

Using High Performance Computing for Online Flood Monitoring and Prediction

Stepan Kuchar, Martin Golasowski, Radim Vavrik, Michal Podhoranyi, Boris Sir, Jan Martinovic

Abstract—The main goal of this article is to describe the online flood monitoring and prediction system Floreon⁺ primarily developed for the Moravian-Silesian region in the Czech Republic and the basic process it uses for running automatic rainfall-runoff and hydrodynamic simulations along with their calibration and uncertainty modeling. It takes a long time to execute such process sequentially, which is not acceptable in the online scenario, so the use of a high performance computing environment is proposed for all parts of the process to shorten their duration. Finally, a case study on the Ostravice River catchment is presented that shows actual durations and their gain from the parallel implementation.

Keywords—Flood prediction process, High performance computing, Online flood prediction system, Parallelization.

I. INTRODUCTION

THERE are many types of natural disasters in the world and many of them depend on geographical location of the observed area. Floods are one of the worst and most recurrent types of natural disasters in our region [1]-[3].

Floods represent a major problem in many regions all around the world [4], [5] and they also frequently affect the population of Central and Eastern Europe. Almost all large rivers in Central and Eastern Europe have experienced catastrophic flood events, e.g. the 1993 and 1995 flooding of the river Rhine, 1999 and 2002 Danube/Theiss Rivers, 1997 Oder River, 2001 Visla River and 2002 Labe River.

The growing number of losses caused by floods in countries around the world suggests that general mitigation of disasters is not a simple matter, but rather a complex issue in which science and technology can play an important role [6]-[8]. Local governments require reliable models for flood simulations and predictions to save on ample funding that must be otherwise invested in post-flood repairs for impacted regions [2]. Therefore, the issue of flood prediction and simulation has been selected as a case for experimental development.

This article was supported by Operational Programme Education for Competitiveness and co-financed by the European Social Fund within the framework of the project New creative teams in priorities of scientific research, reg. no. CZ.1.07/2.3.00/30.0055, by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070) and the project of large infrastructures for research, development and innovation of Ministry of Education, Youth and Sports with reg. no. LM2011033.

S. Kuchar, M. Golasowski, R. Vavrik, M. Podhoranyi, B. Sir and J. Martinovic are with the VSB-Technical University of Ostrava, IT4Innovations National Supercomputing Center, 17. listopadu 15/2172, 708 33 Ostrava, Czech Republic (corresponding author phone: 597-329-588; e-mails: stepan.kuchar@vsb.cz, martin.golasowski@vsb.cz, radim.vavrik@vsb.cz, michal.podhoranyi@vsb.cz, boris.sir@vsb.cz, jan.martinovic@vsb.cz).

Online flood monitoring and prediction systems are one of the means to detect the potential flooding event, warn the emergency committees and provide up to date information for supporting their decisions on which countermeasures should be taken to mitigate the impact of the flood. These systems collect and monitor data from the network of meteorological and hydrological gauges along with the meteorological forecast and periodically execute rainfall-runoff and hydrodynamic simulations to provide short-term prediction of the situation. The need for high accuracy and a short time frame of provided information comes at an expense of high computation demand that can be facilitated by high performance computing (HPC) clusters and supercomputers. We are developing and operating one such online flood monitoring and prediction system called Floreon⁺.

II. FLOREON⁺ SYSTEM

Floreon⁺ is an online system for monitoring, modeling, prediction and support in disaster management primarily developed for the Moravian-Silesian region in the Czech Republic. As it is a web-based platform backed by an HPC infrastructure, its modularity, flexibility and responsiveness allows easy integration of various thematic domains, regions and data. Its main goal is to support operational and tactical disaster management processes by providing and integrating simulations from several domains in near-real time.

The system currently focuses on the domains of flood monitoring and prediction, monitoring and analysis of real time traffic situation and modeling of air pollution caused by dangerous substance leaks

The outputs of the system are provided through a web interface with a map component. This solution consolidates all the integrated sections into one complex geographical framework and therefore delivers detailed overview of the current situation as well as historical events. Simulation results are also provided through web services, which enable their easy integration into other software solutions (e.g. systems for disaster management).

As all executed simulations along with their inputs are being stored, Floreon⁺ is able to analyze historical data and use them to identify patterns in their behavior as well as use the data for statistical analysis of their properties and inaccuracy.

