AGNPS and SWAT Model Calibration for Hydrologic Modelling of an Ecuadorian River Basin under Data Scarcity

D. Matamoros*, E. Guzman*, J. Bonini* and P.A. Vanrolleghem**

* Facultad de Ingenieria Maritima, Escuela Superior Politecnica del Litoral, P.O Box 09-01-5863, Guayaquil, Ecuador. (E-mail: dmata@goliat.espol.edu.ec)
** Department of Applied Mathematics, Biometrics and Process Control. Ghent University, Coupure Links 653, B-9000 Gent, Belgium. (E-mail: Peter.Vanrolleghem@ugent.be)

Abstract Chemical fate assessment in a river basin requires flow and sediment characterization that can be performed by continuous field measurements or modelling. Once this step is completed, the environmental modelling can be started. In current assessments done in Ecuador, field data are not easily available and are also scarce. Therefore, the use of models is a challenge. Two models, AGNPS and SWAT, were tested to evaluate their potential applicability under data scarcity. After gathering all necessary data, the models were calibrated from a hydrological point of view. Some inverse modelling techniques were applied to overcome the lack of necessary data. This paper shows the results of the hydrological calibration of both models.

Keywords AGNPS, Model Calibration, SWAT, Watershed Assessment.

Problem Definition

The Flemish Interuniversitary Council (Belgium) and the Guayaquil Polytechnics School (Ecuador) are currently evaluating the potential environmental impacts that could occur in a river basin with intensive banana production. Two models able to simulate the impact of pesticide usage in a river basin were compared: AGNPS and SWAT. The selected watershed (Chaguana river) is located at the most south western part of Ecuador (Matamoros et. al, 2002). Before model usage, all necessary data were gathered, processed and converted in the appropriate format. Among the data collected, precipitation and flow information were the main concern in view of the hydrological calibration of both models.

The flow data was collected from three gauging stations (Chaguana and Zapote). The flow measurements represent average monthly flows and cover a period of 4 years (1979, 1980, 1982 and 1983). There is no available measurement after 1983. The location of gauging stations is depicted in Figure 1.1.

Although there was information from 8 surrounding weather stations, reliable precipitation data were only obtained from four of them (Machala, Pagua and Pasaje stations). Figure 1.1 also shows the location of weather stations. Precipitation data represent total monthly values and also the maximum 24-h precipitation in every month. The recording period is from 1979 to 1999.

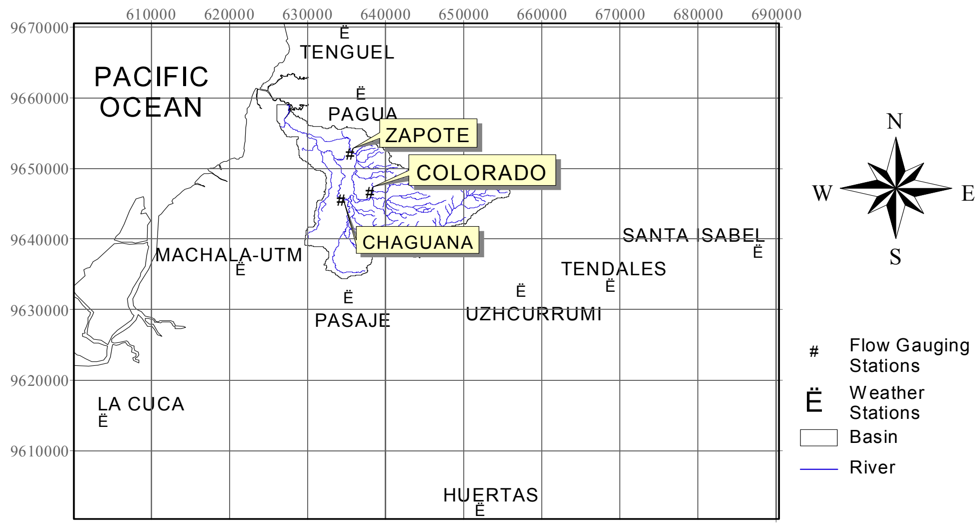

Figure 1.1 Location of weather stations and flow gauging stations for Chaguana basin.

