50	98.94
150	321.09	4.27	6.0	1790,84	12.02	980.64	98.68

Much more information can be gained from the statistical output, considering the restricted extent of this study we could not strive for it. We encourage you to complete the analysis of the simulation results.

References

1. Altiok, T. (2007): Simulation Modeling and Analysis with Arena. Elsevier Inc. ISBN 13: 978-0-12-370523-5.
2. Arena Professional Reference Guide, Rockwell Software Inc., 2000.
3. Bratley, P.,-Fox, B. L.-Schrage, L. E. (1987): A Guide to Simulation. Springer-Verlag, New York, NY.
4. Benkő, J. (2009): Modeling Kanban Systemic Production with Arena Simulator. GépGyártás, Vol. XLIX, Nr. 4.
5. Devore, L. (2003): Probability and Statistics for Engineering and the Sciences. Wadsworth Inc, Belmont, CA.
6. Fishman, G. S. (1978).: Principles of Discrete Event Simulation. John Wiley.
7. Kelton, W. D., Sadowski R. P., Sturrock, D. T. (2004): Simulation with Arena. Mc Graw Hill Higher Education, ISBN 0-07-285694-7.

Farm Management Information systems for farm Power and machinery management

Imre KOVÁCS, István HUSTI
Department of Applied Management, Institute for Engineering Management

Abstract

Management Information Systems (MIS) are an integral part of the overall management system in a purposeful organization and form parts of tools such as, enterprise resource planning (ERP) and overall information systems (IS). The use of information technology tools and methods in agricultural mechanization accelerated. The application of ERP systems widespread in many areas of economy is not yet generally used in the agriculture. The current paper is analyzing the link between the tasks of agricultural mechanization management and the FMIS (Farm Management Information System).

Keywords

farm machinery management, agricultural technical development, ERP, FMIS.

1. Introduction

The information technology developed very quickly last 10-15 years. Thanks to this development there are available more foolproof and heavy-duty computers, quickly and more trusted local reds not only for the industrial and commercial but if the agricultural users.

One of the most eye-catching changes is the wide permeation of the Enterprise Resource Planning (ERP) and the fast development of the integrated, customizable and ready to use information systems. These modern systems are able to influence positively the organizations and give excellent tools to their managements.

The ERP systems map the economic processes of the enterprises with the tools of informatics. These systems work with separated but functionally related modules. There are two types of mapping. The first one is the type of one's own developing. This is very expensive; therefore it is used when the economic process is very individual. In the case of the agricultural mechanization it is recommended to use the second type, with standard software. In this case a software developing company makes modules for consolidated economic processes and these are available for license-fee.

The Fig. 1 shows a possible structure of the standard software with its modules. The different fields can function as autonomous modules, but the external interfaces permit of a real-time data flow among the modules. [5]

The goal of the present paper is to introduce – on the basis of tasks of the farm machinery management – the linking points and possibilities, which support the management with existing IT applications. We would like to give a base that we can integrate efficiently the tasks of the farm machinery management into the ERP systems.

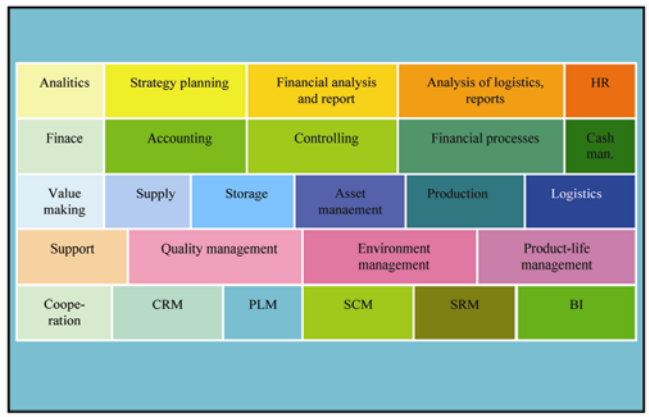


Figure 1. The most important modules of ERP
(Source: Hetyei, 2004)

2. Methodology

The paper deals with the mechanization of agricultural enterprises. Although this activity is important, it is only a part of the farming, therefore it cannot be operated separately [4]. Its main tasks are the planning, the execution and the management of the optimal machinery stock for the production. The withdrawal of machines from the production and theirs secondary reutilization is also an important task of the machinery management. All the above mentioned tasks take part strongly in the execution of the goals of the enterprise, and also some classical management functions are recognizable in them.

These tasks can be realized more efficiently with the ERP systems, which give the methodological background of our paper.

2.1. The way from ERP to FMIS

If farmers are to prosper in this turbulent economic environment they must manage their productive resources more efficiently and become more effective business managers. Innovative computer-based management tools have the potential to increase the quantity and quality of information available for decision making. Used in conjunction with modems, computers will soon provide the opportunity for remote farm businesses to access new sources of management information through connection to the Information Communication Technology (ICT).

Information and decision making process are inseparable. A system for providing information is vital to a business decision making process. Farm decision makers use information from a wide range of sources, but, one of the most valuable sources of specialized information about the farm operation is provided by a farm record system (FRS). A FRS can include financial and production record types. It may be as simple as a basic cash book, or so large and complex that it requires the processing capabilities of a computer to maintain it efficiently.

A computer-based system is conceptualized primarily as a management tool. It is a system of hardware and software elements capable of supporting a FRS and performing analysis on the data. The adoption and use of a computer enables farm businesses to operate a larger number of record types, and perform more extensive and complex analysis than would be possible using manual procedures alone.

The farm management information system (FMIS) is conceptualized as a FRS that provides information to support farm business decision making. If a FRS only provides information to individuals or organizations outside the farm business, then it is not a FMIS.

Farm management information systems (FMIS) are an integral part of the overall management system in a purposeful organization and form parts of tools such as enterprise resource planning (ERP) and overall information systems (IS).

FMIS differs from regular information systems because the primary objectives of these systems are to analyze other systems dealing with the operational activities in the organization. In this way, FMIS is a subset of the overall planning and control activities covering the application of humans, technologies, and procedures of the organization.

Fig. 2 shows the conceptual decomposition of the different management systems in an organization.

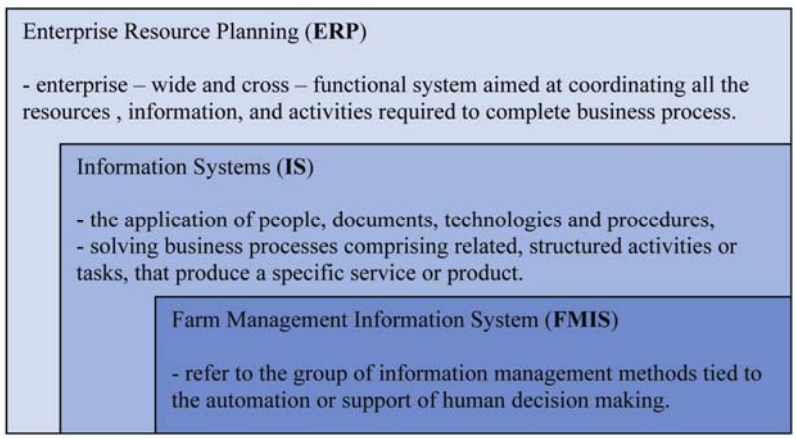


Figure 2. Concept of farm management information systems
(Source: Lewis, 1997)

By following this conceptual framework and notation, a FMIS is defined as a planned system for the collecting, processing, storing and disseminating of data in the form of information needed to carry out the operations functions of the farm. [7]

The Fig. 3 shows the perceived boundaries of the proposed farm management information system.

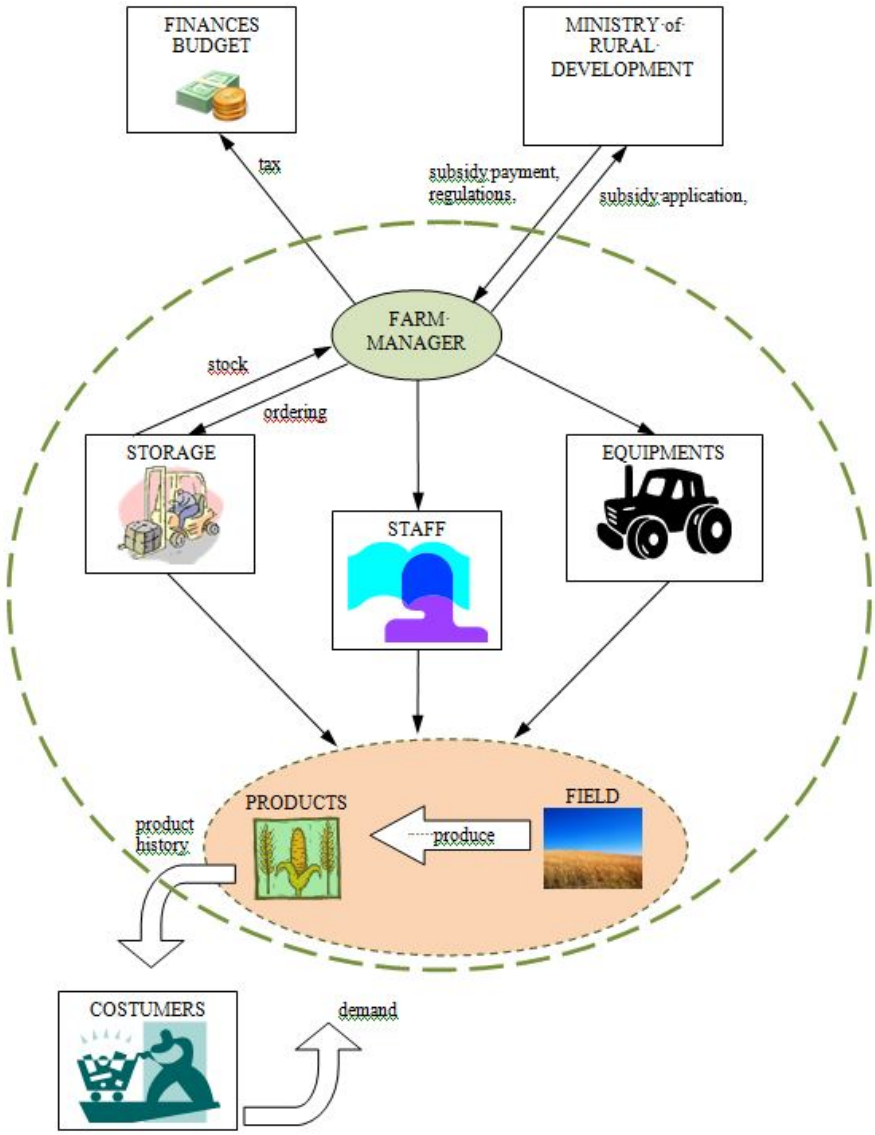


Figure 3. The components of the farm management information systems *(Source: Sirensen, 2010)*

The central entities in the proposed system include the farm manager, the fields, the products and production input. The Government and the consumer as external components are included since the system is very much influenced by these entities. [1].

3. Results

3.1. Phases of planning and decision making

The first stage of the asset management is the phase before the supply. From the functions of the machinery management the main accent is focused to the planning and decision making. [5] During the planning we put the goals. We plan also the future needs of machinery. We can decide about the next steps on the basis of the comparison of the needs and the existing production capacities.

a) Determination of the needs of tangible assets

The needs of tangible assets mean the quantitative and qualitative consistence of the tools, which make possible the maximal utilization of the natural and economic capabilities of the enterprise. To give an exact definition is quite difficult. Even so all enterprises should know what would be the optimal quantitative and qualitative consistence of the tools in their conditions. This planning task receives the information from the data of production. With their help we are able to plan the operations of the production technology. We can get the information from the Farm diary and the Field recording. In this stage of the planning it occur the selection, the modification and the adaptation of the operations, the accumulation of the plans of the different productions. With the accumulation of the above mentioned plans we can plan the supply of materials and the immobilization of capacities. There are many solutions to the execution of the technological operations. But we have to take into the consideration, that every solutions have different conditions.

b) Determination of supply of tangible assets

The supply of tangible assets means the proportion of the effectively extant and the theoretically wanted assets. Essentially it refers to the relation between the extant assets and the theoretically optimum of needs of assets. [4] The goal of the present analysis is to define in which fields are lack of assets or over-mechanization. The registry modules of assets have the needed information about the parameters of the tools (in the stock taken) and their relation to the workers.

c) Review of the possibilities of the fulfillment of need of assets

In view of supply of assets it can be transpired how many and what kind of assets are needed to reach the optimal stock of tools. In this phase those possibilities must be taken into accounts which are really available. Generally it is true, that not the ownership of tools is important for the enterprises but their

availability in the required quality and quantity, just in time. The associations of traders and wage-workers of machines help us with their databases to calculate the conditions of the use of the occurrent lease-work. The most important parameters for the calculations are the type, the needed quality, the place, the time and the volume of the tasks. After the calculations can we receive offers from the suppliers about the conditions of the lease-work (price, type of machine etc.).

3.2. Functions of organization and leading

The above mentioned phases belong to the preparation and decision-making. The organization means the assignment of the means to the processes. The investment of machines means that technical-economic process, which is manifested in the organization. The investment alone cannot guarantee the fulfill of tasks planned to the machine. To operate the machines is the job of the employees. The rational use of the machines is the precondition of the rise of efficiency. For this it is needed an adequate leading-style, which is described in the leading-function. In this function we have to emphasize not only the leading-styles, but if the adequate methods of motivation and information-management. The module of human-resource management help us to manipulate the employees-database, to link the workers with the machines, to record the finished jobs, and to calculate the wage.

d) Investment and supply of tools

The investment means the technical-economic activity to establish or enlarge the stock of assets. We use it in the process of asset-management, if we need new or used assets on the basis of an analysis of different alternatives. The supply of agricultural tools is quite wide if we need new or used machines. We can search them in the database of machine-traders, but we should to narrow them by trademark, type, age, capacity etc.

The financing and the analysis of investment-thrift play a very important role in the decision of the investment. This information is available from the traders too. During a tender we can use the calculations of different credit, support and leasing possibilities.

e) Use of assets

We can define an asset as an aggregation of potential features too. During the use we can see how these features can help to reach the goals of the enterprise. From the other side we have to focus to the rational use of tools to have more profit.

During the modern machine-use the GPS applications – like the Agromodul – give us a support of informatics. These systems are able to use RF card identification, so we can monitor all events on- or offline by recording the data of the machine-operator, the used materials and the finished works. There is also the possibility to record the data of the time-use, so we can monitor the time of working, maintenance and transport. [6] There is also a Maintenance module,

where all the maintenance and repair operations are monitored. The materials and components of the maintenance and repair are monitored by the Logistics module.

3.3. Leading function

The main task of the leading is to coordinate the tasks to reach the goals. To know the absolute and relative costs of the tasks is very important, because this information gives the basis for the needed interventions. The received information during the use of machines helps us to calculate the optimal time of the machine-use, its rejection and its secondary reuse.

The necessary calculations are able to automate with integrated systems of informatics.

f) Amortization of tools, determination of amortization period

The main task of the amortization politics is the monitoring of the amortization of the machine during its use. The aging of the assets is expressed by the amortization. The amortization means the accounting of the loss of value because of aging. The original gross value of the machine put in practice is decreasing proportionately with the use, so the machine has an amortization (gross value= amortization + net value). The annual amortization is defined by % of value of the asset. We can calculate the amortization period if we divide the gross value (100%) of the asset with the annual amortization (%/year). There are different types of amortization. We can use a linear method, or a progressive and digressive method to accelerate or decelerate the amortization in the proportion of time or capacity.

g) Determination of the time of rejection

The worn-out or superseded asset will be taken away from the production or rejected (not always after the amortization to 0). The time of rejection is influenced by many factors. There are cases when the technical reliability of the amortized asset is still appropriate. In these cases it is worth to use further, because we don't have to calculate with the amortization and so the operation get cheaper. It is also true, that in these cases don't generate fund of amortization, which is a good basis for the developments. We plan the time of rejection individually. We can decide about the rejection of a machine on the basis of the profitability data received from the Controlling module.

h) Determination of the optimal time of use

The optimal time of use define the period as long as reasonable the use of the asset. Examples prove that the relative costs of the use of asset during its life-period first decrease then after reaching a minimum value increase. We should use an asset on the basis of economic calculations until a unit of the operational costs (e.g. Ft/hour, Ft/km) starts increasing. [2] In the practice we should first define the use of the asset and then calculate the amortization and rejection. From the other side the amortization must be harmonized with the optimal use of

the asset and after then we should reject the machine. Unfortunately in the real practice it happens very rarely because of the economic force or inaccurate cost calculations.

These necessary calculations are able to automate with integrated systems of informatics.

i) The secondary reuse of assets

We should find a solution in the enterprise that the unnecessary or rejected assets don't go to pot. We know well, that an asset accounted to 0 has much more value than to dispose of it. The secondary reuse means cost sparing or income for the enterprise. Moreover many times the need of assets can be supplied by buying used machines. On the Fig. 4 we can see the classical asset management model wedged in the modules of the FMIS components.

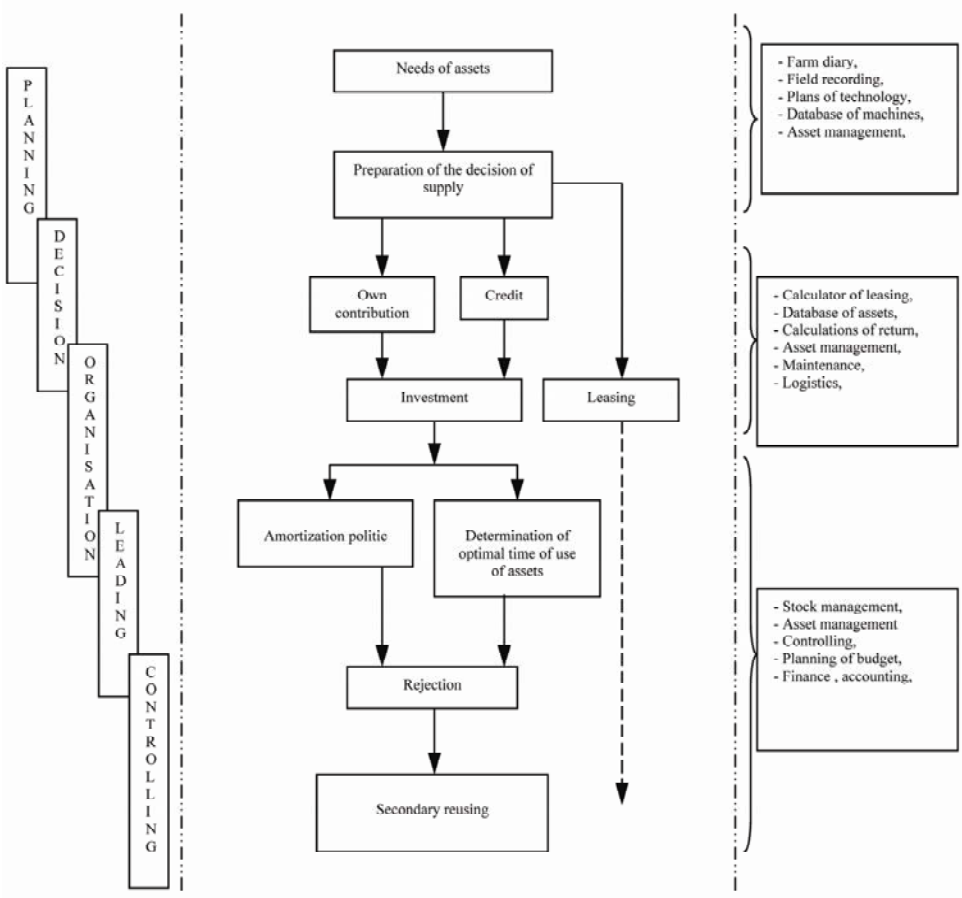


Figure 4. Model of the asset management from the view of the management functions and information modules
(Source: Husti, 1999)

4. Conclusion

In the study we have been examined the most important tasks of the mechanization management on the basis of the management functions. It is concluded that the basic functions (planning, decision making, organizing, management and control) can be supported by farm management information systems. But instead of isolated solutions we need to link them. In this paper we gave the theoretical base of this.

References

[1] C.G. SIRENSEN, S. FOUNTAS. et. al.: Conceptual model of a future farm management information system. Science Direct, Computers and Electronics in Agriculture 2010. Netherlands, p.:37-47.

[2] D. HUNT: Farm Power and Machinery Management, Wiley- Blackwell, 2001. ISBN 0813817560

[3] HETYEI J.: ERP rendszerek Magyarországon a 21. században. Computer Books, Budapest, 2004. ISBN 963618318X

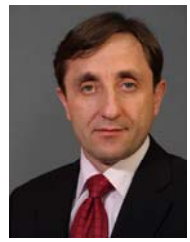
[4] HUSTI I.: A mezőgazdasági gépesítés ökonómiája és menedzsmentje. Szaktudás Kiadó, Budapest, 1999. ISBN 9633562651

[5] KOCSIS J. ET. AL: Menedzsment műszakiaknak Műszaki Könyvkiadó, Budapest, 1994. ISBN 963160392X

[6] SULYOK D.: Leíró és döntéstámogató informatikai rendszerek (szoftverek) alkalmazásának előnyei. Agrárágazat, Budapest, 2007. április.

[7] T. LEWIS: Evolution of farm management information systems Science Direct, Computers and Electronics in Agriculture 1997. Netherlands, p.:233-248.

Institute for Mechanics and Machinery



Professor Dr. István SZABÓ,
Director of the Institute

Dear Reader,

The Institute for Mechanics and Machinery (in short IMM) has been known for integrating fundamental engineering disciplines in the fields of classical mechanics, design, computer aided technologies, and machine constructions. Today it consists of three scientific and educational units:

- Department of Mechanics and Technical Drawing
- Department of Machinery
- Department of Agricultural and Food-industrial Machines

The activities and the scope of scientific interest of the last Department had been expanded last year with integrating system engineering and machinery management into the educational and research programs run by IMM.

IMM has got two external departments : Department of Biotechnics hosted by Oncotherm Ltd., and department of Farm Machinery set up by the National Institute of Agricultural Engineering.

Courses provided by our staff covers a wide range from basic engineering areas (e.g. computer aided drawing, statics, strength of materials etc.) to more specialized domains (biomechanics, computer aided analyses, machine design, bio-systems engineering, theory of engineering systems, communal machines, digital image processing, machinery management etc.) at all levels of studies (Post secondary, BSc, MSc, PhD, postgraduate courses).

IMM focuses on several advanced research areas, running projects financed by the European Union, the Hungarian National Office for Research and Technology and also by industrial partners.

Activities are concentrated around the following key areas of interests:

Biomechanics

Applied engineering numerical methods

Computer based modeling and analyses

Advanced methods in engineering design

Energy purpose biomass production

System Engineering and machinery management

Papers published in the following pages may give the Readers some impression of the scientific activity and interest of the researchers working at the Institute for Mechanics and Machinery.

Determination of granular assemblies' discrete element material parameters by modelling the standard shear test

István KEPPLER[1], István OLDAL[1], Attila CSATÁR[2]
[1] Szent István University, Hungary
[2] Hungarian Institute of Agricultural Engineering, Hungary

Abstract

Soil and agricultural products interact with the tools and equipment used to manipulate and store them. This interaction causes the load of the tools and of the agricultural product. The practicing engineer has to know how these types of materials behave so as to be able to examine and control their mechanical behaviour. To fulfil this aim we have to find a general model of these materials and apply this to the specific cases. A relatively new model of granular assemblies is the so called discrete element based model. In this model we describe the granular assembly as the collection of large number of small rigid bodies, and the modelling process of the assemblies' behaviour is based on solving the equations of motion of this large number of particles directly. The question that arises from the practical use is how we can determine the parameters which affect the interaction between the particles: the coefficient of static- and rolling friction, coefficient of restitution, Young modulus and Poisson's ratio of a given (in some cases very small) particle. The direct measurement of these quantities is of course impossible for really small particles. We developed a new method based on performing the standard shear test of the given materials and by modelling the shear test using discrete element method. If the discrete element parameters of the numerical model are properly set, the output data coming from the measurement will be in good agreement with the data coming from the discrete element model.

Keywords

mechanics of granular materials, discrete element method, shear test

1. Introduction

Engineers working on the field of agriculture, food- or pharmaceutical industry or in the architecture frequently met problems arising from the special properties of granular assemblies. The most important properties determining their mechanical behaviour are their material- and failure models.

