INTEGRATION OF INFORMATION TECHNOLOGY SYSTEMS FOR FLOOD FORECASTING WITH HYBRID DATA SOURCES

M. C. Fava[1]; G. Santana[2]; D. A. Bressiani[1]; A. Rosa[1]; F. E. A. Horita[2]; V. C. B. Souza[3] and E. M. Mendiondo[1]

[1.] *Escola de Engenharia de São Carlos - Universidade de São Paulo (University of São Paulo-USP)*
[2.] *Instituto de ciências Matemáticas e da Computação - Universidade de São Paulo (University of São Paulo - USP)*
[3] *Centro de Tecnologia da Universidade Federal de Alagoas (Federal University of Alagoas - UFAL)*

ABSTRACT: This paper presents a new methodology for flood forecasting integrating volunteered geographic information – VGI and wireless sensor networks – WSN. The methodology was applied to São Carlos city urban basin, São Paulo state, Brazil and showed promising results compared to the forecasted ones using only sensor data. The major contribution of this methodology is the high capacity of adaptation to different kinds of data, increasing the amount of data and reducing the forecasting uncertainty. Also, the proposed methodology can simulate scenarios of predicted precipitation using weather radar and two other methods - Precipitation forecasting using weather data and forecasting using statistics techniques. This work is still under development and the main objective of this paper is to present the current status of the research and the proposed methodology. The preliminary results showed that the basin model still needs improvements before conducting calibration.

Key Words: Flood Forecasting, WSN, VGI, SWMM

1. INTRODUCTION

According to the United Nations (UN) Brazil is among the countries with the highest incidence of floods. The total economic direct and indirect costs related to flood incidents reach 3% of the GDP and the damages are equivalent to 11% of tax revenues. Brazil has shown annually hundreds of deaths by drowning, flooding and accidents in flood valleys (Mendiondo *et al*., 2013).

The increase in floods and torrents over the last decades has motivated researchers to develop more accurate forecasting models. Although useful, it is noted that the existing models of hydrometeorological prediction have limitations. However, these models have features that can be complemented. The rainfall-runoff models, for example, are commonly used for creating flooding scenarios only with probabilistic rainfall (Alvisi *et al*., 2013; Hughes, 2013). On the other hand, recently studies of precipitation forecasting models based on radar data have presented good results (Austin and Bellon, 1974; Löwe *et al.*, 2014). The integration of these two distinct phases of the hydrological cycle (rainfall-runoff model and radar based precipitation forecast) can be a real contribution to the current systems.

Another technology that can be added is the wireless sensor networks (WSN). Using sensors among technologies of transmission of real-time information has higher robustness compared to the traditional data acquisition techniques (Hughes *et al*., 2011). This system, however, has the disadvantage of high costs and difficulty in ensuring protection of the field equipment. Recently, a new concept has emerged - the VGI (Volunteered Geographic Information) - which consists in getting data through information provided by volunteer citizens. Unlike the WSN, the VGI has low cost and easy maintenance (Goodchild, 2007; Poser and Dransch, 2010; Degrossi *et al*., 2014; Fava *et al.,* 2013).

Regarding the information which serves as input data for model prediction, these two methods can be used together, complementing and increasing promisingly the amount of information and quality of model prediction.

The aim of this study is to propose the integration of diverse systems that directly or indirectly assist in flood forecasting using hybrid data sources. Therefore, this system is applied to the urban watershed of São Carlos - SP.

The proposed flood forecasting hybrid system: Hydrological Alert Model with Participatory Basis (HAMPB) was implemented through various modules; each responsible for executing a task. The details of integration and data exchange between different applications through the use of management systems database was discussed. This work is still in progress, therefore the main contribution is the proposal of a new methodology and the partial results obtained.

2. DESCRIPTION OF THE HAMPB SYSTEM

The general scheme of HAMPB model is shown in Figure 1. The main program ("Brain") centralizes and manages all modules. In the following sections details of each module are explained. Field data are obtained from different data sources and standardized and formatted for HAMPB.