The official database, where the flow data were obtained, did not contain measurements of suspended sediments to perform the model calibration. Therefore, thirteen monitoring points, including the place of gauging stations, were set along the river to determine suspended sediments and flow rates during four sampling campaigns between 2001 and 2002 (14 November 2001, 30 March 2002, 5 July 2002 and 11 November 2002). Due to accessibility issues, only 46 km out of 320 km of the river system were sampled. Each sampling campaign represents a single event for modelling purposes. Figure 1.2 gives the location of the sampling points along the Chaguana river basin.

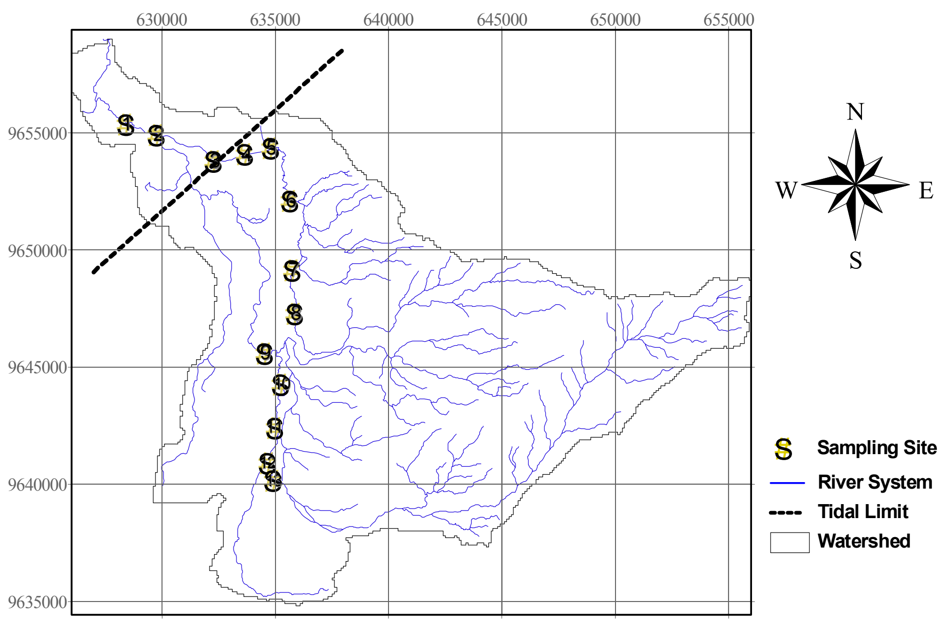

Figure 1.2 Location of sampling points along the Chaguana river basin.

Methodology

Basin Division

The research team that evaluated the AGNPS model divided the Chaguana river basin in 192 AGNPS cells which drain into 78 river reaches (see figure 1.3.a). An explanation of the meaning of an AGNPS cell can be found in the AGNPS User's manual and in Matamoros et al (2004). The resulting cell areas vary between 1 and 829 Ha.

The second research team (Bonini and Guzman, 2003) evaluated the SWAT model by dividing the basin in 44 sub-basins, mainly based on the locations of the three existing gauging stations (Colorado, Zapote and Chaguana gauging stations), as shown in figure 1.3.b. The SWAT division is based on the concept of Hydrological Response Units (HRUs) applied to each sub-basin.

