

Assimilation method for hydrology models: FEWS Kolubara case study

Luka Stojadinović¹[0000-0002-5312-6577], Miloš Milašinović²[0000-0002-3296-5224], Darko Janjić³
and Nikola Milivojević¹[0000-0001-5330-1219]

¹ Jaroslav Černi Water Institute, Jaroslava Černog 80, Belgrade, Serbia

² University of Belgrade - Faculty of Civil Engineering, Bulevar kralja Aleksandra 73, Belgrade, Serbia

³ PWC Srbijavode, Bulevar umetnosti 2a, Belgrade, Serbia
st_luka@hotmail.com

Abstract. The role of the flood early warning system is to provide information on forecasted water levels along river reaches obtained using mathematical models. Accordingly, it is necessary to ensure adequate initial conditions of mathematical models for the forecast period. Initial conditions are provided through the assimilation process, where corrections to input data, states, or model parameters are made based on the difference between observed and modeled values in the pre-forecast period. This paper presents the initial testing results of the proposed assimilation method for a hydrological-hydraulic model within the Kolubara Flood Early Warning System, based on solving an optimization problem using a genetic algorithm.

Keywords: flood forecast, assimilation, hydrology, genetic algorithm.

1 Introduction

In the past two decades, we have witnessed large-scale flooding worldwide. Changes in the frequency and intensity of precipitation increasingly lead to flooding events on rivers, where the time available for taking defense measures from the moment of warning is short. Reducing risk and enhancing resilience and preparedness for an effective response to flood events can only be achieved if timely and reliable information about the danger is provided. Therefore, the implementation of hydrometeorological systems for early warning and flood alerts is essential to minimize damage and take appropriate flood defense measures in time [1].

One such early warning system is being developed for the Kolubara River basin in central Serbia. The role of the flood early warning system for Kolubara basin (Kolubara FEWS) is to provide information on forecasted water levels along river reaches to issue potential warnings. In fulfilling this role, the early warning system encompasses numerous activities, primarily the collection of meteorological and hydrological data, as well as the use of the collected data through mathematical models that generate forecasted water levels along the modeled river sections [2].

The mathematical model used in Kolubara FEWS is a continuous, coupled hydrological-hydraulic model. This model is utilized for forecast calculations; however, to obtain the most accurate hydrological forecasts using the model, it is essential to ensure an adequate initial model state at the beginning of the forecast calculation. The initial model states can be obtained by updating the model state using observed input data (recorded precipitation and temperatures). Due to the mathematical simplifications of complex physical processes that need to be modeled, as well as insufficient knowledge of the spatial distribution of the basic physical quantities that serve as model inputs (particularly precipitation distribution), it is inevitable that, in some cases, the model results will differ from observed values. Since the continuity of the model propagates errors into the forecast, it is necessary to introduce additional model corrections through the process of assimilation. The assumption is that the initial state of the model for forecast calculation is adequate if the differences between observed and modeled water levels/flows in the period prior to the forecast are minimal. Therefore, the assimilation process involves running the mathematical model for the period immediately before the forecast—the assimilation period—during which the necessary corrections are made.

In the Kolubara FEWS, an assimilation process has been introduced with the goal of spatial correction of precipitation, as it is assumed that due to point measurements at precipitation stations, spatial distribution of precipitation is insufficiently known. In addition to spatial correction of precipitation, the initial soil moisture reservoir states are also corrected, thereby compensating for errors arising from the mathematical simplifications of the model.

Spatial precipitation correction is performed by introducing virtual stations at which precipitation series are assumed. The assumption is that using both recorded precipitation at actual precipitation stations and assumed precipitation at virtual stations will result in a more accurate spatial distribution of precipitation across the entire basin. A similar approach is used for correcting the initial states, by assuming corrections at the virtual station points, which are then spatially distributed throughout the basin.

Therefore, the task of assimilation is to determine the precipitation values at virtual stations and the corrections of the initial states that will result in minimal differences between observed and modeled water levels during the assimilation period.

