Nonlinear temporal and spatial forecasting: modelling and uncertainty analysis (NoTeS) – MASIT20

Risto Ritala
Tampere University of Technology, Institute of Measurement and Information Technology
P.O. Box 692, FI-33101 Tampere
Tel. +358 40 8490 922
Email: risto.ritala@tut.fi

Esa Alhoniemi
University of Turku,
Department of Information Technology
FI-20014 Turku
Tel. +358 2 3338 692
Email: esa.alhoniemi@utu.fi

Tuomo Kauranne
Lappeenranta University of Technology,
Department of Mathematics
P.O. Box 20, FI-53851 Lappeenranta
Tel. +358 5 6212 825
Email: tuomo.kauranne@lut.fi

Kimmo Konkarikoski
Tampere University of Technology, Institute of Measurement and Information Technology
P.O. Box 692, FI-33101 Tampere
Tel. +358 3 3115 2470
Email: kimmo.konkarikoski@tut.fi

Amaury Lendasse
Helsinki University of Technology, Adaptive Informatics Research Centre
P.O. Box 5400, FI-02015 TKK
Tel. +358 40 7700 237
Email: lendasse@hut.fi

Miki Sirola
Helsinki University of Technology, Adaptive Informatics Research Centre
P.O. Box 5400, FI-02015 TKK
Tel. +358 9 451553
Email: miki.sirola@hut.fi

Abstract

NoTeS project is developing a generic toolset for spatio-temporal forecasting and forecast uncertainty analysis through analyzing five widely different test cases: forecasting energy consumption, diagnosing quality variations at a pulp mill, estimating and controlling paper quality on the basis scanning sensors, supporting operational decisions at nuclear power plant, and segmenting medical images. An abstract set of functions common to all these applications – with possible exception of image segmentation – has been identified, and the corresponding structures for data, models and estimates specified. The project is in stage in which the test cases have demonstrated practical relevance and economic potential and in which the specification work for generic toolset is stabilizing.

Keywords: Forecasting, uncertainty, nonlinear modelling, variable selection, Kalman filtering, SOM, segmentation

1 Project background and goals

It is common to industrial production processes, economic systems, and multi-imaging that they have complex dependence structures, internal stochastic mechanisms, and that the data obtained from these systems are uncertain. In all the cases the goal of analyzing the data is to support decision making – either in human or automatically optimized form – through inferring the dependence structures as non-linear models and then to use the models in forecasting (temporal structures), in object enhancement (spatial structures) or in object forecasting (spatio-temporal structures). The spatial (imaging), temporal (system dynamic) and spatio-temporal (system dynamic of functionally related observations) modelling are strongly linked in that the mathematical theory of non-linear time series analysis to large extent carries over to non-linear spatial series and spatio-temporal analysis simply be changing the dimensionality of the support set from 1 to 2.

The technical objectives of this project are:

1. To integrate the pre-existing methodologies of the partners (SOM, time-series analysis, functional time series analysis, distribution dynamics, data-assimilation, and Kalman filtering; see References) into a coherent approach to nonlinear modelling, analysis and forecasting/filtering for dynamic/spatial/spatio-temporal systems with inherent stochasticity and measurement uncertainty. Forecasting provides information about the uncertainty/probability density functions (pdf) of the predictions and about their development over the forecasting horizon. Such uncertainty information is necessary when the forecasts are fused with other data in decision making.

2. To implement this methodology in a form that can be readily used in **practical data analysis** either autonomously or as packages easy to integrate into present systems of the companies partly financing the project.

3. To demonstrate the applicability of the methodology in three widely differing application areas: in analysis of industrial process of papermaking and nuclear energy production (temporal and spatio-temporal analysis), in analysis of energy markets (temporal analysis) and in analysis of medical multimodality imaging of healthy and disease state Imaging analysis). Altogether project addresses five application cases, three of which are in the area of industrial processes.

2 Project work programme

Table 1. Basic information of the project.

