International Congress on Environmental Modelling and Software

6th International Congress on Environmental Modelling and Software - Leipzig, Germany - July 2012

Jul 1st, 12:00 AM

Global sensitivity analysis for the flow components of a conceptual rainfall-runoff model

Carolina Massmann

Hubert Holzmann

Massmann, Carolina and Holzmann, Hubert, "Global sensitivity analysis for the flow components of a conceptual rainfall-runoff model" (2012). *International Congress on Environmental Modelling and Software*. 218.
https://scholarsarchive.byu.edu/iemssconference/2012/Stream-B/218

International Environmental Modelling and Software Society (iEMSs)
2012 International Congress on Environmental Modelling and Software
Managing Resources of a Limited Planet, Sixth Biennial Meeting, Leipzig, Germany
R. Seppelt, A.A. Voinov, S. Lange, D. Bankamp (Eds.)
http://www.iemss.org/society/index.php/iemss-2012-proceedings

Global sensitivity analysis for the flow components of a conceptual rainfall-runoff model

Carolina Massmann, Hubert Holzmann
University of Natural Resources and Life Sciences, Vienna
(carolina.massmann@boku.ac.at, hubert.holzmann@boku.ac.at)

Abstract: This paper deals with a conceptual rainfall-runoff model in which the total flow is obtained as the sum of the flow of individual flow components, such as surface flow, interflow and groundwater flow. In contrast to classical sensitivity analyses, where the sensitivity of the total flow to each parameter is analysed, we show here the results of variance and sensitivity analyses carried out for each flow component. It was observed that the variance for the total flow spans a range of four orders of magnitude. A comparison with the variances of each flow component allows identifying the process with the highest variance at each time step, which can be regarded as the dominant process. With respect to the first order indices it was seen that high values are common when the related flow component has a high variance, while interactions are predominant when the respective process is not as important. These interactions often involve parameters that were designed for describing other processes, illustrating in this way how parameters can have an indirect effect on many processes. It is concluded that such an analysis motivates thinking in terms of the processes. Specifically, it is possible to structure the period into different segments, depending on the most important process and to analyse these segments as a group, facilitating the identification of patterns.

Keywords*:* conceptual rainfall-runoff model, flow components, Sobol's indices, model variance, dominant processes

1 INTRODUCTION

Hydrological models are commonly used tools in water management related activities. However, despite the ubiquity of models and model applications, there are still many controversies in the hydrological community. One issue relates to the development of more and more complex models, with a higher number of parameters that need to be calibrated. While it is true that such models achieve a better fit to the measured data in the calibration phase, they often fail to provide better results than simpler models during validation. That is why some groups advocate for the identification and use of simple parsimonious models which focus on the dominant hydrological processes [Grayson and Blöschl, 2000]. The processes which dominate at a specific location and at a defined time depend on the hydrological regime (e.g. low or high flow), on the environmental conditions (e.g. rainfall intensity and duration, hydrophobicity of the soil surface) and on the landuse of the catchment (e.g. deciduous or evergreen trees, presence of litter or harvest residues). Finally, it must be kept in mind that the processes a model needs to take into account also depend on the objective of the model and on the questions it is expected to answer.

Sensitivity analyses are useful tools that can increase the understanding of models, aid in model calibration and contribute to the identification of dominant processes. Sensitivity analyses are carried out with respect to the final model result (in our case it would be the total discharge) and they indicate how changes in the inputs are reflected on the model outputs.

In contrast to the above described practice, we present here a sensitivity analysis for the sub-processes of the model. There are only a few examples of sensitivity analyses carried out for the sub-modules of a model. For instance, Judd et al. [1974] show the result of a sensitivity analysis for the four submodules of a model that assesses the probability of contamination. The four models are linked sequentially, so that the output for one module is the input for the next one. Hartebrodt et al. [2010] mention in a forestry related paper that “one can use the sensitivity analysis of sub-models to identify causal chains or to reduce the complexity” of a model, however it was not clear if they carried out such an analysis. Finally, we found in Sumner [2010] some examples of the application of a sensitivity analysis to the sub-compartments of biological models.

