DYNamic Identifiability Analysis (DYNIA) – Thorsten Wagener

PUB-IAHS Workshop
Uncertainty Analysis in Environmental Modelling
6th – 8th July 2004

a) Introduction:

DYNIA is a new approach to locating periods of high identifiably for individual parameters and to detect failures of model structures in an objective manner. The proposed methodology draws from elements of the Regional Sensitivity Analysis (RSA), includes aspects of the GLUE approach, wavelet analysis and the use of Kalman filtering for hypothesis testing.

The steps taken in the procedure can be seen in the flow chart in Fig. 1. Monte Carlo sampling based on a uniform prior distribution is used to examine the feasible parameter space. The objective function (support) associated with each parameter set is shown as a dotty plot (Fig. 1[a]) and the cumulative distribution of the the top 10% of the distribution is derived (Fig. 1[b]). Segmenting the range of each parameter (e.g., into 20 containers) and calculating the identifiability measure (gradient) in each container leads to the (schematic) frequency distribution (Fig. 1[c&d]). The highest value, indicated by the darkest colour, marks the location of greatest identifiability (peak) of the parameter.

Calculating parameter identifiability at every time step t using only the residuals for a number of time steps n before and after the point considered, i.e., using a moving window or running mean approach, allows the investigation of the identifiability as a function of time (Fig. 1[e]). The gradient distribution plotted at time step t therefore aggregates the residuals between t - n and t + n, with the window size being 2n + 1. The number of time steps considered depends upon the length of the period over which the parameter is influential.

b) Advantages

- DYNIA combines the ability to analyse parameter variation in time and the ability separate periods of noise and information.
- The algorithm is easy to understand and can be applied to any dynamic mathematical model of a natural (environmental or hydrological) system. It therefore combines parameter estimation with model diagnostic capabilities.
- A minimum number of assumptions does have to be made. The choice of a cut-off threshold is not critical, because it is only important that the shape of the distribution tops is captured.
- DYNIA can be applied to any model in an off-line mode. It only requires a randomly sampled parameter population and the corresponding time-series produced by the model as an input.
- Any appropriate measure of performance (i.e. objective function) can be used.
- Using parameter variation as an indicator of model structural failures assumes, of course, that the specific parameter does not describe characteristics of the catchment that are time variant, for example the leaf area. Variation in good parameter values in those cases would rather corroborate the model structure, and not indicate a failure.

c) Disadvantages

- DYNIA requires that sensible feasible ranges for each parameter be defined and that the number of models (i.e., parameter sets) sampled is sufficient to represent the shape of the response surface.
- A limitation of the proposed measure of identifiability arises if any near-optimal parameter values are remote from the identified peak of the marginal distribution, as the relevance of such values would be diminished. It is therefore important that a detailed investigation of the dotty plots be undertaken to verify periods of high identifiability. So far results are surprisingly stable even with small sampling densities.
- Parameter interdependence is only implicitly considered in this approach, i.e. parameters are analysed as a set. Interdependence should additionally be estimated by the investigation of the response surface or the variance-covariance matrix.
- Care has to be taken when interpreting the DYNIA results of time steps at the beginning and the end of time-series. Here the full window size cannot be established and the result is distorted. This is an effect similar to the cone of influence in wavelet analysis.

d) Assumptions

Uniform random sampling is sufficient to explore the feasible parameter space.

The feasible ranges for all parameters can be defined by the modeler.

Lack of model performance is not dominated by model structural error (in the current implementation of the algorithm).

e) Most appropriate application areas

1. To estimate parameters, i.e. simple model calibration or identification. The approach presented is an attempt to mimic manual calibration more closely than traditional schemes. One requirement of manual calibration is that the proper calibration of a conceptual model should result in parameters that cause model components to mimic processes they were designed to represent (NWS, 2000). This requires the isolation of the effects of each parameter—a task difficult to achieve with manual or single-objective calibration, but in line with the DYNIA approach.
2. To analyse model structures. The awareness of the influence of model structural inadequacies on prediction uncertainty has grown in recent years, and the analysis of parameter variation in time is one possible approach to analyse this.
3. The algorithm relates model parameters and response modes of the natural system. The correct working of the model can therefore be checked. Do the parameters, and therefore model components, actually work as they are supposed to? Are there components that have little or no effect in producing the desired response?
4. To investigate data outliers or anomalies. Model structural error is not always the cause for time-varying parameter values. It can also be caused by erroneous data. Further analysis is usually required to be able to distinguish between the two. For example, data error might reveal itself by just being a one-off misfit between observed and calculated flow, whereas structural error will more probably be a consistent problem during similar response modes in different years.

f) Reading list

Wagener, T., McIntyre, N., Lees, M.J., Wheater, H.S. and Gupta, H.V. 2003. Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis. Hydrological Processes, 17(2), 455-476.

Wagener, T., Camacho, L.A. and Wheater, H.S. 2002. Dynamic identifiability analysis of the transient storage model for solute transport in rivers. Journal of Hydroinformatics, 4(3), 199-211.

g) Software availability

DYNIA is part of a more extensive suite of uncertainty analysis methods combined using a single Graphical User Interface in Matlab®. The tool is called the Monte Carlo Analysis Toolbox (MCAT) and contains other elements from RSA, GLUE, multi-objective analysis etc. It can be obtained from the author (thorsten@hwr.arizona.edu) free of charge.

h) Web links or other information

MCAT can also be downloaded from ewre-www.cv.ic.ac.uk/software/toolkit.htm
or from the SAHRA Hydroarchive at www.sahra.arizona.edu/software.

i) Figures

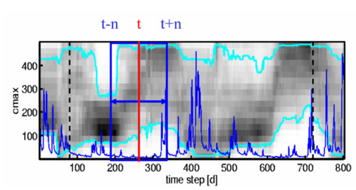

Figure 2: Example of DYNIA output for a parameter of the PDM model. The gray shading in the background indicates the shape of the response surface (i.e. the marginal posterior distribution of parameter cmax) at each time step. Dark areas are peaks in the distribution. The red line shows the a single point in time, while the blue box marks the range over which residuals are calculated. The blue hydrograph is the observed streamflow. The cyan lines indicate the 90% confidence limits. The dashed lines indicate the areas where the full window size cannot be maintained.

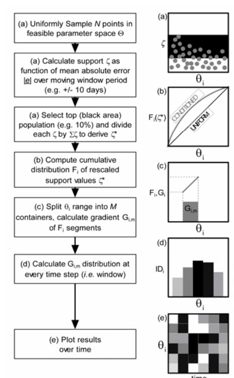

Figure 1: Schematic description of the Dynamic Identifiability Analysis (DYNIA) procedure.

j) Delegates Comments (please add !!)