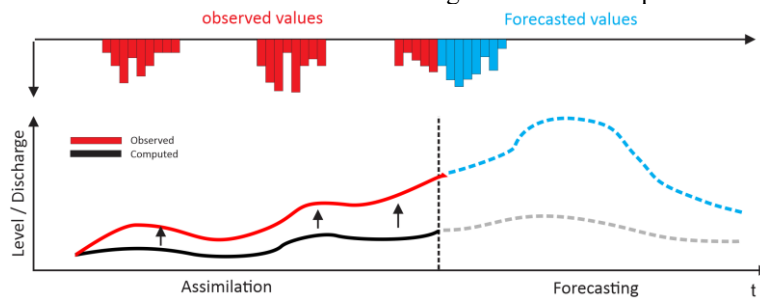


Fig. 1. Illustration of the assimilation process in the flood early warning systems

The previously described assimilation process can be reduced to an optimization problem where the objective is to find parameters that define precipitation at virtual stations and corrections of initial states in such a way as to minimize the function representing the difference between observed and modeled water levels. The assimilation process in the Kolubara FEWS runs automatically within the hydrological forecasting platform.

The following sections of the paper will present the mathematical model for the Tamnava pilot basin (part of the Kolubara basin), where the proposed assimilation method is being tested, a description of the hydrological forecasting platform in which assimilation and forecast calculations are executed, as well as a detailed explanation of the proposed assimilation method. Additionally, initial tests of the assimilation method will be presented, which are part of a broader testing effort currently underway.

2 Methodology

2.1 Mathematical model of Tamnava river Basin

The mathematical model of the Tamnava pilot basin is a coupled hydrological-hydraulic model. It consists of a hydrological model, whose role is to transform forecasted precipitation into runoff from sub-basins, and a 1D hydraulic model, whose role is to simulate wave propagation along the Tamnava and Ub rivers. The purpose of the model is to forecast water levels along the Tamnava and Ub rivers based on precipitation and temperature forecasts within the basin.

The hydrological model was developed using the HEC-HMS (Hydrologic Engineering Center – Hydrologic Modeling System) software package [3]. It comprises of 18 sub-basins with a total area of 724 km², using a 1-hour time step to account for the flash-flood characteristics of the basin. The outputs of the hydrological model are hydrographs from the sub-basins, which serve as input data for the hydraulic model at profiles where boundary conditions are defined.

The one-dimensional hydraulic model was developed using the HEC-RAS [4] (Hydrologic Engineering Center's River Analysis System) software package. The model includes the Tamnava and Ub rivers, with a total modeled river length of 93.5 km.

Calculations within a single sub-basin of the hydrological model can be divided into two parts. The first of the calculations is the vertical water balance. The vertical balance calculation can be described as a sequence of reservoirs between which water exchange occurs. The hydrological model is designed for continuous simulations, where continuity is reflected through changes in the model state (i.e., water volumes in the reservoirs) during the simulation.

The modules involved in the vertical balance calculation include:

- the evapotranspiration module,
- the snow process simulation module,
- the vegetation reservoir module, and

- the soil reservoir module.

The second part involves the calculation of direct and base runoff from the sub-basin, i.e., determining the shape of the output hydrograph. The modules responsible for this part are the effective precipitation transformation module and the baseflow module. The following figure shows the delineation of the Tamnava basin (outlined upper left part of the basin) as well as the river sections simulated using the hydraulic model.

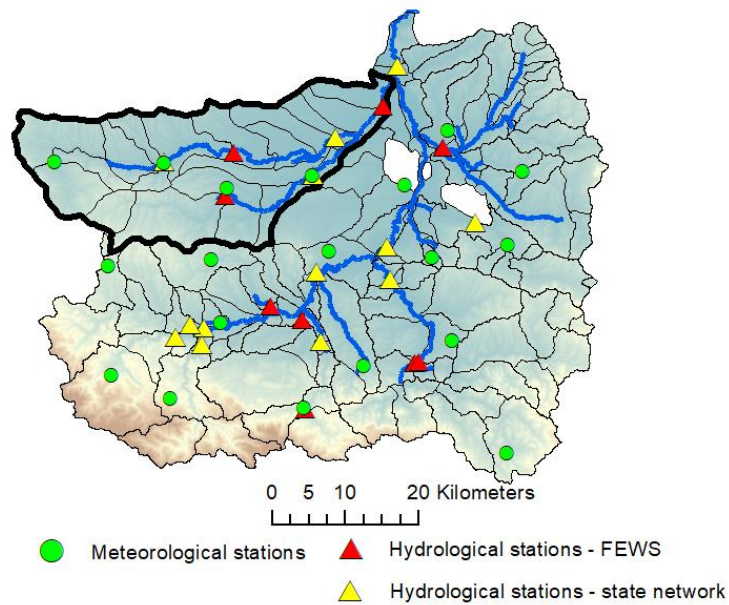


Fig. 2. Tamnava river basin (upper left part) as part of the Kolubara river basin.