To our knowledge there are no previous studies dealing with applications of a sensitivity analysis for the subprocesses of a hydrological model. Therefore, we used a variance decomposition approach for analysing the flow components in a conceptual rainfall runoff model. In the first part the variances of the flow components are calculated and related to the prevalent climatic conditions on the site. In a second step we show two processes in more detail.

2 METHODOLOGY

2.1 Sensitivity analysis methods

Sobol’s Method

This is a method that allows carrying out a decomposition of the variance using Monte Carlo simulations. The variance is a measure of the spread of the model results: periods with a high variance have a high variability in the results depending on the parameter values used, while periods with a low variance show similar values regardless of the values of the considered parameters. This total variance (*Var*) can be allocated to the different parameters on its own and to the different interactions between them:

$$Var = \sum_{i}^{n} Var_i + \sum_{i}^{n} \sum_{j=i+1}^{n} Var_{ij} + \sum_{i}^{n} \sum_{j=i+1}^{n} \sum_{k=j+1}^{n} Var_{ijk} + \ldots + Var_{12\ldots n} \quad (1)$$

where *n* is the number of considered parameters, Var_i describes the variance explained by each parameter individually, and Var_{ij} describes the variance explained by the interactions between two parameters and so on. The first order indices (S_i) describe the proportion of the variance explained by a parameter on its own:

$$S_i = \frac{Var_i}{Var} \quad (2)$$

The total indices (S_{Ti}) describe the proportion of the variance explained by a parameter including all its interactions with other parameters:

$$S_{Ti} = 1 - \frac{Var_{\sim i}}{Var} \quad (3)$$

where $Var_{\sim i}$ stands for the variance not explained by i which is calculated by adding all the terms of Eq. (1) not including the variable i . The difference between the total and the first order indices gives then the proportion of the variance explained by the interactions involving the considered parameter. Sobol's method is a simple Monte Carlo based approach that allows the calculation of these indices. It is based on two equally sized matrices in which each column has values for one parameter (and each row is, therefore, a parameter set with one value for each parameter). The results using the Monte Carlo runs in these matrices are used for calculating the total variance. For estimating the effect of a specific parameter, the same procedure is carried out after the columns for the corresponding parameter are interchanged between the two matrices. More details about the method can be found in Cibin et al. [2010] and Saltelli [2002].

2.2 Modelling

The modelling was done for the Rosalia catchment in Austria, a small 2.35 km^2 basin covered mostly by forests. The model used here consists of a soil and a groundwater storage and is described in more detail in Holzmann and Nachtnebel [2002]. The input to the model is the measured precipitation and the model output is the stream discharge (volume of water passing through the measurement gage per time unit). This discharge is the sum of the discharges generated by the different mechanisms:

$$Dischage = Qsat + Qint + Qgw + \ Qhort + Qsnow \qquad (4)$$

where Qsat, Qint, Qgw, Qhort and Qsnow refer to the discharge produced by saturation flow, interflow, groundwater flow, Hortonian flow and snow melt, respectively. Since this paper deals primarily with the flow components and not the model parameters, we will describe here only the parameters mentioned in the results section:

- *hr1*: describes the size of the soil storage. As can be seen in Fig. 1, saturation flow takes place as long as the water content in the soil storage is above *hr1*.
- *k1*: is the recession coefficient for the saturation flow and in defines the rate with which the water leaves the soil storage through the upper outlet.
- *proz*: is a snow melt related parameter and describes the proportion of melt water routed into the melt water storage, from where it is quickly released into the stream. The remaining part (1-*proz*) infiltrates into the soil and is included into the soil water storage.
- *sk* is also a snow melt related parameter. It represents the recession coefficient for the snow melt storage and describes the rate with which the water is released from here to the stream.