From the mechanical point of view, two different type of material model can be established: the so called discrete element model, where the physical parameters of the interaction between the distinct grain particles (the so called

micromechanical parameters) are modelled, and the continuum model, where the whole granular assembly is modelled as a continuum. The difficulty of the mechanical modelling arises from the problem that the continuum model's (here called as macromechanical) parameters can be determined by measurements developed long ago, but the determination of the micromechanical parameters is difficult and sometimes impossible.

At the department of Mechanics and Engineering Design of Szent István University and at the Hungarian Institute of agricultural Engineering we developed a method for determining the micro mechanical parameters of granular assemblies by carrying out measurements of the macromechanical parameters and by modelling the same measurements using discrete element method.

By presenting a given granular assemblies shear test, we demonstrate the determination of the micromechanical properties of that assembly. The micromechanical parameters of the assembly are determined by modelling the shear test using discrete element method. By changing the micromechanical parameters in the discrete element method it is possible to get different shear diagrams. If the discrete element parameters of the numerical model are properly set, the output data coming from the measurement will be in good agreement with the data coming from the discrete element model. We suppose that the micromechanical parameters describing the shear process with good accuracy are the micromechanical parameters of the granular assembly, and can be used for the discrete element modelling of the given type of granular material.

2. The discrete element method

Granular material is a conglomeration of discrete solid, macroscopic particles. From purely physical point of view, the best way of the modelling of such kind of material is solving the equation of motion for all of the particles assembled, taking into account the collisions and other interactions between them and the other bodies (such as containers, tools). Of course, from the practical point of view, this is impossible for large number of particles. Or it is better to say this has been impossible for all cases until recent years. Nowadays, for suitable number of particles, such kind of modelling is possible.

Figure 1. Granular materials

The method used for this type of modelling is the so called discrete (or distinct) element method. Discrete element method (DEM) is the name of a mathematical method used for describing the kinematical behaviour of particles. As defined by Cundall [1], discrete element methods allow finite displacement, rotation and separation of elements, and new particle contacts are recognized algorithmically as the simulation proceeds.

After Cundall's pioneering work, the computational backgrounds of DEM based research were established. Even the strong hardware needs of such computing can be satisfied for modest problems. Modest means here sufficient number of particles to model real engineering problems arising from the industry. The difficulties of such kind of modelling are the determination of micromechanical parameters influencing the interaction between the different particles. These parameters to be determined prior to starting the DEM modelling are:

1. The shear modulus (G) of the particles.
2. Poisson's ratio (υ) of the individual particles.
3. Density (ρ) of the particle.
4. Coefficient of restitution C_R.
5. Coefficient of static friction μ_s.
6. Coefficient of rolling friction μ_r
7. Cohesive energy density C_s.

The method of finding these parameters is called as the calibration of DEM model. The calibration process is based ono observing and measuring the behaviour of the granular material in given circumstances, and on modeling the same process using DEM. By comparing the results of experiments and their models, the micromechanical parameters can be determined. Our calibration process is based on the so called standard shear test.

3. The standard shear test

The standard shear testing technique for particulate solids is based on the so called Jenike shear cell [2]. The Jenike Shear Cell (fig. 2.) consists [3] of a base (1) shear ring (2) and shear lid (3), the latter having a bracket (4) and pin (5). Before shear the ring is placed in an offset position and a vertical force F_v is applied to the lid and hence to the particulate solid within the cell by means of a weight hanger (6) and weights (7).

A horizontal force is applied to the bracket by a mechanically driven measuring stem (8) which is driven forwards at a steady rate of 1-3 mm/min. This stem is attached to the drive system through a force transducer which measures the shear force F_s. During the shear operation the shear ring moves from the original offset position to the opposite. During shear a shear zoned develops inside the sample, and in this way the shear force vs. time plot can be transformed to a shear force – shear strain plot. Using Jenike shear cell, the

friction between the granular material and the container wall (the so called wall friction) also can be determined [3].

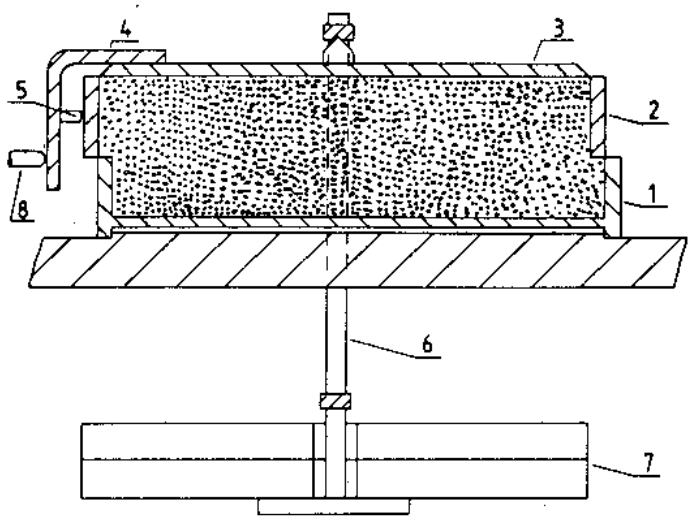


Figure 2. Jenike shear cell [4]

Based on the description found in the literature [4] we developed an automatic shear device making the same shear process described above (fig. 3.).

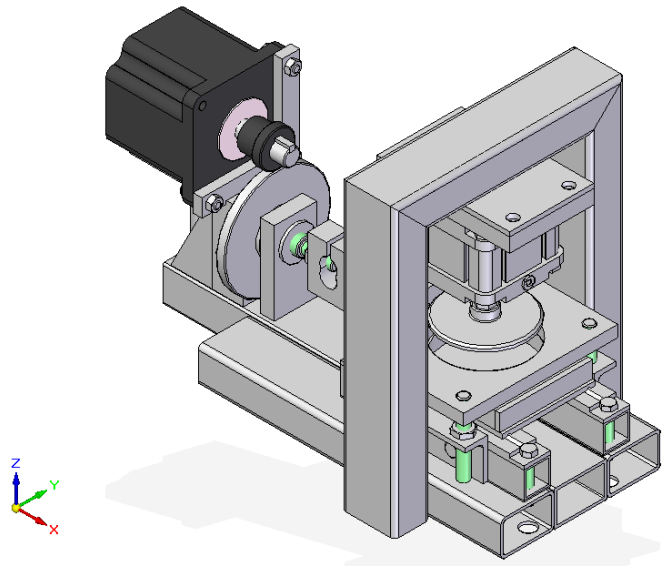


Figure 3. Automatic shear testing device

By using this automatic shear testing device, we determined the yield loci (the envelope of the Mohr circles corresponding the failure at given normal load) of cement powder for demonstrating the usability of our calibration method. First we made critically consolidated samples of the cement powder, critically consolidated means that the shear force – shearing distance diagram has no local maximum point, but saturates within the shearing distance of the cell. A particulate solid is critically consolidated with respect to the state of stress applied, when it yields without change in bulk density at a constant state of stress [4].

By pre-shearing the material sample using the compressive force corresponding to critically consolidated sample of cement, then lowering the compressive force successively to smaller normal forces (fig. 4.), we made the points of the yield loci for cement powder. Using linear approximation, we determined the yield line of the given powder by fitting a line to points determined by measurements (fig. 6.).

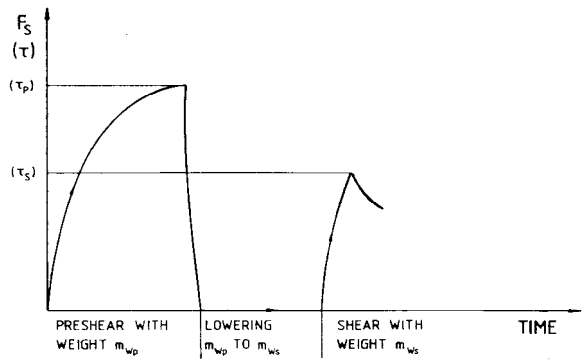


Figure 4. Preshear, lowering and shear [3]

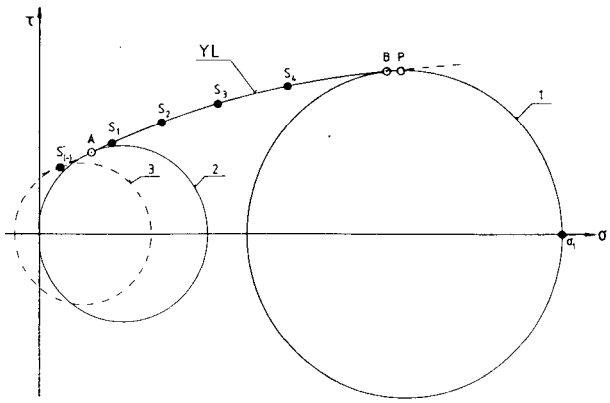


Figure 5. Yield loci

4. Dem model of the standard shear test

By using commercial DEM software, we created a simple shear cell model. The model consisted of two cylinders (having the same size as the standard Jenike cell), an upper plate and a plane dividing these two. We filled the cylinders with spherical particles having radius randomly distributed radiuses varying around r=0,003 m with a variance m (in DEM models, the particle sizes must not be inevitably the same as the sizes of the real particles [5]). The first phase of modelling was the filling of the cylinders with granular materials, then a pre-compression phase followed the filling. The pre-compression was made by the upper plate, as it was forced to move downwards until pressure on it resulted the vertical force needed to reach critically consolidated sample in the virtual tester.

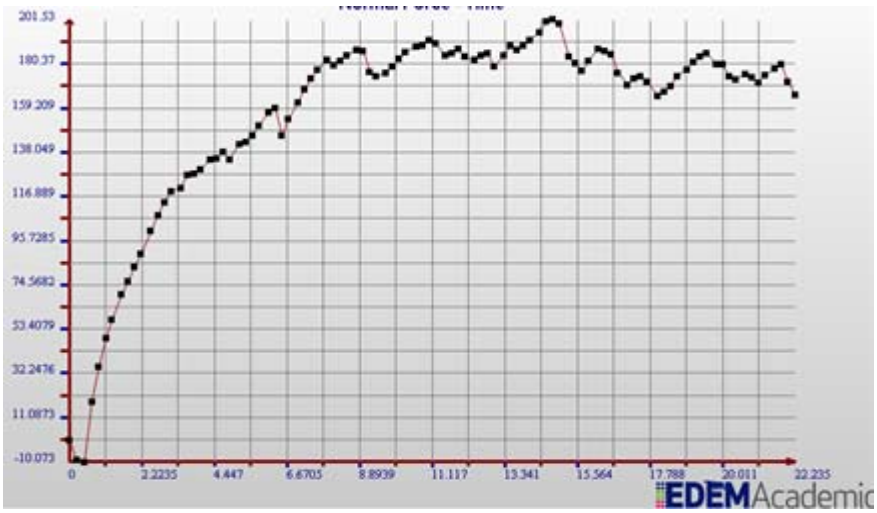


Figure 6. Shear force – time diagram from the DEM model, critically consolidated sample

After reaching critically consolidated state of the sample, the upper cylinder has started to move in the horizontal direction, modelling the shear process. From the pressure arising from the particles in the cylinder, the vertical force acting on the cylinder wall was determined.

By successive modelling of the shear process with DEM using different micromechanical parameters we managed to get yield loci from the „virtual" shear test being in a good match with the yield line determined by measurements (fig. 7.).

On fig. 7 small black dots are the measured yield points of cement powder, red dots denote the average values and blue dots correspond to the yield loci from the virtual DEM model of the given powder. The micromechanical parameters corresponding to the model were: 15000 Jm^{-3} cohesion energy density, 0.25 as Poisson's ratio, 10^8 Pa shear modulus, 3000 kgm^{-3} density, 0.5

coefficient of restitution, 0.001 coefficient of rolling friction and 0.5 static friction coefficient.

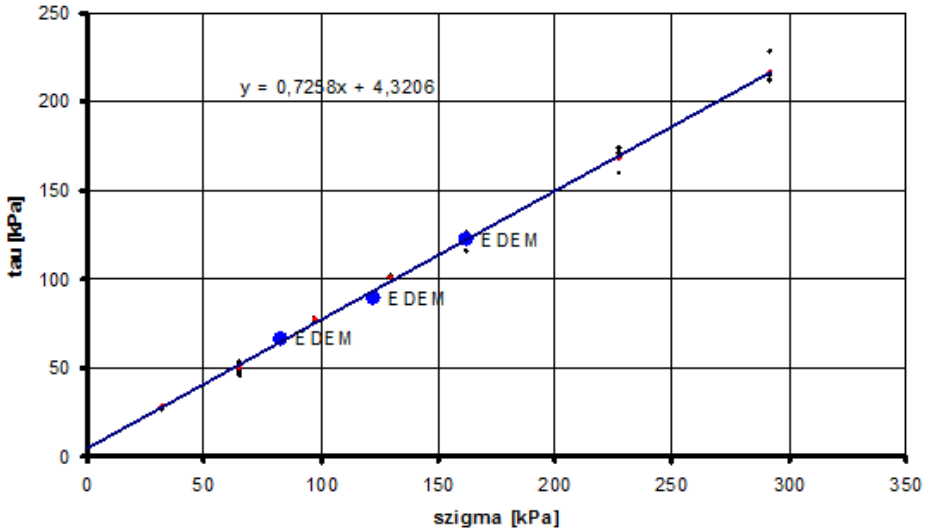


Figure 7. Yield loci in linear case from measurements, and yield points from DEM model

By using the same method described above, the determination of micromechanical parameters is possible for any kind of granular materials.

5. Conclusions

We developed a method for determining the micro mechanical parameters of granular assemblies by carrying out measurements of the macromechanical parameters and by modelling the same measurements using discrete element method. Our calibration process is based on measuring the failure properties of the granular material during standard shear test. We demonstrated the usefullness of this method by measuring the failure properties of cement powder, and by doing the same using the virtual discrete element shear tester.

Further question is arising from these preliminary results as we do not know, wheter these values of micromechanical parameters are unique i.e. are there other set of values for the micro parameters resulting the same failure line during DEM shear tests? Are the results good for the practice meaning wheter using these values as micro parameters for modeling other mechanical processes give results beeing in good agreement with the results of measurements?

For finding the ansver for these questions, we plan to execute sensitivity inspections for the data arising from the shear model, and to do triaxial and oedometric tests on the same material samples and compare them against their

discrete element model based on the micromechanical paramaters determined by the dicrete element shear test.

6. Nomenclature

G	shear modulus of the particle	Nm^{-2}
υ	Poisson's ratio of the particle	–
ρ	density of the particle	kgm^{-3}
C_R	coefficient of restitution	–
μ_s	coefficient of static friction	–
μ_r	coefficient of rolling friction	–
C_s	cohesive energy density	Jm^{-3}
F_v	vertical force	N
F_s	shear force	N
r	perticle radius	m

7. Acknowledgements

The authors would like to acknowledge the support of the Hungarian Academy of Sciences (Bolyai Research Fund).

References

[1] P.A. Cundall, O.D.L. Strack: *A discrete numerical model for granular assemblies,* Geotechnique, 29:47–65, 1979.

[2] Jenike, A. W.: *Storage and Flow of solids*, Bulletin No. 123, Utah Engineering Experiment Station, University of Utah, Salt Lake City, 1964.

[3] Prescott, J. K., Barnum R. A: *On powder flowability*. Pharmaceutical technology october 2000 pp. 60-84. 2000.

[4] The Institution of chemical engineers: *Standard shear testing technique*. Published by the institution of chemical engineers, England, 1989. ISBN 0 85295 232 5.

[5] Bagi Katalin: *A diszkrét elemek módszere.* BME Tartószerkezetek Mechanikája Tanszék, 2007.

Spectral separation of different nutrient levels in winter wheat (Triticum aestivum L.) at characteristic wavelengths

Kornél D. SZALAY, László FENYVESI
External Department of Agricultural Engineering,
Institute for Mechanics and Machinery

Abstract

The rapidly growing demand toward data and information cannot be satisfied any longer by the conventional sampling methods. There has been a need for a measuring technology that provides broad opportunities of evaluating local or global processes or balances according to the various aspects. The dynamic development of the different remote sensing technologies resulted in the hyperspectral imaging spectroscopy, which is one of the most advanced technologies in optical remote sensing. It has greatly improved the efficiency of data utilization and created new perspective for modern information management in precision agricultural production. With the adaptation of the hyperspectral technology different quantity and quality parameters can be measured in a fast, precise and economic way. This technology is adequate to analyse vegetation cover by evaluating average radiance spectra of the electromagnetic radiation reflected from their surfaces. Reflectance spectra of vegetation cover depend on the absorption and transmission characteristics of the surface cover. In this study, we analyse samples of variously treated winter wheat using FieldSpec 3 MAX spectroradiometer in the wavelength range of 350 nm to 2500 nm. The experimental arrangement was set by broadcasting different nutrition levels applied on 10 m^2 replicated field plots, which were studied according to different parameters. Beyond the spectral analysis, we evaluated correlations including additional parameters, such as plant height (cm), ear size (cm), yield (kg/10m^2) and quality parameters to make comparations. According to the preliminary results, certain characteristic spectral intervals of the samples showed reflectance differences due to the various nutrient supply. After clarifying all decisive correlation, the spectral methodology can greatly assist in describing and tracking the current dynamics of nutrient supply and plant up-take, both of which are indispensable to perform precision nutrient-supply being one of the most important factors in assisting in the environment protection or reducing expenses of farming.

Keywords

laboratory spectroscopy, spectroradiometer, winter wheat, nitrogen fertilizer

1. Introduction

In the modern agriculture, beyond the productiveness and profitability, the environment protection should be kept in mind, as well. The inappropriate nitrogen fertilization threatens mankind indirectly by damaging the ecosystem (Nagy et al. 2008) and endangers directly through the nitrate accumulation in plants (Nádasy and Nádasy 2006). According to our experience, beside the climate and weather aptitudes (Erdélyi 2008, 2009, Erdélyi et al. 2009; Kulpács et al. 2010, Szentpétery et al. 2010), the correctly chosen production site (Tarnawa and Klupács 2006) and agro-technology (Balla et al. 2010), or rather the harmony between the individual demands or resilience of the certain variety (Szalay et al. 2010) and the applied nutrient supply are determinative. In our modern culture, people pay more and more attention to their healthy lifestyle, thus, their demand toward healthy foods increases. To produce quality food, correctly chosen production methodology is needed. Wheat is our most important grain. Beyond the vital elements bread contains the most of the vitamins, too. After all healthy and homogenous vegetation can only be produced by satisfying the nutrient demand of the plant (Jolánkai 2004). In order to achieve the optimal plant nutrition and to avoid, or minimize the environment pollution we have to know the exact nutrient supply (Virág et. al 2010) and plant up-take system. In the following trial, beyond the conventional examinations (Szalay et al. 2009) we carried out 'ex situ' spectral analysis to study the nitrogen treated and untreated samples of the 'Alföld 90' winter wheat variety.

With laboratory measurement method we can exclude the natural light environment by which we can gain fundamental results that can be extended with 'in situ' field measurements. Collecting data in laboratory, or in field with spectroradiometer gives the so called ground reference to the airborne hyperspectral remote sensing that is the most developed technology among the optical remote sensing systems. Using this method, we measure the spectral radiance reflected by the evaluated vegetation that depends on their absorption and transmission characteristic. We can obtain quantitative relationships between the environmental or physiological parameters of vegetation cover or the soil quality parameters and the features of reflectance spectra. These reflectance spectra have moreover spectral signatures characteristics of the different surface covers allowing their accurate classification, and identification. Concerning these data, we can gain information according to the examined targets quantity and quality parameters. By using modern remote sensing technologies (Jung et al. 2006a, 2006b) we can gain several times more pieces of information, those would not be visible for the human eye (Kristóf 2005) and makes possible the sampling on large areas and collecting data from different surface processes (Csorba and Jordán 2010) including several monitoring possibilities (Kardeván 2010) according to soil (Máthé et al. 2010), vegetation (Tolner et al. 2010), or climate attributes. The detectors of the spectroradiometer applied in this study expand the limits of the visible light (VIS) – 400 to 700 nm – based RGB technology (Lágymányosi and Szabó 2010) and make possible to carry out examination in near infrared (NIR)

and short wave infrared (SWIR) part of the electromagnetic radiation, too, by detecting the wavelength range of 350 nm to 2500 nm. The coordinates of in-field examinations are recorded so the ground spectrum can be fitted to the adequate pixel of the hyperspectral airborne image that is an important element of subsequent evaluation processes. The number and the quality of in-field measurements determine the final accuracy of the airborne images. This technology is adequate to analysed vegetations in a fast, precise way (Fenyvesi 2008, Milics et al. 2008, Yang et al. 2009, Milics et al. 2010).

2. Materials and methods

The 'Alföld 90' winter wheat variety was studied in agronomic replicated blocks (Figure 1.). Experimental plots were sown on chernozem soil (calciustoll) at Hatvan- Nagygombos, (Central Hungary). The half of the plots received 80 kg ha^{-1} nitrogen fertilizer in form of ammonium nitrate (0-0-36), the others did not received any mineral fertilizer. The plot size was 10 m^2. The experimental plots were evaluated concerning plant height (cm), ear size (cm), yield (kg/plot), and quality parameters – such as wet gluten content (%), protein (%), furthermore thousand kernel weight (g). Beyond these, prior to the harvest, wheat ears from all plots were gathered and were analyzed beside the harvested grain according to their spectral characteristic with an ASD Fieldspec 3 MAX spectroradiometer (Figure 2.) with 'ex situ' method in the wavelength range of 350 to 2500 nm.

Figure 2. Plots at Hatvan- Nagygombos

Figure 2. ASD Fieldspec 3 MAX

Figure 3. Light-isolated measurement cabinet

Figure 4. Laboratory measuring arrangement

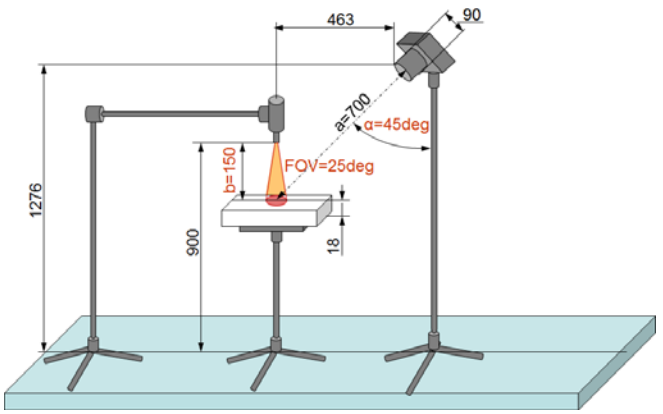


Figure 5. Illustration of the measuring arrangement

The measuring process was carried out in a light-isolated laboratory cabinet (Figure 3). Laboratory condition makes possible to achieve outstanding

precision, since the disturbing environmental light is shielded and the undesirable reflexions from the interior of the measuring cabinet are minimized by the appropriate arrangements of the implements (Figure 4. and Figure 5.). The special material of the sample stage and the cabinet interior result in minimal reflectance over the whole electromagnetic spectrum detected by the spectroradiometer. The 'ex situ' method makes possible to study the collected wheat samples – wheat ear, kernel, straw, leaf – independently from the weather conditions, among steady circumstances. The reflectance of the wheat ear and kernel samples illuminated by special laboratory lamp was recorded.

3. Results and discussion

By forming the plant height to ear size ratio, the data show, that the nitrogen fertilized wheat's ear size and plant height ratio became smaller, than the untreated parallel's. Nitrogen also resulted in higher yield. Studying the quality, such as the protein and the wet gluten content, we found that both parameters increased as a result of the mineral fertilizer. In case of thousand kernel weights, the variants treated with nitrogen also resulted in higher values.

Evaluating the wheat ears with spectroradiometry, we computed the mean reflectance spectra of the treatments. Red colour represents the nitrogen fertilized, while green the unfertilized parallel. According to these curves the spectral characteristic of the different treatments are divergent (Figure 6.), but the deviation of those seems independent according to the differences generated by the mineral fertilizer.

By removing the continuum of the curves (Figure 7.), we found a characteristic interval between 1700 nm and 1800 nm. At the wavelength of 1738 to 1745 nm the red curves run unequivocally below the green ones (Figure 8.).