2.2 Hydrological forecasting platform

The previously described mathematical model is an integral part of the hydrological forecasting platform. The aim of the platform is to automatically issue hydrological forecasts with each new meteorological forecast, i.e., every 12 hours.

The first step in generating a hydrological forecast is obtaining the initial state of the mathematical model at the forecast start time (time "0"). Ideally, the initial model state is obtained by running simulations with observed data (precipitation and temperature from meteorological stations) for the period immediately before the forecast, using an initial state derived from earlier calculations. This period (hereafter referred to as the assimilation period) is determined based on observed precipitation in the basin, such that it includes all rainfall episodes recorded in the last 21 days prior to the forecast issuance. Once the assimilation period is defined, the initial state for the start of the assimilation period (time "-N") is retrieved from the database of assimilated states.

An initial calculation is then performed, during which observed precipitation from stations is mapped onto the sub-basins of the hydrological model using the Thiessen

polygon method. Based on the differences between the recorded water levels at hydrological stations and the results of the hydraulic model, a decision is made on whether additional corrections to input data are needed using the assimilation method. If the differences between observed and simulated water levels are significant, the assimilation process described in the next chapter is carried out.

After obtaining the model's initial state at time "0", forecast precipitation and temperature time series are used to perform a 10-day forecast simulation. The final assimilated state of the mathematical model is then saved in the database of assimilated states for use in future assimilation calculations. However, the state at time "0" is not saved in the database; instead, the state at time "-V" (short for verification), which is 3 days before the forecast start is stored. This is because the state at time "0" has not yet been verified—i.e., it is not yet known whether this state is adequate, due to the lag between rainfall events and the resulting wave peaks at hydrological stations.

All the processes are executed automatically. System users also can influence both the assimilation results and the forecast results through user applications [5] developed specifically for the Kolubara FEWS.

2.3 Assimilation method

Basics of the method. The fundamental assumption of the assimilation method being developed for the mathematical model of the Kolubara FEWS is that potential differences between observed and modeled water levels during the assimilation period are caused by insufficient knowledge of the spatial distribution of actual precipitation across the entire basin, as well as the lack of precise values for potential evapotranspiration.

By spatially mapping precipitation from meteorological stations in the basin onto the sub-basins of the hydrological model, errors can occur in estimating the total amount and dynamics of rainfall that fell on a given sub-basin. This results in incorrect states of certain hydrological model reservoirs, as well as poorly estimated effective precipitation for sub-basins.

Therefore, it is necessary to correct precipitation over the sub-basins during the assimilation period in such a way as to achieve better alignment between observed and simulated water levels (hydrographs) at hydrological station profiles. Since the number of variables that could be corrected is large (hourly precipitation series for several days across all sub-basins), it is necessary to introduce a limited number of global variables that can modify the precipitation distribution across the entire basin.

For this reason, the concept of "virtual" stations is introduced: additional points within the basin used to more precisely determine the spatial distribution and dynamics of precipitation during the assimilation period. Given that the basin is already well covered by meteorological stations due to the construction of dedicated stations for the system, it is assumed that precipitation values at virtual stations do not significantly differ from those recorded at nearby meteorological stations. Thus, in the first step, precipitation at virtual stations is assumed based on data from the nearest meteorological stations.

These precipitation series are then scaled in intensity and mapped to the sub-basins using the kriging method, which allows greater flexibility in station influence compared to the Thiessen method. The next step involves creating a spatial correction layer for precipitation timing. At meteorological stations, precipitation timing is not corrected because the values are measured. However, at virtual stations, timing corrections are assumed. The kriging method is then used to generate a spatial correction map for the entire basin, which is ultimately used to define the precipitation values for the hydrological model's sub-basins.

This approach produces a corrected precipitation distribution that is spatially continuous across the entire basin and, at meteorological station points, matches the observed values.

A similar logic can be applied to corrections needed due to errors caused by unknown values of actual evapotranspiration. These errors manifest through incorrect initial states of the soil moisture reservoir in the hydrological model at the start of the assimilation period. Since the initial soil moisture states of the hydrological model's sub-basins cannot be directly measured, it is necessary to assume corrections to these initial states at both meteorological and virtual stations, and then spatially map them to the sub-basins using the kriging method.