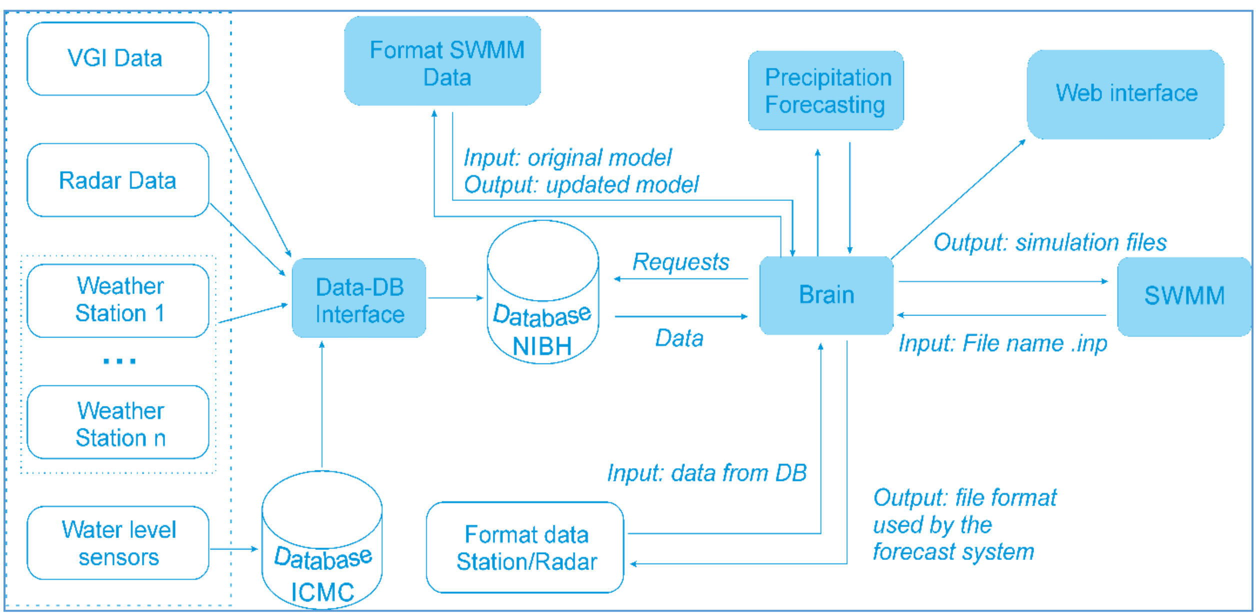

Figure 1: Fluxogram of HAMP Model.

2.1 Data Acquisition

To obtain VGI data, a system developed by Degrossi *et al.* (2014) was used. The data are reported sporadically by volunteers via mobile app, website or both. The user can inform the current river level by observing on the water level rulers or reporting intervals of flood hazard index at channel walls proposed by Rotava *et al*. (2013) (Figure 2a).

The WSN data are obtained through nodes of WSN network consisting of water level sensors installed at three points of the basin (Figure 2-b). These data are transmitted through mobile network to a database at the Institute of Mathematics and Computer Sciences (ICMC), USP.

(a) (b)

Figure 2: (a) Water level and hazard index and (b) Water Level Sensor.

Data from meteorological stations are composed of weather variables as: pressure, relative humidity and temperature. Public data provided by the National Institute of Meteorology (INMET) was used, however, the HAMPB can obtain data from other weather stations e.g. Brazilian National Institute of Space Research (INPE), Natural Disasters Monitoring Center (CEMADEN) or other sources.

HAMPB model has a module capable of receiving data, requiring only data processing to fit the variables used for the prediction system. The HAMPB model can forecast precipitation by three different methods that are described on the next section.

For practical reasons, the HAMPB model does not perform queries on ICMC and National Institute of Meteorology databases all the time. The data is replicated to a local database called NIBH database located at research laboratory “NIBH/SHS/EESC-USP” at USP São Carlos – SP, Brazil. This module is also responsible for collecting any VGI information received via website or mobile application integrating them into NIBH database. Data acquisition module is also responsible for importing the data obtained by the software from meteorological stations. All necessary information to run the HAMPB model is centralized in a single local database, eliminating problems such as lack of communication or latency with other servers.

2.2 Precipitation Forecasting

The second module is called "rainfall forecasting". “Brain” constantly executes this module along with the data acquisition. With the data in the required format, the user can run and get the rainfall graph derived from precipitation forecast model.

The HAMPB model has three models for rainfall forecasting: (i) a prediction model that uses radar data and data from meteorological station, proposed by Gonçalves (2009); (ii) if there is no radar information available a second model predicts only weather data; (iii) and the third model uses a univariate (1D) linear regression of previous rainfalls. The three models are more detailed below.

2.2.1 Precipitation forecasting using radar data

Gonçalves (2009) proposed a model for forecasting rainfall using information generated by weather radar systems, coupled with a hydrometeorological prediction model, followed by a rainfall-runoff model. The precipitation forecast model was an adaptation of the model initially proposed by Georgakakos and Bras (1984) and applied by Andrade (2006).

The precipitation forecast considers 1D (one dimensional) vertical approach of a cloud as a reservoir of condensed water, whose temporal variation of the liquid water mass is expressed by a balance equation of moisture as:

$$\frac{dX(t)}{dt} = I(t) - O_t(t) - O_{b(t)} \tag{1}$$

where $(dX(t))/dt$ is the temporal variation of water storage in the cloud [kg.m^{-2}.s^{-1}]; $I(t)$ is the humidity due to the condensation of the vapor contained in the ascending air entering the cloud [kg.m^{-2}.s^{-1}]; $O_t(t)$ is the output of vapor through the cloud top [kg.m^{-2}.s^{-1}] e $O_{b(t)}$ is output of the moisture through the bottom of cloud [kg.m^{-2}.s^{-1}].