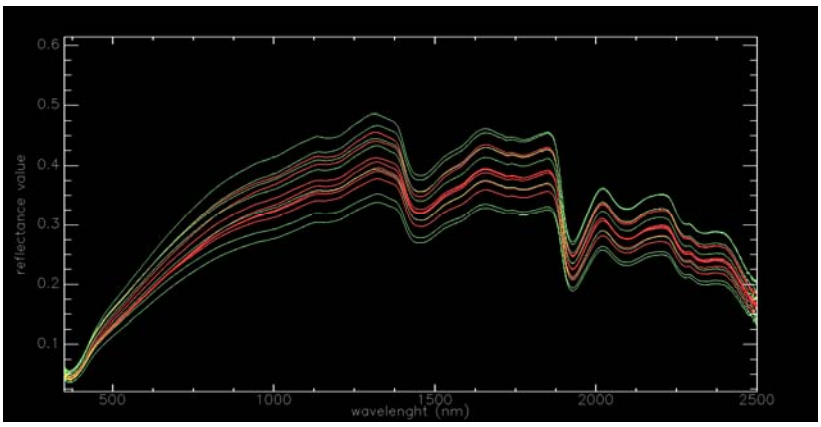


Figure 6. The normal reflectance curves of the wheat ears seem independent from the various nitrogen treatment as red curves are mixed with green ones

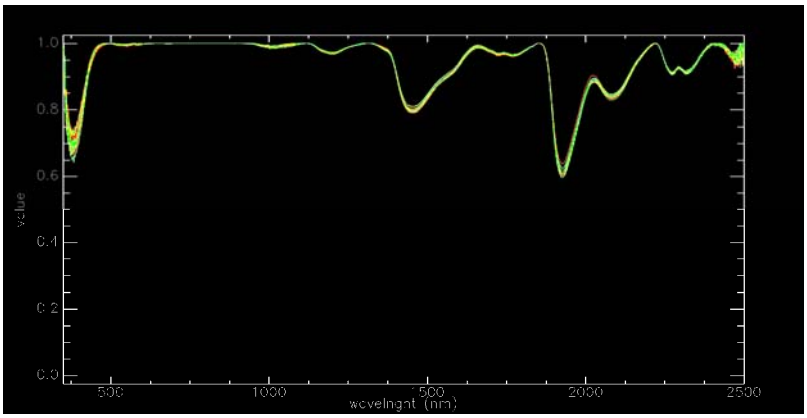


Figure 7. Continuum removed curves of wheat ears after the normalization of the normal reflectance spectra

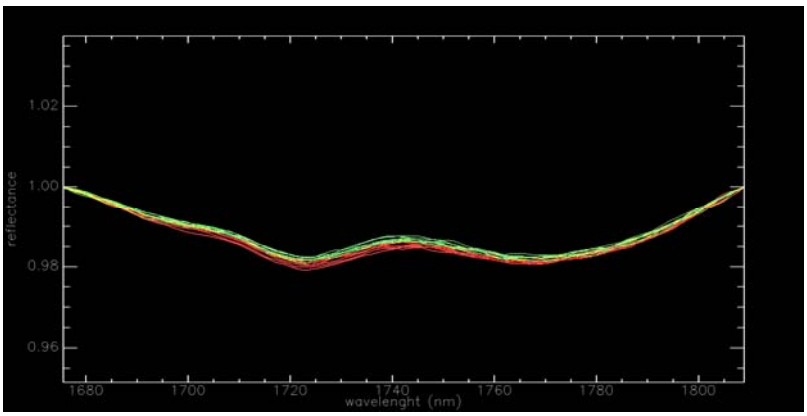


Figure 8. At this zoom level the characteristic interval makes possible to separate the curves of nitrogen treated and untreated wheat ears

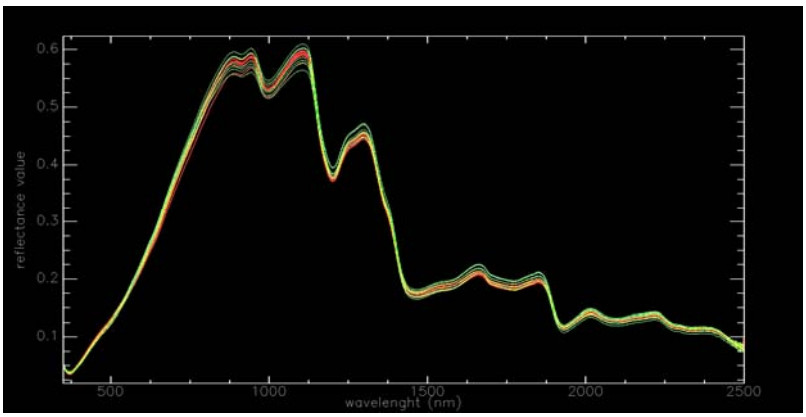


Figure 9. The normal reflectance curves of the kernels are mixed, thus the different treatments cannot be separated

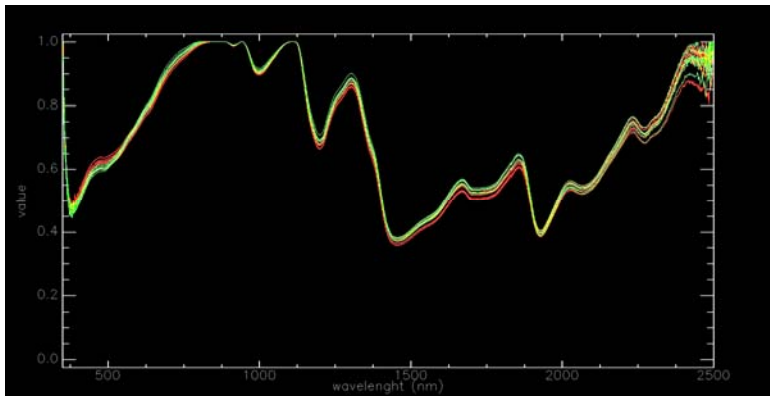


Figure 10. The continuum removed curves of the kernels after the normalization of the normal reflectance spectra

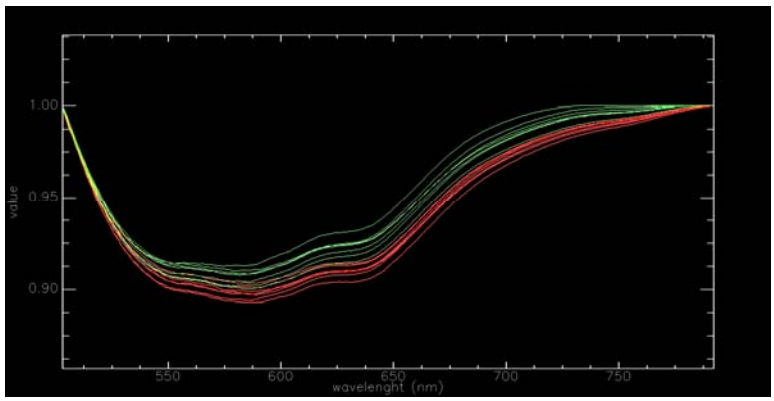


Figure 11. At this zoom level the curves of the nitrogen treated and untreated kernels can be separated in the characteristic wavelength range

As the normal reflectance curves – likewise in case of wheat ears – did not show any consequent correlation according to the nitrogen dose (Figure 9.), we repeated the above mentioned process of the continuum removal (Figure 10.) of reflectance curves.

As a result, curves showed a characteristic range from 500 to 800 nm. Distinct difference in the wavelength of 627 to 775 nm was identified (Figure 11.).

The pre-procession of data were made with ViewSpecPro software and further process steps were carried out with ENVI image analyser software. We used continuum removal to normalize reflectance spectra in order to compare individual absorption features from a common baseline.

4. Conclusions

The differing quantitative and qualitative parameters of the wheat variety generated by the diverse nitrogen fertilizer treatment resulted in spectral

differences which was detectable with spectroradiometer. Among the analysed reflectance curves after removing the continuum of the spectra we found two characteristic intervals. According to the wheat ears at the wavelength range of 1700 nm to 1800 nm and in case of the kernels 500 nm to 800 nm at which the treatments became distinguishable. By evaluating the most important parameters of the winter wheat such as yield, protein, wet gluten content and farinographic value with conventional laboratory technology the interrelation between spectra and nutrition application can be clarified. After the appropriate calibration and validation process the spectral methodology can greatly assist in describing and tracking the current dynamics of nutrient supply and plant up-take in a fast and economic way. The defined correlations based on laboratory 'ex-situ' measurements can be greatly implemented in the 'in situ' in-field and airborne hyperspectral imaginary resulting in time and cost effective, precise sampling method which can exquisitely be applied in the modern agriculture.

Acknowledgement

The authors would like to thank the financial support received from the Szent István University, Gödöllő and the Hungarian Institute of Agricultural Engineering.

References

Csorba, Á. – Jordán, Gy. (2010): Preliminary Results of Airborne and Ground-Based Hyperspectral Mineral Mapping of Acidic Mine Waste in the Recsk Mining Area, Hungary. Contributii Stiintifice in Tehnologii si Echimpamente Pentru Evaluarea si Protectia Mediului. Simpozium National, Arcalia (Bistrita-Nasaud), September 24 - 26, Caiet de rezumate, 33 - 34 pp.

I. Balla, Cs Mátyás., P. K. Gergő, J. Márton, M. Neményi (2010): Site-specific Nutrient Replenishment For Winter Wheat (Triricum aestivum L.). Proceedings of the IX. Alps-Adria Scientific Workshop. Špičák, Czech Republic 629-632 pp.

Erdélyi É., Boksai D., Szenteleki K., Hufnagel L., (2009): The role of biomass in mitigation of global warming. CIGR Symposium . 2009.09.1-4., Rosario, Argentina.

Erdélyi, É. (2009): Sensitivity to Climate Change with Respect to Agriculture Production in Hungary (2009) Precision Agriculture '09 Edited by: E.J. van Henten, D. Goense and C. Lokhorst, Wageningen Academic Publisher, p. 559-567.

Erdélyi, É., (2008): The potential impacts of climate change on main field crops and their yields, case studies in Hungary. "Klima - 21" Füzetek 55 (English Special Edition): 53-79.

Fenyvesi L. (2008): Characterization of the soil – plant condition with hyperspectral analysis of the leaf and land surface, Cereal Res. Com., (Supp 5) 659-663 pp.

I. Virág; M. Csiba; G. Milics; N. Smuk; M. Farouk; M. Neményi (2010): Efficiency of the variable rate fertilization technology int the aspects of agricultural practise. 18th International Poster Day and Institute of Hydrology Open Day. Transport of Water, Chemicals and and Energy in the Soil-Plant-Atmosphere System. Slovenská republika, Bratislava, 2010. november 11., (szerk: A. Čelková) ISBN 978 80 89139 21 7, pp. 612-616.

Jolánkai M., (2004): Szántóföldi növények vetőmag-termesztési technológiája Eds:Izsáki Z, Lázár L. (2004): Szántóföldi növények vetőmagtermesztése és kereskedelme. Mezőgazda Kiadó, Budapest 183 p.

Jung, A., Kardevan, P., Tőkei, L. (2006): Hyperspectral Technology in Vegetation Analysis. Progress in Agricultural Engineering Sciences Vol 2 (No.1), December, 93-115. http://www.akademiai.com/content/276511331ql76777/

Jung, A., Tőkei, L., Kardeván, P. (2006): Application of airborne hyperspectral and thermal images to analyse urban microclimate. Applied Ecology and Environmental Research. Applied Ecology and Environmental Research 5 (1): 165-175 http://www.ecology.kee.hu/

Kardeván P. (2010): A távérzékelési technológia jellemzése. Távérzékelési előadás FVM MGI 2010

Klupács, H. – Tarnawa, Á. – Balla, I. – Jolánkai, M.: (2010.) Impact of water availability on winter wheat (Triticum aestivum L.) yield characteristics. Agrokémia és Talajtan vol. 59 no. 1 (2010), 151-156 pp.

Kristóf D. (2005): Távérzékelési módszerek a környezetgazdálkodásban. Doktori értekezés 9 p.

Lágymányosi A. and Szabó I.: Calibration procedure for digital imaging. Synergy and Technical Development (Synergy2009), Gödöllő, Hungary, 30. August – 02. September, 2009. CD-ROM Proceedings

Máthé, L. – Pillinger, Gy. – Kiss, P. (2010): Vályogtalaj mechanikai jellemzőinek vizsgálata a nedvességtartalom és ülepedettség függvényében. XV. Fiatal Műszakiak Tudományos Ülésszakának konferencia kiadványa, Kolozsvár, Románia, 2010, pp. 201-204. ISSN 2067-6 808

Milics G., Burai P., Lénárt Cs. (2008): Pre-Harvest Prediction of spring barley nitrogen content using hyperspectral imaging. Cereal Research Communications, Akadémiai Kiadó, Volume 36, 1863-1866 p. Proceedings of the VII. Alps-Adria Scientific Workshop. Szlovákia, Stara Lesna. 2008. április 28. – május 2.

Milics G.; Virág I., Farouk M. A., Burai P., Lénárt Cs. (2010): Airborne hypersrectral imaging for data collection for resilient agro-ecosystems. 9 th Alps-Adria Scientific Workshop. Növénytermelés. Špičák, Czech Republic,

2010. 04. 12-17., Edited by M. Harcsa. Akadémiai Kiadó, Vol. 59., pp. 593-596.

Nádasy E., Nádasy M., (2006): Some harmful or useful environmental effects of nitrogen fertilizers. Cereal Research Communication 34. (1): 49- 52 pp.

Nagy V.; Štekauerová V.; Šútor J.; Milics G. (2008): Felszín alatti vízkészletek – a talajnedvesség mint megújuló energiaforrás. A fenntartható fejlődés és a megújuló természeti erőforrások környezetvédelmi összefüggései a kárpát-medencében c. nemzetközi konferencia. (lektorált konferenciakötet) Pécs. 37-46. p. Kiadja az MTA Regionális Kutatások Központja. ISBN 978 963 9899 05 6

Szalay D. K., Farkas I., Szalay D. (2009): Evaluation of nutrient supply as abiotic stressor on winter wheat Triticum aestivum L. performance. Cereal Research Communication 37. (2): 21- 24 pp. Neum, Bosnia-Herzegovina

Tarnawa, Á. – Klupács, H.: (2006.) Element and energy transport model for an agricultural site. Cereal Research Communications, Vol. 34 No. 1 pp 85-89,

Tolner L., Vágó I., Kovács A., Tolner I., Füleky György (2010): Energy forests for the environmentally compatible nutrient management. 40318 Konferencia kiadvány Gyöngyös

Zs. Szentpétery, A. Máté, I. Farkas (2010): Climatic Stress and Resilience Related to Winter Wheat (Triticum aestivum L.) Yields in Different Cropyears. Proceedings of the IX. Alps-Adria Scientific Workshop. Špičák, Czech Republic 207-2010 pp.

Szalay D. K., Jassó J., Szalay D., (2010): Evaluation of resilience threshold concerning precipitation and the Hagberg-falling number. Crop Production, Volume 59. Špičák, Czech Republic 397-400 pp.

Yang C., Everitt J. H., Bradford J. M., Murden D. (2009): Comparison of airborne multispectral and hyperspectral imagery for estimating grain sorghum yield, Transaction of the ASABE, 641-651 pp.

Function diagnostics of complex systems by fluctuations

Oliver SZASZ, Gyula VINCZE, Andras SZASZ
External Department of Biotechnics, Institute for Mechanics and Machinery

Abstract

The power-density function of noise spectrum of open and complex systems changes by the power of frequency. We show that the fluctuation origin and the noise-powered description are equivalent to describe the coloured noise power-density. This observation offers an invariant quantity characteristic of the complex systems to specify the forecast of failure, the timing of regular corrections and/or the assessed operation life of the system. These considerations are well applicable for living systems and their preventive care.

Keywords

power-density, complex-systems, regular-maintenance, system-invariant, fluctuation-diagnostics

1. Introduction

Test of proper function of complex systems during their long-term service, their ageing, wearing and tearing is of vital importance in the respect of preventive maintenance and operation life. The problem is rather complex and complicated, as we have to conclude on the properties of the given system from the results of its test: we have to specify the characteristic values of the type in question by using the operation data of several single parts. This problem can be practically traced back to a single root: the used systems are open, and connected to their environment through a number of interactions; so, it can not be considered as closed, even for the duration of a measurement. They are definitely open from energetic points of view, (energy exchange with the environment). Interactions indispensable for the operation (on which the effect of the system is directed, retroactive effects); influences of environment (environmental loads, e.g. temperature, contamination, pressure, rain etc.), as well as the effect of user's habits and conditions (e.g. early morning usage, usual usage order, effects of usual intensity, direction etc.) are all effecting the ageing of the actual system. A permanent control and maintenance services need to keep the complex function active. We will show, that the characteristic values of energy input, (feeding and take out); as well as the non-deprivable substantial characteristics, are able to check the actual process in general.

Measurements of any dynamic effects are always noisy. The desired signal (electrical, mechanical, etc.) and the actually measured one differ. The

measurement clarity is characterised in these cases by the signal/noise ratio. Dynamical effects and changes could be noise-free only in the case of very simple and reversible cases (at the energetically closed systems). This is practically a theoretical idealization, because in the reality the noise is always present as the random or systematic fluctuation of the given signal (measured, set, used etc.), Robinson (1974), Freeman (1958).

The noise source is composed of many-sided interactions; the continuous energy and entropy/information exchange of open dynamical systems, the mutual dependence of the single subsystems, and the actual noise spectrum are formed in a synergetic way, Reif (1965). Consequently, the desired effect is accompanied in every real case by the noise spectrum composed of the specific features of the dynamical systems. Thus, the noise is a certain degree of appearance of parameters, processes, dynamical behaviour etc. arising always, but not directly involved in the given examination.

In the course of the usual wearing tests and quality examinations, each element of the system is examined separately by using several sensors, and during this measurement one tries to eliminate or minimize the noise. Consequently, the aim at these measurement procedures is to filter the noises and create the best possible signal-to-noise ratio in order to get the most exact information possible regarding the given partial system.

In the case of open, dissipative systems (basically, every occurrence realizing not spontaneous thermo-dynamical changes, e.g. heat engines, biological systems, electromagnetic radiators etc.) the reduction of noise is impossible by fixing the interactions, because the open, dissipative feature assumes the definite interaction with the environment. For this reason, at the real, irreversible dynamical systems we may consider only the second possibility, namely, we have to reckon with noise anyway, and – at the most – the chosen dynamical methods may suppress the noise and bring out the „useful" signal as far as possible.

The noise, however, provides information on the interactions (inside and/or outside the system) of the examined system. In this case the measured signal is not the useless noise, but the fluctuation properties, which could carry the systemic changes in the complex system.

Our objective to obtain information on the whole dynamics of the complex systems in order to be used for the planning processes and qualitative examinations. Our conception is based on the recognition that the whole dynamics is included in the noise, and practically all those dynamical variables appear therein the interactions having a share in the creation of the given (desired/useful) signal. Moreover, the noise spectrum gives account of the correlations within the system. Therefore, the operation of the whole system can be analysed from its noise spectrum. All the failures arising because of wearing, tearing and fatigue processes (in general through stochastic changes) results in the continuous change of noise spectrum. Therefore, the assumption of noise spectrum allows the prediction of the wearing and tearing (fatigue etc.) processes.

2. Simple derivation and description of coloured noises

In spite of the fact that a complex system has a great exchange of information with their environment, these can be characterized by a stationary state in general, that is, by a state in dynamic equilibrium. Therefore, the dynamic equilibrium characterizing the appropriate operation can describe the time-dependent effect (H(x,t)) by a fluctuation around the average, that is:

$$H(x,t) = \langle H(x,t) \rangle + \delta(H(x,t)) \tag{1}$$

where $\langle H(x,t) \rangle$ denotes the averaging, and $\delta(H(x,t))$ is the actual deviation from the average (fluctuation). (Of course, there are also dynamical and non-equilibrium systems (e.g. explosives), however, their effect measurable as a rough average (e.g. relative destroying effect measurable e.g. in the dynamite equivalent) is specified also by a fluctuation around the average.) Because, later on, we are going to examine the time behaviour of the process with a specified (fixed) x, therefore the variable x will not be indicated hereafter.

The process is random if the variable is stochastic, and in this case the power density function of process H is

$$S_H(f) = \frac{\left| H(f) H^*(f) \right|^2}{\Delta f} \tag{2}$$

where Δf is the effective band-width of the Fourier integral, * denotes a conjugate and

$$H(f) = \int_0^\infty \exp(-2i\pi ft) H(t) dt \tag{3}$$

In other wording, Voss (1993):

$$S_H(f) = \mathrm{Re}\left\{ \int_0^\infty \exp(-2i\pi ft) C_H(t) dt \right\} \tag{4}$$

where $C_H(t) = \langle \delta(H(t)) \cdot \delta(H(0)) \rangle$ is the autocorrelation (pair-correlation) function of process H between two points of time, that is:

$$C_H(t) = \left\langle \left(H(t) - \langle H(t) \rangle \right) \left(H(t_0) - \langle H(t_0) \rangle \right) \right\rangle = \langle H(t_0) H(t_0 + t) \rangle - \langle H(t_0) \rangle^2 \tag{5}$$

The functions $S_H(f)$ and $C_H(t)$ are naturally not independent, as in addition to (4), on the basis Wiener-Khintchine relationship, Robinson (1974), Reif (1965) the following is valid:

$$C_H(t) = \mathrm{Re}\left\{\int_0^\infty \exp(-2i\pi ft) S_H(f)\,df\right\} \tag{6}$$

If the $C_H(t)$ correlation function decays by the time constant τ (of course, this is a requirement in the majority of real cases), namely:

$$C_H(t) = \exp\left(-\frac{t}{\tau}\right) \tag{7}$$

then

$$S_H(f) = \frac{\tau}{1+(\omega\tau)^2} \qquad [\omega = 2\pi f] \tag{8}$$

If $(\omega\tau)\ll 1$ (this is valid in the case of fast decay of correlation or on low frequencies), then S(f) is constant, the noise is independent of the frequency and we get the so-called white noise.

If multiple fluctuations exist with randomly fluctuated time constants of the correlation (7), and $D_H(\tau)$ denotes its distribution for the given process H, then

$$S_H(\omega) = \mathrm{Re}\left\{\int_0^\infty \int_{\tau_{min}}^{\tau_{max}} \exp(-i\omega t)\exp\left(-\frac{t}{\tau}\right) D_H(\tau)\,d\tau dt\right\} = \int_{\tau_{min}}^{\tau_{max}} \frac{\tau}{1+(\omega\tau)^2} D_H(\tau)\,d\tau \tag{9}$$

If the $D_H(\tau)$ function is scale invariant, namely, if for example, van der Ziel (1950)

$$D_H(\tau)\,d\tau = \frac{d\tau}{\tau} \tag{10}$$

then we get the so-called pink noise (Flicker noise, 1/f noise, etc.):

$$S_H(f) = \frac{1}{f} \tag{11}$$

In the given multiple fluctuation case the relationship (9) can not be normalized, so

$$\int_0^\infty D_H(\tau)\,d\tau \neq const. \tag{12}$$

therefore a frequency cut-off shall be carried out at the high and low boundaries. We can demonstrate as well (Shlesinger (1987)) that the lognormal distribution of $D_H(\tau)$:

$$D_H(\tau) = \frac{1}{\pi\tau\sigma}\exp\left(\frac{-\left[\log\left(\frac{\tau}{\langle\tau\rangle}\right)\right]^2}{2\sigma^2}\right) \tag{13}$$

results 1/f noise, like it is shown in (11). In this case, there is not any normalization problem and the cut-off is uncertain. We may demonstrate as well (Shlesinger (1987)) that the distribution of the product of random distribution variables is always lognormal, thus it results every time in (11). However, if there is not any dominant lognormal distribution, but the system can be described by a wide range of distributions, then

$$S_H(f) = \frac{1}{f^{\alpha}} \tag{14}$$

where basically we can not fix the value of α because it can considerably depend also on the general parameters of the system (e.g. temperature, pressure etc.), Shlesinger (1987). In this way, the slope of the log-log scale representation of (14), we get directly the character of the noise of process H.

$$\alpha = \frac{\log[S_H(f)]}{\log[f]} \tag{15}$$

The changing noise of the dynamical variables – as the spectrum characterizes the system – assumes the existence of a certain order and self-organization in the examined system. The self-organization comes into existence by the mutual determinacy and during the complex operation of partial systems building on one another, requiring the work of others and determining the dynamics of the other ones by causality. The complex systems with many-sided connections to their environment are in a non-equilibrium, non-stationary state, and have a high-level hierarchical structure. The subsystems forming the structure are connected to each other in many ways through their physical, chemical processes or other information network. The amplitude of physical, chemical, information signals generated by various individual subsystems, their characteristic time or other properties can change in a wide range. E.g., the most simple biological systems show a great variety of processes on the individual characteristic time scales and are connected to each other by scaling, Vicsek (2001), Brown (2000), Calder (1996). As we showed, the noises carry dynamical information about the operating systems, and may give information on the wearing phases of the given structure. In general the noise of any open dynamical systems could be scaled by $1/f^{\alpha}$ spectrum, Li (1989).

