Dynamic Data Driven approach to improve the performance of a river simulation

Adriana Gaudiani[1] and Emilio Luque[2]

[1] Universidad Nacional de General Sarmiento, Instituto de Ciencias, Buenos Aires, Argentina
agaudiani@campus.ungs.edu.ar,

[2] Universitat Autònoma de Barcelona, Dept. de Arquitectura de Computadores y Sistemas Operativos, 08193, Bellaterra(Barcelona), España

Abstract. In this research we incorporate the contributions of the dynamic data driven systems development that is based on the possibility of incorporating data obtained in real time into an executing application, in particular a simulation. This paper reports on the first phase of our research in which we have used this idea to enhance the simulation quality of a river flow simulator by dynamic data inputs during the computational execution. We had presented an optimization methodology of this simulator model in previous works but in this opportunity, we could handle those time periods when a sudden level change takes place in the river and we could improve the forecasting prediction. These results are the path towards the development of an automatic calibration framework fed with real-time data.

Keywords: simulator optimization, dynamic data driven, model calibration, real-time data

1 Introduction

A simulation system is continually influenced by real time data for better analysis and prediction of the system under study but uncertainties is inherent in modeling studies. It is vital that these models should be equipped with robust calibration and uncertainty analysis techniques, as explained in [4]. The accuracy of computer simulation depends largely on having reliable data.

In previous researches, we dealt with uncertainty in the values of the input parameters of a river basin model and the impact of this uncertainty in estimating daily water height and forecasting reliability. We proposed an optimization via simulation methodology for enhancing the simulation quality looking for the best set of parameter values to calibrate the computational model [2]. To carry out this research work we used an hydrodynamic model, Ezeiza, for the Paraná River, which is located in Argentina. This system domain is characterized by a large number of parameters and finding the optimal set is a computationally intractable problem. To deal with this issue we implemented a heuristic in two-phases in order to reduce the search space of the parameter values, as we explain in detail in [3]. In order to make the optimization method clearer, we show an outline with its main idea in Fig. 1. We achieved a reduction of 13-19% in the

simulation errors compared to the classical simulation. In more recent work we apply this optimization scheme tacking advantage of the inherent continuity to the simulated system as we explain in [6].

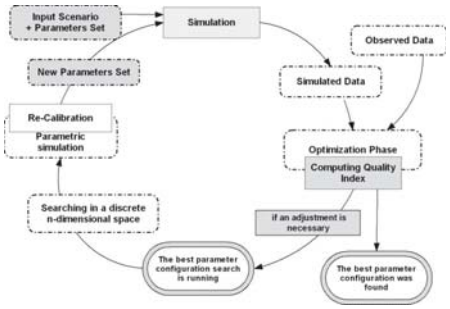


Fig. 1: Calibration process of the hydrodynamic model

In this paper we focus on enhancing our previous results and achieving better simulation quality. The improvement is based on the fact that the simulated system is a dynamic system and the adjusted parameters values of the model may change over time during the simulation. On the other hand, we try to extend this profit to most of the stations located in the domain of the Paraná River. Our main objective is to detect when the simulator needs a new calibration in a decision-making process. This process is carried out in a dynamic way and what we mean is that this detection occurs while simulation is running.

To address this problem we present a Dynamic Data Driven (DDD) approach to improve the Paraná River basin simulation based on our previous two-phases enhancing scheme. The simulation shows that the new method can give a 50% reduction in prediction error in Rosario city compared with our previous approach. One obstacle to implement successfully this new scheme is the rapid growth of computing time required to complete the simulation tuning. High performance computing provides the infrastructure and tools to handle this problem.

2 Dynamic Data Driven technique for a better prediction

In general, traditional simulations separate from real systems by using a few real-time data as say [1]. DDD can be considered as a new simulation paradigm aimed at achieving a better prediction of the behaviour of the system under study. As Xialun Hu says in [7] DDD Simulation conects Real-Time data with simulation and incorporating real-time data into a running simulation model has the potential to significantly improve simulation results. Many authors present their jobs in this line [5]

We propose to consider the assimilation of the real data, that is, the heights of the water in the river bed measured at each monitoring station. Without assimilating these observed data into the real system, the difference between the simulated data and the

real data is likely to grow on undesirable values. This is the main idea that motivates directing our research in this direction.

Incorporating real-time data from the system under study we are able to incorporate dynamic data into a running simulation model in order to greatly improve simulation results by computing the simulation error at each time step. This study prompted us to define a conceptual framework that dynamically detects the simulation error rate and makes the decision to start a new calibration. This approach requires predefining a threshold that will determine when a new calibration process is necessary. This threshold value was determined based on the knowledge we acquired of the system during the optimization work prior to this research.

The results obtained with the experimentation carried out to verify the effectiveness of the presented technique encourages us to begin the development of a conceptual environment that dynamically adjusts the simulation. Below we present the proposed methodology.

3 Methodology

We proposed a methodology guided by a continuous assimilation of the real data during the simulation. This research is based on our previous work and we need to deal with several challenges.

A dynamic assimilation function run at the same time with the river simulation which process the real time data and determining in a dynamic way the quality of the simulation, in order to forecast the water wave propagation as accurately as possible. It is therefore necessary that simulated and real data be fed into the DDD function and this process is responsible for stopping the simulation each time a new calibration is needed.

Figure 2 we can see the steps of the DDD assimilation module running together with the river simulator. In this graph, we show the successive calibration steps which will perform the search of a new set of adjusted parameters.

Each time that a search of the adjusted set of parameters is required it is performed as an optimization problem which minimize the quality index (QI) of each simulation scenario by using a parametric simulation. This optimization methodology take place in two phases which combines an optimization heuristics and a simulation analysis. We had to reduce the search space to make the optimization problem a computational treatable problem, as we described in [3].