The following weather variables are measured on the surface: temperature (T_0), pressure (P_0) and dew point temperature (T_d). These variables are used as input data for the model. The portion of the model relative to the mass of liquid water (X) for a given time *t* is the estimated content of vertically integrated liquid water (*VIL*) calculated by weather radar (Gonçalves, 2009).

Besides the use of VIL information for the echo top determines the height of the precipitating systems. Providing an information of Echo Tops, *VIL* in the next interval (t +1), and applying the information to estimate precipitation proposed Georgakakos and Bras (1984) gives the value of the expected rain for each time interval. More details about the rainfall model prediction using radar data can be obtained at Gonçalves (2009).

2.2.2 Precipitation forecasting using weather data

The precipitation forecast based on pure, without radar information, weather data was initially proposed by Georgakakos and Bras (1984) and applied by Andrade (2006). A weather station provides values of temperature, pressure and dew point temperature to predict precipitation one step forward. Future temperature, pressure and dew point are estimated through correlation of each variable.

2.2.3 Precipitation forecasting using statistics techniques

In order to perform the prediction of rainfall for the next hour, the methodology proposed by Georgakakos and Bras (1984) is used, which consists on relating data from earlier rainfall through a univariate linear regression of lag 1 to the variable to be forecast (Equation 2):

$$y_{t+1} = \bar{y} + (y_t - \bar{y}).\rho_y + \varepsilon_1.\sigma_y.\sqrt{(1-\rho_y^2)} \tag{2}$$

where y_{t+1} is the forecasted precipitation value at *t + 1*, y is the current precipitation, $\bar{y}$ is the average of y, ε_1 is a random number between 0 and 1, σ_y is the standard deviation of y, and ρ_y is the autocorrelation coefficient of lag 1 given by Equation 3:

$$\rho_y = \frac{\sum_{t=1}^{n-1}(y_t-\bar{y}).(y_{t+1}-\bar{y})}{\sum_{t=1}^{n}(y_t-\bar{y})^2} \tag{3}$$

Where *n* is the number of intervals considered in the autocorrelation.

2.3 SWMM Interface

Having the acquired water level data (VGI and WSN) through the data acquisition module and the predicted precipitation obtained from the prediction module, hydrological simulations were assessed with the Storm Water Management Model (SWMM) developed by the U.S. Environmental Protection Agency (Rossman, 2010). SWMM is a rainfall-runoff model capable of simulating single or extended period of events, simulating quantitative and qualitative aspects. The model has been widely used in academia due to its high performance, free, open source and large support Internet group of users (Rossman, 2010).

SWMM Interfacing Module proposed in this methodology (Figure 1) is responsible for all actions involving the use of SWMM. It is responsible for editing the input file for SWMM model of a desired basin and automatically inserting the rainfall data obtained from the prediction made in the previous step. In addition, this module also uses the information obtained from the levels of WSN networks and VGI, converting them into flow values and updating the information of the initial and final flow at each link on SWMM or even updates level values directly without necessarily turning them into flow values. This module is responsible for making the updated file for the simulation. These data can also be modified in a simpler way, in cases where only one of the points of the simulated basin need a new value or receive a new information, it is possible to only change the values at this point of interest.

2.4 Web Interface

The module “Web Interface” is responsible for using simulated data and comparing with the actual data for the error calculations and display the graphics display. These results can be published on a website for real-time early warning.

2.5 Brain

“Brain” is responsible for running each module in the correct order. “Brain” module receives data from various hybrid sources and match these information so that with each new round of information the model is updated, providing online forecasts.

Figure 3 shows three possible different monitoring configurations for using HAMPB. Situation 1 indicates VGI, WSN and the observation point (simulation) in three distinct places. The situation 2 shows a scenario where the three points are located in the same place. And the third situation shows a place of observation and monitoring point (WSN) complemented with a VGI point in other place.