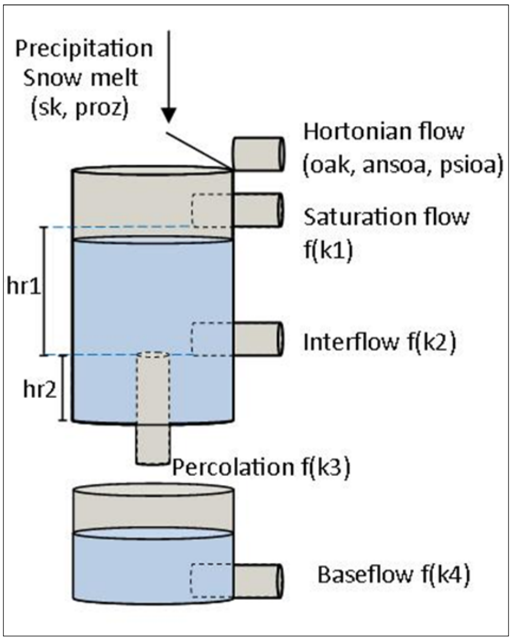

Figure 1: Structure of the rainfall runoff model used.

3 RESULTS AND DISCUSSION

3.1 The variance of each flow component

The variances of the flow components are shown in Fig. 2. It is seen that Hortonian flow, saturation flow, and snow melt related flow show the largest variability, while groundwater flow and interflow have the smallest ones. This can be understood if we consider that snow melt takes place only when there is snow and energy supply is high enough, that Hortonian flow is only observed when the rainfall intensity is larger than the infiltration capacity of the soil and that saturation flow can be only observed when the soil is saturated (e.g. after a rainfall of long duration). Groundwater flow, on the other hand, is observed continuously during the whole period and interflow also results from a mechanism which is observed during longer time periods.

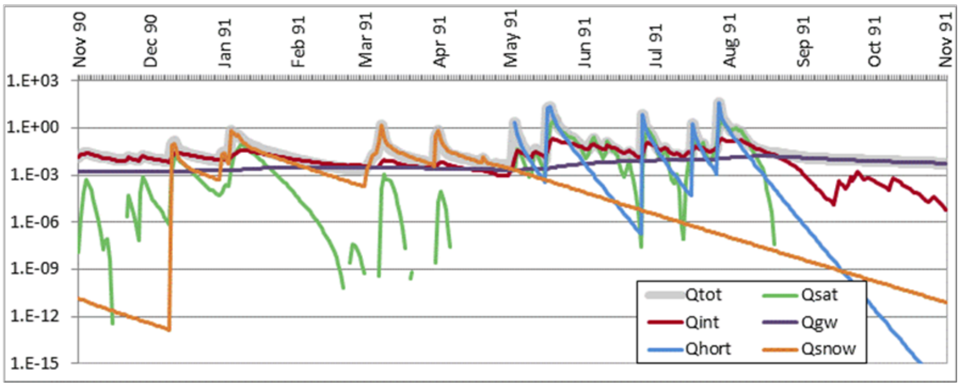

Figure 2: Total variance and variances for each flow component from November 1990 to October 1991.

The rainfall and temperature during the considered period is shown in Fig. 3. A comparison with the variances shows that there is a close relationship between the variances of the flow components and the climatic conditions. Snow melt related flow shows large variances between December and May, specifically when the temperature rises again after a period in which it was below zero and during which precipitation fell. Hortonian flow is observed between May and August, the period concentrating most of the rainfall. It is also seen that all variance peaks coincide with peaks in the measured discharge. Finally, it is observed that the snow melt variance peaks are of similar magnitude than the Hortonian flow variance peaks (compare peaks for both processes in Fig.2), while the measured discharge is much higher when Hortonian flow is important than when snow melt related flow is important. This suggests that the parameter ranges assigned to the snow melt parameters might be too wide.