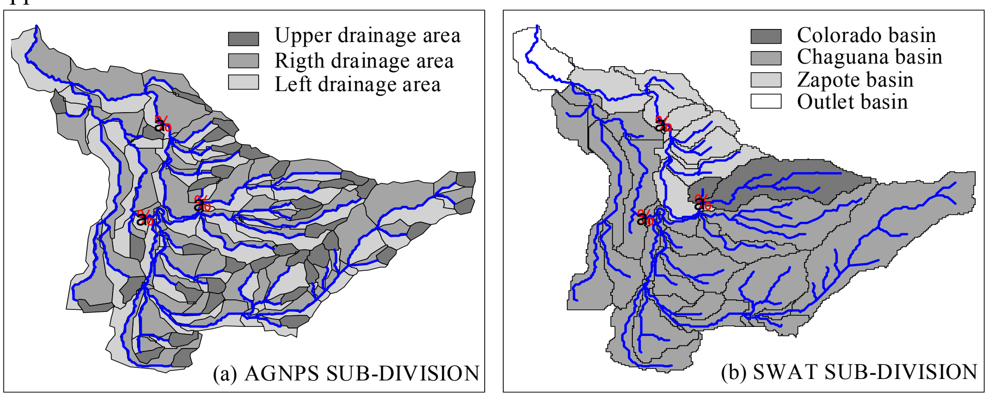

Figure 1.3 Chaguana basin division based on the applied model.

Fit Criteria

The whole calibration procedure was done with data collected from the three available gauging stations, and sediment samples gathered during the four sampling campaigns. Three statistical parameters were used to determine the goodness of fit of the predicted values related to the measured values:

- The Coefficient of Determination (r^2) is the square of the Pearson's Product Moment Correlation Coefficient, and it varies from 0.0 (poor model) to 1.0 (good model).

$$r^2 = \frac{\sum_{i=1}^{N}(O_i - O_{avg})(P_i - P_{avg})}{\left(\sqrt{\sum_{i=1}^{N}(O_i - O_{avg})^2}\right)\left(\sqrt{\sum_{i=1}^{N}(P_i - P_{avg})^2}\right)}$$

Where

O i , P i — Observed and Predicted value for each modelled event

O avg , P avg — Observed and Predicted average value for the evaluated range of data

- The Coefficient of Efficiency (E), developed by Nash and Sutcliffe (1970), ranges from minus infinity (poor model) to 1.0 (good model).

$$E = \frac{\left(\sum_{i=1}^{N} (O_i - O_{avg})^2 \right) - \left(\sum_{i=1}^{N} (O_i - P_i)^2 \right)}{\left(\sum_{i=1}^{N} (O_i - O_{avg})^2 \right)}$$

- The Index of Agreement (*d*) developed by Willmott (1981) presents the same range of values as the coefficient of determination.

$$d = \frac{\left(\sum_{i=1}^{N} \left(\left| P_i - O_{avg} \right| + \left| O_i - O_{avg} \right| \right)^2 \right) - \left(\sum_{i=1}^{N} (O_i - P_i)^2 \right)}{\left(\sum_{i=1}^{N} \left(\left| P_i - O_{avg} \right| + \left| O_i - O_{avg} \right| \right)^2 \right)}$$

For further information, Legates and McCabe (1999) have written a complete discussion on these three statistical coefficients normally used in hydrological and climatic model evaluations. In addition, a relative bias was estimated for every pair of measured and predicted values. Then, an average was estimated for all the sampled values based on the following equation.

$$Bias_{AVERAGE} = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{O_i - P_i}{O_i} \times 100 \right)$$

Flow Calibration

The first calibration step adjusted the flow predictions as close as possible to the flow measurements. As written before, three gauging stations, known as Chaguana, Zapote and Colorado gauging stations, were used. They are mainly located in the middle course of the river basin.

The models evaluated by the research teams are mainly runoff-based. One of the most sensitive parameters for flow calibration in runoff-based models is the SCS Runoff Curve Number, which is an indicator of how much water is running off from the soil surface. The higher the Curve Number, the higher the estimated runoff that is obtained. Therefore, flow calibration was conducted by adjusting the Curve Number for each land use type involved in the basin assessment.