III. ONLINE FLOOD PREDICTION PROCESS

Currently, the main part of the Floreon⁺ system is the flood monitoring and prediction domain. This part contains software modules for automatic and manual execution of the flood

prediction process along with other auxiliary tools (e.g. tool for importing and updating schematizations, monitoring tools, etc.). The online flood prediction process is presented in Fig. 1 and described in the following subsections.

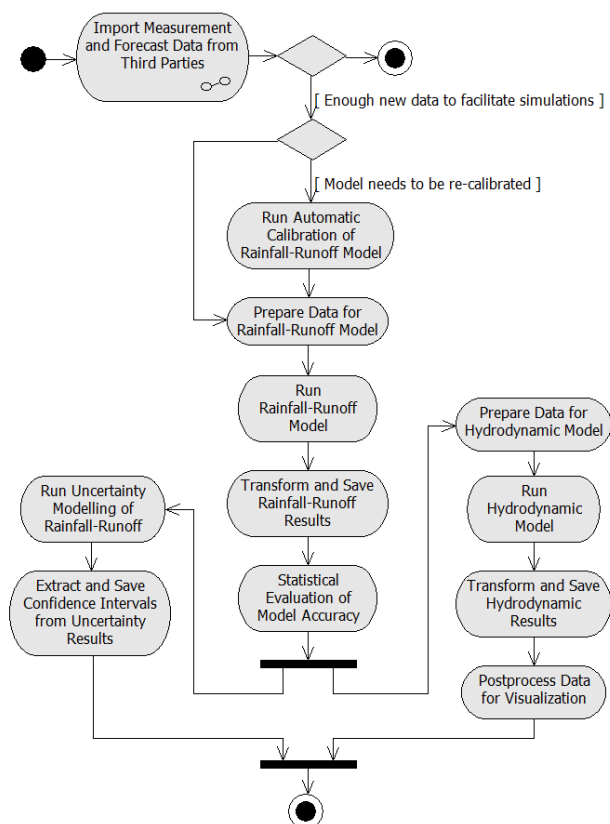


Fig. 1 Online flood prediction process

A. Importing Measured and Forecast Data

The online flood prediction process is started automatically every time new measured or forecast data arrive from the third party providers.

Oder catchment management office sends data from the network of hydrologic gauges every 10 minutes. These data contain current observed water levels, discharge values, precipitation rates and temperatures.

Academy of Sciences of the Czech Republic provides the meteorological forecast from their MEDARD model, which is based on the WRF model, every 6 hours. These forecasts contain precipitation rates, temperatures, wind velocities and wind directions and are issued for the next 72 hours.

Not every successful import starts a new batch of flood prediction simulations. Currently, simulations are executed at the start of every hour as the standard time step in the simulations is set to one hour.

B. Automatic Calibrations of Rainfall-Runoff Models

As the hydrologic parameters of the modelled catchments change continuously based on the current meteorological situation, it is difficult to accurately predict the response of the river flow to a rainfall event without very frequent calibration

of the model. We therefore created a software module for automatic calibration of the rainfall-runoff models in the system (for more information see [9]). This automatic calibration uses optimization methods to minimize the error between the measured and simulated values.

Currently, the calibration method is in an experimental state and is not deployed in the operational environment, but we are planning to run it either once a day or before every simulation depending on the current flooding situation.

C. Rainfall-Runoff Simulations

Modeled domain of the Floreon⁺ system covers the whole Moravian-Silesian Region, the northeastern regional administration unit of the Czech Republic. Hydrologically, it is composed of four catchments of the rivers Opava, Oder, Ostravice and Olza. Hydrological axis of the domain is the river Oder, which is the first order river and all other modeled rivers are its main tributaries. All modeled catchments represent the uppermost part of the whole catchment of the river Oder.

Rainfall-runoff (RR) models used within the Floreon⁺ systems are HEC-HMS (developed by USACE) and Math1D (own in-house model). Both models are semi-distributed event RR models. Physical parameters of the catchments are described by the semi-distributed catchment schematization derived from the DEM and other static geographical data, such as river network, water bodies, hydrological group of soils, land use, land cover etc. Dynamic data inputs are the observed and predicted hydro-meteorological data that are imported from third party providers (mainly precipitations and discharges). Hydrological loss of causal precipitation is solved by the SCS CN method and Green-Ampt model. Transformation of the rainfall to the outflow is modeled using the Clark unit hydrograph and the baseflow is calculated using recession method. Routing of the water in the river channels is solved by kinematic wave model.