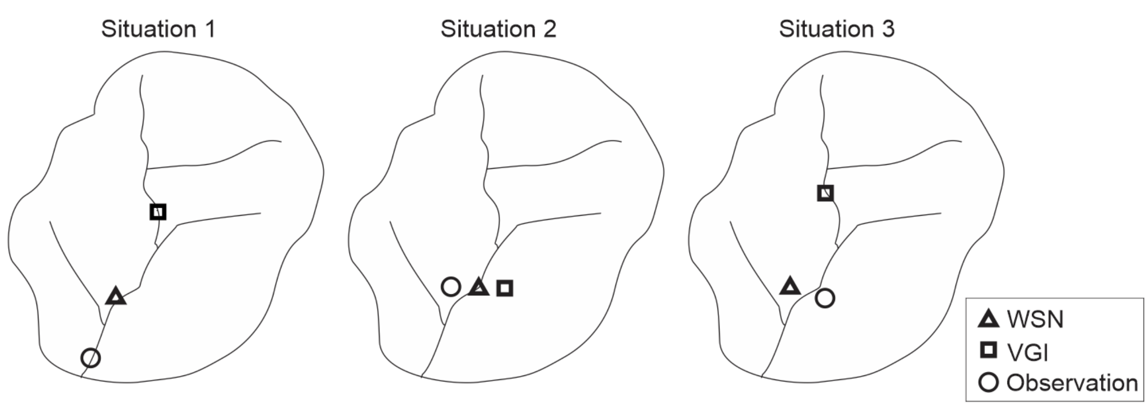

Figure 3: Distinct monitoring configurations for the basins.

3. MODEL CALIBRATION

The calibrator used was a multi-site calibrator developed by Shinma and Reis (2014). The optimization technique used in the calibration of the models of nested basins have shown robustness compared to the use of sequential calibration of each sub-basins (Shinma and Reis, 2014)

The calibrator system calibrates 8 parameters, defined into valid physical intervals, the parameters and its intervals are listed in Table 1.

Table 1: Lower and upper limits of the calibration parameters (Shinma and Reis, 2014).

Parameter	Acronym	Lower limit	Upper limit
Width factor	mW	0,8	1,2
Slope multiplier	mS	0,8	1,2
Manning roughness coefficients for impermeable areas	NI	0,01	0,03
Manning roughness coefficients for permeable areas	NP	0,1	0,25
Horton's Maximum infiltration rate (mm/h)	I0	8,0	44,14
Horton's Minimum infiltration rate (mm/h)	Ib	2,5	22,0
Horton's Decay Constant (h^{-1})	K	2,0	7,0
Manning's roughness coefficient multiplier	Mn	0,8	1,2

3.1 Application

The calibration was conducted for the Monjolinho urban watershed, São Carlos - SP; a river basin with a large number of tributaries that contributes to a single drainage unit. The basin was modeled in SWMM through 156 nodes and 57 sub-basins with drainage areas ranging between 0.07 to 5.1 km². The total drainage area of Monjolinho watershed is 76 km^2. The basin has a population density of 194.53 inhabitants/km² according to the municipal planning secretary (PMSC, 2005). It has an average altitude of 856 meters in relation to sea level and the soil is considered highly permeable. Land use and land cover, topography and channel characteristics (Figure 4) were considered for modeling. Currently there are 15 water level meter installed in different parts of the basin and in three of these points there are also a range of hazard index. The calibration system uses genetic algorithms to perform the optimization and needs a set of parameters. The parameters used for the genetic algorithm were an initial population of 100; with 100 iterations: 100; mutation probability of 0.07; and crossover probability of 0.7.

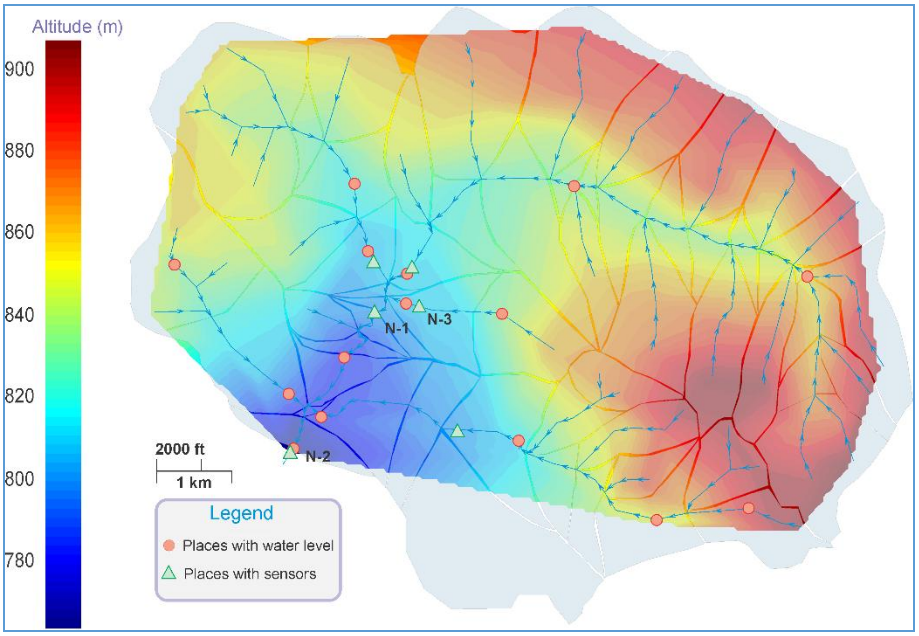

Figure 4: Monitoring points of São Carlos Basin.

