CHAPTER 4. PREDICTIVE MODELLING OF PESTICIDE FATE IN RIVERS

Tendales Town located at the lower course of the Chaguana river, Ecuador

Parts of this chapter were published in

Matamoros D. and Vanrolleghem P.A. (2001) Pesticide assessment of the banana sector in an Ecuadorian watershed. In: Proceedings 53rd International Symposium on Crop Protection. Gent, Belgium, May 8 2001. Med. Fac. Landbouww. Univ. Gent, 66(2b), 863-872.

Matamoros D., Guzmán E., Bonini J. and Vanrolleghem P.A. (*Submitted*). AGNPS and SWAT Model Calibration for Hydrologic Modelling of an Ecuadorian River Basin under Data Scarcity. *Submitted to:* Water Science and Technology.

Matamoros D., Van Biesen L. and Vanrolleghem P.A. (*Submitted*). Development of a Geographical Information System for pesticide assessment on an Ecuadorian watershed. *Submitted to:* Water Science and Technology.

4.1. INTRODUCTION

The first step in the study reported in this chapter was to estimate a gross distribution of the environmental concentrations caused by the pesticides most used in the Ecuadorian banana sector. Two screening models were evaluated to get an overall environmental distribution of the evaluated pesticides. This gross distribution helped to take decisions regarding the sampling campaigns and the use of more complex models to determine the impacts in the surrounding environment. The screening models used were:

- Environmental Quality Criteria Model - EQC, version 2.02 (after Mackay *et al.* 1997)
- Exposure Assessment Modelling System - EXAMS, version 2.98.01 (Burns 2000)

Chapter 2 contained a brief description of both models, explaining the characteristics and requirements of each model. However, it will be necessary to expand a little bit further regarding other aspects to assess the Chaguana basin. After estimating the gross environmental distribution of pesticides with screening models, the research pointed its analysis towards more complex models such as the integrated models that use GIS techniques to handle input data. Within the VLIR-ESPOL Project, the main goal was to establish the potential environmental impacts that could occur in a river basin with intensive banana production. In the present research, two models able to simulate the impact of pesticide usage in a basin were applied: AGNPS and SWAT. They use the same type of data whose accuracy varies depending on user's choice.

Regarding temporal resolution, both models can work with daily and monthly information. On the other hand, spatial resolution depends on the way the basin is divided. Basically, these models are semi-distributed and lumped models where all data are aggregated into each unitary subdivision of the basin.

In the current research, two independent studies were conducted on the same watershed. This Ph.D. thesis mainly focused on the use of the AGNPS model by dividing the watershed in 192 AGNPS cells on the basis of sub-basin divisions (left-, right- and upstream drainage areas). Another team, under the author's guidance, used the SWAT model and divided the watershed in 44 sub-basins with their corresponding Hydrological Response Units (HRU) on the basis of land use and soil data.

4.2. SCREENING MODELS AS TOOLS FOR PLANNING SAMPLING CAMPAINGS

4.2.1. MODELS STRUCTURE

Although the EXAMS and EQC models consider the environment as compartments, they actually do not perform calculation processes in the same way, i.e. EXAMS runs only in an aquatic system and EQC is used in a multi-compartment environment. Table 4.1 shows a summary of differences and similarities between the models. However, both models can lead to similar results when only the aquatic portion of the environment is considered as shown in the next sections.

Table 4.1. Comparison between EXAMS and EQC characteristics

Characteristic	*EQC*	*EXAMS*
Computer Environment	Windows interface	DOS
Type of Model	Deterministic (steady-state)	Deterministic (steady-state)
Compartment layout	Air, soil, water and sediment (Unique compartments)	Water and sediment (multiple linked compartments)
Compartment geometry	Constant	Can be changed
Chemical data	• All asked data is necessary to run the model • Only one chemical at a time	• The model can estimate some lacking data • Until 5 chemicals + 6 ion species for each chemical
Transport Processes	Advection	Advection + Dispersion
Environmental data	Not site-specific at all	Site-specific

In addition, the way the environment is subdivided in compartments differs from one model to another. While EQC uses one big unitary world with unique compartments representing every component of the environment, EXAMS can use more than one compartment representing the same environmental component (water or sediment). Figures 4.1 and 4.2 shows how the environment is subdivided in both models.

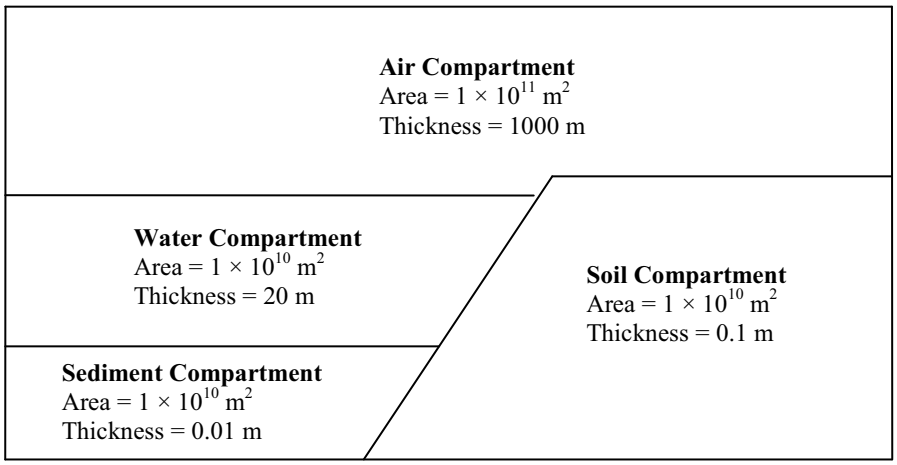

Figure 4.1. Environmental division performed by the EQC model

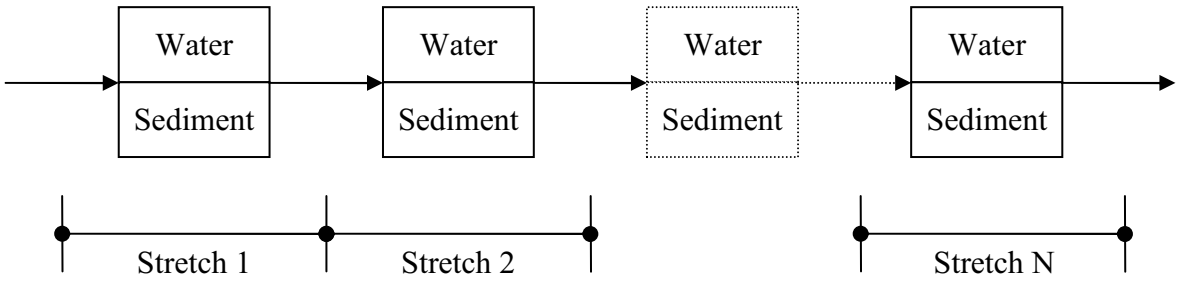

Figure 4.2. River subdivision performed by the EXAMS model

4.2.2. ANALYSIS OF RESULTS

Based on the information collected in the previous chapters, imazalil, thiabendazole and propiconazole can be considered to represent the most used pesticides in the Ecuadorian banana sector. The chemical data needed for running the models were obtained based on the average of several data found in the literature (Linders *et al.* 1994; Tomlin 1997). Table 4.2 shows the values used in the assessment for each chemical. Figures 4.3 and 4.4 show the results as charts obtained from the EQC and EXAMS runs respectively.

