Use of Performance Criteria in Calibrating Methods for Modeling and Simulating the Pollution Phenomena of Surface Waters

Lorand Toth1*, George Artur Găman1, Daniel Pupăzan1, Cristian Nicolescu1, and Marius Kovacs1

1National Institute for Research and Development in Mine Safety and Protection to Explosion – INSEMEX Petroșani, 32-34 G-ral Vasile Milea Street, Petroșani 332047, Romania

Abstract. The release of certain substances into surface waters (lakes, rivers, estuaries and oceans) to the point where they interfere with beneficial use of water or with natural functioning of ecosystems defines the phenomenon of pollution. When stating aspects of pollution modelling, we refer to constitutive equations of the model, which may involve different values, so that the shape of equation is flexible while maintaining its structure. Quantifying the phenomenon of water pollution through simulation and spatial-temporal modelling requires the use of hydrological models that use characteristic parameters such as: bathymetry, hydrodynamic roughness, speed, Model Boundary Conditions, etc. The current paper is driven by lack of clear performance guidelines for pollution models for software users trying to demonstrate to customers and/or end users that a model is fit for purpose. Thus, common problems associated with data availability, errors and uncertainty as well as model examination will be addressed, including its calibration and validation on a case study materialized on a watercourse located in the Jiu Valley, Romania. The scientific article is intended to be a point of reference both for software users (numerical modelers) and for specialists in charge of interpreting the accuracy and validity of results from hydrodynamic models.

1 Introduction

The important issues of river pollution and spread of pollutants must be addressed with the help of prediction tools (models for pollutants transport) to develop systems for assessing pollution, counteracting it and making the right management decisions. This field has seen tremendous progress, especially in the last 15 years, materialized through monitoring campaigns that benefit from high-performance technological support, data processing tools that are able to process impressive amounts of data, new methods for calculating parameters with a better accuracy and, last but not least, modelling methods that have been improved. However, there are still important issues that require special attention from researchers: for example, parameter estimation techniques; wider applicability as well as transferability of mathematical models between case studies.

Consultants and teachers use a wide variety of different modelling practices and frequently pay insufficient attention to potential errors associated with the measured (and modelled) data used to calibrate and validate the model. This can lead to poor model performance and uncertain model predictions. Without an agreed methodology and a performance standard for calibration and validation of models, there is a risk of variation amid approaches, efforts being lost because of inefficient or inadequate calibration methods. Inconsistencies amid methodologies will make model intercomparing problematic.

Because the accuracy of model calibration critically depends on the calibration data used, high attention is paid to some of the most common issues associated with quality of data.

From the outset, it is important to define terms commonly used by numerical modelers: (a) calibration - a process that requires the adjustment of certain model parameters to achieve the best model performance for specific locations and applications; (b) verification - correct implementation of the model and assumptions used; and (c) validation - establishing agreement between forecasts and observations [1]. Validation is performed by running the model using data covering an alternate period and/or another location, without making any additional adjustments to the model parameters [2].

If a comparison of measurements and results of the model suggests that model predictions are close to measurements, then the implemented model is assumed to be both a verified implementation of the assumptions and a valid representation of the system being modelled. The paper builds on practical modelling experience and extends to previous and limited guidance on model calibration and validation, which focuses on criteria based on Eulerian points that define model performance [3, 4, 5].

2 (Theoretical) Aspects regarding data, calibration and validation

As an illustration of typical calibration and validation processes applied in most pollutant dispersion models in water bodies, a schematic diagram of steps to be followed for a hydrodynamic model is shown in Figure 1. In the case of the first model simulation, model parameters are
set to values recommended by the modelling software guide (i.e. "initial settings"). Critical parameters affecting performance of the model (for example, roughness of the course’s bed) are then adjusted to obtain the optimum between model predictions and the set of measurements. The fact that values set for the parameters involved in simulation are physically adequate must be considered [6,7]. Performing a good calibration of the model and using it in the wrong conditions is as bad as using a poorly calibrated model [8].