Project partners	Tampere University of Technology; Inst. of Measurement and Information Technology (coordinator, Bayesian information dynamics, uncertainty and measurements) Helsinki University of Technology; Adaptive Informatics Research Centre (nonlinear time series modelling, variable selection, SOM) University of Turku; Department of Information Technology (multi-imaging, segmentation) Lappeenranta University of Technology; Department of Mathematics (Kalman filtering and other estimation methods)
Duration of the project	1.4.2006 – 31.12.2007 (NoTeS), 1.1.2008 – 31.12.2008 (NoTeS2)
Funding of the project	Tekes, Keskuslaboratorio (KCL), Metso Automation, Nordkalk, Process Vision, Teollisuuden Voima (TVO) Total funding 684 000 €, Tekes share 602 000 € (NoTeS) Total funding 400 000 €, Tekes share 360 000 € (NoTeS2) Varsinais-Suomen sairaanhoitopiiri (VSSHP) is a non-funding partner providing data for the project

The process has focused on five test cases and inferred their commonalities into a specification of an integrated system. Limited prototypes of the generic toolset were constructed and some of the test cases demonstrated with these prototypes. This work will then be continued in the sequel project NoTeS2 which tackles the five original test cases, an additional one, and continues with further prototyping of the generic toolset. The project runs generic method integration and five test cases in parallel. The first project had nine Work Packages scheduled so that both in mid-term and at the end of the project the experiences from test cases are generalized into specifications and prototypes for a generic toolset servicing all the test cases. The sequel project has five work packages with similar structure.

The test cases span a large range of estimation and forecasting applications. Three of the NoTeS test cases are purely temporal: one for energy demand prediction (TC1), one for pulp mill quality variation (TC2) and one for nuclear power station decision support (TC4). The additional NoTeS2 test case on quality management in mineral production (TC6) belongs to this category as well. The temporal cases differ in that the energy demand prediction deals with selection of variables for non-linear system models, whereas the pulp mill case and the minerals case deal with variable delay, and the nuclear power plant case applies SOM in a quasistatic way. Furthermore, the nuclear power plant case analyzes the Man-Machine Interface and visualization, not tackled extensively in the other applications. The energy demand prediction case is being further developed into spatio-temporal forecasting by applying functional data analysis. The fourth test case on quality variations in paper web (TC3) is a spatio-temporal estimation and forecasting problem with severe constraints on state measurement information. Hence the problem of measurement optimization is also addressed, together with other projects dealing with optimal measurement (Tekes funded Tuotanto2010 –project, and Academy of Finland funded

SyDeMIS project). The fifth case analyses segmentation in multi-imaging (TC5) and is thus purely spatial.

3 Project results

3.1 Generic nonlinear temporal and spatial forecasting

In order to develop a generic toolset for all the five test cases, the commonalities of the problems have been analyzed in detail. The goal of the generic toolset development is to define a set of generic functions that operate on a generic data structure.

At least four of the applications – excluding test case 5 – can be built with generic functions for model selection, model parameter identification, model validation, state estimation (filtering), state forecasting and model updating. The four test cases are optimal estimation/forecasting tasks and thus forms of Kalman filtering, the term understood in its widest sense. However, the test cases focus on quite different steps of the generic approach as follows:

- TC1 has the focus on model selection (variable selection) which is tightly coupled to model parameter identification and model validation. The state estimation and state forecasting are rather simple, and the model updating is handled mostly by redoing the model selection/parameter identifications steps.
- TC2 has two approaches, the one not using system specific information with focus similar to that of TC2. The approach with system specific information uses pre-processing of the data into nondynamic problem with an external model not identified as part of the task but selected based on considerations of the first principle. After pre-processing the approach is similar to TC1.
- TC3 also has two approaches. In the first one, system specific information about system model is utilized and then the focus is on rather complex state estimation and forecasting steps. In this approach, the measurement system optimal scheduling is also considered as an integral part of the problem. The other approach in TC3 does not use system specific information but identifies the model dynamically, i.e. starts with an initial – and in general poor – guess but updates the model rapidly to an efficient estimation forecasting element. Both approaches in TC3 deal with estimation and forecasting uncertainty explicitly and dynamically.
- TC4 is a diagnostics task in which a Self-Organizing Map (SOM) is the model identified. The approach deals explicitly with prediction errors of discrete states – rather than continuous states in the other test cases.

TC5 is rather different from the other cases. It is a non-dynamic spatial estimation and classification problem. It was included in the project in order to extend the genericity of the toolset to be developed as far as possible. However, it has turned out rather difficult to integrate to other cases. It is imperative to continue this case also and explore the connections to generic toolset, although eventually it may turn out to be too different from the other to be supported by the generic toolset.

From the considerations above, the generic toolset has been prototyped. The main approach has been to provide the strongest support to TC1 and then analyze, which features of the TCs 1-4 cannot be supported, the toolset data structure and functions are then expanded.

At the end of the first phase of NoTeS project, the generic toolset is still at its early stage, but the analysis to develop a prototype of toolset by the end of NoTeS2 is quite mature.