Table 4.2. Physico-chemical data for evaluated pesticides

Pesticide	*Imazalil*	*Thiabendazole*	*Propiconazole*
CAS Number	35554 – 44 – 0	148 – 79 – 8	60207 – 90 – 1
Molecular Mass (g/mol)	297.2 (Tomlin 1997)	201.3 (Tomlin 1997)	342.2 (Tomlin 1997)
Solubility (mg/l)	180 (Tomlin 1997)	< 50 (Montgomery 1993)	100 (Tomlin 1997)
Vapour Pressure (Pa)	1.6×10^{-4} (Tomlin 1997)	2.7×10^{-8} (Tomlin, 1997)	5.6×10^{-5} (Tomlin 1997)
Melting Point (ºC)	52.7 (Tomlin 1997)	297 (Tomlin 1997)	< $T_{ambient}$ (Tomlin 1997)
Log K_{OW}	3.82 (Tomlin 1997)	2.69 (Montgomery 1993)	3.72 (Tomlin 1997)
K_{OC} (l/kg)	2081 – 6918 (Van Leemput *et al.* 1986)	512 at pH 5 – 12 (Montgomery 1993)	650 – 720 (Tomlin 1997)
Half-life in air (h)	2400 (Tomlin, 1997)	-	288 (Tomlin 1997)
Half-life in water (h)	1350 (Tomlin 1997)	Stable in aqu. suspens. (Tomlin 1997)	600 – 2040 (Linders *et al.* 1994)
Half-life in soil (h)	2880 – 4560 (Van Leemput *et al.* 1984)	792 – 9672 (Tomlin 1997; Wauchope *et al.* 1992)	2304 – 5496 (Tomlin 1997)
Half-life in sediment (h)	3240 (Tomlin, 1997)	-	5064 (Tomlin, 1997)
Photolysis rate (h^{-1})	1.9×10^{-2} (Van Leemput *et al.* 1988)	Stable to light (Tomlin 1997)	7.77×10^{-3} (Tomlin 1997)

To perform the comparison between both models, some assumptions had to be made accounting for model differences:

1. ***EQC Assumptions:*** The assessment was conducted in the Level III mode (explained in Chapter 2) with a hypothetical loading rate of 1000 kg/h applied to specific compartments depending on the way pesticide is handled on the farm. For imazalil and thiabendazole, the loading was applied to the water compartment. The Propiconazole loading was applied to the air compartment. There is no possibility to change the dimensions of the unique environmental compartments.
2. ***EXAMS Assumptions:*** The assessment was conducted in Mode 2 (explained in Chapter 2) with the same hypothetical loading rate as the EQC model. Due to the different subdivision process, the loading can be applied only to the water and sediment compartments. For comparison purposes, the water compartment was selected as the one receiving the loading. EXAMS also differentiates the type of loads entering the compartment. For imazalil and thiabendazole, a stream load type was selected (direct discharge). On the other hand propiconazole represents a drift load type (aero-fumigation). For comparison purposes, only one water and sediment compartments were defined with the same dimensions as the EQC model.

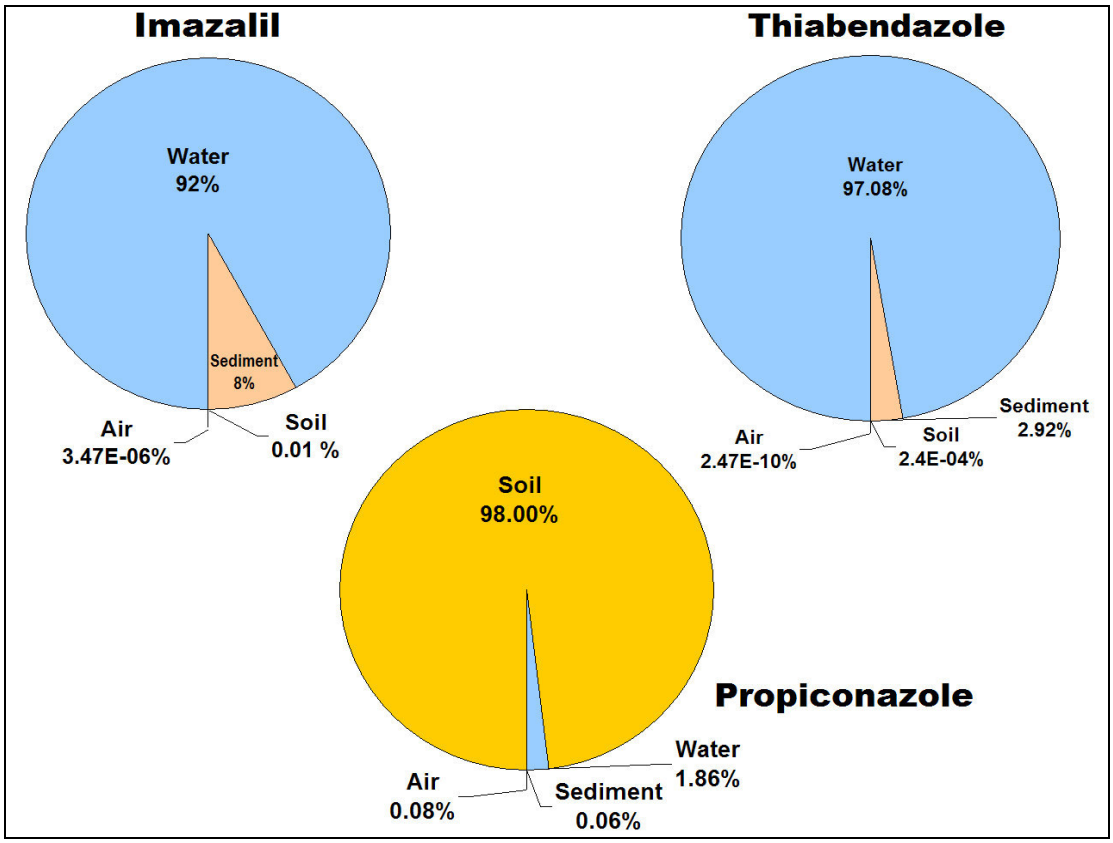

Figure 4.3. Results from EQC running

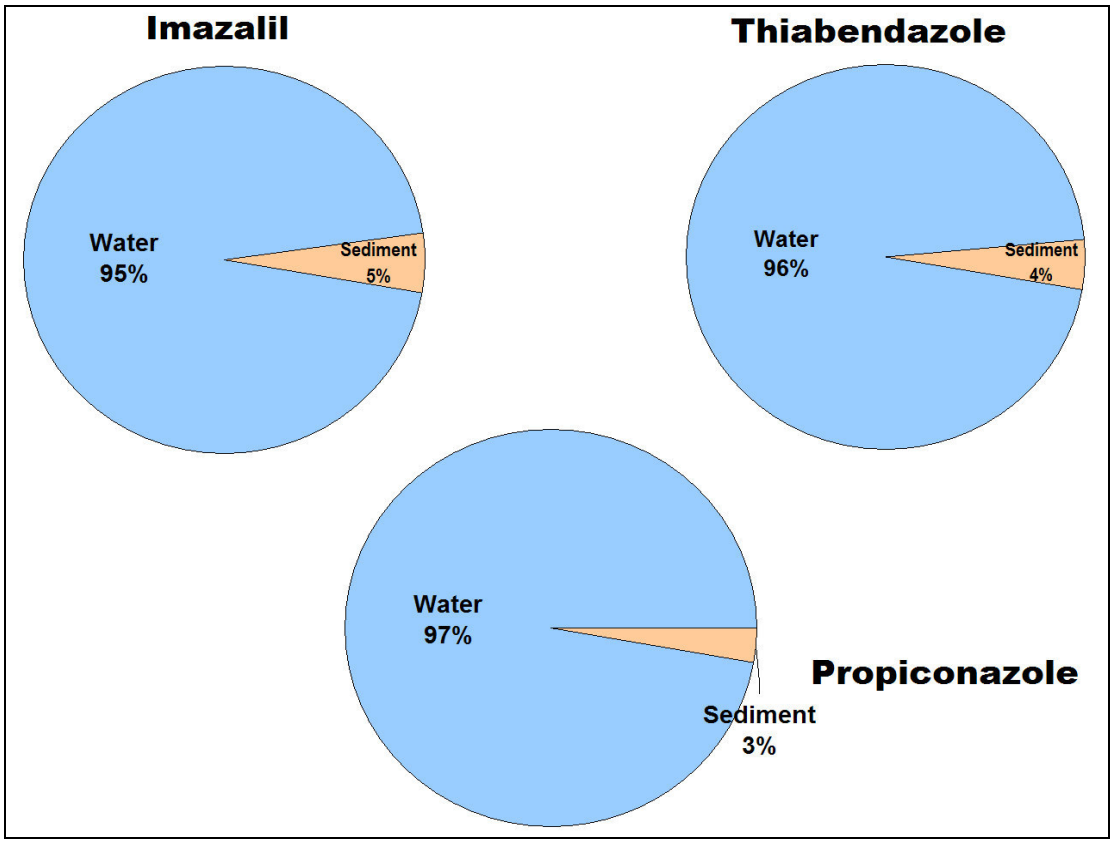

Figure 4.4. Results from EXAMS running

Based on the EQC results, propiconazole tends to distribute more in the soil compartment (around 98% of applied pesticide). Imazalil and Thiabendazole tend to distribute more likely in the water compartment (around 95%). On the other hand, EXAMS results show a significant distribution of these 3 pesticides in the water compartment. Propiconazole, Thiabendazole and Imazalil present a similar distribution between water (around 95%) and sediment (around 5%).