3. Fluctuation of diagnostic quantities

Every complex system can be decomposed into numerous simple subsystems the state of those can be described by some physical parameters that characterises the subsystems.

This means that the state of the whole system is known, if we know the state of its every subsystem. Let us denote by $\overline{X}$ the vector made of the state parameters of subsystems, hereinafter called the micro-state of the system.

In terms of diagnostics the limited number F_i values are of measurable quantities and characteristic of the macro system, on the base of which we are able to judge if the functioning of the system complies with the requirements or not.

These quantities are called macroscopic diagnostic state parameters, and the $\overline{F}$ vector made of them is the so-called diagnostic state vector.

As the number of these vectors is significantly fewer than the number of state parameters serving for the description of the micro-state, therefore, from microscopic point of view the system description by applying the diagnostic state parameters is not complete.

Let us assume that the functions of diagnostic state parameters can be described as a function of micro-state, and seeing that the equipment interacts with the external parameters denoted by the vector $\overline{Y}$, thus:

$$F_i = F_i(\overline{X}, \overline{Y}), (i = 1,2,...,n) \tag{16}$$

As the number of microstates large ($\dim(\overline{X}) >> n$), therefore, with the knowledge of diagnostic state parameters we may predicate not more than statistical statements regarding the micro state of the system characterized by the diagnostic state vector, since many kinds of micro states may belong to the same macro-state, and these micro-states can quickly change in time.

This means that we can specify at most the probability that the micro-state falls into the $(\overline{X}, \overline{X} + \overline{dX})$ interval at time t with the probability of $w(\overline{X}, t)$, that is:

$$P(\overline{X} < \overline{\xi} \le \overline{X} + \overline{dX}) = w(\overline{X}, t)dX \tag{17}$$

As the micro-state of the system may rapidly change in time, therefore, the diagnostic state parameters of () fluctuate in time, consequently, they are stochastic variables.

Such variables can be characterized in the simplest way by the

$$\langle F_i \rangle = \int_{(X)} F_i(\overline{X}, \overline{Y}) w(X, t) dX, \quad (i = 1,2,...,n) \tag{18}$$

mean value and the

$$\sigma_{F_i} = \sqrt{\left\langle \left(F_i - \left\langle F_i \right\rangle\right)^2 \right\rangle}, \quad (i = 1,2,...,n) \tag{19}$$

(16) mean-square deviation.

In accordance with the Tshebyshev theorem the probability of $\left|F_i - \left\langle F_i \right\rangle\right| > a$ is

$$P\left(\left|F_i - \left\langle F_i \right\rangle\right| > a\right) \leq \frac{{\sigma_{F_i}}^2}{a^2} = \frac{\left(F_i - \left\langle F_i \right\rangle\right)^2}{a^2} \tag{20}$$

if σ_{F_i} is very small, then we may conclude from the above inequation that the probability of the deviation is small, in this way, F_i and the average of (18) coincides in practice.

If the above case is not true, we may choose the procedure of characterizing the

$$f_i = F_i - \left\langle F_i \right\rangle, \quad (i = 1,2,...,n) \tag{21}$$

functions expressing the stochastic fluctuations. In the engineering practice these are characterized by the power-density spectrum.

Additional advantage of this description is that we may conclude from the distortion of power spectrum density to the occurrence of some future error, even if on the basis of the average of diagnostic state vector the system can be considered as adequate.

4. Stochastic description

Let us suppose that the fluctuation introduced before can be divided into the sum of quasi-periodic stochastic processes of different time scale that are statistically independent. It is clear that the quasi-periodic stochastic processes of different time scale have different frequency scale as well.

We assume that every component process like this is statistically self-similar.

The $X(t)$ stochastic process is memory-less if the increment of

$$X(t + dt) - X(t) \tag{22}$$

can be expressed in the form of

$$X(t+dt)-X(t)=\Theta[X(t),t,dt] \tag{23}$$

In general, this is a Markov process (Jaynes (2003)).

Let us assume that $\Theta[X(t),t,dt]$ is a smooth function of the X,t,dt variables and $X(t)$ is continuous, then:

$$\lim_{dt\to 0} X(t+dt)=X(t) \tag{24}$$

The stochastic process is self-similar in the sense of Gillespie (1996) if the difference can be divided into the sum of statistically independent increments. Then, they have normal distribution within the interval. Here we may see also the Markov character: memory-less and recursive.

$$\begin{aligned} X(t+dt)-X(t)=\Theta[X(t),t,dt]=\sum_{i=1}^{n} X\left(t+i\frac{dt}{n}\right)-X\left(t+(i-1)\frac{dt}{n}\right)= \\ =\sum_{i=1}^{n}\Theta\left[X\left(t+(i-1)\frac{dt}{n}\right),t+(i-1)\frac{dt}{n},\frac{dt}{n}\right] \end{aligned} \tag{25}$$

Since dt can be chosen as arbitrarily small, therefore, the $t_{i-1}=t+(i-1)\frac{dt}{n}$ times can approach t arbitrarily by choosing a suitable high value for n. Therefore, we get from our above equation for adequately high n by utilizing the continuity that

$$\begin{aligned} &t_{i-1}\to t, X(t_{i-1})=X(t), \\ &\Theta[X(t),t,dt]=\sum_{i=1}^{n}\Theta_i\left[X(t),t,\frac{dt}{n}\right] \end{aligned} \tag{26}$$

Here, we may consider the $\Theta_i\left[X(t),t,\frac{dt}{n}\right]$ expressions as the representations of $\Theta\left[X(t),t,\frac{dt}{n}\right]$ variables. These are statistically independent because the process memory-less. Since n is arbitrarily high, therefore, we may conclude from the central limiting distribution theorem that $\Theta[X(t),t,dt]$ is the sum of n statistically independent $\Theta_i\left[X(t),t,\frac{dt}{n}\right]$ stochastic variables. That is,

the stochastic variable has normal distribution. In accordance with the above, this is true also for the $\Theta\left[X(t),t,\frac{dt}{n}\right]$ stochastic variables.

We may conclude to the following properties from the properties of stochastic variables with normal distribution.

$$\begin{aligned}\left\langle\Theta[X(t),t,dt]\right\rangle &= n\left\langle\Theta\left[X(t),t,\frac{dt}{n}\right]\right\rangle\\ \sigma^2\left(\Theta[X(t),t,dt]\right) &= n\sigma^2\left(\Theta\left[X(t),t,\frac{dt}{n}\right]\right)\end{aligned} \tag{27}$$

where $\langle\rangle$ denotes the mean-value and $\sigma^2()$ is the mean-square deviation. Solution of the functions equations:

$$\begin{aligned}\left\langle\Theta[X(t),t,dt]\right\rangle &= A[X(t),t]dt\\ \sigma^2\left(\Theta[X(t),t,dt]\right) &= D[X(t),t]dt\end{aligned} \tag{28}$$

where A and D are smooth functions of X and t, and D is positive. By taking into consideration the normality of $X(t+dt)-X(t)=\Theta[X(t),t,dt]$ and the above results, we get from (23) that

$$\begin{aligned}&X(t+dt)-X(t)=\Theta[X(t),t,dt]=\mathbf{N}[A(X,t)dt,D(X,t)dt]=\\ &=A(X,t)dt+D^{\frac{1}{2}}\mathbf{N}(0,1)dt^{\frac{1}{2}}\end{aligned} \tag{29}$$

where $\mathbf{N}(0,1)$ is a normal distribution stochastic process of zero average and unitary mean-square deviation:

$$N(x)=\frac{1}{\pi\sigma x}\exp\left(-\frac{\ln^2\left(\frac{x}{\langle x\rangle}\right)}{2\sigma^2}\right) \tag{30}$$

If we change over to the differential equation, we get the

$$\frac{dX}{dt}=A(X,t)+D^{\frac{1}{2}}(X,t)\Gamma(t) \tag{31}$$

non-homogeneous equation, where

$$\Gamma(t) = \lim_{dt \to 0} \mathbf{N}(0, dt^{-1}) \tag{32}$$

is a normal distribution white noise. This is a generalized Langevin equation.

Let us take the most simple one from these stochastic processes:

$$\frac{dX}{dt} = -\frac{1}{\tau} X + D^{\frac{1}{2}} \Gamma(t) \tag{33}$$

describes the Ornstein-Uhlenbeck stochastic process, Ornstein-Uhlenbeck (1930). The mean value decays exponentially, and there is a white noise thereon that drives it. This equation describes the noise of a system comprising an energy accumulator (e.g. mass, revolving mass, capacitor, inductivity) and a linear attenuation (e.g. resistance of medium, ohmic resistance) excited by white noise.

On the basis of simple consideration that the power spectral density of the Ornstein-Uhlenbeck process:

$$S(\omega, \tau) = \frac{D\tau^2}{1 + (\omega\tau)^2} \tag{34}$$

where τ is the time constant of the system, and the spectral density is similar to (8). At the same time, this can be considered as the natural time scale of the stochastic process. Let us introduce a frequency scale by applying the definition:

$$\lambda = \frac{1}{\tau} \tag{35}$$

To characterize the stochastic processes, let us take that $G(\lambda)d\lambda$ is the number stochastic processes falling into the frequency interval of $(\lambda, \lambda + d\lambda)$, then for the the energy spectrum of the stochastic processes falling into the (λ_2, λ_1) interval of frequency scales we have that

$$S(\omega, \lambda_1, \lambda_2) = \int_{\lambda_1}^{\lambda_2} \frac{DG(\lambda)}{\lambda^2 + (\omega)^2} d\lambda \tag{36}$$

If the distribution is uniform, namely, if

$$G(\lambda)d\lambda = \frac{d\lambda}{\lambda_2 - \lambda_1} \tag{37}$$

we get that

$$S(\omega, \lambda_1, \lambda_2) = \int_{\lambda_1}^{\lambda_2} \frac{DG(\lambda)}{\lambda^2 + \omega^2} d\lambda = \begin{cases} D, ha \quad 0 < \omega << \lambda_1 << \lambda_2 \\ \dfrac{D\pi}{2\omega(\lambda_2 - \lambda_1)}, ha \quad \lambda_1 << \omega << \lambda_2 \\ \dfrac{D}{\omega^2}, ha \quad \lambda_1 << \lambda_2 << \omega \end{cases} \tag{38}$$

This is a well-known result with the effect of white, pink and Wiener's noise in the first, second and third interval, respectively.

We can choose a time interval from a representation where the noise is similar to the original one, and within this we may choose an interval where the noise is similar to the noise of the interval from where we carried out the previous selection. Mathematically, this means that we are able to carry out the scaling of the frequency of the chosen component noise in such a way that it will be similar to a noise component of other frequency scale. It follows that the distribution function can be scaled in a self-similar way. (Of course, this can not be applied to every scale, but we can find a scale whereon the distribution function of the system can be scaled.)

$$G(\lambda) = \frac{\gamma}{N} G\left(\frac{\lambda}{N}\right) \tag{39}$$

This means that the distribution function can be overlapped with the distribution function taken on the λ scale by the uniform enlarging of ordinate values on the $\frac{\lambda}{N}$ frequency scale.

We may see easily that the solution of the above functions equation takes the form of

$$G(\lambda) = \frac{A(\lambda)}{\lambda^{1+\alpha}}, \quad \alpha = \frac{\ln \frac{1}{\gamma}}{\ln N} \tag{40}$$

where $A(\lambda)$ is the periodical function of the scale whereon the distribution function is self-similar. Namely,

$$A(\lambda) = A\left(\frac{\lambda}{N}\right) \tag{41}$$

For the sake of simplicity, let us suppose that this function is invariable, and calculate again the energy spectrum of the stochastic processes falling into the (λ_2, λ_1) frequency interval. From this we get that

$$\begin{aligned} S(f, \lambda_1, \lambda_2) &= \int_{\lambda_1}^{\lambda_2} \frac{DG(\lambda)}{\lambda^2 + \omega^2} d\lambda = \int_{\lambda_1}^{\lambda_2} \frac{DA}{\left(\lambda^2 + \omega^2\right)\lambda^{1+\alpha}} d\lambda \\ &= \frac{DA}{\omega^{2+\alpha}} \int_{\lambda_1/\omega}^{\lambda_2/\omega} \frac{1}{\left[\left(\frac{\lambda}{\omega}\right)^2 + 1\right]\left(\frac{\lambda}{\omega}\right)^{1+\alpha}} d\frac{\lambda}{\omega} \end{aligned} \tag{42}$$

The integral can be expressed by using the hyper-geometric functions; however, it is not easy to find out a descriptive meaning. For this reason, let us perform the integration for the $(0, \infty)$ interval. With the exception of the pink noise, the result will be a finite constant. We get the expected result by using this approach.

$$\begin{aligned} &S(\omega, \lambda_1, \lambda_2) = \\ &= \frac{DA}{\omega^{2+\alpha}} \int_{\lambda_1/\omega}^{\lambda_2/\omega} \frac{1}{\left[\left(\frac{\lambda}{\omega}\right)^2 + 1\right]\left(\frac{\lambda}{\omega}\right)^{1+\alpha}} d\frac{\lambda}{\omega} \approx \\ &\approx \frac{DA}{\omega^{2+\alpha}} \int_{0}^{\infty} \frac{1}{\left[\left(\frac{\lambda}{\omega}\right)^2 + 1\right]\left(\frac{\lambda}{\omega}\right)^{1+\alpha}} d\frac{\lambda}{\omega} \propto \\ &\propto \frac{1}{\omega^{2+\alpha}} \end{aligned} \tag{42}$$

Consequently, the self-similar function is the condition for getting the

$$S(\omega) \propto \frac{1}{\omega^{\beta}} \qquad (\beta = 2 + \alpha) \tag{43}$$

power spectral density.

How can these distribution functions be considered as universal? In order to prove the universality of this result we need to carry out a universality test. We are going to show that this can be derived from any distribution function converging to zero in the case of high λ frequencies.

Let us take indeed that $g(\lambda)$ is a function like this, namely, this is a fully general, not necessarily lognormal function, where

$$\lim_{\lambda\to\infty} g(\lambda) = 0 \tag{44}$$

In the dynamic events of a collective system, the movements by time are built strictly on each other, the whole course can be derived only in a recursive way (causality principle). Philosophically, this means that the time of a given process or system can be described only by the superimposing order of interactions (not by the order of side-by-side existence). This means that the interactions are built on each other and set off the wearing (progress of time) process. Therefore, the time (on single-variable and causality basis) is composed of superimposing and successive recurrent steps. The progress of time corresponds to the process of wearing. This – at the same time – corresponds to the recursive self-organization resulting in the formation of Mandelbrot set. Consequently, we get a specific self-organization characteristic of the system, which makes the relevant system specific, and in a certain respect distinguishable from the other ones. In this manner, this self-similar solution corresponds to the thermodynamic notion of entropy.

According to the recursive causality idea (Harney (2003) Jaynes (2003)) Let us generate a distribution functions using recursive method, as the process was described above:

$$\begin{aligned} g_i(\xi_i) &= \frac{\gamma}{N} g_{i-1}\left(\frac{\xi_{i-1}}{N}\right), \quad (i = 1,2,,) \\ g_0(\xi_0) &= g(\lambda) \end{aligned} \tag{45}$$

By using these functions let us generate the

$$\begin{aligned} G(\lambda) &= (1-\gamma)\sum_{i=1}^{\infty} g_i(\xi_i) = \\ &= (1-\gamma)\left[g(\lambda) + \frac{\gamma}{N} g\left(\frac{\lambda}{N}\right) + \left(\frac{\gamma}{N}\right)^2 g\left(\frac{\lambda}{N^2}\right) + \ldots + \ldots \right] \end{aligned} \tag{46}$$

distribution function. It is easy to show that this complies with the

$$G(\lambda) = \frac{\gamma}{N} G(\frac{\lambda}{N}) + (1-\gamma) g(\lambda) \tag{47}$$

functions equation. In accordance with our limitation for high λ-values (see (44)), the value of $g(\lambda)$ is tending to zero, so we get the functions equation

$$G(\lambda) = \frac{\gamma}{N} G(\frac{\lambda}{N}) \tag{48}$$

which expresses exactly the self-similar property.

4. The generation of coloured noise is not univocal

We derived above the $1/f^{\alpha}$ coloured noise from the Orstein-Uhlenbeck process. Now, we are going to show that the coloured noise could be derived from the Lorentz process as well; expecting that in this case the distributions of the individual time constants will be different.

It follows from this that, contrary to the literature (Schesinger (1987)), the lognormal distribution is not a significant demand for the $1/f^{\alpha}$ noise. In order to prove this, let us take the other most simple process from among the self-similar stochastic ones:

$$\begin{aligned} \frac{dX}{dt} &= -\frac{1}{\tau} X + D^{\frac{1}{2}} \Gamma(t), \\ D &= \frac{D_0}{\sqrt{\tau}} \end{aligned} \tag{49}$$

This is called the Lorentz stochastic process. Here, as we saw earlier, $\Gamma(t) = \lim_{dt \to 0} \mathbf{N}(0, dt^{-1})$ is a white noise of normal distribution. We have seen on the basis of simple consideration that the power spectral density of the process is as follows (similar again to (8)):

$$S(\omega, \tau) = \frac{D_0 \tau}{1 + (\omega\tau)^2} \tag{50}$$

In this case, τ is the time constant of the system generating the stochastic signal. This can be considered as the natural time scale of the stochastic process as well to the effect that this gives information on the character of change of the two-point correlation function of the stochastic process.

Indeed, we know that the power spectral density of the signal equals to the Fourier transform of its correlation function. We get from this and (50) that the two-point correlation function:

$$C_{XX}(\vartheta) = F^{-1}\left[S(\omega,\tau)\right] = F^{-1}\left[\frac{D_0\tau}{1+(\omega\tau)^2}\right] = D_0 e^{-\frac{\vartheta}{\tau}} \tag{51}$$

where F^{-1} denotes the inverse Fourier transformation. Therefore, the degree of correlation decreases exponentially by the τ time constant. Because of this property τ is called the time-correlation length. Let us suppose that $G(\tau)d\tau$ is the number of statistically independent stochastic processes falling into the interval of $(\tau,\tau+d\tau)$ time-correlation length, thus the resultant energy spectrum falling into the $(0,\infty)$ interval is:

$$S(\omega) = \int_0^\infty \frac{D_0\tau G(\tau)}{1+(\tau\omega)^2} d\tau \tag{52}$$

If the distribution is scale-invariant, and if we require only the self-similarity than the probability (e.g. for the density function) is scale independent:

$$G(\tau)d\tau = G(\alpha\tau)d\alpha\tau \Rightarrow \frac{d\alpha\tau}{\alpha\tau} = \frac{d\tau}{\tau} \tag{53}$$

so

$$G(\tau)d\tau = \frac{d\tau}{\tau} \tag{54}$$

Hence by using the

$$\int_0^\infty \frac{1}{1+(\tau\omega)^2} d\tau = \frac{\pi}{2}\frac{1}{\omega} \tag{55}$$

improper integral, from (52) we get the expected result:

$$S(\omega) = \int_0^\infty \frac{D_0\tau G(\tau)}{1+(\tau\omega)^2} d\tau = D_0 \int_0^\infty \frac{\tau\frac{1}{\tau}}{1+(\tau\omega)^2} d\tau = \frac{D_0\pi}{2}\frac{1}{\omega} \propto \frac{1}{f} \tag{56}$$

In more general, if we suppose that

$$G(\alpha\tau) = \alpha^{\beta} G(\tau) \tag{57}$$

then

$$G(\tau) = \tau^{\beta} \tag{58}$$

If we require only the self-similarity, we get from (52) and (58) that the noise spectrum of signals falling into the $(0,\infty)$ interval is:

$$S(\omega) = \int_0^{\infty} \frac{D_0 \tau G(\tau)}{1+(\tau\omega)^2} d\tau = \int_0^{\infty} \frac{D_0 \tau^{\beta+1}}{1+(\tau\omega)^2} d\tau \tag{59}$$

Because of the physical representation, it is advisable to convert the integral into the following form:

$$S(\omega) = \int_0^{\infty} \frac{D_0 \tau^{\beta+1}}{1+(\tau\omega)^2} d\tau = \frac{D_0}{\omega^{\beta+2}} \int_0^{\infty} \frac{(\omega\tau)^{\beta+1}}{1+(\tau\omega)^2} d(\omega\tau) \tag{60}$$

For this integral we are not able to give a general solution. Fortunately, in our case, the improper integral can be expressed in a closed form if $0 < \beta + 2 < 2$, Namely,

$$\int_0^{\infty} \frac{(\omega\tau)^{\beta+1}}{1+(\tau\omega)^2} d(\omega\tau) = \frac{\pi}{2\sin\left(\frac{(\beta+2)\pi}{2}\right)} = A \tag{61}$$

from (60) and (61) we obtain:

$$S(\omega) = \frac{D_0}{\omega^{\beta+2}} \int_0^{\infty} \frac{(\omega\tau)^{\beta+1}}{1+(\tau\omega)^2} d(\omega\tau) = \frac{D_0 A}{\omega^{\beta+2}} \tag{62}$$

Consequently, the self-similar function is the condition for getting the

$$S(\omega) \propto \frac{1}{\omega^{\alpha}} \tag{63}$$

power spectral density. Advantage of the applied Lorenzian process instead of the Ornstein-Uhlenbeck one is its well fixed boundary conditions for integration, no arbitrary cut-off is necessary for finite integrals.

The foregoing can be generalized to a great extent. Namely, if we depart from the stochastic process described by the

$$\begin{aligned} \frac{dX}{dt} &= -\frac{1}{\tau}X + D^{\frac{1}{2}}\Gamma(t), \\ D &= \frac{D_0}{\tau^{\gamma}} \end{aligned} \tag{64}$$

equation instead of (49), where $\Gamma(t) = \lim_{dt\to 0} \mathbf{N}(0, dt^{-1})$ is the normal distribution white noise.

Then, on the basis of simple consideration we can see that the power spectral density will have the

$$S(\omega,\tau) = \frac{D_0\tau^{2-\gamma}}{1+(\omega\tau)^2} \tag{65}$$

form. If we require only the self-similarity, we get from (65) and (58) that the noise spectrum of signals falling into the $(0,\infty)$ interval is

$$S(\omega) = \int_0^{\infty} \frac{D_0\tau^{2-\gamma}G(\tau)}{1+(\tau\omega)^2}d\tau = \int_0^{\infty} \frac{D_0\tau^{\beta-\gamma+2}}{1+(\tau\omega)^2}d\tau \tag{66}$$

Because of the physical representation, it is advisable to convert the integral into the

$$S(\omega) = \int_0^{\infty} \frac{D_0\tau^{\beta-\gamma+2}}{1+(\tau\omega)^2}d\tau = \frac{D_0}{\omega^{\beta-\gamma+3}}\int_0^{\infty} \frac{(\omega\tau)^{\beta-\gamma+2}}{1+(\tau\omega)^2}d(\omega\tau) \tag{67}$$

form.

In our case, the improper integral can be expressed again in a closed form if $0<\beta-\gamma+3<2$. That is:

$$\int_0^{\infty} \frac{(\omega\tau)^{\beta-\gamma+2}}{1+(\tau\omega)^2}d(\omega\tau) = \frac{\pi}{2\sin\left(\frac{(\beta-\gamma+3)\pi}{2}\right)} = A \tag{68}$$

Now, from (60) we have that

$$S(\omega) = \frac{D_0}{\omega^{\beta-\gamma+3}} \int_0^{\infty} \frac{(\omega\tau)^{\beta-\gamma+2}}{1+(\tau\omega)^2} d(\omega\tau) = \frac{D_0 A}{\omega^{\beta-\gamma+3}} \tag{69}$$

Therefore, from the self-similarity we get again the

$$S(\omega) \propto \frac{1}{\omega^{\alpha}} \tag{70}$$

power spectral density of the coloured noise!