3.2 Test case 1 (HUT; Process Vision)

Prediction of electric load or electricity consumption in power grid is becoming more and more important due to the liberalization of the energy market in Europe. In many time series prediction problems, linear models provide accurate enough predictions. However, the behaviour of electric time series is rather nonlinear. Furthermore, even a small decrease of the prediction error may largely decrease the operational costs of the grid and then the use of environmental resources.

Even though the short-term and the mid-term predictions have been studied a lot, the accuracy of the mid-term prediction is not yet satisfactory. One of the reasons is the difficulty due to the uncertainty of the external variables (for example, the weather forecast). Thus we have focused on the mid-term forecasting (several weeks ahead) based on hourly values. Our results can be classified into two main categories:

1) Tools for the selection of the best variables for the mid-term prediction have been developed. Variables are selected from the past values of the time series to be predicted itself and from exogenous variables such as temperatures.

Firstly, by reducing the number of input variables we aim to avoid the curse of dimensionality, giving the possibility of increasing the regression generalization performances. Secondly, a reduced set of variables allow in real applications an easier interpretation of the relationship between features and outputs.

A new algorithm for variable selection which was introduced for the problem of short-term selection has been extended to the problem of mid-term prediction. Furthermore, the problems due to the uncertainty of some variable, (for example variables of the weather forecast) have been investigated.

2) The problem of the computational time needed for the prediction has been investigated.

Feed-forward neural networks are often found to be rather slow to train, especially on important data-sets related to the data mining problems of the industry. Therefore nonlinear models are not used as widely as they could be, even considering their overall good performance. The slow training of the networks result from few and simple reasons; many parameters have to be tuned, by slow algorithms, and the training phase has to be repeated many times to make sure the model is proper and to be able to perform model structure selection (number of hidden neurons in the network, regularization parameters tuning).

A new algorithm for the determination of the weights of the hidden neurons called Extreme Learning Machine (ELM) has been introduced in 2004. This algorithm decreases the computational time required for training and model structure selection of the network by at least couple of orders of magnitude. Furthermore, the algorithm is rather simple, and thus easy to implement.

We have proposed a new methodology called Optimally-Pruned ELM (OP-ELM), based on the original ELM. The OP-ELM methodology is accurate; comparing to well-known methodologies the OP-ELM achieves roughly the same level of accuracy. The main advantage of OP-ELM is that the computational time is orders of magnitude less than with traditional methods.

3.3 Test case 2 (TUT, UTU; KCL)

In test case 2, the aim has been to study methods of analyzing the factors causing unwanted variations in a measure quality of an industrial process. The goal was to develop a set of methods which are able to reveal the causes for detected quality problems. The approach has been two-fold: firstly to analyze the data without any system specific information thus obtaining widely applicable and quick to use diagnostics methods, and secondly to use system specific information, such as process dynamic simulation, to obtain an accurate time warping of the signals.

The test case data encompassed 90 continuous and 70 laboratory measurements gauged in various steps of a pulp bleaching process. Each measurement contained from a hundred up to 9000 time points. The data was given in a normalized form and no specific information of the structure or the function of the system was provided along with the data. This was particularly limiting for the second approach in which system specific information is utilized.

We discuss first the approach without system specific information and then the methods developed for time warping with dynamic process simulation models.

The analysis without system specific information proceeds as follows. First the data is pre-processed in three steps:

1) Pruning the candidate variables by discarding identical, constant and the downstream measurements and the variables having too many missing values.

2) Separating the unwanted time points based on user specified criteria.

3) Estimating the lags between the measurements by utilizing the information of the order of the detectors in the system and the maximum delays between different process steps.

The variable selection is applied in two steps:

1) Fitting a PLS (partial least squares) model using SFFS (sequential floating forward selection).

2) Finding similar shapes in other variables in the vicinity of a time window containing interesting variation in the target variable. Several distance functions were tested to measure the similarity: correlation, partial correlation, and Euclidean distance of original data, Fourier transformed segments and warped segments.

The approach without system specific information provided results as follows. The set of variables chosen by SFFS/PLS method varied according to chosen training data, however, a set of variables were chosen virtually in every run. Although most variables picked by SFFS/PLS were logical choices, the resulting model was not sufficient to describe the given data. The model was able to estimate the main trend of the observed quality, but a closer look would reveal defects in reproducing the minor variation. Possible causes for the observed flaws include 1) erroneous lag estimates, 2) badly chosen training data, 3) lack of laboratory measurements and 4) nonlinear nature of the dependencies cannot be expressed with a linear model such as PLS.

Even more alternation of chosen variables was observed in the lists built by the segment based method. Furthermore, only few of the variables chosen by SFFS/PLS were selected by this method. An obvious reason for the differences is that this method uses the chosen variable as the sole source of deviation, while the SFFS/PLS method considers also the interaction between the chosen variables.