In the models, flow estimates are based on single or continuous daily events (rainfall), so it is necessary to have daily data to calibrate the model. In the present research, the gauging stations only had average monthly flows. In addition, weather data were also limited to total monthly values, the number of rain events and the maximum 24-hour precipitation fallen during every month. Therefore, it was necessary to generate daily precipitation data for the recorded period of the gauging stations by considering the following assumptions:

1. There is only one maximum precipitation event in every month corresponding to the recorded 24-hour precipitation at a specific reported day.
2. As a first approach, the rest of the monthly precipitation is equally distributed among the recorded number of rainy days in a month.
3. The model is run for every estimated daily event for that month. A mean monthly flow is obtained by averaging the resulting daily flows. That average monthly flow is compared with the reported flow in the corresponding gauging station.
4. If the statistical parameters are still showing "poor" fit, then the daily events in the month are rearranged by always keeping the maximum monthly precipitation and the number of rainy days in mind. This process is repeated until the flow predictions fit the measured values.

Suspended Sediment Calibration

Both models predict suspended sediments based on precipitation events (single or continuous). However, there was no information regarding the precipitation that fell during the sampling campaigns. Therefore, it was necessary to use the flow prediction to estimate the unknown inputs (rain events) which are producing the sediment outputs for calibration purposes. This method is known as Inverse Modelling, and it has been used in several applications. Basically, Inverse Modelling is the use of a model output to estimate a model input.

The models were only run in the river reaches that showed measurements of suspended sediments during one of the sampling campaigns. The sediment calibration was performed by adjusting the two parameters that mainly contribute to the sediment yield and do have more uncertainty in their estimated values: the cover management factor (C) and the practice management factor (P).

Results and Discussion

Flow Calibration

Basically, the Colorado gauging station was calibrated first because it has a smaller drainage area and only two land cover types. The Zapote gauging station was calibrated second because it is located immediately downstream the Colorado station. And finally, the Chaguana station was calibrated by adjusting the curve number for each of its drainage basins. Unfortunately, there was no gauging station at the outlet of the basin. Table 1.1 shows the estimated statistical coefficients of fit in all three gauging stations for both model runs.

Table 1.1 Coefficients of fit during flow calibration for both models in the three gauging stations.

Gauging Station	Model	r^2	E	d	Bias (%)
Colorado	AGNPS	0.83	0.37	0.88	90.31
	SWAT	0.85	0.72	0.90	10.00
Zapote	AGNPS	0.85	0.53	0.91	26.79
	SWAT	0.88	0.77	0.94	-23.76
Chaguana	AGNPS	0.87	0.73	0.93	0.83
	SWAT	0.90	0.80	0.95	-8.79

The AGNPS model could predict the flows in the Chaguana gauging station fairly well (r^2 = 0.87, E = 0.73, and d = 0.93). For the Colorado and Zapote stations, the model showed lower values of goodness of fit (E = 0.53 for Zapote and 0.37 for Colorado). It is concluded that the AGNPS model usually fails to predict flows that occur in very small drainage areas with very low precipitation events, and this is mainly because the output results are restricted to three decimal place positions. Therefore, any predicted flow below 0.001 m^3/s (1 litre per second) is reported as zero. In addition, the lack of data for flow validation is critical; the data mainly represent extreme events (an “El Niño” event occurred during 1982 and 1983).

The outcome of the SWAT model was more accurate than the AGNPS model. The efficiency (E) of the flow calibration process was above 0.7 for all gauging stations. Although predictions for the Zapote and Colorado stations were improved, they are still below the ones for the Chaguana station.

The main reason for the result improvement is that the SWAT model uses the concept of Hydrologic Response Units (HRU) to couple land cover and soil information within each sub-basin. As described in the SWAT user's manual, Hydrologic Response Units are portions of a sub-basin that possess unique landuse, management and soil attributes. Although this concept is similar to the attribute of an AGNPS cell, the main difference is that a SWAT sub-basin can have many HRUs within it. On the other hand, AGNPS sub-basins can only contain a maximum of three AGNPS cells. Therefore, the information loss in SWAT can be reduced significantly when aggregating data in a sub-basin. This improvement results in better predictions because of a better characterisation of the evaluated basin. However, the improvement on the accuracy could be jeopardised if soil or land cover information is not accurate too.