From this, we may draw the conclusion that the self-similarity can be considered as a fundamental property in the generation of coloured noises, and the existence of the self-similarity alone is a satisfactory condition for its presence; nor the under-lying stochastic processes nor the distributions have role to generate this phenomenon.

5. Connection between the fluctuation and the induced noise theory

Above we derived the $1/f^{\alpha}$ noise from the noise spectrum of system driven by white noise, while van der Ziel (1950), Shlesinger (1987) and Milotti (2002) derived the coloured noise from the fluctuations. Next, we are going to show the equivalence of these two methods; namely the white noise powered and the fluctuation gained systems are both have coloured noise spectrum.

The thermo-dynamical fluctuations can be characterized by macroscopic fluctuation quantity. The field range in which the fluctuation is generated is not uniform regarding the fluctuation quantity, however, it is in a state of equilibrium in every point. This latter means that among the field ranges the exchange of extensive quantities characteristic of the fluctuation can be neglected during the relaxation time of equilibration. Additional characteristic of the thermo-dynamical fluctuations is that the fluctuation lasts for a finite duration, and the rate of change of the individual a_i, $(i = 1,2,...,n)$ extensive parameters during the fluctuation can be expressed by the extensive quantities participating in the fluctuation.

$$\frac{da_i}{dt} = f(a_1, a_2, ..., a_n), \quad (i = 1,2,..,n) \tag{71}$$

Let us take an extensive the relaxation time of which is much longer than that of the others, then the fluctuation can be described by this sole extensive. Let us

suppose that the equation (71)is linear and the system returns into its equilibrium, then the equation describes a completely deterministic and noiseless fluctuation process of one variable:

$$\frac{da}{dt} = -\lambda a \tag{72}$$

The solution of this equation will be as follows:

$$a(t) = a(0)e^{-\lambda t} \tag{73}$$

Then, the correlation function is

$$f_{aa}(\tau) = \langle a(\tau)a(0)\rangle = [a(0)]^2 e^{-\lambda|\tau|} \tag{74}$$

and the power spectral density of this:

$$S(i\omega) = \int_{-\infty}^{\infty} f_{aa}(\tau)e^{-i\omega\tau}d\tau = [a(0)]^2 \frac{\lambda}{\lambda^2 + \omega^2} \tag{75}$$

How do we have the noise from deterministic conditions? Deterministic process generate a fluctuation and its spectrum is deterministic. The expected noise is not deterministic, and it is not a fluctuation, it is a noise. To make comparison, we suppose that the deterministic fluctuation signal randomly repeats itself. The constructed noise is the series of randomly repeating deterministic signals. If we introduce a white noise function into the deterministic equation (the result is the Langevin equation), then the amplitude of the white noise spectrum can be chosen in such a way that it corresponds to the noise spectrum generated by the deterministic and random repetition frequency. The same is true also to the correlation function. This is a white noise and $1/\omega^2$ Brownian noise for small and high ω values, respectively.

The power spectral density of the random series of such fluctuations differs only in one proportionality coefficient, as we explained in Freeman (1958). If the distribution of λ frequencies is uniform, then the resultant spectrum will have white noise, 1/f and Brownian noise at the first part, in the middle and at the tail, respectively. Schlesinger also departs from this type of fluctuation, but he writes the equation (72) in the equivalent form:

$$\frac{da}{dt} = -\lambda a = -\frac{1}{\tau}a \tag{76}$$

In this case, instead of (75), the spectrum is:

$$S(i\omega) = \int_{-\infty}^{\infty} f_{aa}(\tau) e^{-i\omega\tau} d\tau = [a(0)]^2 \frac{\tau}{1+(\tau\omega)^2} \tag{77}$$

Supposing (like we did before in part No. 1.) that the probability density function of the time correlation length is lognormal, we get the resultant noise spectrum of $1/f^{\alpha}$. We may conclude, that the deterministic nature of this process is not an essential request to get coloured noise spectrum; if we suppose that there is a random series of such fluctuations, in the same way as in (50).

Introducing e.g. a stochastic exciting term into (76):

$$\frac{da}{dt} = -\frac{1}{\tau} a + q(t) \tag{78}$$

We state: the spectrum of the signal shall correspond to the power spectral density of fluctuation (77); and this condition can be always fulfilled. In order to prove this, let us calculate the Fourier transform of the equation (78). Then we get:

$$\left(i\omega + \frac{1}{\tau}\right) a = q(\omega) \rightarrow a(\omega) = \frac{\tau}{1+(i\omega\tau)} q(\omega) \tag{79}$$

From this, we have the power spectral density:

$$S(\omega) = \frac{\tau^2}{1+(\omega\tau)^2} |q(\omega)|^2 \tag{80}$$

We may see that if

$$q(\omega) = \frac{a(0)}{\sqrt{\tau}} \tag{81}$$

is chosen, it leads us to the expected result. Consequently, if q(t) is a white noise of $\frac{a(0)}{\sqrt{\tau}}$ amplitude, then the noise spectrum of the signal is equivalent to the noise spectrum of fluctuation! Moreover, two stochastic processes are equivalent if their noise spectra are the same.

Present results are applicable for such a complex systems as the living organisms. The fractal physiology controls a living system by the time-fractal

analysis, West (1990), Bassingthwaighte (1994), Musha (1994), which is equivalent the above noise/fluctuation approach. The analysis of the normally functioning living organism shows that these noises are self-similar according to their time scale. As it is shown, Wagner (1995), Butler (1993), it is able to discover the abnormalities very early, and/or it is able to check the ageing-status of the human body, Walleczek (2000). We had investigated the method also theoretically earlier, Szendro (2001/a), Szendro (2001/b).

6. Conclusion

Our present study shows a possibility to measure a system-dependent invariant (scale-independent) parameter, which characterizes the actual status of the whole complex system, informs about the interactive "harmony" of the system, makes possible to check the proper function of the system as a complex unit. We observed that the noise contains the entire dynamics and practically every dynamical variable of the whole system, the interactions of which contribute to the generation of the given (desired/useful) signal. Therefore, we may examine the system as a whole, and analyze the operation of system on the base of its noise spectrum. All the faults arising from wearing, tearing, fatigue processes (in general through stochastic changes) will result in the continuous change of noise spectrum. Therefore, the measurement of noise spectrum allows the forecast of wearing/tearing (fatigue etc.) processes. This information facilitate to control the given system among the concretely functioning conditions including its evolutionary trend, predicting the possible failures or life-time thresholds in time of the proper function without statistically determined system-independent data.

References

Bassingthwaighte J.B., Liebovitch L.S. and West B.J. (1994). Fractal Physiology. Oxford Univ. Press, New York – Oxford

Brown J.H., West G.B. (2000) Scaling in Biology. Oxford University Press

Calder III W.A. (1996). Size, Function and Life History. Dover Publications Inc. Mineola, New York

Butler G.C., Yamamoto Y., Xing H.Ch., Northey D.R. and Hughson R.L. (1993). "Heart rate variability and fractal dimension during orthostatic challenges". J. Appl. Physiol. 75: 2602-2612

Freeman J.J. (1958). Principles of Noise. John Wiley & Sons, Inc.

Gillespie D.T. (1996). "The mathematics of Brown motion and Johson noise". Am. J. Phys, 64(3):225-240

Harney H.L. (2003) Bayesian Inference, Springer Verlag, Berlin Heidelberg NewYork

Jaynes E.T. (2003) Probability Theory, The logic of Science, Cambridge University Press, Cambridge

Li W. (1989). "Spatial 1/f Spectra in Open Dynamical Systems". Europhys. Lett. 10:395-400

Milotti E. (2002) $1/f^{\alpha}$ noise: a pedagogical rewiev, arXiv preprint, physics/0204033, E-GLEA-2, Buenos Aires, Sept. 10-14, 2001

Musha T., Sawada Y. (1994). Physics of the Living State. IOS Press, Amsterdam

Uhlenbeck G. E., Ornstein L. S., (1930) On the theory of Brownian motion, Phys. Rev. 36:823—841

Reif F. (1965). Statistical and Thermal Physics. McGraw Hill, New York

Robinson F.N.H. (1974). Noise and Fluctuations. Clarendon Press, Oxford

Schesinger M.F., West B.J. (1988). 1/f versus $1/f^{\alpha}$ noise. In Random Fluctuations and Pattern Growth. Experiments and Models (Edited by H. E. Stanley and N. Ostrowsky) Kluwer Academic Publishers Dordrecht-Boston-London

Shlesinger M.F. (1987). "Fractal time and 1/f Noise in Complex Systems". Ann. New York Acad. Sci. 504:214-225

Szendro P., Vincze G., Szasz A. (2001/a). "Pink-noise behaviour of biosystems". Eur. Biophysics, 30:227-231

Szendro P., Vincze G., Szasz A. (2001/b). "Bio-response on white-noise excitation". Electro-and Magnetobiology, 20:215-229

van der Ziel A. (1950). "On the Noise Spectra of Semi-Conductor Noise and of Flicker Effect". Physica, 16(4):359-372

Vicsek T. (2001). Fluctuations and Scaling in Biology. Oxford University Press

Voss R.F. (1993). "1/f Noise and Fractals in DNA-base Sequences". In: Crilly AJ, Earnshaw RA, Jones H, (Eds.): Application of Fractals and Chaos, The Shape of Things. Springer-Verlag, Berlin, Heidelberg, pp. 7-20

Wagner C.D., Mrowka R., Nafz B. and Persson P.B. (1995). "Complexity and "chaos" in blood pressure after baroreceptor denervation of conscious dogs". Am. J. Physiol. 269: H1760-H1766

Walleczek J. (2000). "Self-organized Biological Dynamics and Non-linear Control". Cambridge Univ. Press

West B.J. (1990). Fractal Physiology and Chaos in Medicine. World Scientific, Singapore, London

Invited Papers

1. Klaus GOTTSCHALK, Csaba MÉSZÁROS and Janine ELLNER
A model for surface drying observation of fruit

The authors' team is research partner of the faculty in the field of energetic research and process modelling.

2. Moira MIRANDA, András EDELMAYER, Sándor MOLNÁR
Federated filtering: classical approaches in new approximations for distributed systems estimation

The authors are research partners of the faculty in control and systems theory and process modelling. They focus both on practical applications and new theoretical approaches to control engineering problems.

3. István Péter SZABÓ, Gábor SZABÓ, Péter SZENDRŐ
Accumulation Temperature of an Experimental PCM Solar Tank

The authors' team is research partner of the faculty in the field of agricultural energetics and solar energy.

A model for surface drying observation of fruit

Klaus GOTTSCHALK[1], Csaba MÉSZÁROS[2] and Janine ELLNER[1]
[1]Institut für Agrartechnik Potsdam-Bornim e.V. (ATB)
[2]Department of Physics and Process Control,
Institute for Environmental Engineering Systems

Abstract

During ventilation of the fruit directly with air, the influences of the developed air flow on the evaporation process around single fruit bodies or clusters of fruits can be observed by using an infrared (IR) thermographic camera. The effect of surface drying is determined by using the evaporation cooling effect along the surface. The fruit surface temperature is visualized by image analysis of the thermographic images to discriminate local flow and transfer effects. Respecting the boundary conditions, the distribution of the air humidity, vapor flow from the surface of the produce to the ambient flowing air and the evaporation process can be calculated. It is demonstrated from the model that the infrared thermographic camera is not measuring the wet bulb temperature from a drying surface as expected but the transient temperature of the real non-adiabatic evaporating process.

1. Introduction

Moist on surfaces of fruit and potatoes may imply risks in development and spreading of microbial organisms which affect storability and shelf live of the produce. Therefore, it is evident to dry off surface moist prior storage or remove condensed water on the surface. Surface moist exist if a film of liquid water is present on the surface of the produce. The moist should be removed or evaporated to the ambient air as quick as possible (transpiration). The transport mechanisms during transpiration are mass diffusion and heat conduction in air as well as convection in flowing air.

The significant process during drying is the mass transfer, i.e. the transition of water vapour from the surface of the individual fruit passing the boundary layer to the surrounding moving air. The heat transfer is similar to this process.
Important parameters to the drying process are therefore the mass transfer coefficient and the heat transfer coefficient, as well as the parameters for the determination of the thermodynamic condition of the ambient air, like temperature, humidity and air flow velocity.

2. Transpiration model

The primary effects on the moisture loss are the diffusion processes in the ambient air and in the produce material. The environmental conditions can easily

be obtained under test. The analytical model gives a tool to investigate the mass loss for the transpiration process of the produce. The model described here is an alternative approach to the studies of the transpiration processes that were worked out before, e.g. by Gottschalk et al., 2007. The modelling method described there is a method to determine the environment influences on the moisture loss.

The moisture loss is experimentally determined by (Linke, 1997a, 1997b)

$$E = \frac{m_0 - m_1}{A \cdot (t_1 - t_0)} \tag{1}$$

where m_0 is the total mass of the fruit with surface area A at start of experiment a time t_0, m_1 is the total mass of the fruit after experimental time t_1, (Linke, 1997b).

This phenomenon follows the diffusion equation

$$j = - D \text{ grad } Y \tag{2}$$

with D as diffusion coefficient $D_{air} = 25.6 \cdot 10^{-6}\ m^2\ s^{-1}$ for diffusion of water vapour through dry air, and specific humidity (moisture content) Y defined as $Y = m_{humid\ air} / m_{dry\ air}$.

The specific mass loss flow E is identified with the mass flow density j into the ambient air at the surface: $E \equiv j$.

Respecting the boundary conditions for the mass flow density j at the surface of the fruit and for Y at a distance r_A from the surface with $r_A >> R_0$ gives the solution of the diffusion equation

$$\partial Y/\partial t = D_{air} (2/r\ \partial Y/\partial r + \partial^2 Y/\partial r^2) = 0 \tag{3}$$

for a sphere of radius R_0 (in spherical coordinates r) which is for stationary state (Figure 1)

$$Y(r) = Y_{sA} \cdot rH \cdot (1 - R_0/r) + Y_{s0} \cdot R_0/r \tag{4}$$

with rH relative humidity of the ambient air. In this model the fruit is idealized to an object with spherical shape.

Using

$$\rho = (1+Y) / (R_A + Y\ R_V) \cdot p/T \tag{5}$$

the mass (vapour) flow density in dry air for stationary state can be calculated using equ. (2):

$$j = - D_{air}\ d\rho/dr = - D_{air} \cdot d\rho/dY \cdot dY/dr \tag{6}$$

(R_A, R_V gas constants for air and vapour, resp.; p air pressure; T temperature)

The solution represents the mass flow from the surface of the fruit into the ambient air at any time.

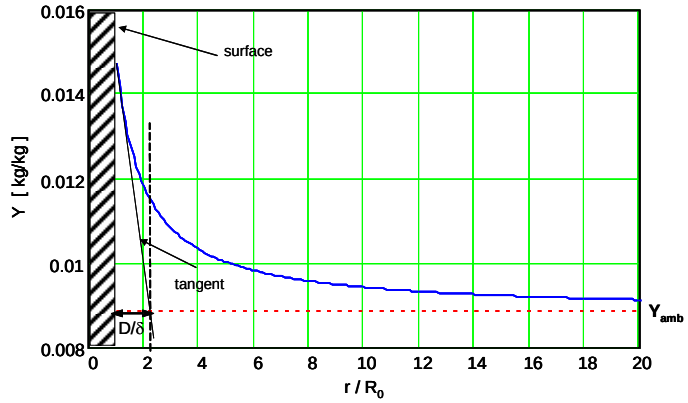


Figure 1. Simulation example of a course of air humidity Y in air apart from an apple surface in units of R_0. Surface and ambient temperature 20 °C, ambient air rel. humidity rH=60 %

The course of the air humidity at distance from the fruit surface (Figure 1) gives an approximation for the boundary layer thickness δ at very low or zero air flow velocity, i.e. at pure diffusion. The air humidity (Y_A) far apart from the surface can be measured easily at a sufficiently far distance r_A, while Y_{R0} close to the surface can be assumed as in the saturation point state at surface temperature (Y_{s0}) as long as the surface is wet.

3. Air flow

During convection still the speed of moving air passing the surface is of importance apart from the air temperature, the surface temperature and the humidity. Air speed can be measured e.g. by means of a hot wire anemometer in the proximity of the surface (boundary layer) or be computed by flow simulation by means of numeric fluid mechanics (CFD = Computational Fluid Dynamic).

Dimensionless numbers are used for better comparison of results. Air flow velocity w (in m s^{-1}) is replaced by the Reynolds number ($Re = w \cdot 2\,R \cdot \nu^{-1}$), with R radius of the spherical produce (for example: R = 35mm), ν kinematic viscosity of air, $\nu = 15.58 \cdot 10^{-6}$ m^2 s^{-1} at 25°C, 1 bar.

The Reynolds number is therefore proportional to the air flow velocity but not constant along the path of the air flow. The fruit is assumed as of spherical shape. A ‘locally variable’ Reynolds number can therefore be defined. For spherical or cylindrical objects the flow profile is symmetrical around the body within an even laminar parallel flow.

At the stagnation point (windward side of the object) the flow velocity is zero, increases around the body, reaches a maximum value close in the zenith and decreases in the leeward side again. In the leeward side either a laminar von Kármán vortex trail or a turbulent flow is developing dependent on the basic flow velocity (or Reynolds number).

The air flow velocity is measured for validation using a hot-wire anemometer, from which the Reynolds number Re is calculated (the velocity w is implied in the Reynolds number.

For estimating the boundary layer thickness $\delta_\beta=\delta$ for mass transfer the boundary layer theory for a plate is taken. A basic flow with velocity w is passing the flat plate along x, Figure 2. The boundary layer (B. L.) thickness is calculated as (Baehr, 2006):

$$\delta_\beta = \frac{5 \cdot \sqrt{x}}{\sqrt{Re_x}} \tag{7}$$

with 'local' Reynolds number defined as $Re_x = w \cdot x \cdot \nu^{-1}$.

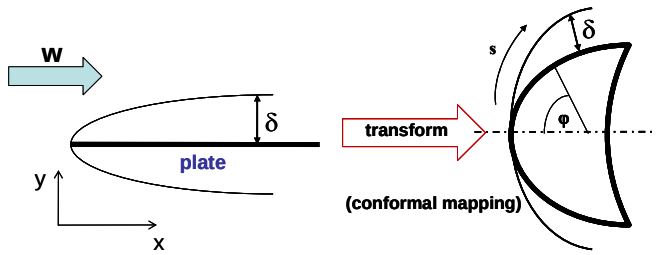


Figure 2. Boundary layer thickness δ approximated from the film theory for a flat plate and transformed to spherical shape

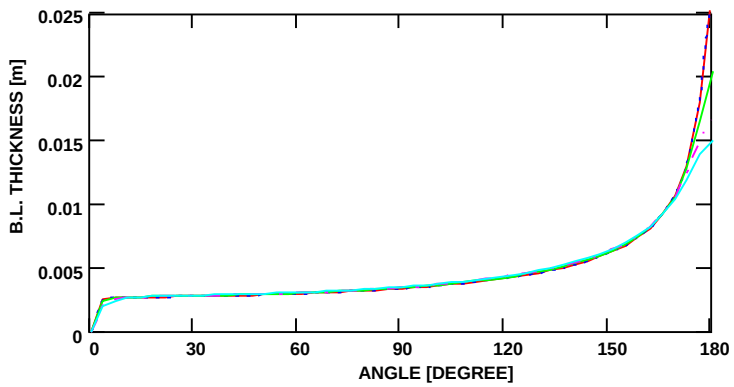


Figure 3. Calculated boundary layer thickness δ along air flow path around half of a fruit of spherical shape

The shape of the plate is transformed to a sphere surface and the B.L. thickness δ re-calculated for a flow passing the spherical surface along the path s = s(φ) with increasing angle φ, Figure 3. As expected, the boundary layer thickness increases significantly along leeward side, while it remains almost constant below 90 degree.

4. Mass transfer

It is still of interest how the air flow is developing around a fruit body when ventilating directly the body and how the transpiration and heat exchange process is influenced by the air flow pattern. Transpiration is a process on which diffused water to the surface is evaporating. As a result, the surface temperature decreases significantly and can be measured with an IR-camera. The evaporated water is diffusing from the surface of the fruit to the ambient air penetrating a boundary layer of humid air. The thickness of the boundary layer is dependent on the velocity of the air passing the fruit surface. The amount of vapour passing the boundary layer is expressed as mass flow density or mass flux *j*, i.e. the flow of mass (water vapour) per time unit and traversed area (units: kg s^{-1} m^2).

The course of the mass flow density (Figure 1) gives an approximation of the boundary layer thickness δ at air flow velocity equal to zero, i.e. if pure diffusion takes place. Hence, the convection term is zero but increases when forced air flow develops. At air flow velocity w>0 the boundary layer thickness δ decreases, which causes the mass transfer coefficient β to increase. This dimensionless definition leads to the Sherwood number Sh

$$Sh = \beta \cdot \delta / D \tag{8}$$

from which the mass transfer coefficient β can be determined if $\delta = \delta_\beta$ is known or can be approximated using the ratio $D/\delta = \beta/Sh$; $D = D_{air}$ the diffusion coefficient (for vapour in air).

The definition for β is

$$\beta = -D_{air} \cdot (\partial Y/\partial r)_{surface} / (Y_{s0} - Y_\delta) \tag{9}$$

(Y_{s0} air humidity at saturation point on the surface, Y_δ at saturation point at the boundary layer distance) from which it can be seen that the ratio D / δ can be read out from Figure 1.

For the example stated here the ratio $D / \delta = \beta / Sh$ is approx. $(1.20…1.22) \cdot R_0$ when neglecting free convection; (R = fruit radius).

For laminar air flow around a sphere, the Sherwood number is found (Baehr, 2006) to be

$$Sh = 0.644\ Re^{1/2}\ Sc^{1/3} \tag{10}$$

using the Reynolds number, the Schmidt number Sc = ν / D and ν the kinematic viscosity. By knowing the velocity field w(s) around the object, the mass transfer coefficient can be determined.

Using the air-flow velocity distribution around the fruit (sphere), measured or modelled, e.g. with CFD, we can obtain the (local) mass transfer coefficient β(φ) for the angle φ along the path half around the fruit $0 \leq \varphi \leq 180°$ (Figure 4 and 5). With this method and applying eq. (7), β was found as approx. $1.8 \cdot 10^{-3}$ m s^{-1} at the zenith of the fruit ('local', at φ=90°), while measured in the range of $2.5 \ldots 3.3 \cdot 10^{-3}$ m s^{-1}, (Linke, 1997a, 1997b).

The mass flow density j is calculated with (compare equ. 2)

$$j = \beta \cdot \rho_{air} \cdot (Y_{s0} - Y_{\delta}) \tag{11}$$

The humidity Y_{δ} at the boundary layer thickness can be approximated with the humidity Y_{amb} at a sufficient high distance from the surface (thus in the ambience); Y_{s0} is the moisture in the saturated area close to the wet surface at surface temperature. These values can be determined from a psychrometric chart (Mollier's h-x chart) for water vapour in air by using the temperature and the relative humidity, which can be easily measured.