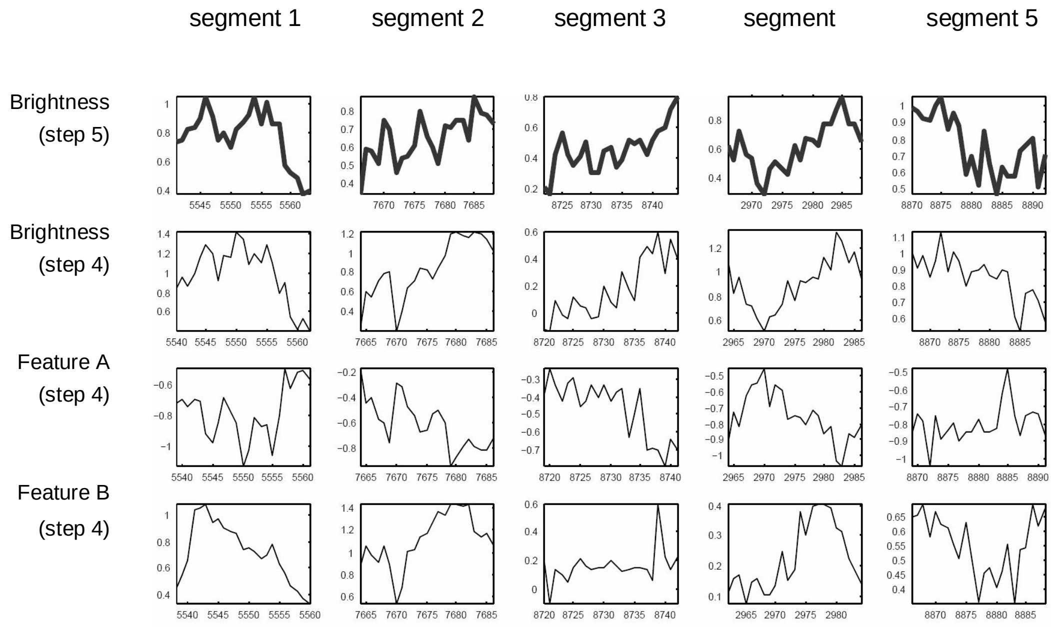

Figure 1. Finding similar shapes in observed signal and the other variables using correlation as the distance measure. Similarity between the brightness in step 4 and step 5 was striking. The effects of self-evident factors should be compensated if they are not considered interesting in the researcher's point of view. Earlier influencing factors can be further tracked down by setting the best explaining factor as the new target for the analysis.

The approach with system specific information was based on impulse response and residence time functions. In particular, tank dynamics with time varying tank volumes were studied. It was shown that accurate time warping methods can be based on tank dynamics simulation, but this – as all model based methods – is depending on high quality model. The computational tools for such analysis were implemented in Matlab/Simulink. The method could not be directly applied to test case data, as it requires absolute value volume and flow information whereas the test case data were normalized. However, in any practical process diagnostics task, the absolute information is available and thus the methods can be readily applied. In practical applications the main problem would be that the tank dynamics may be quite complex and even time varying due to material clogging and other phenomena.

With time warped data with high number of variables the problem of variable selection must also be solved. As this is done for data without dynamic interactions, the methods are simpler. As part of the study some well-known linear regression variable selection methods were implemented in Matlab: stepwise regression, non-negative garrotte, and least absolute shrinkage and selection operator (LASSO).

3.4 Test case 3 (TUT, LUT; Metso Automation)

Paper quality is measured on-line with scanning sensors travelling over the web from one edge to the other in 20–30 s. The paper web is travelling continuously at speeds 20–30 m/s so that during one scan the web has advanced 400–900 m. Paper quality is controlled separately with time-wise controller (MD control) and spatial controllers (CD control). Therefore the signal from the scanning sensor needs to be separated into MD and CD components for the controller. This test case analyzes the estimation of CD and MD variations in the web, in the presence of variations that are not entirely in either CD or MD. In particular, the effect of scanning speed and selection of sensor path have been addressed.

The approach in estimation has been that of Kalman filtering. Various forms of systems models have been applied. In particular the following cases have been analyzed in detail

- random walk disturbance model in which CD position stochastic processes are independent
- random walk model with controller action effects included
- both random walk models with an additional random walk disturbance that is common to all positions in CD (provides a tuning parameter in CD/MD separation when interpreting deviation from prediction into CD and MD components)
- effect of additional but uncertain pure MD variation
- ARMA/Fourier models for the 2d variation.