Using the described method, the assimilation task becomes an optimization problem, where the objective is to minimize the differences between observed and modeled hydrographs during the assimilation period by adjusting the following parameters:

- Correction coefficients for precipitation intensity at virtual stations
- Kriging interpolation parameters for intensity
- Corrections to precipitation timing at virtual stations
- Kriging interpolation parameters for timing corrections
- Corrections to initial soil moisture states at meteorological and virtual stations
- Kriging interpolation parameters for initial soil moisture correction.

Implementation on the Tamnava Basin. For the Tamnava basin, the positions of 4 virtual stations have been selected, as shown in the **Fig. 3**. Accordingly, the total number of optimization parameters amounts to 22 ($4 + 2 + 4 + 2 + 8 + 2$, according to the previously listed set of parameters).

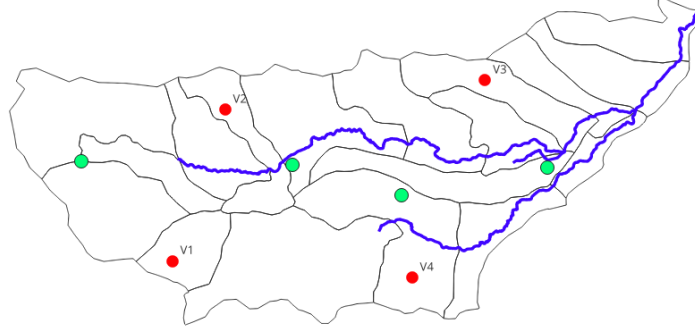


Fig. 3. Location of the virtual (red) and meteorological (green) stations in the Tamnava Basin

The objective function of the optimization process is given by the following equation:

$$RMSE_A = \frac{\sum_1^k RMSE_{HS}^j}{k} \quad (1)$$

Where $RMSE_{HS}$ is the error of the hydrograph at each hydrological station, and k is the number of hydrological stations. The hydrograph error at a hydrological station is calculated using an RMSE equation that has been modified in such a way as to assign greater weight to the hydrograph fit during the period immediately preceding the forecast, i.e., during the verification period:

$$RMSE_{HS} = \sqrt{\frac{\sum_{i=1}^v p_1 (Zobs_i - Zsim_i)^2}{v} + \frac{\sum_{i=v}^n p_2 (Zobs_i - Zsim_i)^2}{n-v}} \quad (2)$$

Where $Zobs_i$ is the observed water level on the hydrological station, $Zsim_i$ is the simulated water level, n is the number of hours of assimilation, v is the number of hours of verification, p_1 is the weighting coefficient for the verification period and p_2 is the weighting coefficient for the period between the start of the assimilation and verification period. These coefficients are set to the values of 0.7 for p_1 and 0.3 for p_2 , where the verification period is 3 days long.

The previously described optimization problem is solved using the NSGA-II algorithm, which belongs to the class of genetic algorithms [6]. The following parameter search boundaries have been adopted for optimization:

- Rainfall intensity correction coefficients: from 0.7 to 1.3
- Shift in rainfall timing: from -6 to +6 hours
- Correction of initial soil moisture state: from -20 mm to +20 mm
- Kriging interpolation parameters: range from 0 to 25 km, and nugget from 0 to 1.

The adopted parameter limits are conservative in the sense that they do not allow for excessive changes to the state of the mathematical model. This is because the continuous model is executed automatically every 12 hours, and allowing overly large corrections could lead to instability and the accumulation of errors in the model states.

Assimilation method testing. Before adopting the previously described assimilation method, extensive testing must be conducted, considering its intended purpose. Accordingly, the testing of the assimilation method can be divided into two groups.

The first group involves testing with synthetically generated examples, where it is assumed that the spatial distribution of precipitation across the entire basin is completely known, that the initial states of the mathematical model at the beginning of assimilation are known, and that the model results at hydrological station profiles are “perfect” — meaning that the known precipitation in each sub-basin and the known initial states will produce exactly the water levels and flows simulated by the mathematical model. The goal of this testing is to verify the configuration of the assimilation method (e.g., testing the method's stability, verifying optimization algorithm parameters, etc.). The second group consists of tests using real data recorded within the basin. This paper will present two tests based on hypothetical scenarios: a sensitivity test and a stability test of the assimilation method.

