Ecological Modelling 332 (2016) 83–93

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

# Implementation of robust statistics in the calibration, verification and validation step of model evaluation to better reflect processes concerning total phosphorus load occurring in the catchment

Mieczysław S. Ostojski, Joanna Gębala*, Paulina Orlińska-Woźniak, Paweł Wilk

*Institute of Meteorology and Water Management, National Research Institute, Podlesna St. 61, 01-673 Warsaw, Poland*

ARTICLE INFO

*Article history:*
Received 8 October 2015
Received in revised form 29 February 2016
Accepted 5 April 2016
Available online 22 April 2016

*Keywords:*
Mathematical modeling
Outliers data
Robust statistic
Calibration
Verification
Validation

ABSTRACT

Obtaining a high quality of environmental modeling in the calibration, verification and validation processes causes problems for modelers all over the world. Uncontrolled variability of environmental conditions (diversity of natural phenomena, meteorological and climatic conditions occurring within the drainage basin), inaccuracy of measurement equipment, averaging the data obtained and the errors of human work vitiates data population to a greater or lesser extent. These factors impede the evaluation of modeling, described as the "goodness of fit" of modeling results to that of observational data.

Therefore, from the point of view of the correct interpretation of the results of calibration processes, and verification and validation of the model, it is essential to find a statistical tool which would allow limiting the influence of strongly outlying data on the final result of the correspondence of the model to the actual conditions. Such a tool may be constituted by robust statistics as presented in the article, or the L-estimators to be more precise, which appeared to be effective in the case of an environmental data set, which are characterized by certain measurement unreliability, and in addition, a limited number of result populations. In the research Marcomodel DNS was used, with a SWAT module, for two catchments located in Northern and Central Poland. The evaluation of modeling quality was conducted by the use of classical statistics and robust statistics. The results of the classical statistics were based on three statistical measurements: coefficient of determination ($R^2$), were compared with percent bias (PBIAS) and Nash Sutcliffe efficiency (NSE), and with the results of these measurements after the process of winsorization. In the robust estimates calculation the strength of outliers was decreased. It may be a beneficial phenomenon worth considering in the evaluation of quality modeling of surface waters.



## 1. Introduction

Models allow us to describe processes occurring in the catchment area, not only as a simple regression dependence, but taking into account real relations between the individual processes. Such tools have made it possible to optimize the system of interaction in order to achieve the intended purpose. The quality of the results obtained by the models depends on the proper carrying out of the process of checking the modeling results and determining if they fit the monitoring data from direct observation. For this purpose, a number of statistical indicators for measuring the degree of reality by mapping the representation of the model calibration, a verification and validation process of the model, were used (Maréchal and Holman, 2004; Moriasi et al., 2006; Sarma et al., 1973; Alansi et al., 2009; Chiang et al., 2012). However, even monitoring data may be burdened with errors caused by averaging obtained data, rounding data, inaccuracies of measuring equipment as well as the uncontrolled variability of environmental conditions (Huber, 1981).

No model is entirely accurate despite the fact that it may be a very precise reflection of reality. In the model calibration process, the iterative improvement method is used. In the gathering of modeling results for a population it should be more probable for minor errors and less probable for major errors to appear. To avoid major error accumulation, a stochastic model of the investigated reality should be diagnosed and identified. In order for that to be accomplished, procedures which are sensitive yet resistant to disruptions should be used. If the investigated reality is assigned with an inappropriate model, it may happen that the assigned modeling method, insensitive to outliers, will conceal the incorrectness of the model.

* Corresponding author. Tel.: +48 225694264.
*E-mail address:* joanna.gebala@imgw.pl (J. Gębala).

http://dx.doi.org/10.1016/j.ecolmodel.2016.04.004
0304-3800/

The paper is to state the hypothesis in the discussion if the robust environmental model is fitted with the same accuracy to the observation of non-extreme and extreme periods of natural phenomenon. The collection of environmental data is a specific population assumedly burdened with certain errors. It should be considered if the model which reacts correctly to input data, such as meteorological data, in non-extreme situations, is incorrect in case of outlying observations. Does the model calibration aimed at fitting extreme data cause a general decrease in modeling quality as a general fitting to the probability distribution or family of distributions, by use of which we are trying to describe the natural phenomenon? Perhaps, when constructing the optimal method of estimation, the observed data should be treated as values assumed by certain random variables. The approach presented in the paper, to some extent, will treat the theory of estimation in a purely mathematical way, in order to present a formal statistical study.