The properties of these methods have been tested with a simulator developed during 2006 in the project. The effects of scanning modes on the quality of MD/CD separation and the 2d estimates have been studied with the simulator. This work has led to understanding the opportunities and limitations of irregular scanning paths and useful information for practical systems has been obtained. The main conclusion is that the present commercial systems for CD/MD separation cannot handle irregular scanning but the separation/estimation method must be tailored from kalman filtering approach and with proper choice of noise models. If the scanner is completely stopped as part of the path, a switching between noise models is needed.

Fast MD waves and “tilted waves are known to cause major CD/MD separation problems when occurring at frequencies close to harmonic or subharmonic to the scanning frequency. The project has shown that by varying the scanner speed and analyzing with Fourier transforms the forecasting error, detection methods for fast MD variations and “tilted waves” can be set up.

The optimal scanning path problem has been formulated both for optimal controller performance and for optimal quality of information. The both formulations lead to complex dynamic programming problems. However, their solutions in some simplified cases have been constructed.

A major part of the work in TC3 during 2007 was devoted to a unique data acquisition task. On a board machine the raw data of the scanning sensors was collected together with the entire machine reel of length of some 3 km. With the help of a partner outside the project the machine reel was measured on-line with high resolution (web travel during one scan was less than 1 m to be compared with the on-line scanner in which web travel is several hundreds of meters). The on-line scanner data and the off-line data were aligned to better than 10 m (MD) accuracy. As the off-line data gives the true 2d-variation, the performance of on-line scanner could be analyzed in detail. Furthermore, with the off-line data any scanner path could be simulated and compared with the performance of regular scanner. Such data has not ever before been available and thus provides an excellent test bench for analyzing estimation and forecasting methods. Control Systems 2008 Conference to be held at Vancouver in June 2008 has accepted four papers on this data set and the estimation methods developed and analyzed in TC3.

3.5 Test case 4 (HUT; TVO)

Early fault detection with data-analysis tools in nuclear power plants is one of the main goals in test case 4 of NoTeS project. The industrial partner in this subproject is Teollisuuden Voima Oy, Olkiluoto nuclear power plant. Data analysis is carried out with real failure data, training simulator data and design based data, such as data from isolation valve experiments. A control room tool, visualization tools and various visualizations are under development.

A toolbox for data management using PCA (Principal Component Analysis) and WRLS (Weighted Recursive Least Squares) methods has been developed. Visualizations for e.g. trends, transients, and variation index to detect leakages are used. Statistically significant variables of the system are detected and statistical properties and important visualizations are reported. Data mining methods and time series modelling are combined to detect abnormal events.

X-detector tool based on feature subset selection has been developed. The idea is to do real-time monitoring and abnormality detection with efficient subsets. Measuring dependencies and cluster separation methods are used in variable selection in this visualization tool.