The second step in the calibration process is to reduce uncertainty in model predictions. Normally, it uses carefully selected values for the model input parameters and compares model predictions with observed data, for the same conditions. This is often done iteratively, without any fixed rules and is guided by user’s experience and knowledge of the processes to be modelled. The third stage in the calibration process involves validating production of the model of interest (water level, water speed, flow rate, direction, etc.). Validation involves running a model using parameters that were determined during the calibration process and comparing predictions obtained with observed data that were used in the calibration. Use of automated techniques for model calibration is now widespread. Typically, self-calibration procedures are based on the Monte Carlo method or other sampling schemes to estimate the best choice of values for several input parameters. While self-calibration can provide a powerful, labour-saving tool that can be used to substantially reduce subjectivity that frequently characterizes manual calibrations, more attention must be paid when using these approaches to ensure theoretical limits for each specific input parameter. Frequently, post-calibration evaluation of numerical model’s results is subjective and is based only on specialized interpretation of graphical results. However, increasing complexity of model functionality and their use by end-users, requiring information on model accuracy to reduce risks, has led to an increasing need for guidelines regarding quantification and evaluation of model performance.

2.1. Performance guidelines

Bathymetry - One of the most common problems associated with calibrating a hydrodynamic model is underlying errors in bathymetric data. As standard practice, analysis of bathymetric data should ensure (through a review of data from the study area) the use of the latest bathymetry survey data. Features and contours must be checked on previous maps and diagrams. Appropriate mesh dimensions reflecting the spatial distribution of bathymetric data must be determined and, if data are already in grid form, interpolations and / or reductions on the model networks must be performed by reference to the original data.

The summary of bathymetric and topographic data requirements for hydrodynamic models is presented in Table 1. Here, a distinction is made between types of applications, the most accurate being associated with shore regularization projects.

These distinctions are used in other tables and are useful because they define accuracy of key data needed to build a model for other applications. The correct use of the most appropriate data for a given application saves time and effort. While Table 1 reflects bathymetric and topographic data of requirements for modelling a
watercourse segment, including specifications for average distances between survey positions, minimum acceptable distance of sections between banks, survey age, they provide equally useful guidelines, also for other types of water bodies.

**Table 1.** Bathymetric and topographic requirements for different models (A - design; B - evaluation; C - strategy)

| Model type | Average distance between researched depths (m) | Minimum distance of shore section (m) | Elapsed time between probes (years) |
|------------|-----------------------------------------------|--------------------------------------|------------------------------------|
| Design (A) | ≤ 2                                           | 5                                    | ≤ 1                                |
| Evaluation (B) | 2-10                                      | 10                                   | 1-5                                |
| Strategy (C) | > 10                                          | > 10                                  | > 5                                |

Thorough checks of horizontal and vertical probe’s database should always be performed before any running of the model and models should always aim to use a common reference database. Usually for vertical positions, national or mean sea level (MSL) benchmarks are widely used. However, while national benchmarks are useful in local-scale models, MSL is more useful in larger regional models in all geographical locations.

Riverbed roughness - hereinafter referred to as "bed roughness" is a primary calibration variable for all water body models. It is also essential for accurate modelling of other processes, such as pollutant dispersion and sediment transport. Regardless of the method chosen to define bed roughness, values are manipulated iteratively by the user within limits reported by literature. In the absence of data to accurately define bed thickness, these "typical" values are often used in model applications. However, in many cases, this is an oversimplification and as much information as possible about riverbed characteristics should be obtained, so that appropriate roughness values can be allocated.

Water level - A model calibration for water level should include examination of amplitude, phases and asymmetry. Specifically, the test should examine: rising differences in maximum and minimum area, root mean square error (RMSE), bias and scattering index (SI). It is recommended that performance of the minimum required model to be ± 0,10 m.

Current Speed - In 2D depth average hydrodynamic models, current speed predictions must be examined for amplitude, phase, direction and asymmetry. Specifically, the test shall examine differences in maximum flow rate, average flow direction, root mean square error RMSE, bias and SI scattering index. Corresponding values of velocities at average depth must normally be obtained from measurements of points at a certain reference water column height or from currently measured profile data. Data resulted from statistical analyses of the model's performance must be carefully interpreted. Value of the root mean square error (RMSE) indicator provides a quantitative measure of how well the model fits into the data average. It is recommended that the bias be less than 0.2, SI < 0.5 and RMSE < 0.2 in order to indicate accurate statistics.