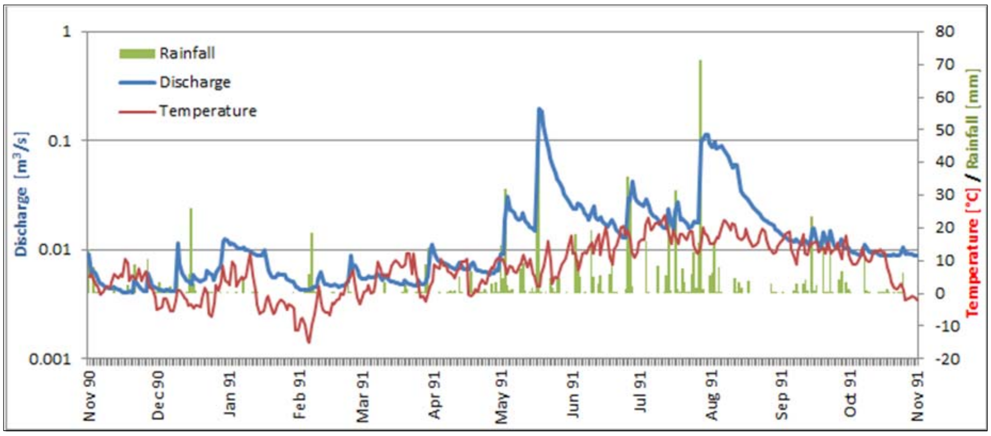

Figure 3: Measured temperature, rainfall and discharge between November 1990 and October 1991.

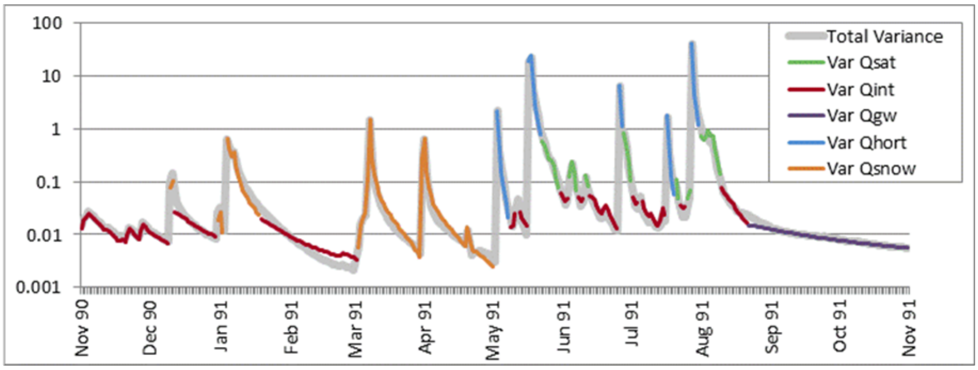

Figure 4: Dominant Processes

As a next step the process with the highest variance at each time step was identified and plotted together with the total variance (Fig. 4). This process with the highest variance can be regarded as the dominant process at each time step, since it is the process that is responsible for the highest variations in discharge, depending on the considered parameter set. It is seen that all processes are important at some time: the snow melt related flow and interflow during the first half of the period; groundwater flow during the last two months and Hortonian flow, saturation flow and interflow between May and September.

Such a plot provides information on the periods the modeller should focus when calibrating specific processes. More specifically, each parameter should be calibrated by comparing the observed and modelled variables during the periods in which the parameter in question is the dominant parameter. However, it also raises the question if there is enough information for calibrating all these processes. Two aspects need to be considered in this respect: the amount of data and its information content. While both aspects are related, since longer time series usually provide more information, there are additional characteristics which define how informative the data is. One of these characteristics is the representativeness of the data series, which should present samples of all typical situations that occur. Berthet et al. [2010] found, for example, that for obtaining representative values for the root mean squared error, the data series need to be much longer than 10 years, sometimes even several decades long. The other important feature is that the data should not have disinformative periods, which is, however, very difficult to assess [Beven and Westerberg, 2011].