In this paper, we present the performance of VGI, WSN systems for rainfall-runoff events occurred at the São Carlos basin on the following days: 27/12/2004, 25/01/2005, 31/12/2005 and 27/01/2006. The time step used was registered every 5 minutes (Δt = 5 min).

We calibrated the model using node N-1 as showed in Figure 4. The nodes N-2 and N-3 are used to simulate flow after the model is updated using the HAMPB methodology (Figure 4).

Figure 5 shows preliminary results obtained for a precipitation event. The simulation obtained from the calibrated model (red line) was close to the real values (blue line) for most of the time series. However, the simulation was notsatisfactory for the flood peaks, which are the most important for flood forecasting. The calibration can be improved by modifying the objective function of the optimization method, in order to the minimize the errors on the peaks.

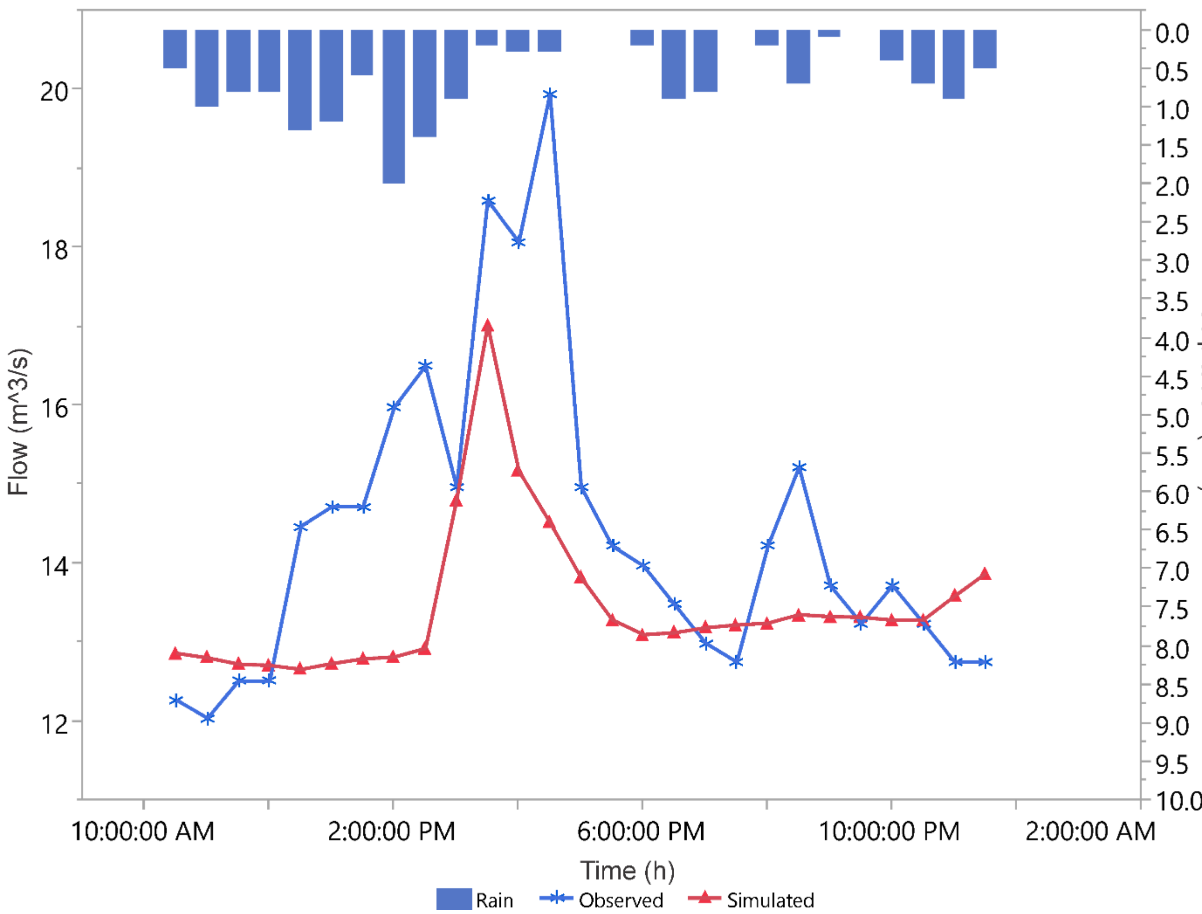

Figure 5: Calibration results in a monitoring point.

After the hydrological model calibration we performed rain forecast using weather data (temperature, pressure and dew point temperature) presented in Figure 6. The blue bars are the predicted rainfall while the green bars represent field measurement. After obtain the predicted precipitation, SWMM model was duplicated; the first copy was updated with the field data and the second copy was updated with the predicted rain, simulations were performed using the two new models – the green line shows the simulation using observed precipitation and the blue line shows the simulation using the predicted precipitation.