Mean pollutant concentration - Successful calibration of the pollutant dispersion model can be assessed visually by comparing the average concentrations measured and predicted, over a given period. For most applications, the aim should be to perform a model calibration of ± 20% of the measured mean concentrations. In areas where a number of pollutant concentration measurements are available from several sites, a calibration level of ± 30% of the average concentration in most sites would be considered a good calibration level. If there are only discrete values of pollutants in water samples, experience shows that calibration of only ± 40% is feasible, as discrete measurements are subject to higher levels of uncertainty.

2.2 Error, accuracy, and uncertainty of model calibration data

In engineering and environmental modelling studies, use of quantitative model evaluation methods is perceived as providing a more objective, consistent and reproducible validation and evaluation of the model. However, it goes without saying that identification of systematic or random errors in the model results can be quickly detected by the human eye. In practice, model evaluation is most effective when both qualitative and quantitative approaches are used. These statistics can provide additional useful information on spatial coherence, correlations and coherence, and will often indicate explanations and the origin of possible differences between model results and measurements [9].

Because it defines the metric against which the model's performance will be evaluated, evaluating errors, accuracy, and uncertainty in data used for calibrating the model is an important step in the modelling process. Therefore, it is essential to quantify the error, accuracy and uncertainty by understanding the instrumentation, the method of implementing the instrument and its location, as well as any post-processing issues.

It is necessary to distinguish between systematic and random measurement errors. All measurements are prone to systematic errors resulting from, for example, imperfect instrument calibration (zero error) and changes in environmental conditions. Similarly, random errors are usually present in a measurement or other observations and result from inherently unpredictable fluctuations in readings of a measuring device or in their interpretation. The error can be quantified by comparing several measurements and reduced by mediating several measurements. Systematic errors cannot be detected in this manner, because they always "push" the results in the same direction. However, when identified, they are easier to remove from a data set using trend removal techniques (e.g., regression analysis). Taking as an example the calibration of a 2D depth-mediated hydrodynamic model using Acoustic Doppler current profiler (ADCP) data, the first necessary thing is to obtain the average depth current from the measurement. These data processing steps bring errors that are difficult to quantify. Moreover, if a given
measurement footprint is within a highly turbulent field, then the accuracy of the average flow measurement will be influenced by the sampling time and may lead to significant errors if the flow is not correctly sampled at that location. Given this, increased attention should be paid to spatial and temporal inconsistencies that could lead to calibration and/or bias errors.

2.3 Sensitivity analyses

Sensitivity analyses are used for studying how the uncertainty in the output of a model can be distributed in different sources of uncertainty in its input. Sensitivity analyses are performed by varying the input parameters (in a certain range), and examining the model’s response. Sensitivity analyses can be useful for a number of purposes, including (a) testing the robustness of the model leading to uncertainty, (b) increasing the understanding of the relationships between input and output variables of a model, (c) identifying model errors by finding unexpected relationships between inputs and outputs; and (d) simplifying models by identifying model inputs that have no effect on output or identifying and removing redundant parts of the model structure. Sensitivity analyses can also help reduce uncertainty by identifying model inputs that cause the greatest uncertainty in the output, thus allowing adjustments to increase robustness of the model. Sensitivity analyses are therefore a vital part of assessing whether a model is appropriate for the purpose and time must be allotted to achieve a credible model of sensitivity study.

One area of sensitivity analysis that requires special attention refers to sensitivity of a given model to errors in the input data (e.g. bathymetry, water level and average speed). This is especially important when there are errors and/or uncertainties in multiple input data sets, which can lead to complicated errors in the model's output data.