Outlying observations, on the other hand, will be considered from the statistical point of view. The assumption was presented that the approach is not incorrect in the case of environmental quality phenomena. The data concerning that phenomena suggest that they reflect distributions with so called fat tails (5–10%) better than they do with normal distributions. The use of robust estimation enables the protection of parameter estimations against the emergence of unexpectedly large errors. The issue of outlying observations of environmental data has often been described in literature (Lange et al., 1989; Rossi et al., 1992; Chris, 2007; Yi and Myong, 2008; Marazzi and Randriamiharisoa, 1993; Bosch et al., 2011). In scientific research using mathematical models as a source or complement of existing data, the elimination of the influence of outlying data is essential. The development of measurement techniques and possibilities contributes to providing longer sequences and a larger quantity of data. It is connected with a higher frequency of anomaly in populations of data in most scientific disciplines. It is widely known that all large collections of information contain outlying data. To decrease their influence on model calibration results, at least a few methods are used. In literature the term appears Robust Probabilistic Multivariate Calibration (RPMC) in the framework of latent variable models. RPMC models deal with outliers by probabilistic description instead of by simply removing them (Lange et al., 1989; Verboven et al., 2012; Yi and Myong, 2008; Krause et al., 2005). Most of these methods are attempts to improve the robustness of common multivariate regression techniques, such as principal component regression (PCR) (Chatterjee and Price, 1977) and partial least squares (PLS). Considerable attention has been given to replacing the no robust least squares estimate with a robust estimate. Notable examples of these robust estimates include M-estimates (Huber, 1981), the Stahel–Donoho estimate, least median of squares (Rousseeuw, 1984), and S-estimators (Davies, 1987; Lopuhaa, 1989). A recent overview of these robust estimates has been given by Møller et al. (2005).

In the paper, the evaluation of modeling with the use of robust statistics and L-estimators was carried out. Modeling was done in Macromodel DNS/SWAT for two Polish catchments. Poland is a country characterized by variability of weather phenomena and large hydrological diversity. With these natural causes of outliers in the observational data collection, there is the uncertainty of individual measurements resulting from the methods of measurement, changes in measurements in the case of very long measurement sequences, or other reasons. As a result, for the majority of the catchments in Poland we are dealing with data burdened with outliers.

Therefore, to properly evaluate the fitting of the modeling results and observation results, statistical tools such as robust estimation become useful. They enable assigning appropriate strength to outlying observations and thus decrease their negative impact on the assessment of results of the model fitting assessment to the observations using statistical measures (Jureckova and Kumar Sen, 1996; Marazzi and Randriamiharisoa, 1993; Mokhtar bin, 1990; Staudte and Sheather, 1990; Wilcox, 2012; Zaleski, 2004; Stehr et al., 2008).

## 2. Aim of the work

In environmental modeling the presence of outliers hindering the overall assessment of the quality of the modeling is observed. During calibration, verification and validation, the modeling is evaluated with the use of statistical measures of fitting the simulation with observation. The aim of the work is to use a statistical tool, which is a robust estimation, and more specifically $\alpha$-winsorized estimators belonging to the group of L-estimators of the location parameter. This paper attempts to answer the question of the use of robust statistics to describe the results of the environmental calibration model as more reliable than the methods of classical statistics.

For this purpose, the results of modeling in a Macromodel DNS (Nutrient Discharge Sea) with SWAT (Soil and Water Assessment Tool) module – Macromodel DNS/SWAT (Ostojski, 2012) for the flow and general phosphorus was used.

In the paper two pilot catchments, i.e. the Middle Warta and the Rega were analyzed. For flow and total phosphorus loads, the processes of calibration, verification and validation using robust statistics were carried out. This allows for the best reflection of real phenomena and processes occurring in the catchment by the model.

## 3. Materials and methods

### 3.1. Outliers in calibration, verification and validation of environmental modeling

Currently used mathematical models (Ambrose et al., 1998; Dąbrowski et al., 2011; Lechowicz and Wrzosek, 1995; Ostojski et al., 2014; Sokolov, 2002; Vollenweider, 1998) are useful tools for the analysis of the phenomena occurring in the catchment, including analysis of the hydrological regime of the river, the impact of climate change on water resources of the catchment area, the impact of hydraulic structures on the size of nutrient loads, the volume of surface runoff, and the impact of anthropological pressure on water quality or value of the river absorption.

The correct analysis of the modeling results is necessary in order to obtain reliable simulations. For this purpose, calibration, verification and validation processes are carried out, and they are defined respectively as:

- adjustment of model parameters in order to obtain the greatest convergence of modeling results and observations,
- checking at measuring points whether the model is a good representation of the conceptual model, performed on independent data during the process of calibration,
- final checking, at different measuring points than calibration and verification points, whether the model is a good representation of reality by comparing the modeling results with observations.

Utilization of statistical measures for the above processes allows us to evaluate the matching of modeling results to observation and the occurrence of overestimation or underestimation of the data. It is recommended to use two or more statistical measures. The most frequently mentioned are: determination coefficient $R^2$, Nash-Sutcliffe model efficiency coefficient NSE and prediction coefficient BIAS (Maréchal and Holman, 2004; Moriasi and Arnold,