Through the graph it is possible to see that there is a delay of 90 minutes between the maximum peak of precipitation and flow. The simulated flow at this point remains above the real values for a few hours (between 12.30 and 16.30), while the observed flow rate immediately decreases with the decrease of peak rainfall showing that the time of concentration and the duration of the rain require adjustments on the modeling that were not yet solved by the calibration.

The flow forecasting of the second point (N-3) is shown in Figure 7. The predicted flow was overestimated by the model, and the same rainfall prediction was used for the entire basin. The prediction results were similar to those found in the first section. Although the flows have presented higher values to the first point (maximum flow of 45 m^3/ s on the first point and 13m^3 / s in the second point), the pattern found in the two figures is quite similar. And, likewise, we can verify the need for adjustments on the model.

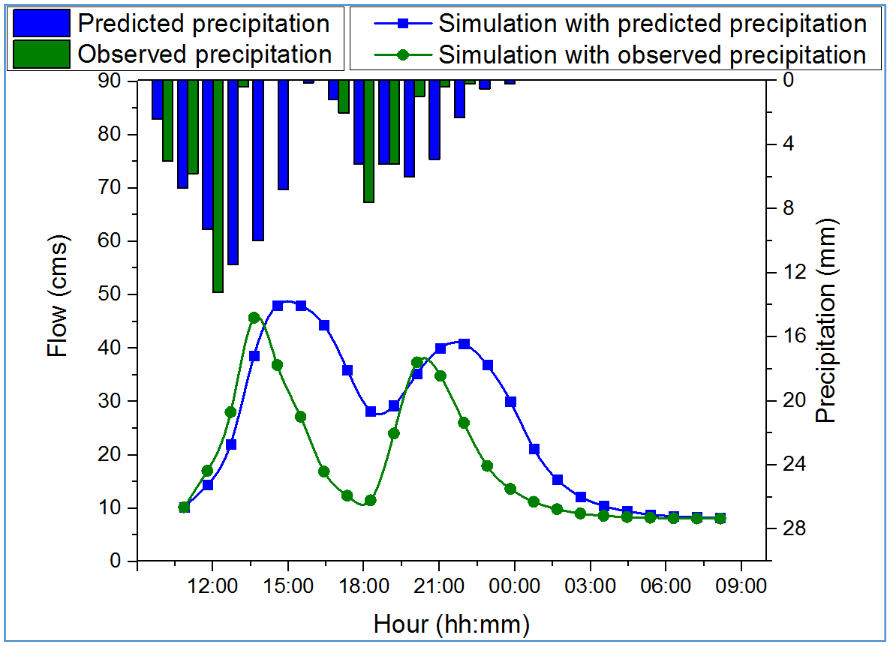

Figure 6: Flow forecasting for node N-2.

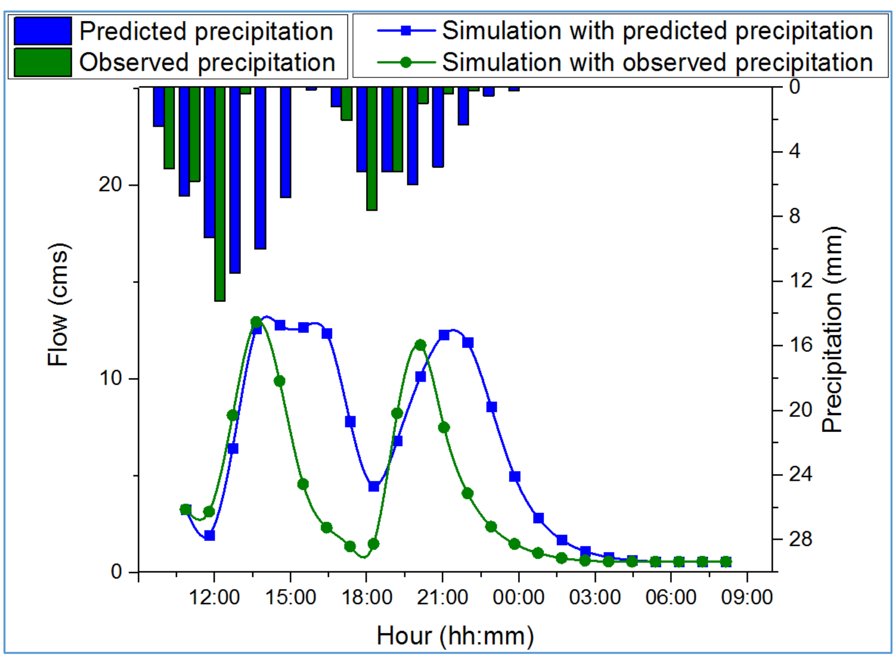

Figura 7: Flow forecasting for node N-3.