Simulations of the models are executed every hour with new observed data and each simulation consists of a 5-day long hydrological simulation with observed data and a 2-day long hydrological prediction with forecast data each with a time step of one hour. The main model outputs are discharges for all modeled gauges that are visualized by the observed and predicted hydrograph overlaid by each model's (HEC-HMS and Math1D) hydrographs (see Fig. 2). Verification of the model outputs is done either visually, using the mentioned hydrographs, or numerically by computing several statistical indicators for the observed and predicted part separately (for more information see [10]).

D. Uncertainty Modeling and Simulations

RR models are usually used for predicting behavior of river discharge and water level and one of the inputs of RR models is the short-term weather forecast. However, weather prediction models are affected by errors that can severely affect modelling results. Magnitude of the error depends on a lot of factors such as wind, topography, temperature or humidity. One of the most sensitive components of the

weather models is the predicted rainfall intensity. Rainfall intensity is a key element of RR process and modelling, therefore we must take uncertainties of rainfall forecast model simulations into account. Uncertainty modelling can then provide additional information to the RR models and enhance the results of these models based on this information.

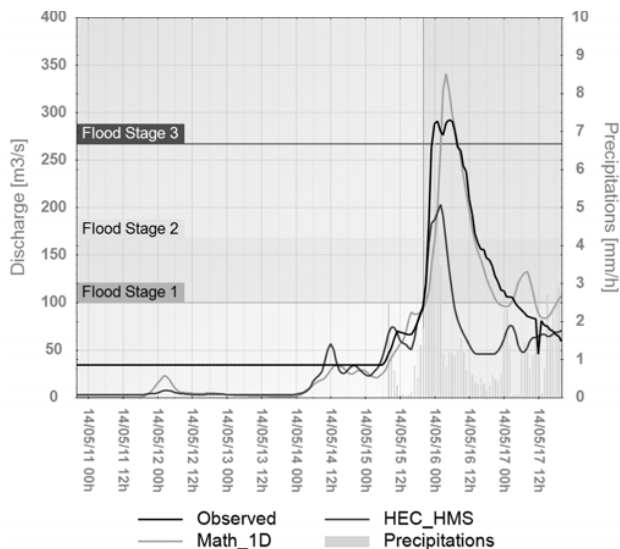


Fig. 2 Example of the rainfall-runoff model results

Such enhancement is very important for decision support in disaster management as it provides probabilistic view on different possible scenarios that can happen in the near future. If the rainfall forecast model underestimates the precipitation rate then the RR results will be very optimistic and the disaster management system will not alert the authorities about the upcoming flood until it is too late.

Our uncertainty modelling method uses Monte Carlo (MC) simulations to compute confidence intervals in which the outputs of the RR models can reside. Such confidence intervals are constructed by coupling respective symmetric percentiles from the results of executed MC simulations. This leads to a number of possible confidence intervals with a well-defined probability of their occurrence (for more information see [10]).

E. Hydrodynamic Simulations

Hydrodynamic (HD) models are commonly used in complex aquatic systems to represent detailed information about water movements. HD models are also used for modeling sediment transport and water quality. The basis of HD models is the set of Navier-Stokes equations that describe the movement of fluids.

Floreon⁺ currently uses HEC-RAS 1D model (developed by USACE) for HD modeling and it uses 1D Saint-Venant equations for computing the simulations. The 1D Saint-Venant equations can be derived from the Navier-Stokes equations.

Sufficient accuracy of modeling results is the main challenge of every HD model. The first step, which must be achieved, is to obtain reliable input data. There is no doubt

that the digital elevation model (DEM) has crucial influence on the HD results. Therefore, the highest available resolution of DEM has to be obtained. Currently, Floreon⁺ uses photogrammetry DEM with the resolution of 10 meters for the entire area of the Moravian-Silesian region and high precision LiDAR DEM with resolution 0.2 meters. Unfortunately, this DEM is available just for a small part of the modeled region (confluence of Olza and Stonavka rivers). Hydraulic data also have a significant influence on the model and include roughness coefficients or hydrographs. A specific subcategory of hydraulic data is represented by the data of geometric description of riverbeds such as information on the number of cross sections, their distribution and location.