Due to the different approaches considered in both models, the overall results cannot be compared directly. EXAMS only performs the assessment in the aquatic portion of the environment (sediment + water). On the other hand, EQC analyses all environmental compartments (air, sediment, water and soil). Therefore, only results from the aquatic portion can be compared between both models. For example, consider the results obtained for propiconazole (figure 4.3 and 4.4):

- The EQC's aquatic (sediment – water) distribution would be 1.923% (0.064% for sediment and 1.859% for water) or 1.923 kg of propiconazole per 100 kg of aquatic mass (sediment + water).

- This total amount of propiconazole entering the aquatic portion could be divided according to the EXAMS results (3% for sediment and 97% for water), and a corrected value for the EXAMS results can be obtained.

$\text{Sediment}_{\text{EXAMS corrected}}$	=	(1.923 kg / 100 kg) × 2.9 %	=	0. 056 kg / 100 kg
$\text{Water}_{\text{EXAMS corrected}}$	=	(1.923 kg / 100 kg) × 97.1 %	=	1.867 kg / 100 kg

- Those distributions now can be compared to the EQC results:

$\text{Sediment}_{\text{EQC}}$	=	0. 064 kg / 100 kg
$\text{Water}_{\text{EQC}}$	=	1.859 kg / 100 kg

- A comparison ratio can be obtained by dividing the corrected EXAM result by the EQC result

$Sediment_{EXAMS\ corrected}$ / $Sediment_{EQC}$	=	0.875
$Water_{EXAMS\ corrected}$ / $Water_{EQC}$	=	1.004

By applying the same procedure to the other pesticides, all comparison ratios are obtained as shown in Table 4.3. In the same table, a comparison for the sediment compartment is also given.

Table 4.3. Comparison between EQC and EXAMS results

Pesticide	$W_{EXAMS\ corrected} / W_{EQC}$	$S_{EXAMS\ corrected} / S_{EQC}$
Imazalil	1.032	0.627
Thiabendazole	0.988	1.389
Propiconazole	1.004	0.875

Note: ***W*** stands for Water Compartment, and ***S*** stands for Sediment Compartment

4.2.3. DISCUSSION OF RESULTS

It is observed in table 4.3 that both models give almost equal predictions for the water compartment. However, the prediction values for the sediment compartment show significant differences for both models. The table shows only the pesticides that were analyzed in the Ecuadorian lab, but other banana pesticides not analyzed at this moment in Ecuador were also evaluated to see the behaviour of both models (glyphosate and tridemorph). Based on the comparison ratio, the differences in sediment predictions are more significant on those pesticides: around 121 for glyphosate and 3 for tridemorph. However, the predictions for the water compartment are still similar (comparison ratio of 0.87 for glyphosate and 0.88 for tridemorph). The main reason that accounts for this difference is the number of processes involved in the calculations for both models. EXAMS considers ionization, complexation and sorption processes of the compound with sediments and biota; while EQC only uses transport, transfer and basic degradation processes in the sediment compartment.

In this case study, the water compartment predictions with EXAMS represents between 0.87 to 1.03 times the predictions for the same compartment using the EQC model. As a preliminary conclusion, both models can be used independently when screening pollution in the water compartment. On the other hand, the sediment compartment is better evaluated by EXAMS than by the EQC model because more interaction processes are considered in the first one.

4.3. SPATIALLY INTEGRATED MODELS

4.3.1. AGNPS STRUCTURE

The AGNPS model is basically a runoff model that needs several types of data to be supplied. In the present research, the model structure was identified in order to optimise the available data sources in the basin. The model needs around 160 input variables to evaluate pesticides in a watershed (USDA-ARS 2002). Input data are distributed in 28 data sections as shown in Figure 4.5. They could be alphanumeric or numeric values, which could be handled with GIS. In the figure, the coloured boxes represent the minimum data sections needed for pesticide evaluation. All necessary information was grouped in primary and secondary data as shown in Chapter 3.

The input data have to be spatially distributed because the model works on a sub-basin basis. All sub-basins in the basin are divided in three zones, as shown in Figure 4.6: upstream, left and right drainage area. In the figure, the dark line represents three river stretches with their corresponding drainage basins (dotted line). Each shaded area in the upper left sub-basin represents what it is called "the cell" in the model.

AGNPS cells are different than raster cells, although many raster cells could be included within an AGNPS cell. The cell division depends on the drainage pattern and a threshold value to form a cell. Therefore, a watershed could be divided as many times as possible in order to capture the data spatial variability. By using GIS techniques, a thematic map can be created showing all generated AGNPS cells. These techniques are based on methodologies mainly developed by Garbrecht and Martz (1993, 1996 and 1999), and later included in AGNPS as a GIS interface. A detailed description of the generation procedure can be found in the TOPAGNPS module manual within the AGNPS documentation.

By overlaying the secondary data with the "*AGNPS cell*" thematic map, the input data for the model was obtained. However, in some cases a data aggregation of many raster maps into a single thematic map had to be done. Aggregation is performed by overlaying the "*AGNPS cell*" thematic map with the corresponding raster map for data extraction. For example, the complete soil information was attached to the general soil taxonomic map (polygons) in order to have a more detailed soil map for AGNPS usage.