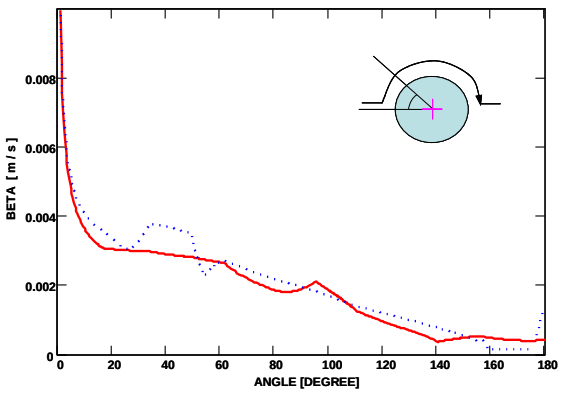


Figure 4. Measured course of the mass transfer coefficient along a spherical fruit

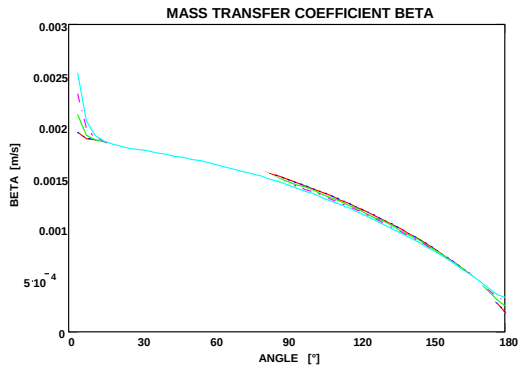


Figure 5. Calculated course of the mass transfer coefficient along a spherical fruit

5. Heat transfer

The heat and mass transfer processes are coupled which is expressed by the Lewis number Le

$$Le = a / D = Sc / Pr \tag{12}$$

(with $a = \lambda_{air} \cdot \rho^{-1} \cdot c_{pg}^{-1}$ thermometric conductivity, D mass diffusion coefficient.). The Lewis-number for water vapour in air is Le ≈ 0.87 and the Prandtl number is Pr ≈ 0.7 for gases, (Baehr, 1996).

The thermal boundary layer thickness δ_T can be approximated by the boundary layer thickness of the mass transfer, thus $\delta_T \approx \delta$.

The dependency of the heat transfer coefficient α from air flow velocity (within the range w = 0.01 to 0.5 m s^{-1}) can be calculated (after Rastovski et al., 1987) as

$$\alpha = 7.8 \cdot \frac{w^{0.6}}{d^{0.4}} \tag{13}$$

with d = 2 R = fruit diameter.

The coupling with the heat transfer can expressed as

$$\beta = Le^{m} \cdot \frac{D \cdot \alpha}{\lambda} \tag{14}$$

with m ≈ 1/3 for gases. With equ. (8) to (10) the mass transfer coefficient β was determined as $\beta \approx 6.4 \cdot 10^{-3}$ m s^{-1} at w = 0.15 m s^{-1}, $D = 26.6 \cdot 10^{-6}$ m^2 s^{-1}, $\lambda_{air} = 0.026$ W m^{-1} K^{-1}.

6. Temperature profile

Thermal imaging (thermometry), using an electronic infrared camera, is a method to determine the drying process when drying off the moisture from a fruit surface (Hellebrand et. al. 2001, Fito et. al. 2004). The temperature distribution along the surface of a wet peach, for example, is shown in Figure 6. The transpiration process of surface moisture lowers the temperature due to evaporative cooling.

The course of the temperature at three different positions (upwind, centre, downwind) along the surface of the moist peach is shown as a thermogram in Figure 7. The evaporation of surface moisture lowers the temperature nearly down to wet bulb temperature. In reality, the heat of the object (fruit body), its

heat capacity and the heat conduction from inside the fruit hinders the surface temperature to drop down to the exact wet bulb temperature.

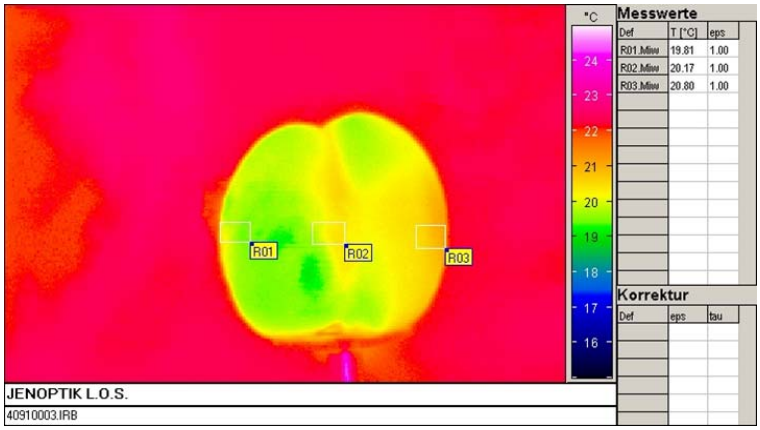


Figure 6. Temperature distribution on the surface of a peach during drying off surface moisture using infrared thermography

As soon as the surface is partially dried off, the temperature increases again to the ambient air temperature value. The movement of a 'drying zone' from the upwind side of inlet air-flow direction to the opposite downwind side can be well recognized, Figure 7 and Figure 8. The temperature distribution development on the surface is dependent on air flow direction and velocity; the drying time is also dependent on air humidity.

During condensation or evaporation of water the surface temperature keeps the permanent *adiabatic* saturation temperature, the so-called 'wet-bulb' temperature θ_s. This temperature can be derived by (Baehr, 2006):

$$\theta_\delta - \theta_s = \frac{\Delta h_v(\theta_s) \cdot \left(Y_s(\theta_s) - Y_s(\theta_\delta) \cdot rH\right)}{c_{pg} + Y \cdot c_{pv}} \qquad (15)$$

with Δh_v latent heat $\Delta h_v(\theta_s) = \Delta h_{v0} - (c_{fl} - c_{pv}) \cdot \theta s$; Y_s humidity at saturation; rH rel. humidity of ambient air; θ_δ temperature at B.L.-thickness; c_{pv}, c_{pg}, heat capacity of vapour (water) and air, resp.; θ_δ is approximated with θ_{amb}. (Remark: T denotes temperatures in K, θ denotes temperatures in °C; T = θ + 273.15).

This implicit equation (15) with respect to the surface temperature θ_{surf} (dry) starting with $\theta_\delta \approx \theta_{amb}$ can be only solved numerically, e.g. with the Regula-Falsi-method. The surface (dry) temperature is increased during condensation or decreased during evaporation according to the psychrometric chart along constant enthalpy Δh_v. The surface (dry) temperature reaches therefore the temperature at the saturation point θ_s.

The temperature inside the fruit is modelled by the classical heat conduction equation for a sphere

$$\frac{\partial T}{\partial t} = \frac{\partial^2 T}{\partial r^2} + \frac{2}{r^2} \cdot \frac{\partial T}{\partial r} \tag{16a}$$

with the boundary condition for the heat transfer between produce surface and air:

$$-\lambda \left(\frac{\partial \mathrm{T}}{\partial \mathrm{r}} \right)_{\mathrm{surf}} = \alpha \, (\mathrm{T}_{\mathrm{surf}} - \mathrm{T}_{\mathrm{air}}) \tag{16b}$$

where λ heat conduction and α heat transfer coefficient; r spherical radial distance.

The classical solution of (16) is well documented (e.g. in Baehr, 2006) and can be reduced to the solution for the surface temperature (dimensionless r =1) for small time changes (dimensionless time<0.2) which may be applicable for our problem:

$$\mathrm{T°}(\mathrm{r}=1,\mathrm{t}) = 1 - \left(\mathrm{erfc}\left(\frac{1}{\sqrt{\mathrm{t}}} \right) - \exp\left(2 \cdot \mathrm{Bi} + \mathrm{Bi}^2 \cdot \mathrm{t} \right) \cdot \mathrm{erfc}\left(\frac{1}{\sqrt{\mathrm{t}}} + \mathrm{Bi} \cdot \sqrt{\mathrm{t}} \right) \right) \tag{17}$$

with T° = dimensionless temperature $\mathrm{T°}(\theta) = \theta_{amb} - \theta / (\theta_{amb} - \theta_{surf})$, Bi = Biot-number; $\mathrm{Bi} = R_0 \cdot \alpha/\lambda$, R_0= Radius of fruit or sphere. To calculate the course of the non-adiabatic surface temperature $\theta_l(t)$ the equ. (15) and (17) are used by replacing θ_{surf} with θ_s. Taking $\theta_l = \theta(\mathrm{T°})$ we have the complete set of equations to simulate nearly the temperature changes of the dry and wet surface temperatures, see below.

7. Results

At *adiabatic* evaporation cooling the wet bulb temperature can be reached on the surface. A thermographic image shows (Figure 7), however, that this temperature is not reached but remains at a higher value. The explanation is that heat is diffusing from inside the fruit to the surface and therefore the real process is *not adiabatic*. Heat diffusion inside the fruit can be calculated with the classical Fourier's method for spheres (see for example Baehr, 2006). Additionally, a heat transport takes place within the boundary layer, whereby the temperature can only remain within the range of the minimum attainable wet bulb temperature θ_s in proximity of the surface (i.e. on the water film) and of the maximal attainable ambient temperature $\theta_{amb} \approx \theta_{\delta T}$, with $\theta_{\delta T}$ temperature at δ_T thermal B.L.-thickness (heat transfer).

Due to the heat exchange processes with the body a measurable temperature is reached in the range of between θ_s and θ_{amb}. The heat transfer is dependent on the air flow velocity according to equ. (13), and therefore variant along the surface because the air flow velocity (or Reynolds number) is variant. The calculation of the surface temperature by the method explained above, esp. after equ. (15) combined with the Fourier's method for spheres (equ. 17), results for this example in solved θ_l (Figure 8).

The temperature distribution development on the surface is dependent on air flow direction and velocity, as well as on temperature and air humidity. The drying time, i.e. the time until the surface moist is totally removed, can be well recognized as the point when all local temperatures on the fruit surface have converged to the ambient or fruit temperature.

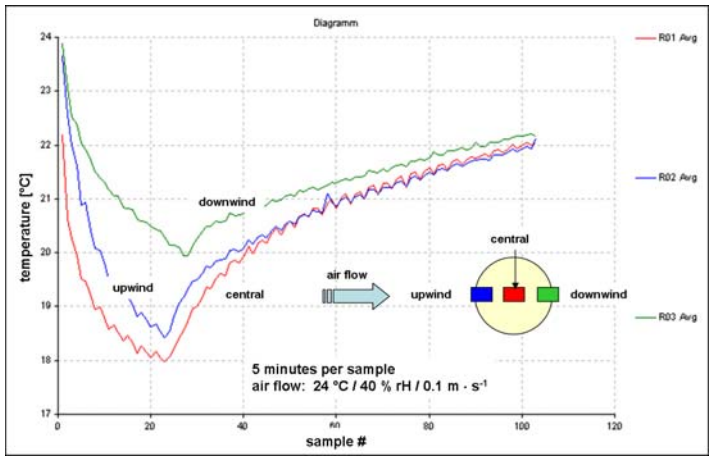


Figure 7. Thermographic measurement of temperature at three local points on the surface of a peach during drying off surface moisture

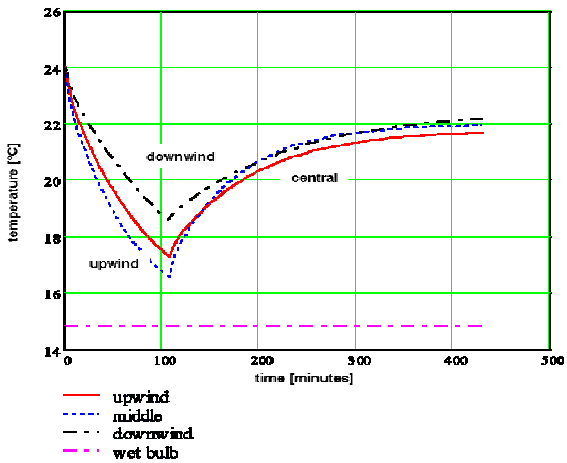


Figure 8. Calculated course of temperature at three local points on the surface of a peach during drying off surface moisture

During surface drying the IR-measured temperature is taken as the temperature $T_{\delta T}$ on the boundary. As soon as the surface is dried off from moisture the temperature $T_{\delta T}$ is converging to ambient temperature and the fruit surface is warming up again. From these calculations (Figure 8) we can see sufficient agreement with the measurements (Figure 7).

In the example (Figure 7 and 8) the surface of a fruit covered with a water film of 80 µm thickness was dried within ½ hrs using air at 24 °C, 40 % rH. The mass loss caused by transpiration was $E = 2.2 \cdot 10^{-5}$ kg s^{-1} m^{-2}, equ. (1). The calculation using equ. (11) resulted in $j = 2{,}4 \cdot 10^{-5}$ kg s^{-1} m^{-2}, which is in good accordance to the experiment (E; equ. 1) (Linke 1997a, b) after assumption of $\beta \approx 6.4 \cdot 10^{-3}$ m s^{-1}, see chapter 4.

8. Conclusion

The primary effects on the moisture loss in a fruit or potato are the diffusion processes in the ambient air and in the produce material. The environmental conditions can easily be obtained under test. The analytical model gives a tool to approximate the mass transfer coefficient for evaporating moist from the fruit surface to the ambient air. The local mass transfer coefficient distribution along the surface and the total mass transfer coefficient becomes possible to calculate. The model can be used for calculating the surface temperature which is in accordance to surface temperature measurements, using thermal imaging. The model in conjunction with IR-measurements gives the possibility to predict heat and mass transfer for local differences of air flow around the surface of a fruit body. It is demonstrated from the model that the infrared thermographic camera is not measuring the wet bulb temperature from a drying surface but the real non-adiabatic evaporating process. Therefore, experimental results from IR-temperature measurements need careful interpretations.

Acknowledgements

The authors acknowledge support of the Hungarian-German Foundation MÖB-DAAD under contract No. P-MÖB /843.

References

Baehr H. D., Stephan, K., 2006: Heat and Mass Transfer. Springer-Verlag, Berlin Heidelberg, 2006.

Fito, P. J.; Ortolá, M. D.; De Los Reyes, R.; Fito, P.; De Los Reyes, E., 2004: Control of citrus surface by image analysis of infrared thermography. Journal of Food Engineering, Vol. 61 (3), 2004, 287-290.

Gottschalk, K.; Linke, M.; Mèszáros, C.; Farkas, I., 2007: Modelling condensation and evapouration on fruit surface. Drying Technology (25): 1237-1242.

Hellebrand, H.J.; Beuche, H.; Linke, M.; Herold, B; Geyer, M., 2001: Chances and Shortcomings of Thermal Imaging in the Evaluation of Horticultural Products. International Conference "Physical Methods in Agriculture - Approach to Precision and Quality", Prague 27-30 August 2001, Proceedings, 112-117.

Linke, M., 1997a: Modelling of Heat Transfer in Individual and Packed Chicory Crops. Proc. of the COST 915 – Copernicus CIPA-CT94-0120 Workshop on Food Quality modelling.

Linke, M., 1997b: Modelling and Predicting the Postharvest Behaviour of Fresh Vegetables. Mathematical and Control Applications in Agriculture and Horticulture (Munack and Tantau, Eds.), Pergamon Press, Oxford, UK, 283-288.

Rastovski, A., 1987: Storage of Potatoes. A. Rastovski, A. van Es et al (eds). Wageningen, Pudoc, 1987.

Federated filtering: classical approaches in new approximations for distributed systems estimation[1]

Moira MIRANDA[1], András EDELMAYER[2], Sándor MOLNÁR[3]
[1]Laboratory of Unit Operations, Department of Chemical Engineering,
Faculty of Engineering, University of Los Andes, Mérida, Venezuela
[2] Systems and Control Laboratory, Computer and Automation Research Institute,
Hungarian Academy of Sciences
[3]Department of Informatics, Institute for Mathematics and Informatics

Abstract

The paper discusses the application of the concept of decentralized filtering to state estimation of distributed systems. A particular decentralized filtering solution, which is based on new approaches of the classical federated architecture, is discussed and investigated. The novel solution addresses the problems of fault tolerance and estimation accuracy by means of the utilization of a reconfigurable structure. The proposed filter is operated in an extended federated architecture in an attempt to adaptively balance estimation accuracy and sensor fault tolerance in the presence of both state and measurement uncertainty when the observations from the system are limited and subject to a variety of sensor faults.

Keywords

dependable systems, distributed filters, enhanced state estimation, extended Kalman-filter, federated filtering, sensor fault tolerance

1. Introduction

Distributed systems and safety critical systems are two major technology trends in advanced systems and control engineering. Due to their lower cost, higher flexibility, simpler installation and maintenance, distributed or decentralized systems are getting increasingly important in a widening array of applications in control, information and communication systems and sensor networks, just to mention a few of the subjected areas. They do not just benefit from a more structured systems design approach, however. The distributed idea may also be required to make effective function mapping techniques available for the design as well.

[1]This work was supported by the project TÁMOP 4.2.1/B-09/1/KONV-2010-0003 at the Széchenyi University of Győr and, in part, by the Hungarian Scientific Research Fund (OTKA) under grant number K-68187, which is gratefully acknowledged by the authors.

Function mapping is an advanced idea of systems and control engineering. Using this idea, overall system functionality is decomposed into a set of partial system functions, which are then implemented by particular physical subsystems relying on advanced software (SW) and hardware (HW) technology. Function mapping makes the production more economical, the implementation more configurable and flexible. The greater the portion of SW in the implementation, the more efficiently the mapping can be done.

The subsystems are individually controlled by one or more dedicated local control units (MPU's). The multitude of MPU's are typically implemented on embedded platforms and interconnected via heterogeneous computer networks, very often wirelessly, for information exchange. Advanced vehicle on-board control systems, including land, marine and aerospace systems, for instance, are increasingly relied on a highly distributed systems architecture and computer networks.

Another important trend of recent development is related with the increasing operational safety and reliability reqiuirements. Due to the high requirements of product quality, increased safety, the minimization of impact on the environment and reduction of manufacturing costs the importance of the utilization of advanced control methods that makes use of dependable technology have became evident in various fields of the industry in the recent years. In order to achieve the system's technical goals the design principles, such as the continuity in operation, the assurance of the continual availability of the primary system's functions for long-term operational safety are becoming the prevailing design factors. This is true not only in the classical application fields of dependable technology, such as in aviation, chemistry and nuclear industry, which are often referred to, traditionally, as safety-critical, but in a widening array of commercial applications as well.

More and more frequently, one of the main objectives of the design is, to create engineering structures in which equipment faults can be detected and removed from the system, quickly and reliably, in such a way that system functionality be continuously maintained over time: fault reconciliation is the central feature in dependable control systems.

Functional distribution and dependability are engineering principles, which are deeply interrelated, however. From the one hand, distributed systems provide more manageable structures, which is the prerequisite for the creation of dependable architectures. From the other hand, elements of dependable solutions can be implemented more efficiently when they are spatially (functionally, logically) distributed.

A fundamental problem of control in distributed engineering structures is to solve detection and estimation problems to supply reliable data for controllers, using scalable algorithms which comply with the stringent performance requirements demanded by embedded applications.

In many technologies, process variables that determine overall performance of the system (e.g., product quality) cannot be measured directly, but can be

estimated/inferred in real time relying on the available measurements. That is to say, given the values of some observable signals (the measurement signal) one would like to estimate (predict) the values of other, not directly measurable signals. In various structures of model predictive control (MPC), for example, which are used in the industry in an increasing variety (see e.g., [9] and [1]), controller operation is mainly based on the estimated values of process variables. In fact, estimation and control are two closely related problems of systems theory.

Considering state estimation in distributed systems, a major difficulty lies in the unreliable data connection (communication) between subsystems and the performance of the distributed sensory system, which is usually implemented over unreliable communication networks. The connections are subject to disturbances and faults, and the network operation is characterized by the often limited or sporadic availability of the measurements. Since the probability of communication failures causing complete data loss for a relatively longer duration of time cannot be excluded, negotiated data in the distributed structures can be easily corrupted. As the problem of estimating the state of distributed dynamical systems based on limited measurements arises in many fields of the industry, novel filtering methods with certain level of fault tolerance to accommodate sporadic communication nonperformance are required.

This paper discusses the problem of dependable estimation and filtering in distributed systems. More particularly, a decentralized filtering solution, which is based on new approaches of the classical decentralized architecture is discussed and investigated. This novel solution addresses the problems of fault tolerance and estimation accuracy in uncertain distributed systems subject to communication losses and limited availability of measurements by means of the utilization of a reconfigurable filter architecture.

The structure of the paper is the following. Section 2 gives a short summary of traditional filtering solutions, generally used in distributed systems. The underlying theory for centralized and decentralized designs is overviewed, and the designs are compared to each other. Section 3 gives special attention to a particular form of decentralized filters, i.e., federated filters, which is then followed by the extension of the classical federated idea in Section 4. The novel solution benefits from the adaptive filter management technique, which is based on the continuous health monitoring of the input channels. This is done to balance estimation accuracy and sensor fault tolerance in the presence of both state and measurement uncertainty when the observations from the system are subject to a variety of sensor faults.

2. Central vs distributed filtering

As a standard solution alternative to the state estimation problem, centralized filters have traditionally been used in the industry for many decades. In this approach a single filter with universal access to all sensors is used for estimating the complete set of the state variables of the system (Fig. 1).

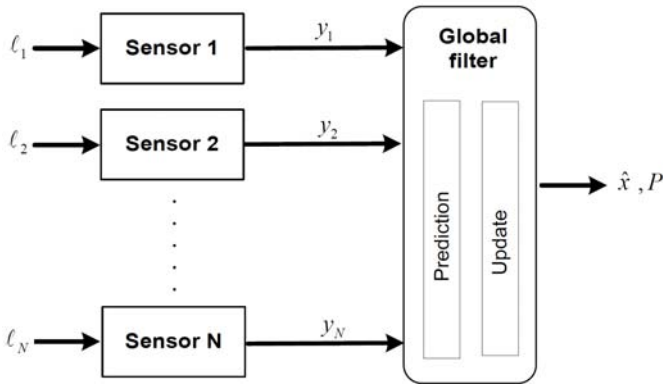


Figure 1. Traditional centralized filter architecture

Centralized solutions, however, especially in large-scale industrial applications, suffer from the size (dimension) of the problem and are prone to both sensor and implementation failures. Reduced order filters are not always reliable, they tend to have poor estimation accuracy, and even instability under certain circumstances. Centralized approaches, moreover, cannot answer the call of the safety and dependability requirements as they lack fault tolerance when faults in the operation of the system or sensors occur. The development of decentralized filtering, therefore, has received increasing attention during the past few years.

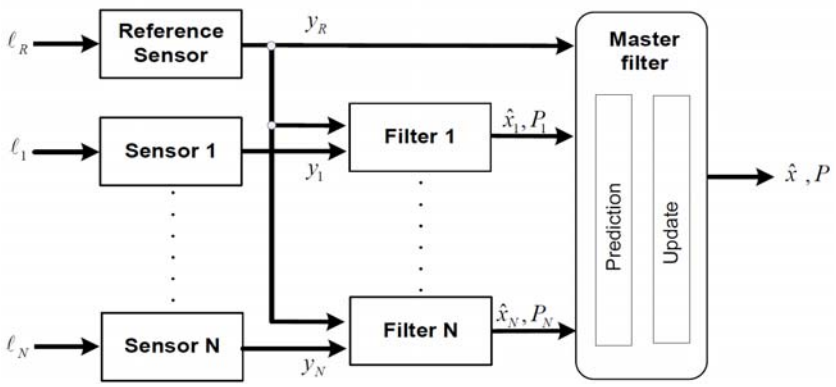


Figure 2. Decentralized filter architecture

It was recognized early that deviding the global estimation problem into smaller, partial problems and the distribution of the overall computation burden among a set of (local) filters have obvious operational advantages in large-scale applications. A solution, which is an alternative to the centralized one, decomposes the global estimation problem into several smaller problems. This is done by accommodating the filter architecture to the physical structure of the

system by assigning one ore more separate, autonomous filter to each subsystems. These filters treat the process information locally at the component or subsystems level using local information of the system dynamics. Then, the estimation of sensor dedicated local filters are fused in a master or fusion filter for composing the local results in a global estimation, see Fig. 2.

Another distinguishing feature of decentralized filtering is the inclusion of reference sensors in the structure. The following two features distinguish the reference sensor from the local sensors:

- The reference sensor acts as a fundamental sensor in the system, and its data is directly observable for both the master filter and any of the local filters. The data from the local sensors is dedicated to the corresponding local filters only.
- The reference sensors are assorted equipments, which are usually more dependable and have higher data rate than others; thus their data is often used for the initialization of some of the local filters. These are accountable to compensate erroneous sensor readings relying on the benefits of the decentralized architecture. The number of reference sensors included in the distributed architecture is a design issue dependent of the particular application.