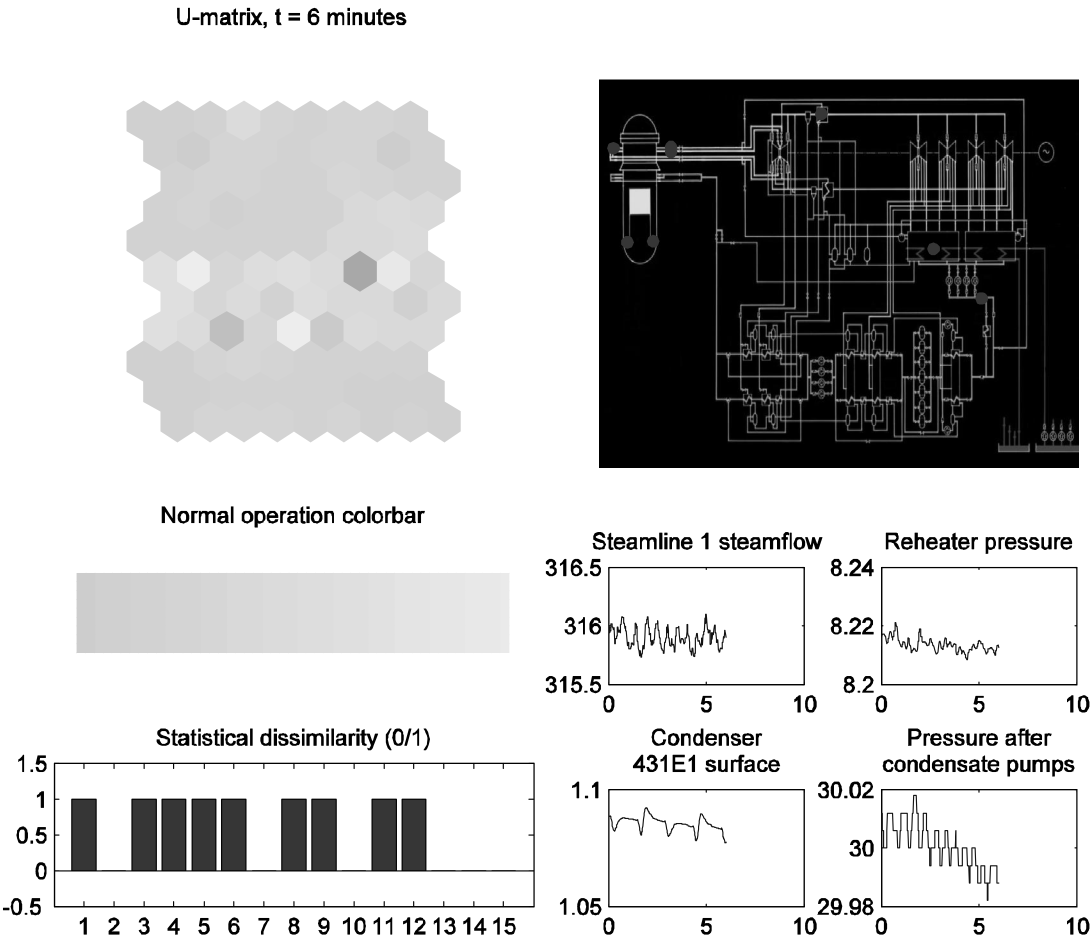

Figure 2. X-detector tool user interface: leakage in the main circulation pump. SOM visualization combined with statistical Kolmogorov-Smirnov test, process flow diagram and selected process variable graphs.

3.6 Test case 5 (UTU; VSSHP)

Magnetic resonance imaging (MRI) is a non-invasive 3D medical imaging method that can be used for detecting structures inside human body. The method is widely used with human brain images in order to acquire information about the anatomy of brain. With different MR scanning parameters it is possible to produce different modalities of images, such as so-called T1- and T2-weighted MR images. For performing later analyses, a physician often wants to segment brain MR images into three regions including the grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF). Magnetic resonance imaging of small children differs significantly from imaging of adults. Due to the smaller head size, a single voxel in an image represents a larger physical region, and because of shorter imaging times and of the possible head movements the data are noisier for small patients. These factors make the segmentation of the infant brain images a difficult task.

In test case 5, a new data-driven method was designed and implemented for segmentation of MR brain images taken of premature infants. A benefit of the new method is that it takes into account the specific characteristics of small children MR imaging. Unlike the customary MR segmentation methods, our method does not make use of the so-called a priori information of the brain structure, since this information is too unreliable and diverse for small children in different stages of their development.

The segmentation of the T1-weighted MR images is launched by a pre-processing step in which the region of the brain tissue is separated from the surrounding scull region. The step utilizes a triangularized 3D mesh which is iteratively adapted to the brain surface. Instead of actual intensity values we apply here the intensity gradients for solving the problems caused by the unwanted variations in the intensities at different regions of the images.

In order to address the variance of the intensity values, the region of the brain tissue is segmented to 3D watershed segments. Then a coarse level segmentation is acquired by clustering the watershed segments into GM, WM and CSF. For clustering, a median value of intensities inside each watershed segment is calculated. After this, the median values are clustered with the EM-algorithm by fitting a three-component Gaussian Mixture Model into the median value data. Finally, the watershed segmentation is used instead of a priori information in the final segmentation of the intensity values of the original T1-weighted MR image.

The brightness inhomogeneity of the brain region is corrected adaptively. This is a demanding task due to the scarcity of image data for premature infants. Here, intensity values in the brain region are first clustered with the K-means clustering method. Then, inhomogeneity correction with the thin plate spline method is applied to the MR image while adjusting 3D-splines so that the inhomogeneity inside the CSF and GM regions is minimized. The same two step process is iterated until the result of clustering remains practically unchanged.

Premature infants go through the so-called myelination process at their early age (< ~2 years). As a result of this, the observed MRI intensity values of the WM change over time. This causes some voxels in the WM region of non-myelinated tissue to erroneously be classified as GM. We address this problem with a special myelination correction step. The T1-weighted image is corrected by utilizing the intensity value relations in the corresponding T2-weighted MR image of the same patient. Although the T2-weighted image is not suitable for the segmentation of CSF, it allows better separation between the intensities from the non-myelinated and myelinated tissue types. This additional information is utilized in the iteration steps of the EM-algorithm when clustering the watershed segments and determining the final WM-GM-CSF segmentation.