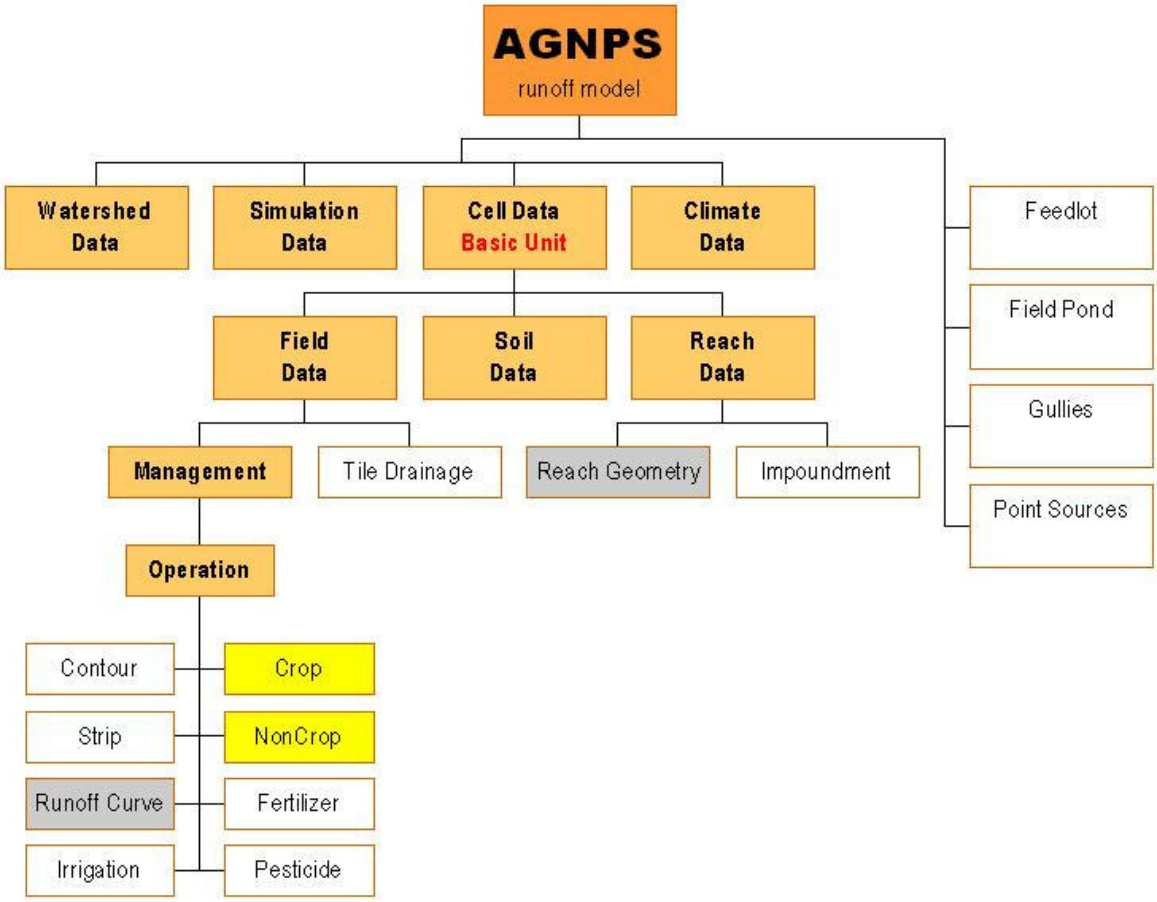

Figure 4.5. **AGNPS model structure after Bosch *et al.* (1998)**

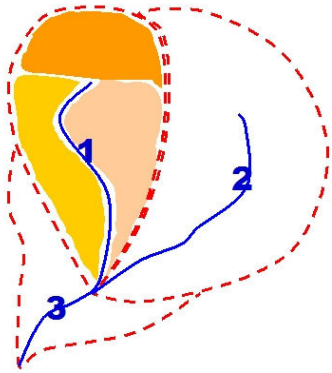

Figure 4.6. AGNPS cells created from watershed subdivision

When performing the overlying procedure, a limitation was observed: an AGNPS cell can only have one soil type and one land use type. However, most of the time more than one soil/land use type can fall into one AGNPS cell. The ArcView interface that comes with the model performs a joint spatial analysis to solve this problem by assigning the soil/land use ID with more surface area to the entire cell. The main drawback of this procedure is the loss of information across the basin. Table 4.4 shows maximum information loss percentages depending on the number of characteristic types falling in a single AGNPS cell.

Table 4.4 Maximum information loss by assigning the object ID with the largest area in a cell

# of ID's falling in a single AGNPS cell	*Maximum information loss based on area*
1	0 %
2	50 %
3	66.67%
4	75 %
n	(n – 1)100/n %

For the analysis, the Chaguana river basin was divided in 192 AGNPS cells which drain into 78 river's reaches (see figure 4.7). The resulting cell areas vary between 1 and 829 Ha. When overlaying the soil thematic map, 80 AGNPS cells presented more than one soil type (table 4.5). Due to the joint spatial analysis, soil information for around 15% of the basin surface area was not included as input data. The same assessment was done with land use as shown in table 4.6. Land use information for around 18% of the basin surface area was not included as input data.

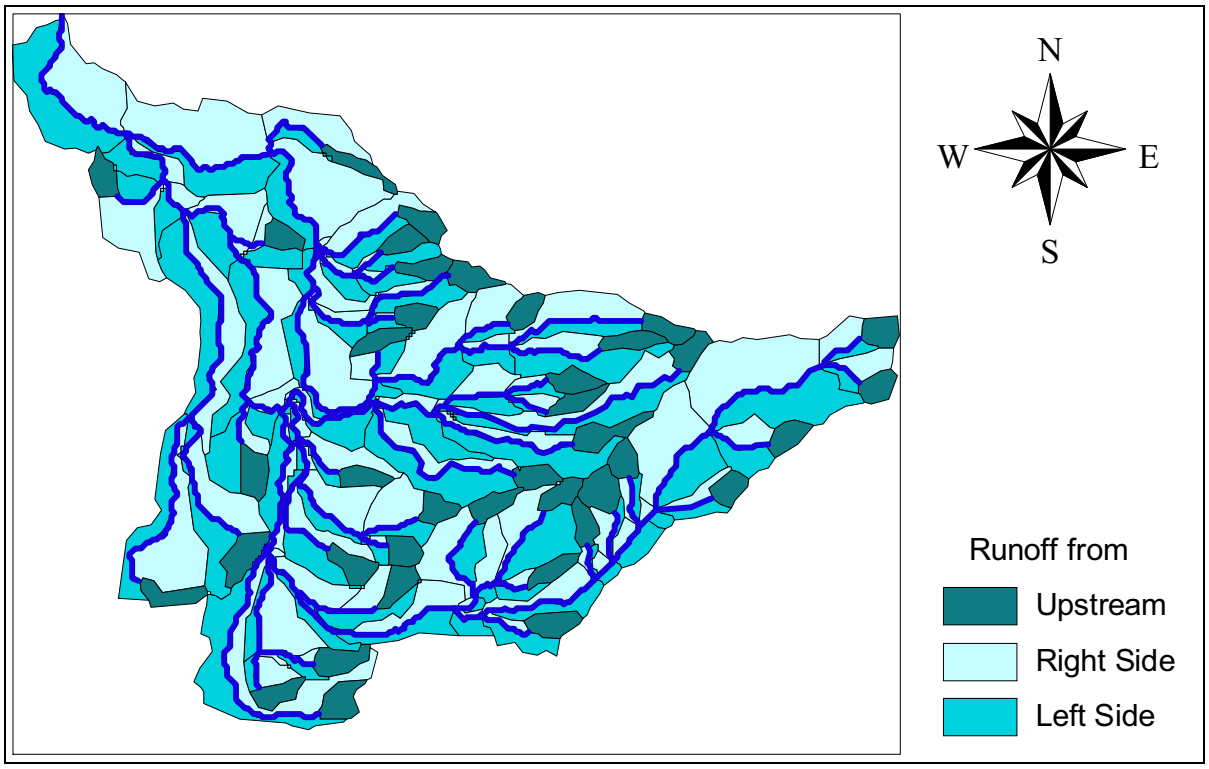

Figure 4.7. AGNPS cell division of Chaguana river basin

Table 4.5. Spatial Joint Analysis done on soil information

# soil types per AGNPS cell	*Number of occurrences out of 192 cells*	*Total area not considered after joint spatial analysis*	*% of total basin area*
2	51 cells	1951 Ha.	6.25 %
3	22 cells	1953 Ha.	6.56 %
4	7 cells	759 Ha.	2.44 %
Total	***80 cells***	***4663 Ha.***	***14.95 %***

Table 4.6. Spatial Joint Analysis done on land use information

# land use types per AGNPS cell	*Number of occurrences out of 192 cells*	*Total area not considered after joint spatial analysis*	*% of total basin area*
2	74 cells	2199 Ha.	7.05 %
3	34 cells	2839 Ha.	9.10 %
4	7 cells	377 Ha.	1.21 %
5	1 cell	123 Ha.	0.39 %
Total	***80 cells***	***5538 Ha.***	***17.75 %***

From the tables, there is more chance to have two types of characteristics falling in one cell than more types. Although it seems that the soil information loss is quite significant (around 15%), the regional soil characteristics are pretty similar in the region (silty sandy soils) and so the loss percentage is acceptable for the assessment. In the case of land use, the loss percentage also could be significant, but only 2% of the not considered area corresponds to cropland. Therefore, also the land use loss is considered acceptable for the assessment. A more detailed watershed subdivision could decrease the information loss. However, the extra effort probably could not be cost-effective. In addition, more data should be needed to get to the more detailed analysis.