Floreon⁺ builds a network of HD models, which are interconnected and use data from the RR models as their inputs. The standard simulations are executed for the current situation and a 2-day prediction with an hourly time step that corresponds with the prediction of the RR models. Nowadays, HD models cover more than 50% of the area of interest (river network of the Moravian-Silesian region) with the main rivers Oder, Ostravice and Olza and their biggest tributaries modeled and monitored.

HD models offer a wide scale of different water results. Results can include an evolution of discharge and water surface elevation, current velocity, temperature as well as sediment transport and a lot of derived statistical values. As these results are rather complex and difficult to quickly evaluate even by experts during flooding events, Floreon⁺ post-processes these results by combining them with geographical information. The result of this post-process is a map layer with flooded areas based on simulated water levels and velocities. This layer is then saved in a geographical database and ready for quick visualization in the system's web interface (see Fig. 3).



Fig. 3 Visualization of the post-processed hydrodynamic results

IV. USING HPC TO SPEED UP THE PROCESS

The main problem with long execution times of the process is not because it contains large models that would take hours to complete, but that it has to run a lot of shorter models. The modeled area of the Moravian-Silesian region consists of four main catchments and each of them is simulated by two RR

models. This means that eight RR simulations have to be performed. Each such run is preceded by automatic model calibration and provides data to the uncertainty modeling simulation and HD simulation. Automatic calibration and uncertainty modeling share the same problem, but on a much higher scale. Each of these methods works on the principle of running multiple simulations with different input parameters that either converge to the configuration with the lowest error in the case of calibration methods, or try to sufficiently cover the stochastic input space in the case of uncertainty modeling. This leads to running hundreds, thousands or even tens of thousands of independent or partially dependent simulations.

HPC as a parallel environment is able to run many hydrological simulations at the same time. This allows the users to use the environment effectively and it shortens waiting time for simulation results even during the high level of demand (e.g. during critical situations).

However, this comes with the implementation cost, because used simulation models are not ready for such simultaneous launching. We had to solve this problem by creating multiple

simulation environments integrated with preparation and finalization code. We named these functional environments Simulators and created one instance for each node and computation core that would be used to perform simulations. Therefore, when the online process starts a simulation based on the imported data, or a user needs to run a simulation manually, the system connects to the HPC cluster and sends all specified parameters and utilizes the HPC environment to find a suitable Simulator instance in the pool of available instances (see Fig. 4). The chosen Simulator prepares the required model and looks into the Floreon Database for imported and simulated meteorological and hydrological data. These are used as inputs to the model and the Simulator starts the simulation. Results of the simulation are saved to the Floreon Database for future use and post-processed. Finally, the Simulator instance is returned to the pool of available instances ready to be used for other simulations. In this way, a specified number of independent simulations can run concurrently and all results can be provided in a short time frame.