Previous efforts on distributed filtering have concentrated on either decentralized filters on centralized or hierarchical topologies or essentially centralized filters on decentralized topologies based on the work [8, 14, 17]. Several decentralized solutions have been developed to decentralize the filter algorithm, topology, and services through tradeoffs of computation and memory that minimize communication. These implementations assume a truly decentralized architecture requiring no central facilities. Since Kalman filter has been an excellent means for exploring decentralized system trade-offs because of its optimality and linearity, the majority of solutions tend to rely on derivations of the linear Kalman filter, see e.g., [16, 12].

As a result of this decentralization technique, the computational load regarding a single filter's implementation, can be significantly reduced. This property makes the solution viable in resource constrained embedded implementations.

3. The federated filter

A special case of decentralized filtering is called federated filtering that was originally proposed [4, 5, 6]. Federated filters are distributed filters which consist of several local filters (LFs) and a master (or fusion) filter (MF). The distinctive property of the federated filter is that information feedback from master to local filters are implemented.

There is a particular LF assigned to each particular subsystem. LFs work in parallel and their solutions are periodically composed by the MF. As the local estimates are usually suboptimal, information feedback from the master filter to each local filter is necessitated. The information which is usually shared among local filters are data about the statistics of the process noise, the filter initial

conditions and some common measurement information (Fig. 3). The feedback relies on the information sharing principle, which was introduced in [4].

The federated filtering scheme is categorized as a decentralized one, but the use of the information feedback via the sharing factors β_i, which determine the distribution of the information provided by master filter around local filters, distinguishes it from the classical decentralized idea. The principle of information sharing is based on the axiom of conservation of information, i.e., $\sum_i^n \beta_i = 1$, where *n* is the number of the sharing filters. This property is discussed further in the following part.

The existence of the feedback, however, has non-desirable side effects on filters' operation. If, for example, an undetected fault, in any sensor loop of the system occurs, due to the closed-loop nature of the federated filtering process, the other filters in the network may be polluted by the fault propagated through the information network of the federated topology. This feature, however, can be used in a variety of ways to improve filters performance through the extension of the classical federated idea, as this is shown in this paper.

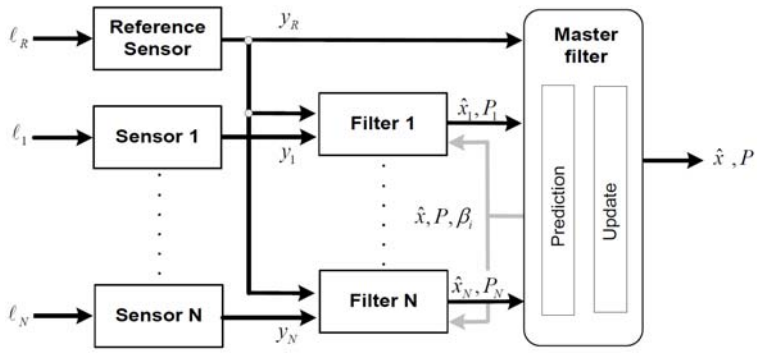


Figure 3. Federated filter architecture

For state estimation in federated architectures, a wide variety of filters can potentially be used and developed. They range from Kalman's original linear formulation, now called the simple Kalman filter (from which the so-called information filter and a variety of square-root filters were derived) to nonlinear problems, such as the extended and unscented Kalman filters (EKF and UKF, respectively). EKF uses linearization techniques and became a standard solution in guidance and navigation. The UKF addresses the nonlinear estimation problem using EKF and is based on a deterministic sampling technique and a nonlinear (the so-called unscented) transformation. These filters provide an optimal solution to the (linear) estimation problem by minimizing the disturbance (noise) effects using a quadratic cost function assuming the disturbances are stochastic processes with jointly Gaussian distribution. There are non-Gaussian extensions of Kalman filtering available too.

If no information about the statistics and distribution of the noise is available, the aforementioned methods are not directly applicable. Still, however, the filtering problem can be posed e.g., as an H_∞ estimation problem and the local, as well as master filters, can be implemented as H_∞ filters.

Whilst the techniques and theory of federated filters (especially those based on the Kalman filter) are relatively well developed for linear systems by now, experience subjecting nonlinear applications are not yet widely available. A rare exception can be found [2]. Experience regarding the use of other advanced filtering methods, such as the above mentioned H_∞ estimators and nonlinear filters are not sufficiently explored. Moreover, decentralized filtering schemes, including those where the filters are aligned in the federated architecture, have been recognized effective in the practice, their potential to enhance estimation performance and fault tolerance have not been widely investigated and fully realized.

In the following part of this paper, a novel idea to avoid filter degradation in various failure modes of the system is proposed. This is done by means of useful extensions of the classical federated idea in nonlinear systems. To maintain filter functionality and performance, an adaptive technique to influence the information sharing property as well as changing the filters' operating mode of the federated architecture is suggested. This is done by means of a sensor redundancy management applied to the fusion rule of the MF. This concept can be considered the extension of the classical federated idea.

The framework that combines sensor fault detection with isolation and fault accommodation through filter reconfiguration is outlined. The challenge is to engineer an operational strategy of the filter's architecture in an attempt to make a tradeoff between best estimation accuracy and sensor fault tolerance in every possible time. For this purpose, the classical federated architecture is extended with a sensor fault detection and sensor management logic.

The basic features and theory of the federated solution are briefly reviewed. The necessary theory of the design of the federated filter is mainly focused on the presentation of the fault reconciliation and fusion algorithm. The federated structure relies on EKF as state estimator, where the core filtering algorithm is considered the nonlinear extension of the standard linear filter. As both Kalman filters and EKFs are considered to be standard techniques of control theory by now, technical details only necessary for understanding the recent problem formulation are mentioned here. For more information, the interested reader is referred to the literature, see e.g., [3] and [10], and also the references therein.

4. The extended federated filter architecture

In the following part the standard nonlinear extension of the well-known linear Kalman filter (i.e., EKF) is used in a particular structure for the solution of the distributed nonlinear state estimation problem.

4.1 Formulation of the nonlinear distributed filtering problem

We are concerned with nonlinear dynamical systems described in the nominal representation by ordinary differential equations subject to noise

$$\begin{aligned} \dot{x}(t) &= \phi(x(t), u(t), w(t)), \\ y(t) &= h(x(t), v(t)). \end{aligned} \tag{1}$$

For a special case, this can be written in state space form, by means of the set of noise-corrupted state and measurement equations of the following form

$$\begin{aligned} \dot{x} &= f(x) + \sum_{i=1}^{m} g_i(x) u_i + w, \\ y_i &= h_i(x) + v_i, \qquad 1 \le i \le p, \end{aligned} \tag{2}$$

where $x \in \mathrm{P}^n$, $u \in \mathrm{P}^m$, $y \in \mathrm{P}^p$ denote the state, the input and the output of the system, respectively, moreover, $w(t)$ and $v(t)$ are the system and the measurement noise, which are independent of each other as well as of $x(0)$.

Assume each sensor s_i measures a signal $y_i(t)$ that is corrupted by measurement noise v_i taken to be a zero-mean white noise. Let Q_i and R_i denote the covariance matrix of v_i and w_i for all i, respectively,

$$Q = \mathbf{E}\{v_k v_k^T\}, \quad R = \mathbf{E}\{w_k w_k^T\}. \tag{3}$$

Let the local estimate and its covariance provided by the i^{th} local filter be represented by $\hat{x}_i$ and P_i $(i = 1,2,...,N)$. The filtering algorithm is considered the extension of the standard linear one. Since the system is not linear, the Riccati matrices that attempt to approximate the *a priori* and the *a posteriori* covariances for each filter are defined, respectively, as

$$\begin{aligned} P_{k|k-1} &\approx \mathbf{E}\{e_{k|k-1} e_{k|k-1}^T\}, \\ P_{k|k} &\approx \mathbf{E}\{e_{k|k} e_{k|k}^T\}. \end{aligned} \tag{4}$$

The filters are initialized with $x_{o|o} = x_o$ and $P_{o|o} = P_o$, and then operated recursively performing a single cycle each time a new set of measurements becomes available.

Each iteration propagates the estimate from the time the last measurement was obtained to the current time. The propagation process consists of two stages:

update and prediction. The update equations are responsible for the feedback, i.e., for incorporating a new measurement set into the *a priori* estimate to obtain an improved *a posteriori* estimate. The *a posteriori* state estimate $x_{k|k}$ is computed as a linear combination of an *a priori* estimate $x_{k|k-1}$ and a weighted difference between an actual measurement y_k and a measurement prediction:

$$\begin{aligned} \hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k\left[y_k - \overline{h}(\hat{x}_{k|k-1})\right], \\ K_k &= P_{k|k-1}\overline{H}_k^T\left(\overline{H}_k P_{k|k-1}\overline{H}_k^T + R_k\right)^{-1}, \end{aligned} \tag{5}$$

where $\overline{H}_k$ is the Jacobian matrix of partial derivatives of $h(x)$ with respect to x, that is

$$\overline{H}_k = \left(\frac{\partial \overline{h}(x)}{\partial x}\right)_{x=\hat{x}_{k|k-1.}} \tag{6}$$

The matrix K_k is chosen such that the filter minimizes the *a posteriori* error covariance. The covariance matrix is updated by

$$P_{k|k} = (I - K_k\overline{H}_k)P_{k|k-1.} \tag{7}$$

The prediction equations are responsible for projecting forward the current state vector and error covariance estimates to obtain *a priori* estimates for the next time step. The state and covariance matrix in the next sampling instant are estimated by

$$\begin{aligned} \hat{x}_{k+1|k} &= f(\hat{x}_{k|k}, u_k), \\ P_{k|k+1} &= \overline{F}_k P_{k|k}\overline{F}_k^T + Q_k, \end{aligned} \tag{8}$$

where $\overline{F}_k$ is the Jacobian matrix of partial derivatives of $f(x)$ w.r.t. x, as

$$\overline{F}_k = \left(\frac{\partial \overline{f}(x,u)}{\partial x}\right)_{x=\hat{x}_{k|k}, u=u_k.} \tag{9}$$

The following characteristics distinguish the synthesizing MF and the preprocessing LF's in the architecture. Each LF is dedicated to the measurements of local sensors. For the update process, the LF's use the update information from its local sensors only and not any others. Conversely, the MF uses the local

filtered estimates $\hat{x}_i, P_i$ as quasi-observables to update the global state vector in a fusion process, sequentially.

The MF is processed at the rate equal to the rate of the LF's, which means that local outputs are subject to fusion on the next stage they are processed. The time updating solution of the MF is represented by the state estimation $\hat{x}_N$ and covariance $\hat{P}_N$, respectively. If all local estimates are uncorrelated, then the global estimate can be given as

$$\begin{aligned} P_f^{-1} &= P_1^{-1} + P_2^{-1} + \ldots + P_N^{-1}, \\ \hat{x}_f &= P_f \left[P_1^{-1}\hat{x}_1 + \ldots + P_N^{-1}\hat{x}_N \right], \end{aligned} \tag{10}$$

where the inverse of the covariance matrix is called the information matrix. Eq. shows that the global information is just the sum of that of the local systems. The global estimate $\hat{x}_f$ is a linear weighted combination of the local estimates with weighting matrices P_f^{-1}, P_i^{-1} $(i = 1, \ldots, N)$.

However, the estimates of different local filters may be correlated. In order to eliminate this correlation, the process noise and state error covariance are set to their upper bounds relying on the technique proposed in [4, 5]

$$Q_i = \beta_i^{-1} Q, \qquad P_i = \beta_i^{-1} P_f, \tag{11}$$

where $\beta_i (\square 0)$ is the information-sharing factor satisfying the rule

$$\beta_1 + \beta_2 + \ldots + \beta_N = 1.$$

For the condition that makes Eq. hold, see [6]. To some extent, this method results in a conservative design because of using the upper bound of the process noise variance matrix instead of the process noise variance matrix itself. Other approaches to the solution of this problem can be found e.g., in the references [13, 15].

For the structure and organization of the filter, see the architecture diagram Fig. 4, where y_R is the reading of a reference sensor, y_i stand for the sensor measurements, $\hat{x}_i$ for the state estimates and β_i for the information sharing variables. P_f denotes the covariance matrix provided by the fusion filter.

4.2 Sensor failure reconciliation

In large distributed systems containing a large number of components, the occurrence of sensor faults which may affect control actions is an increasing concern. Sensor faults may be caused by sensor drifts, step changes, scale factor

errors changing the mean, incorrect calibration, etc. Correspondingly, the degraded modes of sensors can be characterized by the presence of a systematic nonzero mean, in the form of a constant jump bias, a ramp bias, or by an increase in variance of the driving noise.

Federated filtering, by real-time adaptation of the sharing factors, provides a variety of possibilities to compensate these problems and improve filtering performance.

The requirement for fully autonomous filter operation necessitates the inclusion of a sensor management logic. The sensor management is based on a fault detection algorithm, which is responsible for fault accommodation.

To this end, the innovation series of the local filters are subject to continuous monitoring and statistical testing. The blocks Δ_i in Fig. 4 provide residuals ε_i for detecting changes in the statistics of the measurement data, upon which the determination of the measures of the covariance information sharing β_i is relied.

The diagnosis can be based on the analysis of the innovation sequences of the individual filters. To detect failures changing the mean of the innovation sequence a number of statistical methods are available that make use a threshold test on the moving window average of the output of the instruments. These tests commonly known as generalized likelihood ratio tests (GLRT), χ^2-tests and others, are appropriate for online implementation and embedding in the federated structure, according to the general scheme shown in Fig. 4. For further details the reader is directed to Refs. [7] and [10] and also the references cited thereof.

If the system operates normally, the normalized innovation sequence of a filter (ε_i) is a zero mean white noise with unit covariance. A real-time detection of failures affecting the mean and/or the variance of the innovation process triggers a filter reconfiguration action in the sensor management logic of the fusion filter in an attempt to keep the performance of the impaired structure as close to the optimum as possible. This includes the modification of the sharing factors and the switch-over of the reset mode of the filter.

Though the sophistication and implementation of the detection algorithm applied in Δ can be varied (as it was characterized above), for a compact solution of joint detection and reconfiguration, a simple demonstrative algorithm, which is based on the continuous monitoring of the error variance of each particular channel is presented in the following.

Let the median of the estimated states $\hat{x} \in \mathrm{P}^n$ in each particular time be given as $\widetilde{X}_t = \{\hat{x}_i\}|_t$. Let S_i denote the error variance between estimation $\hat{x}_i$ and median $\widetilde{X}$, and define the inverse of the sum of the elements of S_i by

$$\overline{S}_i = \frac{1}{\sum_{j=1}^{n} S_{ij}}. \tag{6}$$

Obviously, the magnitude of S_i characterizes the dependability of a particular local estimation relative to the fused one. Under fault-free conditions measure S_i can be kept below a well chosen threshold limit, meaning that the channel is operating normally.

In case a measurement channel gets faulty, the corresponding S_i will exceed the threshold quite sensitively. Therefore, eq. can be taken as residual generated by each particular LF that may be used for providing information about the appearance of sensor faults (*cf*. Fig 4).

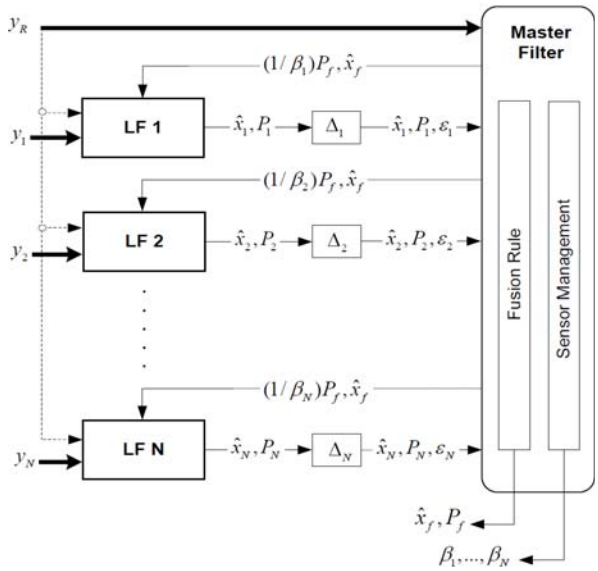


Figure 4. The extended federated filter architecture

Then, based on , information-sharing factors β_i can be calculated in the next step as

$$\beta_i = \frac{\overline{S}_i}{\sum_{i=1}^{N} \overline{S}_i}, \tag{12}$$

where N is the total number of local filters.

Realize that β_i is a decreasing variable with increasing error variance S_i. The application of the median in and is motivated by the use of a limited number of local filters in some applications. This approach may give good results in averaging the estimation error variances in cases when only a small number of

local filters are available (i.e., $N < 10$ as it has been revealed by our investigations). To avoid splitting of the median the side condition $N > 2$ is to be assumed. For the architectures with $N > 10$, however, the application of the statistical mean instead of the median could provide much better results.

4.3 Filter operation

This structure of filters can be operated in two different operating modes, depending on how the fused process data is exploited by the local filters. These modes of operation are, usually, referred to as *reset* and *no-reset* modes. When the *reset* mode is used, the master and local filters are reset by the global solution, i.e., $\hat{x}_i = \hat{x}_f$, and $P_i = \beta_i^{-1} P_f$, for $i = 1, \ldots, N$, meaning that a continuous information feedback from the master to the local filters is present. In *no-reset* mode this information feedback does not exist: each LF keeps its process information ($\hat{x}_i, P_i$) produced locally, thus the MF retains none of the fused data and the global fused estimation ($\hat{x}_f, P_f$) has no effect on any of the local estimations.

Obvious advantages and disadvantages associated with each of the resetting modes are known. The federated filter operated in reset mode is expected to provide better estimation accuracy, while in the no-reset mode a better tolerance of sensor faults, see the extensive simulation study presented in [11].

5. Conclusions

The paper advocates the principle of distributed filtering as the basis for the implementation of dependable filtering solutions in large-scale distributed systems, which are subject to limited availability of measurements due to sensor faults and communication losses. A design strategy with the extension of the classical federated filter for increased accuracy estimation and sensor fault tolerance has been featured. The idea consists of using dedicated local filters to particular components or subsystems and utilizing a master filter for obtaining the global estimation. Information generated locally is divided among the distributed filters in the update process of the filter by a proper choice of the information sharing factors.

Our main motivation was to create, in a novel combination, a distributed filter architecture and a sensor fault management logic to maintain estimation accuracy in every possible time in a structure, which is able to accommodate the effects of sensor faults, effectively.

The key to this solution is the re-organizable architecture represented by the federated idea, as well as the reconfiguration scheme that, with the aid of the utilization of the sensor fault detection and fault management logic, alters the fusion filter law to achieve better performance. Each filter within a federated group is assigned, under the control of the fault management function of the

master filter, to operate in either a full-reset or a no-reset mode. Failure in any input channel can be re-conciliated without the significant deterioration of estimation performance, or, in case the estimation is used in a subsequent control action, without loss of equipment functionality.

References

[1] Sung-Mo Ahn, Myung-June Park, and Hyun-Ku Rhee. Extended Kalman filter-based nonlinear model predictive control for a continuous MMA polymerization reactor. *Ind. Eng. Chem. Res.*, 38(10):3942–3949, 1999.

[2] J. Ali and J. Fang. Multisensor data synthesis using federated of un- scented Kalman filtering. In *IEEE International Conference on Industrial Technology*, pages 524–529, 2005.

[3] R. G. Brown and P. Y. C. Hwang. *Introduction to random signals and applied Kalman filtering*. Wiley, New York, 1992. Second Edition.

[4] N. A. Carlson. Federated filter for fault-tolerant integrated navigation systems. In *Proc. IEEE Pos. Loc. Nav. Symp.*, pages 110–119, 1988. Orlando, FL.

[5] N. A. Carlson. Information-sharing approach to federated Kalman filter- ing. In *Proc. IEEE National Aerospace and Electronics Conf.*, page 1581, 1988. Naecon, Dayton, OH.

[6] N. A. Carlson. Federated square root filter for decentralized parallel processes. *IEEE Trans. Aerospace and Electronic System*, 26(3):517–525, 1990.

[7] Tze-T. Chien and M. B. Adams. A sequential sensor failure detection 14technique and its application. *IEEE Trans. Aut. Cont.*, pages 750–757, 1976.

[8] M. F. Hassan, G. Salut, M. G. Singh, and A. Title. A decentralized computational algorithm for the global kalman filter. *IEEE Trans. Aut. Cont.*, 23(2):262268, 1978.

[9] M. A. Henson. Nonlinear model predictive control: current status and future directions. *Comp. and Chem. Eng.*, 23(2):187–202, 1998.

[10] R. S. Mangoubi. *Robust Estimation and Failure Detection – A Concise Treatment*. Springer-Verlag, London, 1998.

[11] M. Miranda and A. Edelmayer. Enhanced federated filtering for state estimation and sensor fault tolerance with case study of multi-component distillation. Submitted to *Int. J. IET-CTA*, 2011.

[12] Reza Olfati-Saber. Distributed Kalman filtering for sensor networks. In *46th IEEE Conf. Dec. Cont.*, New Orleans, LA, pages 5492–5498, 2007.

[13] H. Z. Qiu, H. Y. Zhang, and H. Jin. Fusion algorithm of correlated local estimates. *Aerospace Sci. and Techn.*, 8:619–626, 2004.

[14] J. L. Speyer. Computation and transmission requirements for a decentralized linear-quadratic-gaussian control problem. *IEEE Trans. Aut. Cont*., 24(2):266–269, 1979.

[15] Shu-Li Sun. Multi-sensor information fusion white noise filter weighted by scalars based on Kalman predictor. *Automatica*, 40:1447–1453, 2004.

[16] Shu-Li Sun and Zi-Li Deng. Multi-sensor optimal information fusion Kalman filter. *Automatica*, 40(6):1017–1023, 2004.

[17] A. S. Willsky, M. G. Bello, D. A. Castanon, B. C. Levy, and G. C. Verghese. Combining and updating of local estimates and regional maps along sets of one-dimensional tracks. *IEEE Trans. Aut. Cont.*, 27(4):799–813, 1982.

Accumulation Temperature of an Experimental PCM Solar Tank

István Péter SZABÓ[1], Gábor SZABÓ[1], Péter SZENDRŐ[2]
[1]University of Szeged, Faculty of Engineering,
Department of Technical and Process Engineering
[2]Department of Machine Construction, Institute for Mechanics and Machinery

Abstract

The one of the most important part of a solar collector system is the solar tank. The relevant type and capacity of the solar tank is a requirement of the good operation of the system.

It is necessary to apply a solar tank because the period of the sunshine does not coincide with the time of the hot water consumption generally. The solar collectors work daytime only, and its power depends on the weather.

In solar systems we have to be mindful of the possible changes of the weather, so we have to apply tanks with bigger heat capacity than in convencional domestic hot water systems. Generally the tank has to store heat capacity for 2 days. It results bigger heat demand. The solution of the problem is the solar tank particularly filled with phase change material. This tank has smaller dimensions and bigger heat capacity than the conventional tanks.

The other important advantage of a PCM solar tank could be the possibility of the operating of the collectors at lower temperature. The efficiency of the solar collectors is lower if the temperature is higher. During the phase changing the PCM-tanks accumulate the heat at constant temperature, so during this time it is not necessary to raise the temperature of the collectors. It could result a higher efficiency of the solar collector system. We have made pre-calculations to study this possibility. According to the calculations this advantageous property does not exist for all of the phase change materials. Some PCM causes higher temperature in the solar tank, so in this case the efficiency of the solar collectors will be lower.

Keywords

solar collector, modelling, solar tank

1. Introduction

The hot water consumption in family houses is generally bigger in the evening and in the morning, so it is necessary to store the utilized energy. The collectors transform but not store the solar energy. The storage is necessary to accomplish in a heat-insulated tank, placed in tempered space.

The temporal difference of energy source and energy needs made necessary the development of storage systems. Except in summer, specially in winter, the temperature of the heat transfer fluid coming from the collector is relatively low (35-60 °C). In this period of time, one way of storage is to use solid-liquid phase change materials. In comparatively small volume the phase change materials have great storage capacity in small temperature interval. [1] Adding PCM (phase change material) modules at the top of the water tank would give the system a higher storage density and compensate heat loss in the top layer because of the latent heat of PCM [2].

There are only a few type of PCM (phase change material) solar tanks are available in commerce.

The PCM tanks have two groups by the construction:

- PCM tanks with inner core,
- PCM tanks with inner balls.