4.3.2. CALIBRATION PROCEDURE

The whole calibration procedure was done with the data collected from three available gauging stations and the measurements of suspended sediments and pesticide concentrations along the river. 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).

$$r^2 = \frac{\sum_{i=1}^{N}(O_i - O_{avg})\,(P_i - P_{avg})}{\left(\sqrt{\sum_{i=1}^{N}(O_i - O_{avg})^2}\right)\left(\sqrt{\sum_{i=1}^{N}(P_i - P_{avg})^2}\right)} \qquad [5.1]$$

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).

$$E = \frac{\left(\sum_{i=1}^{N}(O_i - O_{avg})^2\right) - \left(\sum_{i=1}^{N}(O_i - P_i)^2\right)}{\left(\sum_{i=1}^{N}(O_i - O_{avg})^2\right)} \qquad [5.2]$$

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

$$d = \frac{\left(\sum_{i=1}^{N}\left(\left|P_i - O_{avg}\right| + \left|O_i - O_{avg}\right|\right)^2\right) - \left(\sum_{i=1}^{N}(O_i - P_i)^2\right)}{\left(\sum_{i=1}^{N}\left(\left|P_i - O_{avg}\right| + \left|O_i - O_{avg}\right|\right)^2\right)} \qquad [5.3]$$

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 equation [5.4].

$$Bias_{AVERAGE} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{O_i - P_i}{O_i} \times 100\right) \qquad [5.4]$$

The criterion used to calibrate the model was the trial and error procedure by changing the most sensitive parameters and also the most uncertain values in the model: the curve number (CN), the cover factor (C_M) and the practice management factor (P). Those values influence the model depending on the process to be modelled.

The developers of both models (Arnold *et al.* – SWAT, and Bigner *et al.* – AGNPS) recommend following a step wise procedure in the calibration process. Because both models are mainly runoff-based, flow calibration is the most important step by adjusting the curve number. There are other sensitive parameters affecting the flow such as the snow related parameters, but those can not be applied to Ecuador (a tropical country).

After calibrating the flow, the logical step is to calibrate the sediment yield in the model. This again is done by trial and error procedures. Both models estimate sediment yields based on the USLE, RUSLE or MUSLE approaches. In this approach, all factors (R, K_S, LS, C_M and P) significantly affect the sediment yield estimations. However, only two factors showed to be more uncertain when getting data from the study site: cover and practice management factors. The reason is simple: this approach is not commonly used in Ecuador. Therefore, farmers and engineers do not keep enough records to estimate those values. Even in literature, no information was found regarding tropical crops such as banana, cocoa and citrus. Although there is some degree of uncertainty in the other values, they can be estimated more easily from the gathered information than the C_M and P factors. That was the main reason to use those parameters in the calibration procedure.

The last step in the calibration procedure concerned the pesticides. Again, trial and error was used based on the most uncertain characteristic of the pesticide application in the study area: where was the pesticide applied? and when?

There was no information from the farmers (the degree of collaboration in the project was extremely low). Therefore, a trial and error procedure was implemented to devise the period and location of application. The procedure was to select a cell or basin with banana activity, use the recommended application rate and compared the predicted results with the observed values.

4.3.2.1. 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 (see Figure 3.4), were used, and mainly located in the middle course of the river basin. The recorded period on these stations is limited to only 4 years of measurements (1979, 1980, 1982 and 1983). There is no available measurement after 1983. The flow data represent average monthly values.

The AGNPS model is basically a runoff model that estimates flows based on SCS Runoff Curve Numbers, which are indicators of how much water is running off from the soil surface. The higher the Curve Number, the higher the estimated runoff. In addition, the model can be run in a "warm-up" state which is the possibility to run the model during a certain period of years without giving any result. This process is useful for reaching an appropriate wetness state as similar as the real state of the basin before the predicition phase occurs. The process is mainly done by analysing the variation of water holding capacity during a continuous time-series.

Flow calibration was conducted by adjusting the Curve Number for each land use type involved in the basin assessment. As written before, the basin was subdivided in 192 cells (or 78 sub-basins) with their own runoff characteristic. A way to make the procedure easier is to group the sub-basins in areas with similar characteristics as land use, soil type or geographic region. In table 4.7, the grouped areas for each gauging stations with their own land use distribution are shown. Also the Curve Numbers adopted before and after the calibration procedure are given.

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 4.7. Curve Number adjustment based on land use distribution in drainage basin per gauging station.

Gauging Station	*Drainage Basin*	*Land Use distribution*	*Main Soil Group*	*CN before calibration*	*CN after calibration*
Colorado	Colorado	Pasture (713 Ha / 30%)	B	79	61
		Forest (1665 Ha / 70%)	B	66	55
Zapote	Zapote ([a])	Forest (379 Ha / 9.6%)	B	66	55
		Cocoa (459 Ha / 11.6%)	B	83	70
		Other Crop (683 Ha / 17.2%)	B	83	70
		Banana (1082 Ha / 27.3%)	A	74	67
		Pasture (1357 Ha / 34.3%)	B	79	69
Chaguana	San Jacinto	C and P [(b)] (398 Ha / 3.9%)	B	47	47
		Banana (501 Ha / 4.9%)	B	83	83
		Pasture (1913 Ha / 18.9%)	B	79	69
		Brushes (2645 Ha / 26.1%)	A	68	68
		Forest (4676 Ha / 46.1%)	B	66	55
	Charengue	C and P and F [(c)] (99 Ha / 4.9%)	A	65	65
		Banana (110 Ha / 5.4%)	B	83	83
		Cocoa (555 Ha / 27.4%)	A	74	74
		Brushes (607 Ha / 29.9%)	A	68	68
		Forest (656 Ha / 32.4%)	B	66	55
	La Polvora	Banana (77 Ha / 2.1%)	A	74	74
		Pasture (1591 Ha / 43.8%)	B	79	79
		Forest (1963 Ha / 54.1%)	B	55	55
	Middle Chaguana	Banana (1662 Ha)	B	83	83

Notes:
(a) It does not include the drainage basin that is discharged from the Colorado river gauging station.
(b) C and P = Mixture of crops and pasture
(c) C and P and F = Mixture of crops, pasture and forest

The AGNPS model estimates flows 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 (1979, 1980, 1982 and 1983) 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. AGNPS 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 statistics parameters are still showing "poor" fit, then the daily events in the month are rearranged by keeping always in mind the maximum monthly precipitation and the number of rainy days. This process is repeated until flow predictions fit the measured values.

Figures 4.8, 4.9 and 4.10 show the comparison between predicted and measured flows for Colorado, Zapote and Chaguana gauging stations respectively. Table 4.8 shows the estimated statistics coefficients of fit for all three gauging stations. The model could predict the flows in Chaguana gauging station 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). Based on the model execution it is concluded, 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 more data for flow validation is critical; the data mainly represent extreme events (an "El Niño" event occurred during 1982 and 1983).