The goal of the sensitivity test is to determine the maximum possible range in results (water levels and flows) that can be obtained from different combinations of optimized parameter values. The stability test aims to verify a core assumption: to determine whether minimizing the difference between observed and modeled water levels during the assimilation period leads to improved initial states at the start of the forecast.

Calculations will be performed using hypothetical meteorological scenarios without analyzing the impact of snow (with temperatures above 0°C). Therefore, although temperature series are an input for the coupled model calculation, they will not be explicitly discussed in the following text. Additionally, the potential impact of uneven precipitation distribution within a single sub-basin will not be considered.

Sensitivity testing. Sensitivity testing is conducted to determine the range of results that the assimilation method can produce, both during the assimilation period and the forecast period, using the same set of input data from the meteorological stations. The testing algorithm involves assuming the initial state of the mathematical model at the beginning of assimilation (time -N), assuming precipitation at meteorological stations during the assimilation period, and assuming precipitation over the sub-basins during the forecast period. Mass calculations are then performed by varying the parameters of the method for the assimilation period. The initial states obtained at the end of the assimilation period are then used for forecast computation.

To conduct the test, it is necessary to determine which assimilation parameters will be varied and within what ranges. The total number of parameters that can be varied is 22 — consisting of 6 kriging method parameters and 16 correction parameters. If only the combinations of minimum and maximum values were tested, the number of calculations would be 2^{22} , which is 4,194,304, and this is not practically feasible.

For this reason, several restrictions are introduced in parameter variation:

- Parameters related to kriging interpolation are not varied: the kriging method parameters are adopted so that precipitation mapping from stations to sub-basins yields results similar to the Thiessen polygon method.
- It is assumed that corrections to the initial state are identical at nearby stations

- For the remaining parameters, only the minimum and maximum values are varied.

Using this approach, 4096 different calculations are obtained (2^{12}). The reference calculation used for determining the sensitivity metric will be the one where precipitation is mapped to sub-basins using the Thiessen polygon method.

Stability testing. The fundamental assumption of assimilation is that improving the match between observed and modeled hydrographs during the assimilation period leads to better forecasts. This assumption will be tested in the following way:

1. Using the assimilation method — i.e., by varying assimilation parameters — different precipitation scenarios over sub-basins are generated, all based on the same set of precipitation data from meteorological stations.
2. For each of these scenarios, the mathematical model is run to produce “observed” water levels during the assimilation period and corresponding ideal forecasts.
3. Based on the generated observed values and the adopted precipitation data at meteorological stations, the assimilation process is run for each scenario, resulting in an initial state for the forecast period of each scenario.
4. If the forecasts obtained through assimilation are better than those from the reference calculation (where precipitation is mapped from stations to sub-basins using the Thiessen method), the assimilation is considered **stable**.

The basic input data (initial states and precipitation) for the stability test will be taken from the sensitivity testing. The next step in preparing the input data is the selection of different combinations of corrections that produce the “measured” water levels at stations for both the assimilation and forecast periods.

These combinations (i.e., individuals) will be selected from the completed calculations of the sensitivity test. All calculations from the sensitivity test will be sorted by the maximum flow value at the downstream profile of the model. From the sorted sequence, 11 combinations will be selected that are evenly distributed across the range, along with 10 randomly selected combinations, giving a total of 21 assimilation processes to be tested.

3 Results

The previously described testing of the data assimilation method involves the creation of hypothetical rainfall events at meteorological stations within the catchment during the assimilation period, as well as the generation of rainfall data for the forecast period. For testing the method, an assimilation period of 7 days and a forecast period of 5 days were adopted. The simulation was then performed using the generated rainfall event, for both the assimilation and forecast period. The results of the model are presented in the **Fig. 5**