4 Experimentation

The parameters used to calibrate the computational model are the Manning coefficient (Mn) and the leaves height (Lh). The value of these parameters must be set for each of the 75 sections necessary to discretize the domain of the system, that is, the Paraná River. We want to highlight that these values remain constant while the simulation is running since the simulation algorithm does not modify them. The DDD function is

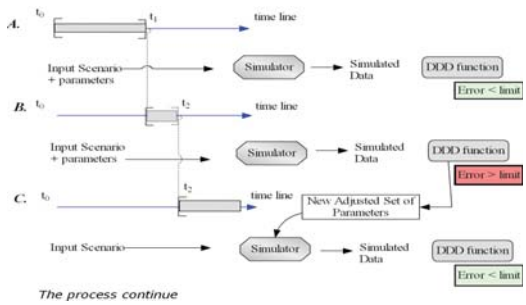


Fig. 2: DDD Optimization scheme and simulator running to calibrate the model dynamically.

continuously active testing the output values to take the decisions. Based on this knowledge, we launched a new simulation stage each time the DDD function detects a QI greater than a chosen threshold determined a priori.

For each simulation day we are receiving the measurements of the water height from 20 monitoring stations located in the river bed. We show in Figure 3 a) the difference between the simulated and real hydrograph for Rosario city and the different times that the simulation must be stopped. The simulation optimized by the new methodology is shown in Figure 3 b) and we want to remark the profit achieved. We compared the improvement achieved in Rosario Station and we obtained the following gain

- The QI in classical simulation is 0.59
- The QI in our OvS initial methodology is 0.51
- The QI in our DDD scheme is 0.29

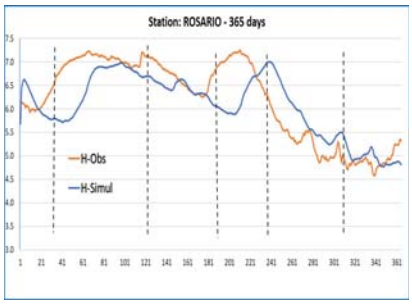
In our previous work we obtained a gain of 13% but now we obtained a gain of 50%.

The whole process is computing time demanding due to the number of recalibration needed. This fact depends on the successive variations in the water level in the real system. We performed 5 calibrations and the whole computing job took 12 hours running on a 16 processors cluster.

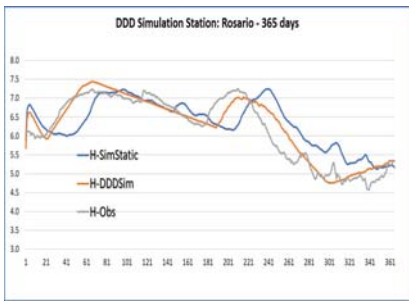
5 Conclusion

We propose to handle the simulation errors arising from those events when a sudden level change takes place, which happens when water level raises or falls more rapidly than the rest of the period. This reflects the underlying model, but it is not our goal to change the simulator kernel and we have focused in a methodology that can deal with these disruptions.

The improvement achieved with our methodology reduced the simulation error about 50% compared with the best results reached in our previous research for Rosario station. Despite we improved forecasting prediction in one station as a pilot test we are encouraged to go on with the investigation. Currently, we are working on a framework to run the complete DDD optimization process and real-time data processing in order to achieve an automatic calibration.



a) We highlight when the Quality Index exceeds the error threshold.



b) Improved results when we use our DDD approach.

Fig. 3: Rosario forecasting improvement

We have to reduce the computing time because we need to improve as many stations as possible and the calibration should not last more than one day. That would be a reasonable time for a real-time flood forecasting taking into account the time it takes for the water wave to travel through the riverbed of the river.

References

1. Erik Blasch, Guna Seetharaman, and Frederica Darema. Dynamic data driven applications systems (dddas) modeling for automatic target recognition. In *Proceedings of SPIE - The International Society for Optical Engineering*, volume 8744, 2013.
2. A. Gaudiani, E. Luque, P. García, M. Re, M. Naiouf, and A. De Giusti. Computing, a powerful tool for improving the parameters simulation quality in flood prediction. *Procedia Computer Science*, 29:299 – 309, 2014. 2014 International Conference on Computational Science.
3. A. Gaudiani, E. Luque, P. García, M. Re, M. Naiouf, and A. De Giusti. How a computational method can help to improve the quality of river flood prediction by simulation. In M. Gomez, Sonnenschein M, and Vogel U., editors, *Advances and New Trends in Environmental and Energy Informatics.*, Progress in IS. Springer, Cham, 2016.
4. Ki Yong Lee, YoonJae Shin, YeonJeong Choe, SeonJeong Kim, Young-Kyoon Suh, Jeong Hwan Sa, and Kum Won Cho. Design and implementation of a data-driven simulation service system. In *Proceedings of the Sixth International Conference on Emerging Databases: Technologies, Applications, and Theory*, EDB '16, page 77–80, New York, NY, USA, 2016. Association for Computing Machinery.
5. E.E. Prudencio, P.T. Bauman, S.V. Williams, D. Faghihi, K. Ravi-Chandar, and J.T. Oden. A dynamic data driven application system for real-time monitoring of stochastic damage. *Procedia Computer Science*, 18:2056 – 2065, 2013. 2013 International Conference on Computational Science.
6. M. Trigila and Luque E. Gaudiani, A. and. Agile tuning method in successive steps for a river flow simulator. *Lecture Notes in Computer Science*, vol 10862:639–646, 2018.
7. Hu X. Dynamic data-driven simulation: Connecting real-time data with simulation. In Yilmaz L., editor, *Concepts and Methodologies for Modeling and Simulation. Simulation Foundations, Methods and Applications*. Springer, 2015.