With regard to the amount of data there are large differences between the processes. For example, interflow dominates during 44 % of the days, while Hortonian flow is dominant only on 6 % of the days (Table 1). When looking at the representativeness of the observations, it is seen that interflow shows dominant periods during the whole year: after snowmelt events in winter and intensive rainfalls in summer. Hortonian flow, on the other hand, is only observed in summer, while it might be possible that some years also have intensive rainfall in spring. Therefore, from the point of view of the amount and representativeness of the observations, it is suggested that interflow might be calibrated more easily than Hortonian flow.

Table 1: Frequencies and proportions of the dominance of each process.

Process	N° days it dominates	%
Saturation flow	33	9
Interflow	159	44
Groundwater flow	71	19
Hortonian flow	22	6
Snow melt related flow	80	22

3.2 Sobol´s indices for snow melt and saturation flow

This section presents the sensitivity indices for two processes: the snow melt related flow and the saturation flow.

Snow melt related flow

Most of the variance of this flow component is explained by the parameters *proz* and *sk* individually, since the sum of the first order indices reaches almost 1 (Fig. 5). The interactions of the parameters describing this process are relatively small, around 14 %, and it is seen that there are no interactions with other parameters, since both curves overlap during the whole period. Another interesting observation is that the parameter *sk* loses influence at the expense of *proz* when snowmelt takes place. The situation is the gradually reverted until the next snow melt event. Similarly, the interactions between the parameters decrease sharply when snow melt is observed, and then start to increase again after that.

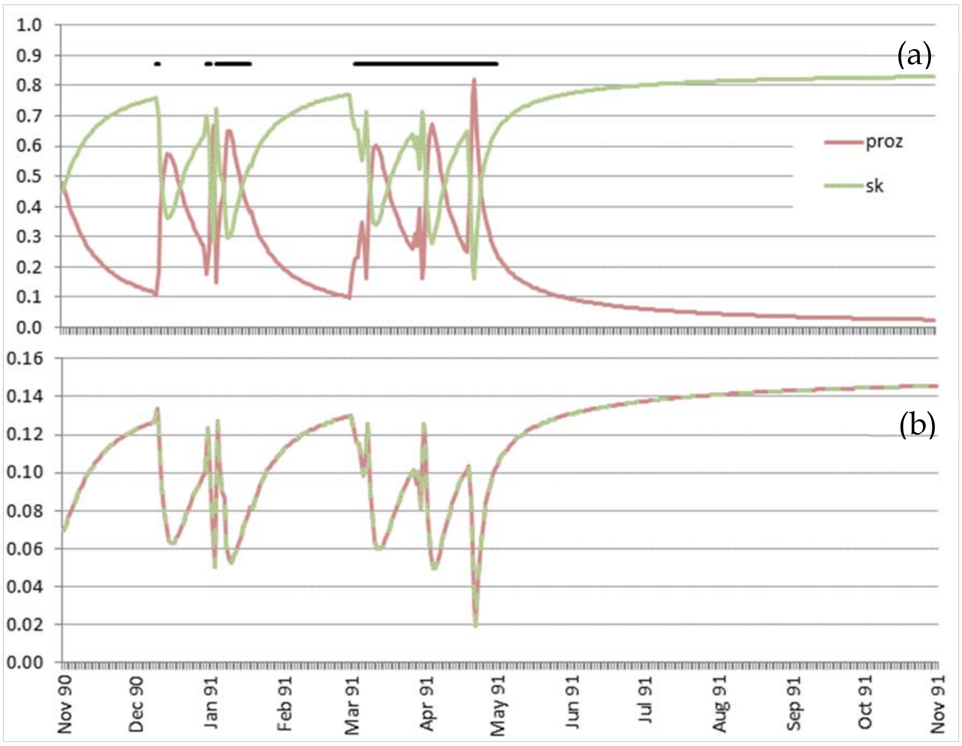

Figure 5: First order indices (a) and interactions (b) for the process ‘snow melt related flow’, in the first and second panel respectively. The black line on top shows the period during which snow melt is dominant.