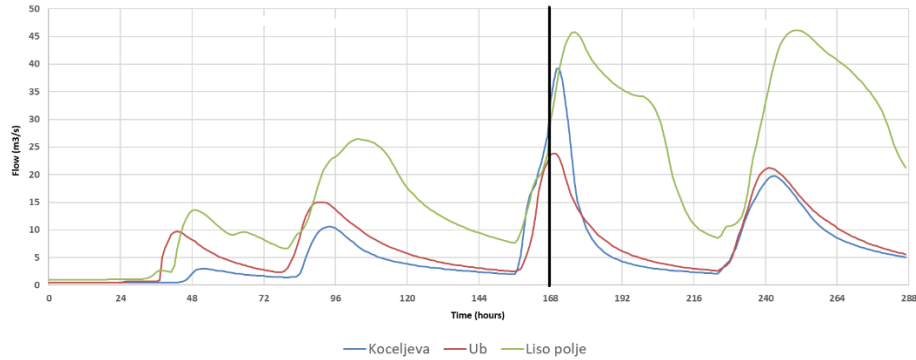


Fig. 4. Results of the mathematical model using the generated hypothetical rainfall event. Black vertical line separates the assimilation and forecast period

A boxplot of the results from 4,096 simulations performed as part of the sensitivity testing of the data assimilation method for the assimilation period (left) and the forecast period (right). The results represent water levels at the profiles of hydrological stations: HS Koceljeva (Tamnava River), HS Ub (Ub River), and HS Liso Polje (Tamnava River downstream of the Ub confluence). For each simulation time step, the median of all sensitivity simulations is shown in orange, while the box at each time-step extends from the first to the third quartile.

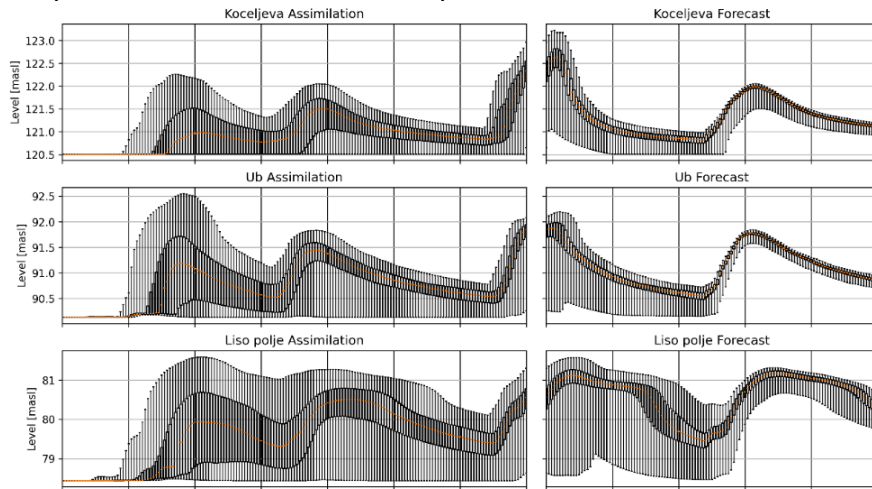
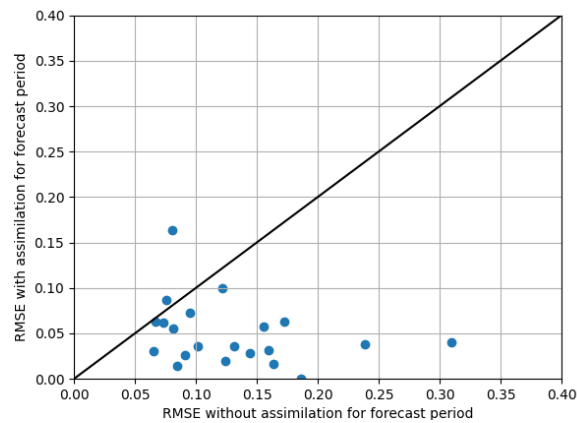


Fig. 5. Box-plot of the sensitivity analysis of assimilation method on the selected hydrological stations

From the results displayed, it can be concluded that the range of results that can be obtained with relatively small corrections to the input data is large, both during the assimilation period and the forecast period.

The testing of the stability of method was performed for 21 hypothetical scenarios. The genetic algorithm was run with 100 generations, each having 90 individuals. The objective function of optimization is the RMSE between the “known” solution and the

computed solution at the previously mentioned profiles. In the next Figure, the RMSE of the forecast for each calculation without using assimilation is shown on the X-axis, while the RMSE in the forecast using the assimilated solution is shown on the Y-axis. All points below the line represent scenarios where assimilation improved the forecast. From the results displayed, it can be concluded that, out of 21 scenarios, 19 cases show an improvement in the forecast compared to the reference results, meaning that the assimilation improved the forecast results.