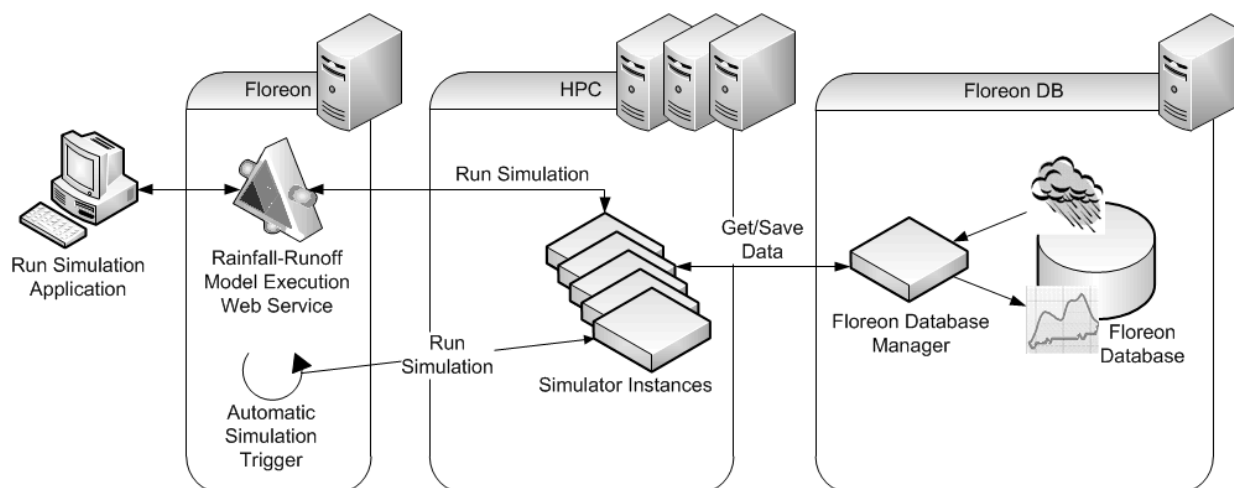


Fig. 4 Running hydrological simulations on HPC in the Floreon⁺ system

V. CASE STUDY

TABLE I

FLOOD PREDICTION PROCESS DURATIONS ON OPERATING HARDWARE

Process Activity	RR Models		HD Models
	Math1D	HEC-HMS	HEC-RAS
Data Preparation	0.8 s	0.7 s	6.3 s
Model Execution	15.2 s	12.8 s	1 min 38 s
Result Post-Processing	2.9 s	3.9 s	12 min 36 s
Simulation Duration	18.9 s	17.3 s	14 min 20.4 s
Time Step Preparation	0.4 s	0.5 s	4.4 s
Time Step Execution	9.8 s	6.4 s	2.6 s
Time Step Post-Process	0.1 s	0.2 s	45 s
Time Step Duration	10.2 s	7.1 s	52 s

To show the exact performance results and the gains of using the HPC environment in the Floreon⁺ system, we have

measured durations of individual parts of the online flood monitoring and prediction process on the Ostravice river catchment. The standard operating environment was used for these measurements – 4 node HPC system each with two 4-core Intel E5420 2.50 GHz processors and 16 GB RAM. Final durations are summarized in Table I where the standard simulation duration is in the upper part of the table and the lower part describes the performance of simulations executed for only one time step of the computation.

The results show that both RR models have comparable short durations around 20 seconds for the standard simulation. Durations for simulations of other catchments in the Moravian-Silesian region are comparable to these results. Therefore if the simulations for all modeled catchments and for both supported models are executed sequentially, it takes around 3 minutes to finish them all. When all of them are executed concurrently (each simulation uses only one CPU

core), it takes only 20 seconds to finish them all.

Standard HD simulations take considerably longer time to finish and it took up to 20 minutes for different modeled catchments. If all possible combinations of input RR model and catchment are executed sequentially, it can take as much as 3 hours to finish. Such long duration is not acceptable in the online system and it only emphasizes the importance of parallelization, as concurrently executed simulations only take 20 minutes to finish. This time is more acceptable, but it can still be too long in critical situations. Fortunately, the parallelization can be extended one step further if steady flow models are used and executed for each time step independently. In this way, a simulation of one time step only takes around 1 minute and all of them can be executed in independent concurrent environments (independent sets of cores on one hardware node or on several different nodes).

Execution time measurements were also performed for automatic calibrations of RR models in the Floreon⁺ system (for more information see [9]). As calibrations are much more computationally intensive than the standard run of the models, they were parallelized using the combination of OpenMP standard and MPI libraries [11]. Experiments with the parallelized versions were executed on computing nodes of the Anselm supercomputer operated by IT4Innovations National Computing Center of Czech Republic where each node is equipped with two 8-core Intel Sandy Bridge E5-2665 2.4 GHz processors and 64 GB RAM. Nodes are connected by a fully non-blocking fat-tree Infiniband network (3600 MB/s).

Measured execution durations of automatic calibrations for different number of CPU cores are shown in Table II.

TABLE II
CALIBRATION PROCESS DURATIONS ON THE ANSELM SUPERCOMPUTER

CPU Cores	Duration
1	11 min 31 s
2	6 min 45 s
4	4 min 49 s
8	3 min 20 s
16	3 min 5 s
32	1 min 33 s
64	49 s
128	47 s

The results show that parallelization of the RR calibration algorithm can significantly shorten the time even though the simulations performed during calibration are only partially independent – multiple concurrent simulations can be performed during one calibration cycle, but several cycles are needed and they have to be executed sequentially. The speedup for the experimental setup is very good for up to 64 CPU cores and it decreases the duration of the calibration from 11.5 minutes to 50 seconds. This makes it feasible to run the calibration process for every simulation to increase its accuracy.

Another computationally intensive part of the process is the uncertainty modeling of RR models. Therefore we also created a hybrid parallel OpenMP/MPI implementation and tested it on the Anselm supercomputer. The resulting durations for

different numbers of MC iterations are presented in Table III.