After the previous step, we perform three simulations to compare the efficiency of the new methodology using the predicted precipitation. The blue line represents the observed water level in the point N-2. The others three lines are the simulated levels incorporating WSN data (red line), WSN + VGI data (orange line) and without changes in the model (green line). The three simulations are very similar until 0 am, then the green line underestimate the water level for subsequent times while the other two methods presents estimated values above the observed data. The forecasting using WSN and VGI was updated with VGI data only in three instants (12pm, 4h30pm and 8pm) using volunteer information from upstream of the point N-2. The forecasting in this instants were slightly better than using only WSN information, however, far from the water level peak. As the VGI data was informed upstream from the N-2, the wave simulation was damped by the mathematical model, thus, correlating all available spatial data can improve the results in all subcatchments.

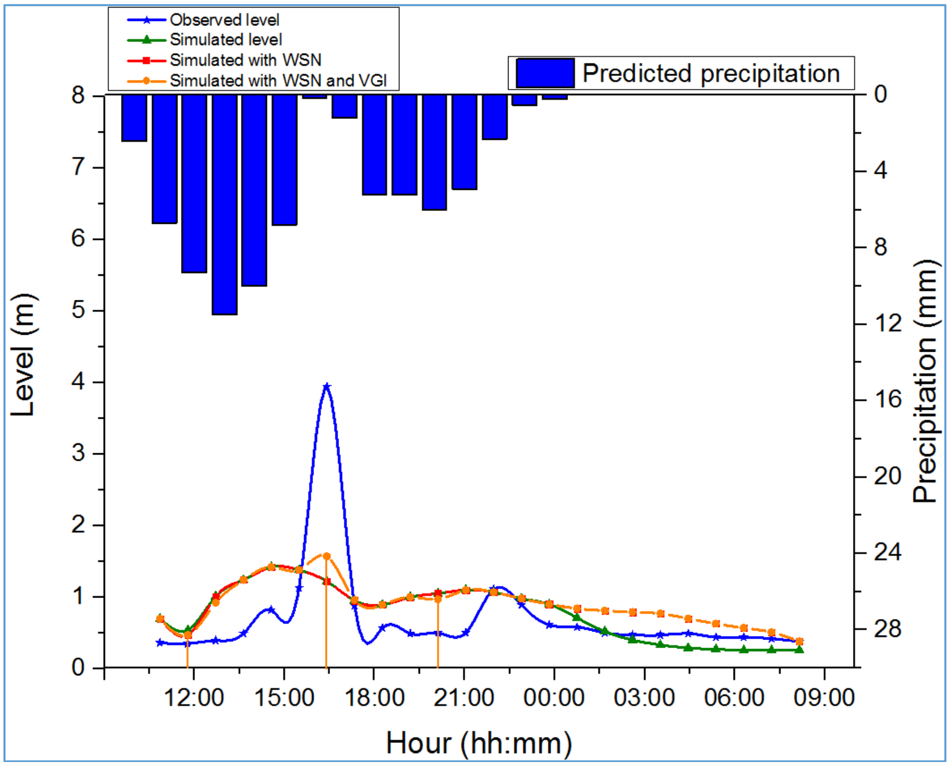

Figure 8: Water level forecasting comparison.

4. CONCLUSIONS

This paper presented a new proposal for hydrometeorological forecasting named HAMPB. This methodology integrates the voluntary prediction models with short-term information, presenting a solution to fill the data spatially in places where there are no monitoring sensors.

The use of data from VGI has as main advantages the improvement and real-time update of the forecasts monitored by WSN sensors and the estimation of levels for points not monitored. The volunteer information's should be treated carefully because they involve great uncertainty. The use of the two technologies has advantages such as for providing a larger amount of information available on the basin for simulations and calibration. Furthermore, the use of VGI technology allow model runs with more accurate data, producing more realistic answers to a point on the basin.

The implementation of new visualization techniques is essential in this study due to the large amount of data obtained on the precipitation forecast module, by monitoring the sensors level, voluntary level information and also the correlations made between the data for each produced forecast horizon.

This research shows the need for more studies in the area and experimental determination of amounts to be integrated into the probabilistic relationships proposed in order to verify the accuracy of the prediction using the proposed model.

The study is under development, improvements on the HAMPB model will be performed, therefore the main contribution of this paper is the proposal of a new methodology and the partial results obtained so far.

The predicted flow is highly dependent on rainfall data. Therefore, improvements on rainfall prediction using radar datasets can significantly improve flow forecasts. Furthermore, analyzing the calibration it was found that the model still requires adjustments in order for the simulations correctly represent the behavior of the basin, reducing the error between the time of concentration and duration of rain. However, it is possible to verify that the proposed model presents theoretical consistency and is a promising method for basins with sparse data.