Suspended Sediment Calibration

To perform the suspended sediment calibration, it was necessary to use flow predictions to assess the unknown inputs (rain events) which are needed to calculate the suspended sediment yield. This method is known as Inverse Modelling, and it has been used in several applications. Basically, Inverse Modelling is the use of a model output to estimate a model input.

In the present case study, the models required the precipitation on the campaign dates to predict the sediment yield. Due to the lack of precipitation data for the sampling days, it was necessary to use the inverse modelling technique. First, a flow graph was obtained from the flow calibration step (figure 1.4) for three locations in the river: the outlet, the Chaguana station and the Zapote station. The estimated rain events for the sampling days were obtained by introducing the monitored flow values in figure 1.4. The obtained rain event intensities represent an average rain event as falling at the same time in the entire catchment area. Table 1.2 gives the estimated rain event for the four sampling campaigns, which are the values used in the model simulation.

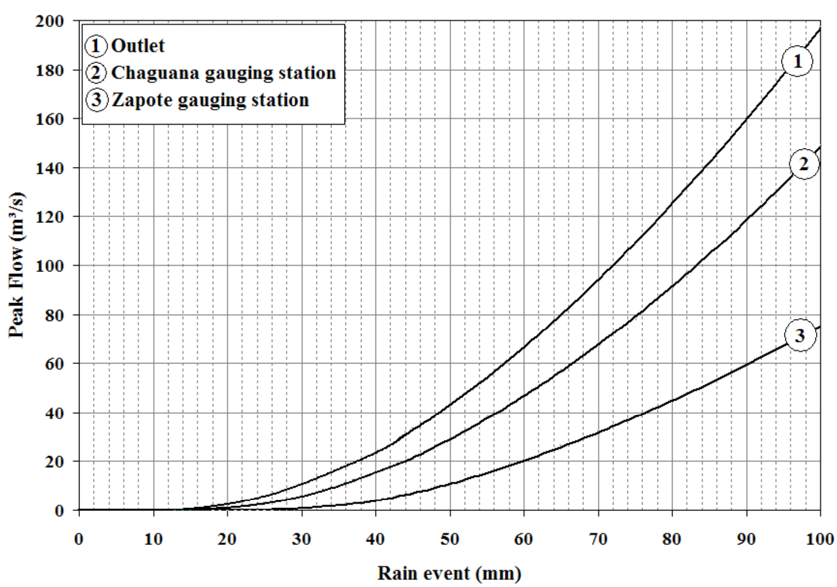

Figure 1.4 Estimated flows based on rain events for three locations in the river system.

Table 1.2 Estimated daily rain event for the sampling days based on the flow calibration graph.

Campaign	Date	Estimated Daily Rain Event
First	14 November 2001	4 mm
Second	30 March 2002	58 mm
Third	5 July 2002	3 mm
Fourth	11 November 2002	5 mm

The suspended sediment calibration procedure was explained in the methodology section. Basically, the cover management factor (C) and the practice management factor (P) were adjusted to fit the predicted values around the observed values. Table 1.3 gives the statistical coefficients of fit in the sampled reaches of the basin for both model runs.

Table 1.3 Coefficients of fit during suspended sediment calibration for both models in the main rivers of the Chaguana system.