Saturation flow

The other process shown here is saturation flow. A glance at the first order indices plot (Fig. 6) reveals that they are much smaller than for snow melt related flow. The most important parameter is *hr1*, which describes the size of the soil storage. With respect to the interactions it is seen that they are much higher than for snow melt related flow, and that for some periods they even surpass the impact of the first order indices. Another characteristic is that the interactions involve much more parameters, for instance the parameter *proz* has a high impact through interactions in May.

Similarly to the snow melt flow, it is seen that when the first order indices are important, the interactions decrease. As seen in Fig 2, these periods with higher first order indices coincide with the periods in which the variance for this flow component is larger.

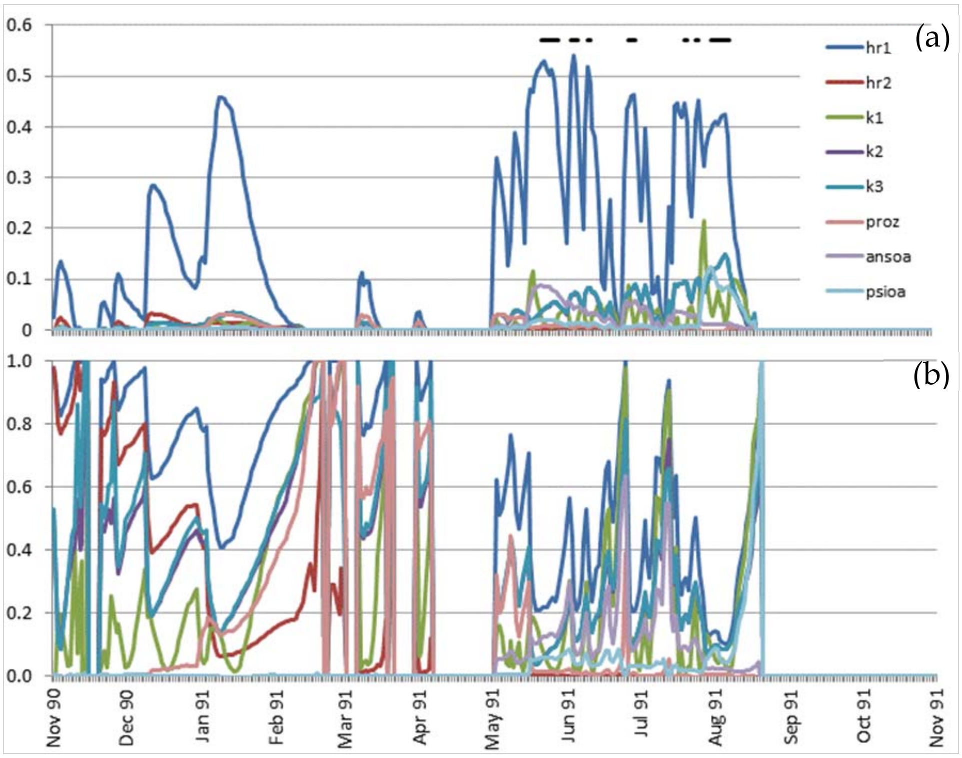

Figure 6: First order indices (a) and interactions (b) for the process ‘saturation flow’, in the first and second panel respectively. The black line on top shows the period during which saturation flow is dominant.

4 SUMMARY AND CONCLUSIONS

A variance and sensitivity analysis for the flow components of a rainfall runoff model was carried out.