2.4 Time series and statistical output

In many cases, showcase of data in time series format helps to reveal the level of confidence between the model and observation data, with gaps between observed and predicted data indicating visual discrepancies between model predictions and calibration data [10]. Calibration should aim to minimize these discrepancies, and statistical analysis should be used to quantify the level of confidence [11]. In addition, it is also informative to compare similar values using a scatter plot that shows both observed and modelled values.

To further quantify the temporal aspect of model calibration, statistical approaches are used for demonstrating that confidence can be placed in model’s performance, at time intervals in a clear and intelligible way [12].

Simple statistics demonstrating the level of confidence between measured/observed data and model’s prediction at a chosen location in the model range, include the mean and maximum differences (often expressed as a percentage) and the standard deviation [13]. There are several quality indices that can be used to demonstrate the statistical congruence between model predictions and observations (Table 2). Within the table, Oi and Si are the measured and predicted values of a given parameter at time ti, respectively, and Ni is the total number of points (data).

Table 2. Statistical criteria used for input data and model prediction

| Quality index | Example | Formulae |
|---------------|---------|----------|
| Accuracy      | Root mean square error | RMSE = \( \sqrt{\frac{1}{N} \sum_{i=1}^{N} (S_i - O_i)^2} \) |
| Bias          | Average bias | Bias = \( \frac{1}{N} \sum_{i=1}^{N} (S_i - O_i)^2 \) |
| Correlation   | Pearson product-moment coefficient | R = \( \frac{\sum_{i=1}^{N}(S_i - \bar{S})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{N}(S_i - \bar{S})^2 \sum_{i=1}^{N}(O_i - \bar{O})^2}} \) |
| Skill         | Brier skill score (BSS): \( X_p \) is the postevent condition predicted by the model, \( X_m \) is the measured postevent condition, and \( X_s \) is the preevent condition. Skill: index of agreement, where \( X \) and \( \bar{X} \) are time series and time average of model and observed values | BSS = \( \frac{\left(1 - \frac{\sum|X_{mod} - X_{obs}|}{\sum(X_{mod} - X_{obs})^2 + \sum(X_{obs} - X_{obs})^2}\right)}{\left(1 - \frac{\sum|S_i - O_i|}{\sum_{i=1}^{N} O_i}\right) \times 100} \) |
| Agreement     | Scatter index | SI = \( \frac{\sum_{i=1}^{N} (S_i - O_i)^2}{\sum_{i=1}^{N} O_i} \times 100 \) |
Accuracy expresses the difference between measured and modelled data defined as:

\[ \text{dif}_i = S_i - O_i \]  

The aim should be to reduce the dif value to lowest possible. Ideally, a minimum difference should not exceed 10%, although this will be highly variable depending on parameters considered and accuracy of calibration data used in the model. Accuracy of modelled data can also be quantified using the mean square error (RMSE). The RMSE value is often expressed as a percentage, where lower values indicate less residual variation and thus better model performance.

The bias expresses the difference between estimator's expectation and actual value of estimated parameter and can be defined as being equal to the average data error. Systematic bias reflects external influences that may affect the accuracy of statistical measurements [14].

The concordance or mismatch between measured / observed data and the model’s prediction time series is frequently quantified using the Pearson product-moment coefficient, R. It is essential to test the statistical significance of the correlation coefficient. In most cases, the Pearson method is appropriate.

3 Results and discussion

In order to establish the dispersion of pollutants in the aquatic environment, in the first stage we delimited the watercourse segment on which pollutant dispersion analysis was performed (Figure 2).

The obtained meshes (Figure 3, 4) did not require triangulation of values [15].

![Figure 2](image-url) Satellite image of the studied stream segment (Maleia)

**a) 2D Mesh Generation**

To obtain the 2D Mesh, the Cartesian coordinates of selected areas, gathered by satellite images were inputed.

![Figure 3](image-url) Obtaining the 2D mesh for the studied stream sector

**b) Triangulating data**

By triangulating the input data (Cartesian coordinates) we obtain the quadratic triangles and linear triangles (Figure 4) necessary for the numerical model used by the RMA2 and ADCIRC subroutines, used by the AQUAVEO SMS software.

Following the triangulation process, triangular elements were obtained outside the limit of the working space.