5. ACKNOWLEDGEMENTS

This research is part of a cooperative research network composite by:

USP/Nap/CEPED-EESC named "Ações colaborativas na redução de vulnerabilidade para desastres hidrológicos usando tecnologias resilientes na drenagem urbana", FAPESP-2008/58161-1 "Assessment of impacts and vulnerability to climate change in Brazil and strategics for adaptation options", FINEP 01.10.0701.00 "MAPLU-Manejo de Água Pluvial Urbana " and CNPq-CEMADEN-USP "Desenvolvimento de sistema de previsão de enxurradas e enchentes para prevenção de desastres naturais".

The authors would like to thanks the grants from CAPES, FAPESP, CNPq PQ 307637/2012-3, BMBF/IB-DLR-CNPq 490148/2011-5, FAPESP 2008/58151-1, FINEP/MAPLU 01.10.0701.00, CNPq-CAPES 552494-2011-9 (Casadinho UFAL+EESC/USP).

6. REFERENCES

Alvisi, S., Bernini, A. and Franchini, M., 2013: "A conceptual grey rainfall-runoff model for simulation with uncertainty." *Journal of Hydroinformatics* 15, no. 1: 1-20. Academic Search Elite, EBSCOhost (accessed April 29, 2014).

Andrade, J. P. M., 2006: *Previsão hidrometeorológica visando sistemas de alerta antecipado de cheias em bacias urbanas*. Dissertação (Mestrado) — Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos.

Austin, G. L. and Bellon, A., 1974: “The use of digital weather radar records for short-term precipitation forecasting”. *Quarterly Journal of the Royal Meteorological Society*, *100*(426), 658-664.

Degrossi, L. C. *et al*., 2014: “Flood citizen observatory: a crowdsourcing-based approach for flood risk management in Brazil”. In: 26th International Conference on Software Engineering and Knowledge Engineering (SEKE). Vancouver, Canada.

Fava, M.C. *et al.,* 2013: “Proposta metodológica para previsões de enchentes com uso de sistemas colaborativos.” In: *XX* Simpósio Brasileiro de Recursos Hídricos. Bento Gonçalves, RS.

Georgakakos, K. P. and Bras, R. L., 1984: “A hydrologically useful station precipitation model: 1. Formulation”. *Water Resources Research*, Wiley Online Library, v. 20, n. 11, p. 1585–1596.

Goodchild, M. F., 2007: “Citizens as sensors: the world of volunteered geography.” *GeoJournal, Springer*, v. 69, n. 4, p. 211–221.

Gonçalves, M. F., 2009: *Previsão de chuva com auxílio de radar de tempo visando a um sistema de alerta antecipado de cheias em áreas urbanas*. Dissertação (Mestrado) — Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos.

Hughes, D. A., 2013: "A review of 40 years of hydrological science and practice in southern Africa using the Pitman rainfall-runoff model." *Journal of Hydrology* 501: 111-124.

Hughes, D. A. *et al*., 2011: “A middleware platform to support river monitoring using wireless sensor networks.” *Journal of the Brazilian Computer Society*, Springer, v. 17, n. 2, p. 85–102.

Löwe, R. *et al.,* 2014: "Probabilistic online runoff forecasting for urban catchments using inputs from rain gauges as well as statically and dynamically adjusted weather radar." *Journal of Hydrology* 512: 397-407.

Mendiondo, E. M. *et al*., 2013: “Ações colaborativas na redução da vulnerabilidade aos desastres hidrológicos usando tecnologias resilientes para a drenagem urbana.” Eixo Temático, In: Yoshizaki, H. (coord.) NAP/CEPED-USP: Núcleo de Apoio à Pesquisa "Centro de Ensino e Pesquisa em Desastres", Pró-Reitoria de Pesquisa/USP, 2013.

Prefeitura Municipal de São Carlos, 2005: Plano Diretor Municipal, Leis e Anexos: Disponível em: http://www.saocarlos.sp.gov.br/index.php/utilidade-publica/plano-diretor.html

Poser, K. and Dransch, D., 2010: “Volunteered geographic information for disaster management with application to rapid flood damage estimation.” *Geomatica*, v. 64, n. 1, p. 89–98.

Rossman, L. A., 2010: *Storm water management model user’s manual, version 5.0.* [S.l.]: National Risk Management Research Laboratory, Office of Research and Development, US Environmental Protection Agency.

Rotava, J., Mendiondo, E. M. and Souza, V. C., 2013: “Simulação de instabilidade humana em inundações: primeiras considerações.” In: *XX* Simpósio Brasileiro de Recursos Hídricos. Bento Gonçalves, RS. 2013.

Shinma, T. A. and Reis, L. F. R., 2014: “Incorporating Multi-event and Multi-site Data in the Calibration of SWMM”*. Procedia Engineering*, Volume 70, 2014, Pages 75-84, ISSN 1877-7058, http://dx.doi.org/10.1016/j.proeng.2014.02.010.