With respect to the variances we found that:

- The variance for the total flow spans a range of four orders of magnitude. The processes responsible for the variance peaks are snow melt and Hortonian flow. When the total variance is lowest, it coincides with the variance of groundwater flow.
- The variances for saturation flow, Hortonian flow and snow melt show large variations, while the variances for interflow and especially for groundwater flow show much less variability when considering the whole period.
- The variances of the more dynamic processes show a clear relationship with the climatic conditions.
- The process with the largest variance in each time step can be regarded as the dominant process, since it is the process explaining most of the variation in the results.
- A quantification of the time steps in which each process is dominant can help in deciding if there is enough information for calibrating the process.

With regards to the sensitivity analysis carried out for individual flow components it can be concluded that:

- High first order indices coincide with periods in which the process has a higher variance, while high interactions are characteristic of periods during which the considered process is not so important.
- The interactions often involve parameters that are used for describing other processes, therefore the effect these parameters have on the discharge is an indirect effect. One example is the snow melt related parameter *proz* which has an important effect on the saturation flow component through interactions.

Complementing a sensitivity analysis for the total discharge with an analysis for each flow component requires additional work and results in more plots and information which needs to be analysed. The advantages of such an approach are:

- The plots are simpler, since only the parameters relevant for the process are depicted.
- It allows thinking in terms of the process. Especially when analysing the differences between modelled and measured outputs, it might be possible to see what part of the process description could be improved.
- It structures the total period into different segments depending on the most important process. These segments can then be analysed as a group, which facilitates the identification of patterns.

ACKNOWLEDGMENTS

This work was funded by the Austrian Academy of Science through the project Domina_HyPro. The data for the Rosalia catchment was provided by Josef Gasch from the Forest Demonstration Centre, University of Natural Resources and Life Sciences, Vienna.

REFERENCES

Berthet L., V. Adreasssian, C. Perrin and C. Loumagne, How significant are quadratic criteria ? Part 1. How many years are necessary to ensure the data-independence of a quadratic criterion value ?, Hydrological Sciences Journal, 55(6), doi:10.1080/02626667.2010.505890, 2010.

Beven K. and I. Westerberg, On red herrings and real herrings: disinformation and information in hydrological inference, Hydrological Processes, doi:10.1002/hyp.7963, 2011.

Cibin, R., K.P. Sudheer and I. Chaubey, Sensitivity and identifiability of stream flow generation parameters of the SWAT model, Hydrological Processes, 24, 1133-1148, doi: 10.1002/hyp.7568, 2010.

Grayson R. and G. Blöschl, Summary of pattern comparison and concluding remarks, IN: Grayson R. and G. Blöschl (Eds.), Spatial Patterns in Catchment Hydrology: Observations and Modelling, Cambridge University Press, Cambridge, 355-367 p, 2000.

Hartebrodt C.H.; Aichholz R. and Braasch M., Analyzing and predicting forestry accountancy network variables with Bayesian Belief Networks as compared to traditional analyzing methods, Small-scale Forestry 10:163-183, doi: 10.1007/s11842-010-9124-0, 2010.

Holzmann, H. and H.P. Nachtnebel, Abflussprognose für mittelgroße Einzugsgebiete - Methodik und Anwendungen [Runoff forecast for medium sized basins – methods and applications], Österreichische Wasser- und Abfallwirtschaft, 54(9-10), 142-153, 2002.

Judd B.R.; North D.W and Pezier J.P., Assessment of the probability of contaminating Mars, Final Report of the Stanford Research Institute for the NASA Planetary Programs Division, Grant NASW-2535, 161 p., 1974.

Saltelli, A., Making best use of model evaluations to compute sensitivity indices, Computer Physics Communications, 145(2), 280-297, doi:10.1016/S0010-4655(02)00280-1, 2002.

Sumner T., Sensitivity analysis in systems biology modelling and its application to a multi-scale model of blood glucose homeostasis, PhD thesis, Centre for Mathematics and Physics in the Life Sciences and Experimental Biology, University College London, 162p., 2010.