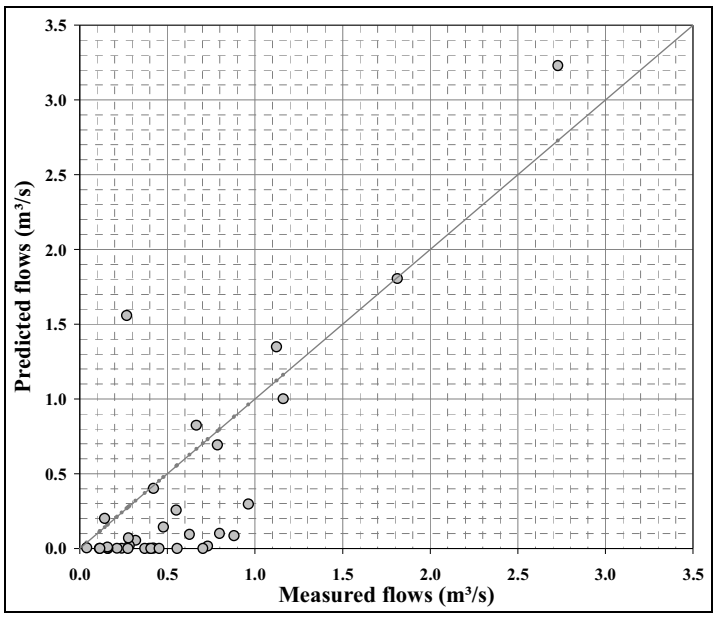

Figure 4.8. Predicted vs. observed values in Colorado river gauging station

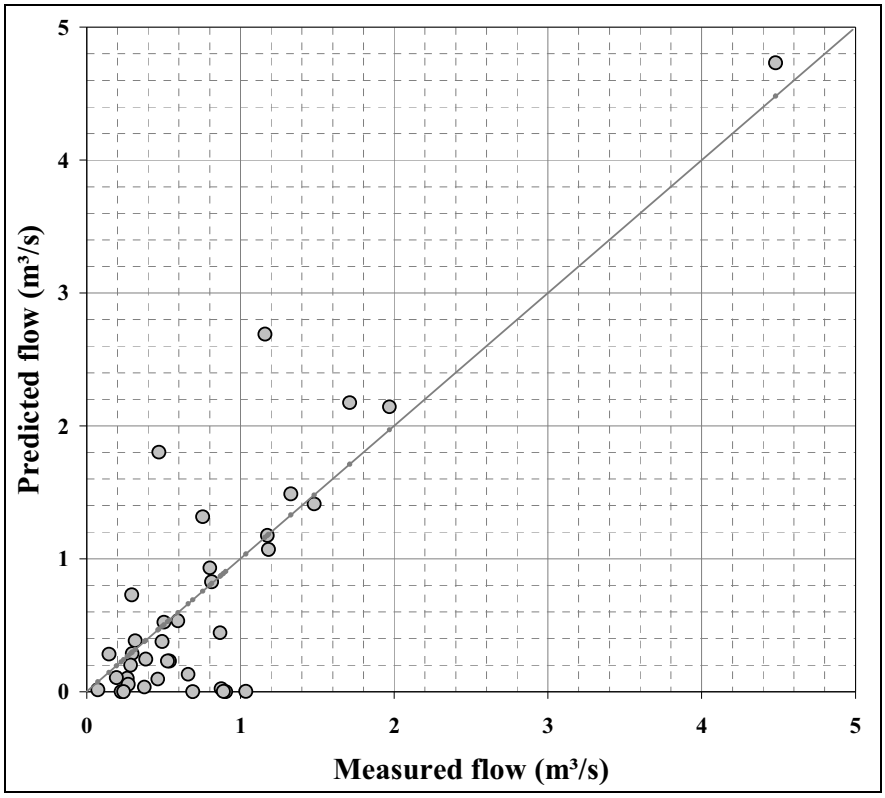

Figure 4.9. Predicted vs. observed values in Zapote river gauging station

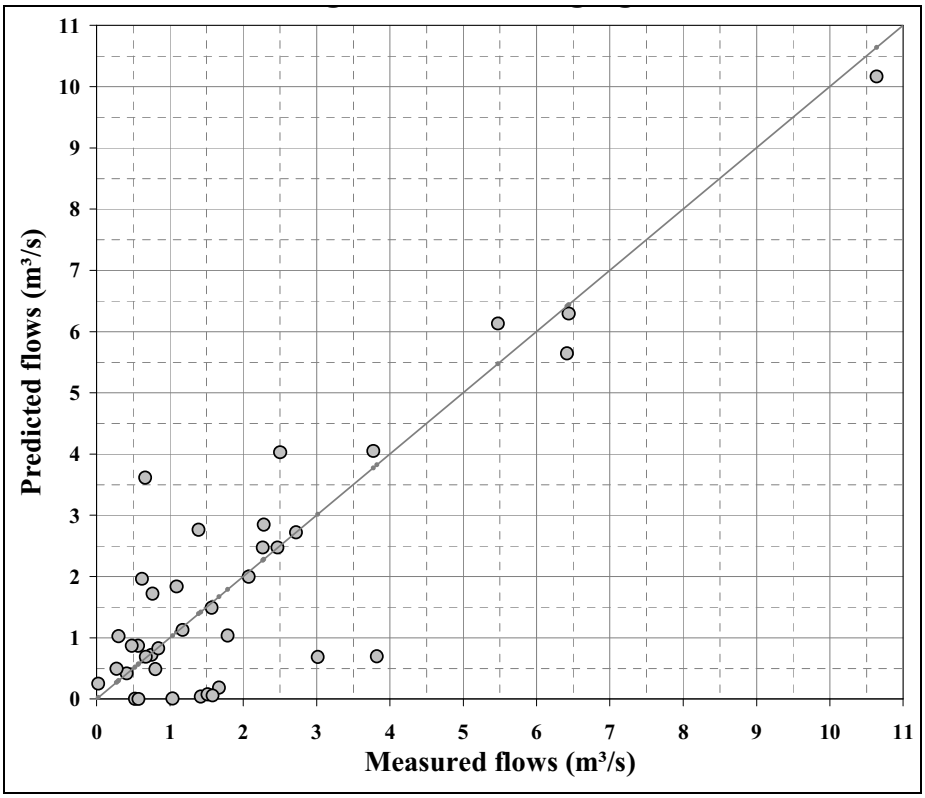

Figure 4.10. Predicted vs. observed values in Chaguana river gauging station

Table 4.8. Summary of measured and predicted (AGNPS flows) statistics for Colorado, Zapote and Chaguana gauging stations

Statistics		*Colorado*	*Zapote*	*Chaguana*
Predicted	**Mean**	0.020 m^3/s	0.284 m^3/s	1.031 m^3/s
	Standard Deviation	0.691 m^3/s	0.977 m^3/s	2.198 m^3/s
	Number of events	35	38	38
Measured	**Mean**	0.421 m^3/s	0.567 m^3/s	1.407 m^3/s
	Standard Deviation	0.527 m^3/s	0.762 m^3/s	2.147 m^3/s
	Number of events	35	38	36
Coefficient of Determination (r^2)		0.83	0.85	0.87
Coefficient of Efficiency (E)		0.37	0.53	0.73
Index of Agreement (d)		0.88	0.91	0.93
Mean relative bias (%)		90.31	26.79	0.83

Based on the previous flow calibration, a relationship between the estimated flow rate and the average precipitation that fell in the basin was developed at three points of interest: the basin outlet, the crossroad over the Chaguana river and the crossroad over the Zapote River (Figure 4.11). The graph can be useful to determine flow rates, and it was used in the sediment calibration step as explained in the next section.

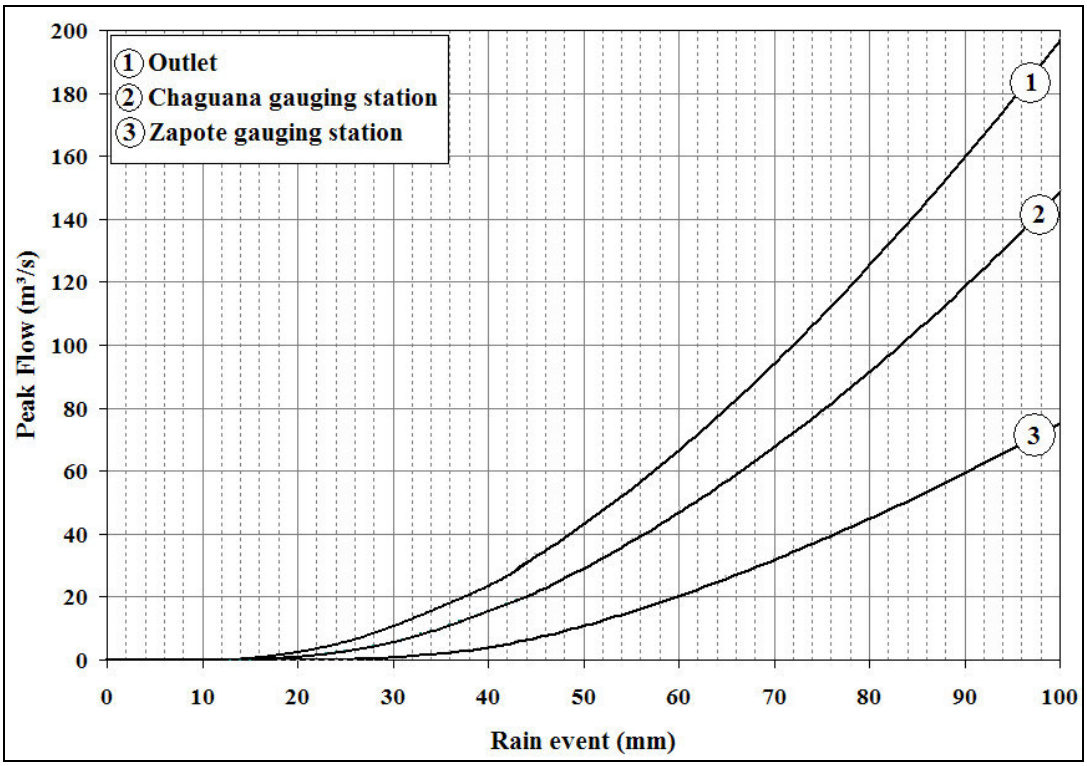

Figure 4.11. Estimated peak flow for three points for an average daily rain event in the Chaguana basin