![Figure 4](image-url) Obtaining the 2D mesh by triangulating the input data

After obtaining the 2D mesh with the triangulated elements, the obtained model was cleaned so that all mathematical equations applied refer only to the water sector considered (Figure 5).

![Figure 5](image-url) The 2D mesh required to apply mathematical models

**c) Configuration of quadrilateral elements**

In order to increase the processing speed of subprograms used, triangles obtained in the previous mesh were further merged, for configuration of the quadrilateral elements (Figure 6).

![Figure 6](image-url) 2D mesh configurated with the help of quadrilateral elements
d) Input data used for the RMA2 subprogram

For running, the RMA2 sub-routine needed the input data measured in situ for the watercourse under study (Table 3):

Table 3. Input data measured in situ for the watercourse under study

| Input data – RMA2                      | Maleia brook |
|---------------------------------------|--------------|
| Widths of the studied river sector:   | 3.58 m       |
|                                       | 3.62 m       |
|                                       | 3.81 m       |
| Water depth (distance between riverbed and water mirror): | 0.37 m       |
|                                       | 0.43 m       |
|                                       | 0.35 m       |
| Length of the river sector:           | 6 m          |
| River flow in the section concerned:  | 1.025 m/s    |
| Average flow rate:                    | 0.73 m/s     |
| Time required to estimate the degree of dispersion: | 60 seconds   |
| Slope of the river on the studied sector: | 5.41 promilles |
| Isotropic value:                      | 25           |
| Roughness:                            | 0.03         |
| Turbulence:                           | 0.045        |

Values for roughness, turbulence and isotropism were selected from literature (Figure 7)

Fig. 7 Entering input data for the RMA2 subroutine

In the RMA2 subroutine, a model check was performed (Figure 9) to detect if all elements necessary for the simulation were entered correctly in order to eliminate certain inconsistencies between the input data [17]. Using the Froude number, the numerical stability of the RMA2 program was checked.

f) Running the RMA4 subroutine

The RMA4 subroutine uses the flow solutions calculated by the RMA2 subroutine to calculate the degree of dispersion for the 2D mesh flow. For the Maleia watercourse, a dispersion simulation was performed for the nitrate indicator because this indicator showed exceedances of pollution limits. Subroutine input data were the minimum concentration of 3.45 mg/l and the maximum determined concentration of 75.2 mg/l. Figure 10 shows the simulation of dispersion degree for nitrate on the Maleia watercourse.

Based on type CFD- computational fluid dynamics numerical simulations (SMS) of water flow in the studied sector, the value of pollutant concentrates (nitrate) was obtained in any point of the dispersion map, thus finding a linear dilution phenomenon of pollutant concentration.
Table 4 shows the statistical guidelines for setting the calibration standards for a minimum level of performance for the studied hydrodynamic model.

**Table 4** - Calibration standards for a minimum level of performance

| Model predictions | RMSE (root mean square error) | Bias | R (Pearson correlation coefficient) | SI (Scatter index) |
|--------------------|-------------------------------|------|-------------------------------------|--------------------|
| Bathymetry         | ±2.5% of the mean water depth | <0.10| >0.95                               | <10%               |
| Water level        | ±0.1 m or 10% of the measured level | <0.10| >0.95                               | <10%               |
| Average current speed | ±0.1 m/s or ±10% to 20% of the measured speed | <0.10| >0.95                               | <10%               |
| Peak current speed | Within <0.05 m/s (very good), <0.1 m/s (good), <0.2 m/s (moderate), and <0.3 m/s (poor) | <0.15| >0.90                               | <15%               |
| Mean pollutant concentration | ±20% of the mean measured pollutant concentration | <0.20| >0.90                               | <20%               |
| Temperature        | ±0.5 °C                       | -    | -                                   | -                  |

The search for the best parameter values can be carried by following a trial and error procedure, which was the most used approach in the past. Accordingly, one makes an initial guess of the parameter value and runs the model therefore obtaining simulated data values that are visually compared with the corresponding observations. If the simulation is not satisfactory the parameter value is changed and the model is running again. The simulation is repeated until a satisfactory solution is obtained. While these guidelines naturally remain open to challenges from software users who require more accurate model performance, the statistical processing of input data adapts the models used to a step closer to calibrating the defined model. Therefore, their use in pollutant dispersion modelling studies is recommended.