While the segmentation algorithm is ready and operates properly, its performance evaluation is currently going on.

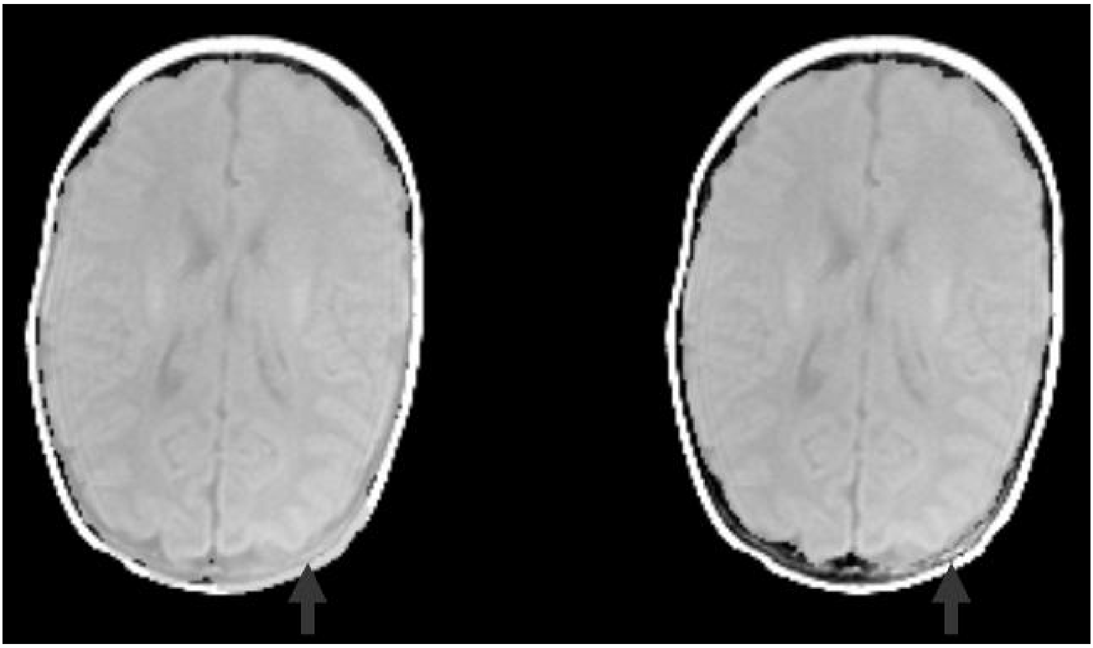

Figure 3. Left: Segmentation of an MR image for small children brain with standard brain extraction method (BET). Right: segmentation with image gradients.