4.3.2.2. Suspended Sediment Calibration

The official database, where the flow data were obtained, did not contain measurements of suspended sediments to perform the model calibration. Therefore, during four sampling campaigns between 2001 and 2002, several monitoring points were set along the river to determine suspended sediments and flow measurements (see chapter 3).

To perform the suspended sediment calibration, it was necessary to use the outputs of the flow prediction to assess the unknown inputs (rain events) which were used in the suspended sediment predictions of those inputs. 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 Chaguana basin case study, the AGNPS model 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 perform an interpolation on the calibrated flow graph obtained from the flow calibration step (figure 4.11). The estimated rain events were obtained by introducing the monitored flow values in figure 4.11. The obtained values represent an average rain event as falling at the same time in the entire catchment area[32]. Table 4.9 gives the estimated rain event for the four sampling campaigns, which are the values used in the model execution.

Table 4.9. Estimated daily rain event for sampling days based on flow calibration

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 AGNPS model was only run in the river reaches that showed measurements of suspended sediments. As shown in table 4.10, the sediment calibration was performed by adjusting the two parameters that contribute to sediment yield and do have more uncertainty in their estimated values: the cover management factor (C) and the practice management factor (P).

[32] In the real world, different rainfall amounts can be measured at the same time in the watershed because precipitation is a spatially distributed variable. However, models sometimes consider only a single rainfall value for a specific runoff assessment (for the whole watershed or for smaller subdivisions)

Table 4.10. C and P factor adjustment based on land use per sampled river.

River	*Sub-basin*	*Land Use*	C_{before}	C_{after}	P_{before}	P_{after}
Zapote	Colorado	Pasture	0.066	0.041	0.990	0.200
		Forest	0.039	0.016	1.000	1.000
	Zapote	Forest	0.039	0.039	1.000	1.000
		Cocoa	0.286	0.170	0.850	0.600
		Other Crop	0.286	0.170	0.900	0.650
		Banana	0.286	0.013	0.400	0.100
		Pasture	0.066	0.003	1.000	0.200
Chaguana	San Jacinto	Crop and Pasture	0.176	0.176	1.000	1.000
		Banana	0.286	0.286	0.250	0.690
		Pasture	0.066	0.003	0.960	0.900
		Brushes	0.066	0.321	1.000	1.000
		Forest	0.039	0.003	1.000	0.900
	Charengue	Crop-Pasture-Forest	0.130	0.130	0.949	0.995
		Banana	0.286	0.286	0.500	0.690
		Cocoa	0.286	0.286	0.750	0.910
		Brushes	0.066	0.039	0.890	0.925
		Forest	0.039	0.100	1.000	1.000
	La Polvora	Banana	0.286	0.286	0.370	0.690
		Pasture	0.066	0.041	0.800	0.861
		Forest	0.039	0.62	1.000	1.000
	Las Juntas	Banana	0.286	1.000	0.300	0.100
		Other Crop	0.286	0.010	0.600	0.100
	Chaguana	Banana	0.286	0.286	0.250	0.690
		Shrimp Farms	0.600	1.000	0.100	1.000

The three statistical parameters used in the flow calibration step were also applied to compare the predicted and the observed suspended sediment values. The comparison was done in two different ways. First, the Zapote and Chaguana rivers were compared separately to determine in which basin the model shows more agreement. Then, an overall comparison was done with all evaluated sampling reaches. Figure 4.12 shows the overall comparison. Table 4.11 gives the statistical values of the performed comparisons.

Table 4.11. Summary of measured and predicted (SS concentrations) statistics for the sampled reaches in Zapote, Chaguana and the whole river system

Statistics		*Zapote*	*Chaguana*	*Basin*
Predicted	**Mean**	36.45 mg/l	31.00 mg/l	34.37 mg/l
	Standard Deviation	13.35 mg/l	57.09 mg/l	48.58 mg/l
	Number of samples	13	28	41
Measured	**Mean**	14.24 mg/l	29.58 mg/l	28.50 mg/l
	Standard Deviation	19.02 mg/l	62.14 mg/l	54.23 mg/l
	Number of samples	13	28	41
Coefficient of Determination (r^2)		0.60	0.94	0.93
Coefficient of Efficiency (E)		0.00	0.88	0.86
Index of Agreement (d)		0.70	0.97	0.93
Mean relative bias (%)		-30.37	-3.64	-6.70

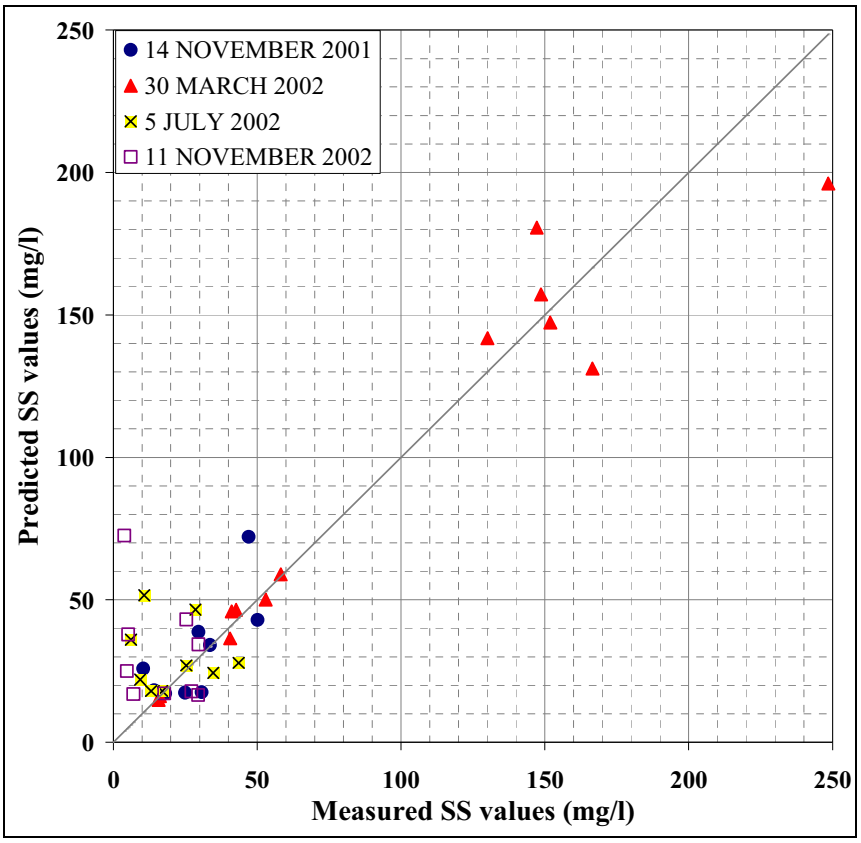

Figure 4.12. Predicted vs. observed SS values in all sampled reaches of the basin

Discussion

The use of inverse modelling techniques represents a useful approach to overcome problems regarding input data. However, this technique also represents a risk because the outcome could 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 values.