Gauging Station	Model	r^2	E	d	Bias (%)
Zapote	AGNPS	0.60	-0.08	0.70	-30.37
	SWAT	0.96	0.86	0.96	-61.89
Chaguana	AGNPS	0.94	0.88	0.97	-3.64
	SWAT	0.97	0.92	0.98	-4.71

Regarding the AGNPS simulation, it is clear that the predictions for the Zapote River are not good enough as the coefficient of efficiency (E) is around zero. However the model can predict the sediment behaviour in the Chaguana river fairly well. There could be many reasons for this difference:

- The Zapote river has less reaches sampled than the Chaguana river during the sampling campaigns. Thus the characterization of this river is quite low.
- As said in the flow calibration, the model has a lower prediction efficiency in cells with small drainage areas. The Zapote river does not have as long a course (drainage area) as the Chaguana river, so the predictions are affected by this difference.
- Perhaps, the most important reason is the information loss that occurred during the spatial data aggregation done by the AGNPS model.

The SWAT model gave a better prediction in both sampled rivers (coefficient of efficiency above 0.8). This behaviour was expected as explained in the flow calibration process because the SWAT model minimizes the information loss by using multiple HRU's in each sub-basin.

Finally, for both models, there were also significant differences between predicted and measured values at specific sampling locations in the Chaguana river for the March sampling period (rainy season) while the rest of the sampling dates showed a good agreement. However, there was some dredging activity at certain locations along the Chaguana river between Reaches 45 and 22. This civil work affected the sediment yield by increasing the suspended solid concentrations on those reaches. For that reason, both models failed to predict the suspended sediment concentrations as the models calculate sediment yield based mainly on soil erosion from runoff. In addition, the last reach, Number 2, still shows predicted values lower than the actual ones. That is because that reach was also influenced by the tidal push at the moment of sampling. When the tide is entering the basin, some sediment is pushed back into the basin. This additional sediment load cannot be predicted by the runoff models. Therefore, both models are no longer applicable for those situations.

Conclusions

There were not enough data available for use in the modelling exercise. Thus, it was necessary to apply inverse modelling techniques to obtain certain unknown inputs. The use of inverse modelling represents a useful approach to overcome problems regarding input data. However, this technique also represents a risk because the outcome may not represent the phenomena to be modelled. For that reason, it is always better to use real data in the modelling process.

In the case study, the lack of daily rainfall data to run the model for the sampling events represented an opportunity to apply inverse modelling techniques. The obtained outputs were within the expected range.

The predicted flows were within acceptable ranges for the Chaguana gauging station, while predictions for the Zapote station are still not good. A potential cause can be that the Zapote river is not as well characterised as the Chaguana river.

Sediment predictions also show a reasonable agreement. However, there are two situations where the models cannot be applied: sediment supplied by civil works conducted in river banks and the sediment push-back caused by tidal influence.

Regarding model performance, the SWAT model with less basin subdivision showed to be more accurate than the AGNPS model. This is explained by the use of *Hydrological Response Units* (HRUs) in the SWAT model, which causes less information loss when doing data aggregation in the processing step.

Acknowledgement

This work has been supported and funded by ESPOL-VLIR project (Component 4). The authors would also like to thank the financial support of the EU Harmoni-CA project EVK1-CT-2002-20003.

References

Bonini, J. and Guzman, E. (2003) Flow and Sediment Estimation in Chaguana River Basin by using SWAT model. B.Sc. thesis, Escuela Superior Politecnica del Litoral, Ecuador (*in Spanish*).

Legates, D.R. and McCabe, G.J. (1999) Evaluating the use of "Goodness of Fit" measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, **35**, 233-241.

Matamoros, D. (2004). Predicting river concentrations of pesticides from banana plantations under data-poor conditions. Ph.D. thesis, Ghent University, Belgium.

Matamoros, D., Bonini, J., Guzman, E., Ramirez, G., and Vanrolleghem, P.A. (2002). Design and implementation of a measuring campaign to model pesticide impacts in an Ecuadorian watershed. *Wat. Sci. Tech.*, **46**(9), 37-45.

Matamoros, D., Van Griesven, A., Van Biesen, L. and Vanrolleghem, P.A. (2004) Development of a Geographical Information System for Pesticide Assessment on an Ecuadorian Watershed. In *Proc. of the IWA World Water Congress and Exhibition*, Marrakech, Morocco 19–24 September (*in press*).