TABLE III
UNCERTAINTY PROCESS DURATIONS ON THE ANSELM SUPERCOMPUTER

CPU Cores	Duration for 10 000 MC Iterations	Duration for 20 000 MC Iterations
32	33 min 1 s	59 min 14 s
64	16 min 17 s	29 min 33 s
128	11 min 25 s	18 min 47 s
256	6 min 49 s	9 min 24 s
512	2 min 47 s	5 min 19 s
1024	1 min 56 s	3 min 23 s
2048	1 min 52 s	2 min 51 s

The results show that the uncertainty modeling is the longest part of the process as it takes 33 minutes to run 10 000 MC iterations even when parallelized on 32 cores (it would take several hours on 1 core). It is possible to decrease the duration to 2 minutes, but it requires a lot of resources when compared to the other parts of the process. It is therefore necessary to manage the number of allocated resources based on the current demand and severity of the situation.

Even though 10 000 MC iterations provide results with high enough precision, sometimes it is necessary to increase the precision. This can be done by increasing the number of iterations but it also increases the duration or the number of required resources of the whole process.

VI. CONCLUSION AND FUTURE WORK

In this article, we described an online flood monitoring and prediction system Floreon⁺ and how the HPC environment can help to provide current and accurate results in a short timeframe. The total duration of the whole flood simulation process for all modeled catchments would take several hours when executed sequentially (or even tens of hours with uncertainty modeling), but it is possible to decrease the time to the order of tens of minutes when the parallelization of calibration and uncertainty algorithms and concurrent runs of multiple independent simulations are implemented.

In the future, we would like to extend the system with additional 1D hydrodynamic models and also integrate support for 2D models that will be even more computationally intensive and will provide additional parallelization challenges. We also plan to experiment with uncertainty modeling of hydrodynamic models and find ways to implement it to the online process while maintaining acceptable duration of the process.

REFERENCES

- [1] P. B. Bedient, W. C. Huber and B. C. Vieux, Hydrology and floodplain analysis, 4th ed., Prentice Hall, London, 2007.
- [2] R. Brazdil, P. Dobrovolny, L. Elleder, V. Kakos, O. Kotyza, V. Kveton, J. Mackova, M. Muller, J. Stekl, R. Tolasz, H. Valásek, Historické a současné povodně v České republice. MU Brno, CHMI, Brno-Prague, 2005.
- [3] E. W. Wohl, Inland Flood Hazards. Cambridge, Cambridge University Press, 2000.

- [4] M. S. Horritt, P. D. Bates, "Evaluation of 1D and 2D numerical models for predicting river flood inundation" in *Journal of Hydrology* 268(1-4), 2002, pp. 87-99.
- [5] M. R. Knebl, Z. L. Yang, K. Hutchinson, D. R. Maidment, "Regional scale flood modeling using NEXRAD rainfall, GIS, and HEC-HMS/RAS: a case study for the San Antonio River Basin Summer 2002 storm event" in *Journal of Environmental Management* 75, 2005, pp. 325-336.
- [6] X. T. Cheng, "Flood and drought evaluation and management" in *General report for question 87. Barcelona: ICOLD, 2006.*
- [7] R. B. Flowers, "Flood management: a key to sustainable development and integrated water resources" in *Proc. 3rd World Water Forum, Kyoto, Stiga, Osaka, Japan, Mar. 2003.*
- [8] H. D. Guo, "Understanding global natural disasters and the role of earth observation" in *International Journal of Digital Earth*, 3 (3), 2010, pp. 221-230.
- [9] R. Vavrik, M. Theuer, M. Golasowski, S. Kuchar, M. Podhoranyi, V. Vondrak, "Automatic Calibration of Rainfall-Runoff Models and its Parallelization Strategies" in *Proc. 12th International Conference of Numerical Analysis and Applied Mathematics, Rhodes, Greece, September 2014.*
- [10] S. Kuchar, T. Kocyan, P. Praks, M. Litschmannova, J. Martinovic, V. Vondrak, "Simulation of Uncertainty in Rainfall-Runoff Models and their Statistical Evaluation in the Floreon System" in *Proc. 11th International Conference on Modeling and Applied Simulation, Vienna, Austria, September 2012, pp.128-133.*
- [11] M. J. Quinn, *Parallel Programming in C with MPI and OpenMP.* McGraw-Hill, New York, 2003.