The benefit of the tanks supplied with the inner core is the easy putting of the phase change material. The disadvantage of the inner core is the small surface. The phase change materials have small coefficient of thermal conductivity, so if the core has too big diameter the process of the melting and the solidification will be slow by the thicker and thicker solid layer on the inner surface of the core.

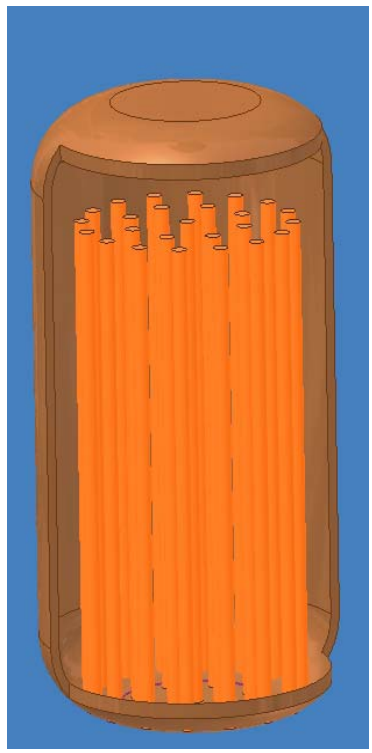


Figure 1. Conceptional model of a PCM-solar tank provided with inner tubes

In the other type of the tanks the balls are filled with the phase change material. The diameter of the balls is very small for the diameter of the tank. The balls are not fixed to the tank. It results a bigger heat exchange surface area between the water and the phase change material. The biggest disadvantage of this construction is the requirement special devices for making, filling and closing of the balls.

Our goal was to combine the advantages of the two different constructions. Additionally the easy manufacturing and the using of adaptable but not expensive materials was important too.

We have designed a tank with inner tubes filled with PCM-material:

This construction has bigger heat exchanging surface, and the manufacturing is simple.

We have made an Excel table to study the operating of the tank. In the table we can change the next parameters:

- diameter and height of the tank,
- diameter, height and number of the tubes,
- thermal properties of the phase change material,
- temperature difference between the PCM and the water,
- the operational parameters of the solar collector system.

2. Phase change materials

These materials can store energy by the melting at a constant temperature. No material has all the optimal characteristics for a PCM, and the selection of a PCM for a given application requires careful consideration of the properties of various substances. Over 20,000 compounds and/or mixtures have been considered in PCM, including single-component systems, congruent mixtures, eutectics and peritectics [3]. The isothermal operating characteristics (i.e. charging/discharging heat at a nearly constant temperature) during the solidification and melting processes, which is desirable for efficient operation of thermal systems [9].

One of the most important aspects during the selecting of the material is the conformable melting point and the high latent heat of fusion. The choice of the substances used largely depends upon the temperature level of the application [8]. Residential, commercial and industrial buildings often have hot water requirements at around 60 °C and bathing, laundry and cleaning operations in the domestic sector generally need it at about 50 °C [4]. The right melting point enables that the phase changing comes off during every usage cycle. Thereby the latent heat could be fully utilized. According to the required temperature of the domestic hot water the melting point should be between 40 and 50 °C. Out of accordance with the conventional solar tanks the temperature of the accumulation of heat is constant. Storage systems using these heat accumulator materials can store the energy from the solar collector at lower temperature level, too in winter. The stored energy can be used for pre-heating the cold incoming water [5].

The value of the latent heat is very important, because the higher latent heat results higher storable heat quantity. According these aspects we can choose from several materials. We have to mind the chemical properties, the thermal expansion and the aspects of safety.

PCMs in the 50 to 100 °C temperature range have been proposed to water heating and off-peak electrical heater applications.

Table 1. Commercially Available Phase Change Materials [6]

PCM	Melting point °C	Heat of fusion cal/g	Density, liquid g/cm^3	
Coolness storage				
$Na_2SO_4 \cdot 10H_2O \cdot NH_4Cl \cdot KCl$	8	29	1,49	gel
PE Glycol	11	24	1,12	
$CaCl_2 \cdot CaBr_2 \cdot 6H_2O \cdot KBr$	12	32	1,78	gel
Heat buffers and sinks				
$CaCl_2 \cdot CaBr_2 \cdot 6H_2O$	15-34	34	1,78	gel
$Na_2SO_4 \cdot 10H_2O \cdot NaCl$	18	35	1,48	gel
$CaCl_2 \cdot 6H_2O$	27	46	1,56	
Moderate temperature				
$Na_2SO_4 \cdot 10H_2O$	32	60	1,48	solid
$CaBr_2 \cdot 6H_2O$	34	28	1,96	
Neopentyl glycol	43	31		
$Na_2S_2O_3 \cdot 5H_2O$	48	48	1,67	
Paraffin wax	50	64	0,76	
Intermediate temperature				
$MgCl_2 \cdot Mg(NO_3)_2 \cdot 6H_2O$	58	32	1,52	
$NaCO_2CH_3 \cdot 3H_2O$	58	54	1,28	
$Na_4P_2O_7 \cdot 10H_2O$	70	44	1,8	gel
$Mg(NO_3)_2 \cdot 6H_2O$	89	39	1,55	
$NH_4Al(SO_4)_2 \cdot 12H_2O$	95	64	1,65	solid
High temperature				
$MgCl_2 \cdot 6H_2O$	114	40	1,45	
Polyethylene	132	48	0,96	solid

1 cal/g = 4,181 kJ/kg

Our first chosen phase change material is the paraffin 5838 (you can see the thermophysical properties in the Table 3.)

The paraffins are waxes at room-temperature. These are hydrocarbons. Increasing the number of C-atoms increases the melting point too. The normal paraffins of type C_nH_{2n+2} are a family of saturated hydrocarbons with very similar properties. Paraffins between C_5 and C_{15} are liquids, and the rest are waxy solids. Paraffin wax is the most commonly used commercial organic heat storage PCM [10].

Paraffin waxes are cheap and have moderate thermal energy storage density but low thermal conductivity and, hence, require large surface area [10].

3. Own conceptional models

The tank with inner tubes – combines the advantages: it has large surface for the quicker heat transfer, and it could be manufactured easily.

The tubes are made of corrosion-resistant steel. The thin walls of the tubes result a very good heat transfer. The tank has 25 tubes with 60 mm outer diameter. Figure 2 shows the arrangement of the tubes and the solar heat exchanger spiral:

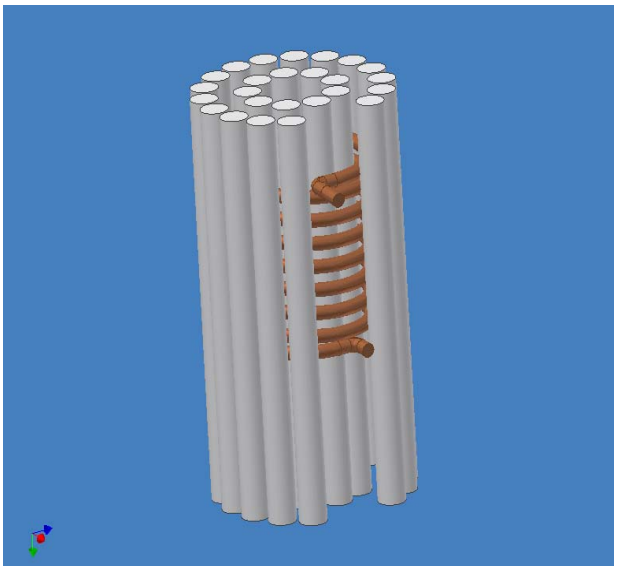


Figure 2. Arrangement of the tubes filled with paraffin and the solar heat exchanger

4. Calculation of the solidification

The main problem of the operating of the PCM tanks is the low coefficient of thermal conductivity of the phase change material. During the discharging of the tank the PCM solidifies to the inner surface of the tube. The thermal flux will be decreased by the thermal insulating effect of the thicker and thicker solid PCM layer. I have calculated the required time of the solidification of the paraffin.

The equation of the thermal conductivity in tubes with two layers (layer 1 is the solid phase of the paraffin, layer 2 is the wall of the tube):

$$\Phi = \frac{2\pi h\left(t_{w3} - t_{w1}\right)}{\frac{1}{\lambda_{PCM}} \ln \frac{d_2}{d_1} + \frac{1}{\lambda_w} \ln \frac{d_3}{d_2}}$$

Legend:

λ_{PCM} – coefficient of thermal conductivity of the paraffin
λ_w – coefficient of thermal conductivity of the tube
d_1 – inner diameter of the solid paraffin layer on the inner surface of the tube
d_2 – inner diameter of the tube
d_3 – outer diameter of the tube
t_{w1} – temperature of the phase change
t_{w3} – temperature of the outer surface of the tube

During the solidification the value of d_1 decreases from d_2 to 0. The t_{w3}-t_{w1} temperature difference depends on the temperature of the water in the tank. With the equation we can calculate the required time of the solidification. The next diagram shows the decrease of the $r_1=d_1/2$ inner radius of the solid phase in function of time. The parameters are according to 60 mm tube diameter, 1 mm wall thickness, the PCM is paraffin, the material of the tube is stainless steel. The difference between the phase change temperature and the temperature of the outer wall of the tube is 2 °C:

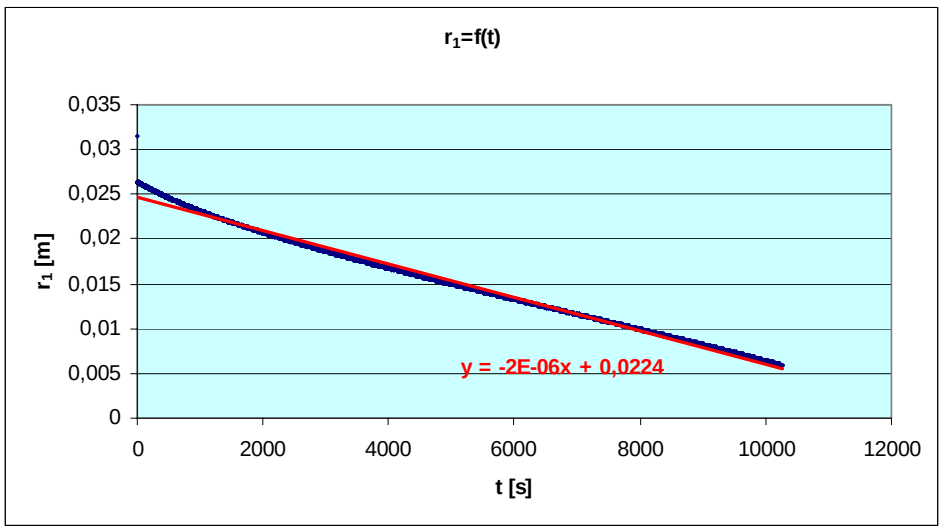


Figure 3. The decrease of the r_1 radius on the boundary of the liquid and solid phase in function of time, during the solidification

The calculated time of the solidification is 172 minutes. Choosing 40 mm tube diameter instead of 60 mm the solidification time decreases to 66 minutes. The next diagram shows the time of the solidification in function of the diameter of the tube:

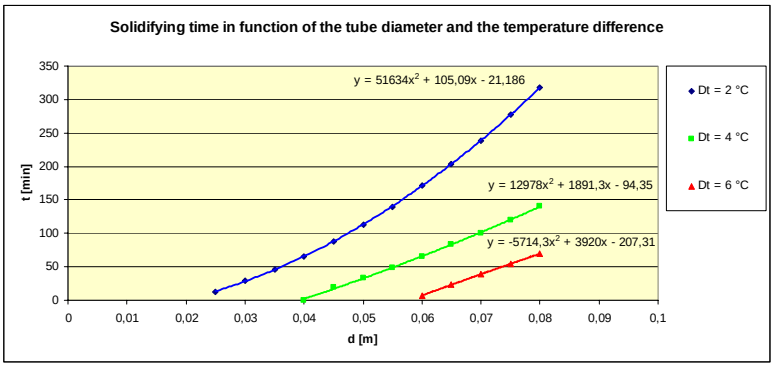


Figure 4. Time of the solidification in function of the tube diameter and the temperature difference between the PCM and the outer wall of the tube

The same amount of the PCM requires more tubes if the diameter is smaller. The specific heat exchange surface is greater, the solidification is quicker, but the volume of the material of the tubes is greater too, so the heat capacity of the solar tank is lower.

The above functions are definable according to the physical properties of the phase change material and the difference between the phase change temperature and the temperature of the outer surface of the tube. We can calculate the maximal tube diameter for a required phase changing time. With these functions we can calculate the ideal parameters for a definite system and operating.

5. Operating at lower temperature

The lower difference between the temperature of the collector and the outer air results higher collector efficiency.

The next diagram shows the irradiation and outer air temperature in a summer day in Szeged:

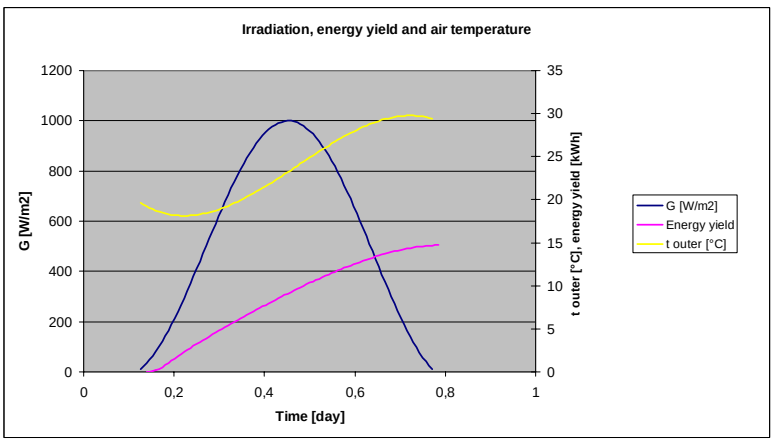


Figure 5. Irradiation, energy yield and air temperature in a typical summer day

The paraffin has smaller specific heat capacity than the water, so it results higher temperature in the PCM-tank, than the conventional solar tank in a summer day:

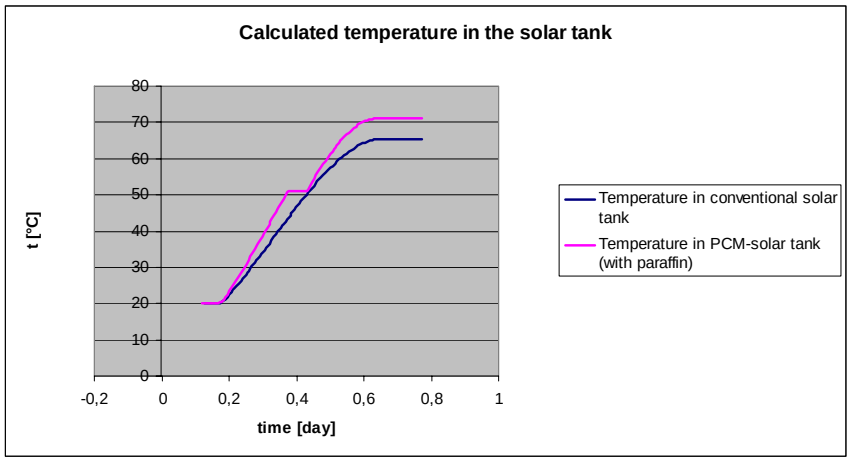


Figure 6. Comparing the temperature in a conventional solar tank and a PCM-tank with paraffin

The diagram above shows the result of a calculation of a PCM-tank with 70 kg water around the tubes and 170 kg paraffin in the tubes. We used for the calculation the efficiency characteristic of our own-designed experimental flat collectors. The equation of the collectors' efficiency:

$$\eta = -6{,}85 \cdot \frac{t_{coll} - t_{amb}}{G} + 0{,}752$$

Legend:

η – collector efficiency

t_{coll} – average temperature of the collectors [°C]

t_{amb} – ambient temperature [°C]

G – solar irradiation [W/m^2]

The temperature of the PCM-tank is higher than the conventional because of the lower heat capacity of the PCM, so the collectors has to operate at higher temperature, and the efficiency is lower because of the heat loss from the collector to the air. If our goal is the higher efficiency we have to choose another PCM with lower melting point.

The melting point of the Glauber's salt (Na_2SO_4 $10H_2O$) is 32 °C [6], and the latent heat is 250,9 kJ/(kgK). The heat capacity is 0,894 kJ/(kgK). The melting point is lower and the latent heat is much higher than the paraffin's. As we can see on the next figure the stored energy could be much higher than with paraffin:

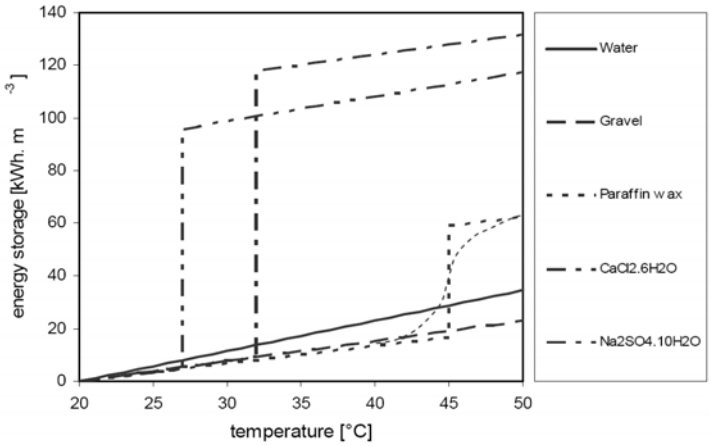


Figure 7. The compare of PCM and classic materials accumulator's charging [7]

The next table shows the properties of some PCM:

Table 2. Physical properties of Paraffin 5838, Glauber's salt and pentahydrate

		paraffin	Glauber's salt	Na_2SiO_3•$5H_2O$ (pentahydrate)
latent heat	J/kg	145000	250900	267000
specific heat, solid	J/(kgK)	2100	894	3830
specific heat, fluid	J/(kgK)	2400	894	4570
thermal conduct., solid	W/(mK)	0,2	0,64	0,1155
thermal conduct., fluid	W/(mK)	0,15		
density	g/cm3	1,412	1,2	1,45
melting point	°C	50	32	48

The next figure shows the calculation for the same day with conventional water tank and PCM tanks with different materials:

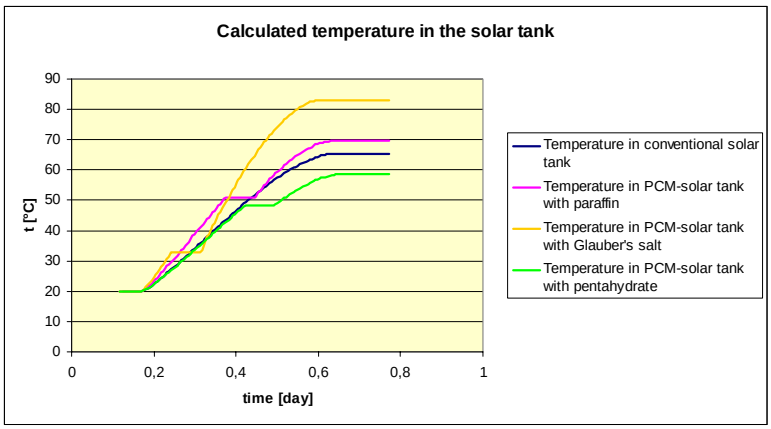


Figure 8. Comparing the temperature in a conventional solar tank and a PCM-tank with different materials

In our every calculations the same summertide weather conditions result lower temperature and accordingly higher collector efficiency with the conventional tanks, so the main advantage of the PCM-tanks is the lower space demand.

As we can see, the last horizontal part of the functions at the maximal temperature shows as the efficiency of the collectors reach the 0, and the temperature in the tank stagnates. The next diagram shows the collector efficiency in function of time:

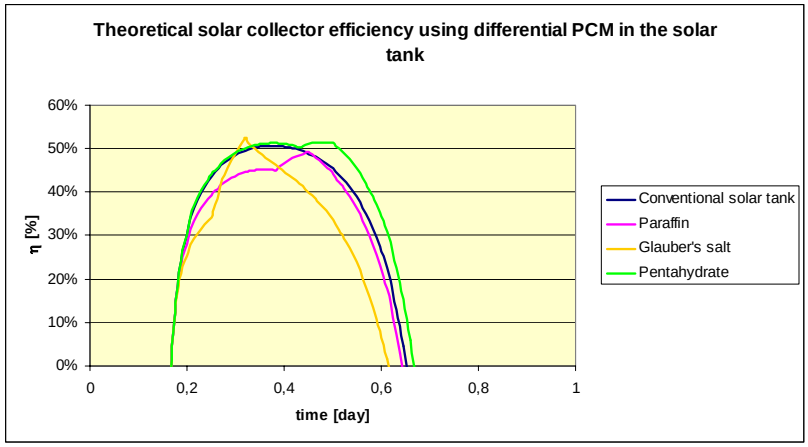


Figure 9. Comparing the values of the collector efficiency using a conventional solar tank or a PCM-tank with different materials

The calculations made for operating without heat removing from the tank. The results of the calculations show that beside of the latent heat the one of the most important thermal property of the PCM is the specific heat. Higher specific heat results higher collector efficiency. Out of the three PCM the pentahydrate has the highest specific heat. It results the highest collector efficiency in the afternoon.

6. Summary

Tank proposed by us with inner tubes combines the advantages of the existing types of tanks. These inner tubes are filled with the PCM.

The cost of the manufacturing of the tank is lower than the conventional tanks in trade with the same heat capacity and the space demand is much lower too. The other advantage of the PCM tank is the constant temperature during the phase changing period of the heat accumulation. This constant temperature could be lower than the temperature of the water tank with the same heat capacity. The lower temperature of the heat accumulation permits the higher efficiency of the collectors at low external temperature, but it depends on the type of the PCM.

The lower heat capacity of the PCM could result higher temperature in the solar tank, so the higher collector temperature results lower collector efficiency. We have made calculations for the operation of the solar system. With our PCM-tank we can test different phase change materials to measure the efficiency.

References

1. Lovász A; Bajnóczy G; Gagyi-Pálffy E; Prépostffy E.: Domestic Hot Water Pre-heater Utilizing Solar Energy. Periodica Polytechnica Chem.Eng. 50/1 45-53. 2006.
2. Muhsin Mazman, Luisa F. Cabeza, Harald Mehling, Miquel Nogues, Hunay Evliya, Halime Ö. Paksoy: Utilization of phase change materials in solar domestic hot water systems. *Renewable Energy 34 (2009) 1639–1643*
3. Dinçer I, Rosen MA.: Thermal energy storage, systems and applications. *Chichester (England): John Wiley and Sons; 2002. ISBN 0-471-49573-5*
4. Dharuman C, Arakeri JH, Srinivasan K.: Performance evaluation of an integrated solar water heater as an option for building energy conservation. Energy and Buildings 2006;38:214–9.
5. Bajnóczy G; Gagyi Pálffy E; Szolnoki L; Prépostffy E.: Solar Energy Storage by a Two - Grade Phase Change Material. *Periodica Polytechnica Chemical Engineering 51/2 (2007) 3–7. doi: 10.3311/pp.ch.2007-2.01. web: http://www.pp.bme.hu/ch*
6. Eric C. Guyer, David L. Brownell: Handbook of applied thermal design. *McGraw-Hill Book Company*
7. Ing. Ivo BĚHUNEK, Doctoral Degree Programme (1) Dept. of Electrical Power Engineering, FEEC, BUT E-mail: ivo.behunek@post.cz
8. M. Ravikumar, Dr. Pss. Srinivasan: Phase change material as a thermal energy storage material for cooling of building. *Journal of Theoretical and Applied Information Technology. 30th June 2008 Vol. 4 No. 6 pp. 503-511*
9. Mithat Akgün, Orhan Aydın, Kamil Kaygusuz: Experimental study on melting/solidification characteristics of a paraffin as PCM. *Energy Conversion and Management 48 (2007) 669–678*
10. Garg HP, Mullick SC, Bhargava AK.: Solar thermal energy storage. *Dordecht: D. Reidel Publishing Company; 1985. ISBN 90-277-1930-6*

The Project named „TÁMOP-4.2.1/B-09/1/KONV-2010-0005 – Creating the Center of Excellence at the University of Szeged" is supported by the European Union and co-financed by the European Regional Fund.

www.nfu.hu www.okmt.hu