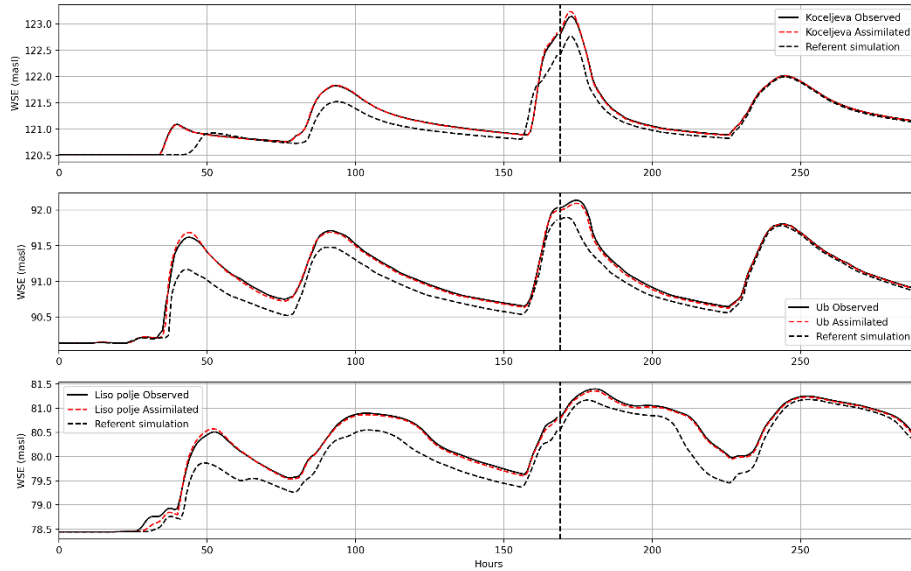


Fig. 7. Detailed results for one of the scenarios in stability analysis.

4 Conclusion

This paper presents an assimilation method developed within the RNU Kolubara system, which is currently being tested on the Tamnava River basin. The method is based on correcting the spatial distribution of precipitation as well as adjusting the soil moisture state of the hydrological model using “virtual” stations. The following are conclusions drawn from the initial testing results of the method.

The range of allowable corrections in the method was chosen to prevent excessive adjustments that could lead to incorrect model states. Accordingly, tests were conducted to verify whether these limited parameters could still influence the model’s results. Based on the testing outcomes, it can be concluded that the specified parameter ranges of the evaluator are adequate, that is, they affect the results and can produce different solutions.

The stability of the method is reflected in whether the forecast period results are improved compared to those without using assimilation. Test results showed that in most cases, the results are better, indicating that the assimilation method is stable.

Further testing of the assimilation method will require additional experiments using hypothetical scenarios, primarily focusing on testing the parameters of the genetic algorithm used to solve the assimilation optimization problem, as well as parameters of the assimilation configuration itself (such as the number and placement of virtual stations, length of the assimilation period, length of the verification period, and similar). After conducting these tests and applying potential method corrections based on the results, the method can then be tested using real data recorded in the basin.

References

1. Cools, J., Innocenti, D., & O'Brien, S. (2016). Lessons from flood early warning systems. *Environmental science & policy*, 58, 117-122.
2. Divac, V. B., Stojadinović, L., Milovanović, M., Vojt, P., & Marisavljević, I. Concept of flood early warning systems in Serbia.
3. U.S. Army Corps of Engineers. (2023). *HEC-HMS Hydrologic Modeling System (Version 4.11)* [Computer software]. Hydrologic Engineering Center. <https://www.hec.usace.army.mil/software/hec-hms/>
4. U.S. Army Corps of Engineers. (2021). *HEC-RAS River Analysis System (Version 6.0)* [Computer software]. Hydrologic Engineering Center. <https://www.hec.usace.army.mil/software/hec-ras/>
5. Milivojević, N., Ćirović, V., Stojadinović, L., Radovanović, J., & Milivojević, V. (2023, March). Technology Platform for Hydroinformatics Systems. In *Conference on Information Technology and its Applications* (pp. 267-275). Cham: Springer Nature Switzerland.
6. Ivanovic, M., Simic, V., Stojanovic, B., Kaplarevic-Malistic, A., & Marovic, B. (2015). Elastic grid resource provisioning with WoBinGO: A parallel framework for genetic algorithm-based optimization. *Future Generation Computer Systems*, 42, 44-54.