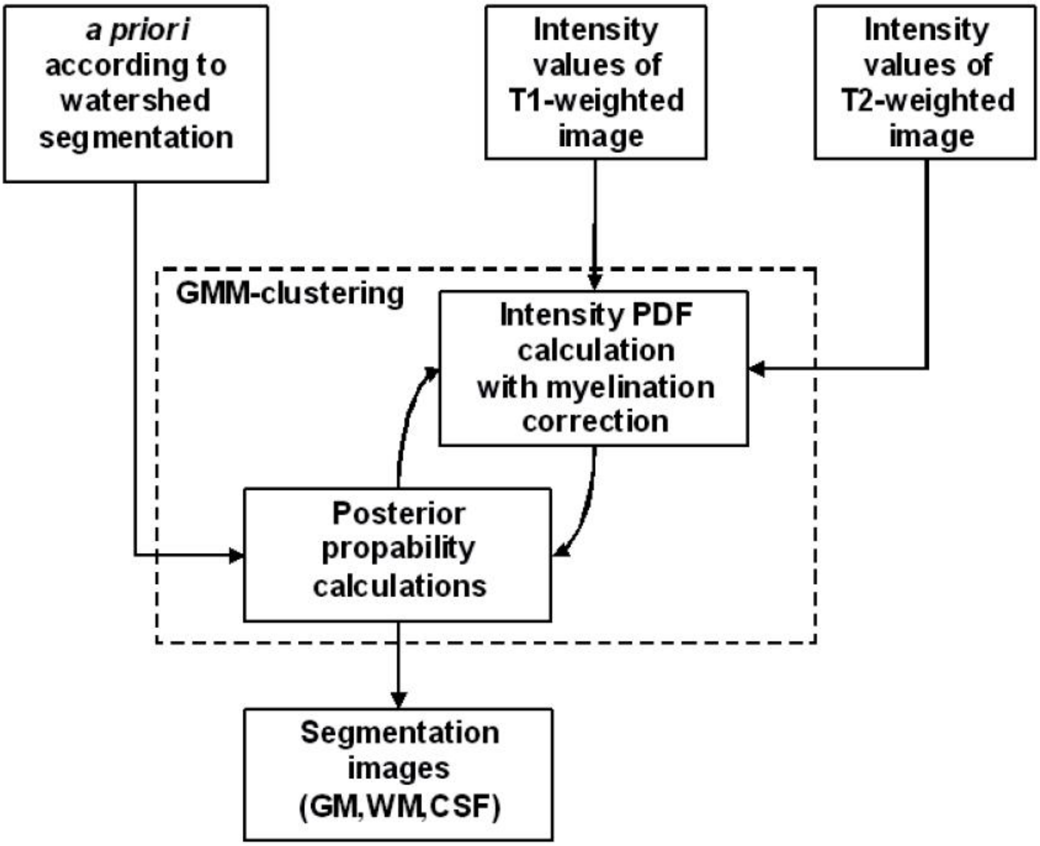

Figure 4. Schema of final brain segmentation to GM, WM, and CSF regions with watershed segmentation as a priori information.

4 Impact of the results

The research problem addressed in NoTeS is highly relevant for MASI programme since non-linearity and non-stationarity are common characteristics in most of the data available from industrial and economic systems. NoTeS project is focused on adding value to data by using advanced modelling techniques. Although linear methods and many existing nonlinear methods provide useful information out of such data, the results of NoTeS increase the amount of information achieved and, in particular, increase the understanding of the validity/uncertainty of the results. The accompanying uncertainty analysis is the key when applying the results in practical decision making through formal statistical decision making theory.

The impact of the results obtained is two-fold. Firstly, the development of the generic toolset opens up wide possibilities for industrial advanced spatio-temporal forecasting applications in long term, and secondly, each of the test cases has direct short-term industrial relevance and these results can be integrated into products of the funding partners. Obviously, successful case studies are also important in building confidence towards the usefulness of the advanced spatio-temporal forecasting methods, and therefore test cases support the uptake of the generic toolset.

Test case 1 has provided forecasting methods for the difficult medium-term optimization energy production under non-stationary conditions. Such methods have tremendous value for the energy producer and opens up strong business opportunities for companies providing software based on these models. Within the funding consortium of the project Process Vision operates in this business.

Test case 2 has provided diagnostic tools for dynamic process and quality variations, in particular for pulp production. Since pulp is produced in large quantities and Finland is an important player in the pulp making business, the reduction of quality variations has both commercial and national value. KCL provides diagnostic services for pulp mills and is expected to be the first exploiter of the methods. The methods develop can be applied in much wider context of process and quality variations in any process industries.

Test case 3 has opened up a new control degree of freedom in paper quality control: the mode of sensor scanning. This has been shown to be an important tool in paper quality management and some of the results are expected to be product features in near future. Metso Automation, a funding partner of the project, is expected to be the main avenue for productisation and commercialisation of TC3 results.

The results of test case 4 strongly support process diagnostics at nuclear power plants. Furthermore, education of students both with various methodologies and application knowledge is an important impact. These benefits can be readily achieved at Teollisuuden Voima, a funding partner of the project.

Medical imaging is an important area both from healthcare and business point of view. Image segmentation of premature infants has been a long standing image interpretation problem to which the NoTeS projects have provided an essentially improved solution. Hence the results of test case 5 have a high impact in healthcare and medical imaging business.

5 Future plans

The project is continued as NoTeS2 for 1.1.–31.12. 2008. All the five test cases from NoTeS are continued and an additional one for mineral processing is added. The main goals of the test cases are two-fold: to further improve and widen the methods developed in the first phase of the project, and to bring the test cases into common framework to support the development of the generic toolset. The further development of test case specific methods will be driven by considerations of practical relevance: the goal in each of the applications is to bring the methods to a level at which they will have immediate impact in commercial products, services and/or production activities.

The development of a full scale prototype of the generic toolset is the main goal for NoTeS2. The work in test cases have laid a strong basis for such a development and the limited prototypes of the toolset, completed during NoTeS, have led to profound understanding about the structure of the tool set. The main challenge is to bring the ideas coming from the mathematical structure of the non-linear estimation/ forecasting problem and software structures together. The project partners will analyze in workshops the software structures required from the point of view of applications. However, as NoTeS2 is only for one year – instead of the initial proposal for a two-year project – the outcome will be a prototype rather than fully tested software.