4 Conclusion

The most frequently used calibration procedure is through the optimization of model performances, which is carried out by comparing observed and simulated data. Expert calibration if of course potentially subjected to a relevant uncertainty, but it might be a very good solution to resolve real world applications. The similarity or dissimilarity of the parameter values in different calibration procedures provides an indication of the reliability of the estimates. Before finally applying the model, it is advisable to finally get the parameter values to be used in practice by performing a calibration by using the whole data sample, to reduce the uncertainty of the parameter estimates as much as possible. Validation is always appropriate in engineering design, in view of the uncertainty affecting hydrological models.

**References**

1. J. G. Arnold, D. N. Moriasi, P. W. Gassman et al., “SWAT: model use, calibration, and validation,” Transactions of the American Society of Agricultural and Biological Engineers, vol. 55, no. 4, pp. 1491–1508, (2012)
2. M. W. van Liew, J. G. Arnold, and D. D. Bosch, “Problems and potential of autocalibrating a hydrologic model,” Transactions of the American Society of Agricultural Engineers, vol. 48, no. 3, pp. 1025–1040, (2005)
3. Burgess C., Complete Valid analytical methods and procedures, Royal Society of Chemistry, Cambridge, United Kingdom, pp 32-43, (2000)
4. Tang W., Fore A., Yueh S., Validating SMAP SSS with in situ measurements, Remote Sensing of Environment, Volume 200, Pages 326-340 (2017)
5. http://www.european-accreditation.org-EA Guidelines of expression of uncertainty in quantitative testing, (2004)
6. Andrew J., Measurement validation via expected uncertainty, Measurement, Volume 30, Issue 3, October (2001)
7. * * * - SR EN ISO/CEI 17025: 2018 - General requirements for the competence of testing and calibration laboratories
8. Eurachem Guide – the Fitness of Purpose of Analytical Methods (Laboratory Guide to Method Validation and Related Topics), ISBN: 978-91-87461-59-0, (2014)
9. Defra/Environment Agency, “Protocols for minimum standards in modelling–flood warning management system phase 2a,” Report prepared for Defra/Environment Agency FCERM management R&D Programme, Tech. Rep. W5C-021/2areport. View at Google Scholar (2005)
10. A. Lefebvre and A. P. Lyons, “Quantification of roughness for seabed characterization,” J. Coast. Res., Spec. Issue 57, 781-785, (2011)
11. H. V. Gupta, K. J. Beven, and T. Wagener, “Model calibration and uncertainty estimation,” in Encyclopedia of Hydrological Sciences, John Wiley & Sons, Ltd., Hoboken, USA, (2005)
12. L. Cea and J. R. French, “Bathymetric error estimation for the calibration and validation of estuarine hydrodynamic models,” Estuarine, Coastal and Shelf Science, vol. 100, pp. 124–132, (2012)
13. N. Huybrechts, H. Smaoui, A. Ouahsine et al., “Prediction of the bed friction coefficient using either
14. N. G. Plant, K. T. Holland, and J. A. Puleo, “Analysis of the scale of errors in nearshore bathymetric data,” Marine Geology, vol. 191, no. 1-2, pp. 71–86, (2002)

15. Toth L., Calamar A., Irimia A., Simion S., Simion A., Research on aquatic pollution level of Maleia river by simulation in Computational Fluid Dynamics, Environmental engineering and management journal 18(4):935-944 (2019)

16. D. P. Dee, “A pragmatic approach to model validation,” in Quantitative Skill Assessment for Coastal Ocean Models, vol. 47, Coastal and Estuarine Studies, pp. 1–13, (1995)

17. R. H. van Waveren, S. Groot, H. Scholten et al., Good Modelling Practice Handbook, STOWA Report 99-05, RWS-RIZA, Utrecht, Netherlands, (1999), http://waterland.net/riza/aquest