After running the model, it is clear that the predictions for the Zapote River are not good enough as the coefficient of efficiency (E) is equal to 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 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.
- Another reason is that information loss occurred during spatial data aggregation.

In the Chaguana river, there were also significant differences between predicted and measured values at specific sampling points only for the March sampling period (rainy season) while the rest of the sampling dates showed a good agreement. There was some dredging activity at certain points 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 as seen in Figure 4.13. For that reason, the model failed to predict the suspended sediment concentrations as the model calculates sediment yield based mainly on soil erosion from runoff. Figure 4.13 shows the predicted and measured concentrations for around 28 km of monitored Chaguana river during the sampling campaign conducted on March 2002.

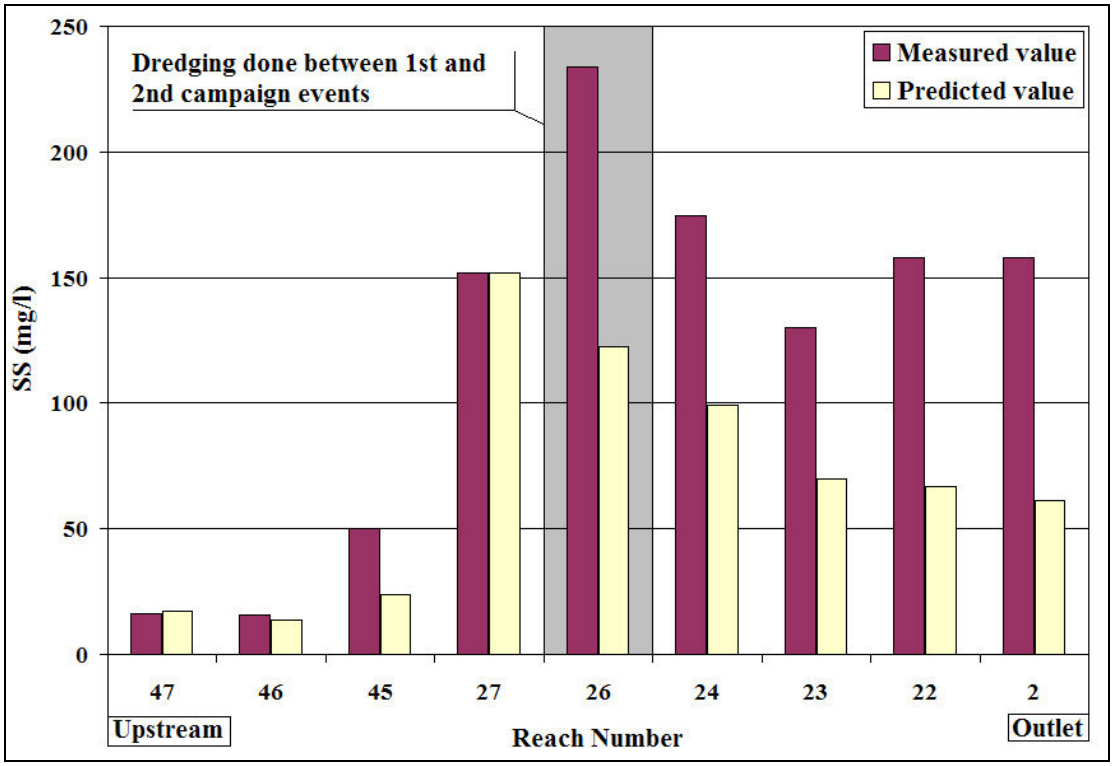

Figure 4.13. Sediment yield impacted due to unexpected changes (dredging) in March 2002

In addition, the last reach, Number 2, still shows values lower than the actual ones. That is because that reach was influenced by the tidal push at the moment of the 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 model. Therefore, the model is no longer applicable for those situations.

4.3.2.3. Pesticide Concentration Calibration

The pesticide calibration was performed based on the sampling dates, the amount of pesticide applied per farm and the number of days from the sampling date up to the last pesticide application in the field. On this step, inverse modelling techniques were also applied due to lack of information.

There were four sampling dates, as written in chapter 3, which roughly represent the climate conditions throughout the year (14 November 2001, 30 March 2002, 5 July 2002 and 11 November 2002). Due to the lack of daily rainfall data, precipitation on those dates were estimated in the previous section.

Regarding the pesticide application rate, the model was initially run with the recommended dose of 100 grams of propiconazole active ingredient per hectare. However, this amount was adjusted based on the following reflections:

- Farms do not apply pesticides on the same date. There was no information regarding application days for all existing farms in the Chaguana basin, but from only one farm outside the basin (used as a reference in the current research)
- The pesticide application rate varies from farm to farm depending on several conditions (degree of sigatoka presence, economic condition, planting dates, etc.)

For simplicity of the assessment, all AGNPS cells with the banana landuse were considered as a banana plantation. Forty out of 192 cells were considered as banana plantations representing around 8128 Ha (16 % of surface area in the Zapote basin and 84% in the Middle-Chaguana basin). First, a unique pesticide management was assigned to each of these cells (dose and application day through the year).

Because the sampling campaigns usually did not coincide with pesticide application, the detected pesticides in the river system were assumed to be the result of previous applications performed during certain days previous to the campaign on certain farms in the basin. By a trial and error procedure, the span of days since the last application of propiconazole in the banana sectors of the basin was determined (see figure 4.14).

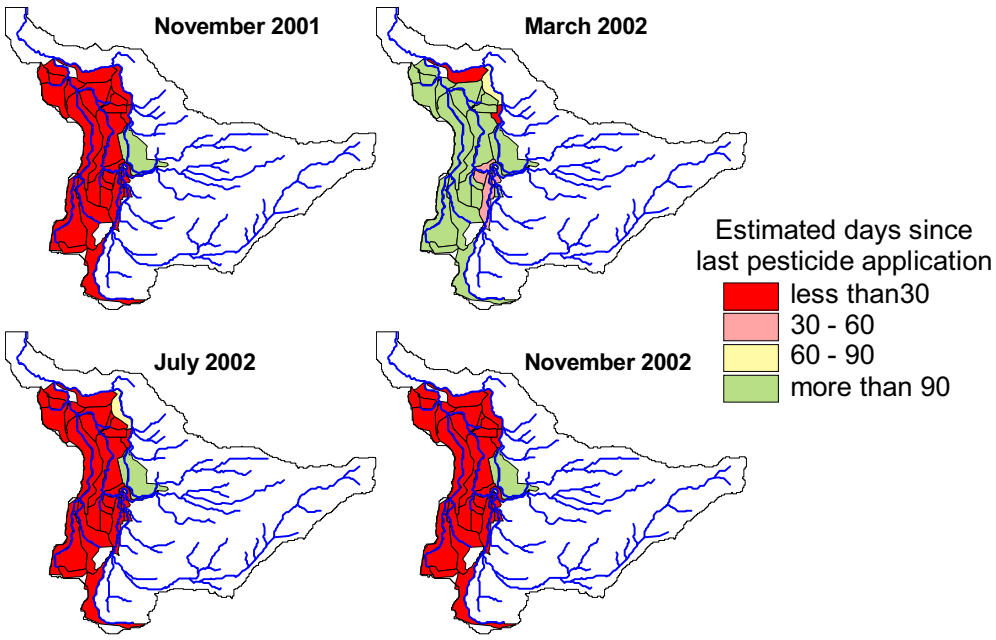

Figure 4.14. Estimated days since last application of propiconazole in the basin farms

The trial and error procedure consisted in running the model several times with different configurations of application sites and application dates until the predicted concentrations fit the measured values on the specific campaign date. This procedure could be avoided by producing monthly predictions. However, only event (daily) concentrations were measured in the sampling campaigns. Based on the findings of figure 4.14, the model was run and calibrated for each sampling day with estimated application dates. It can be seen that the farms located in the Zapote basin did not apply propiconazole on a more frequent basis than the farms located in the Middle-Chaguana basin.

Table 4.12 gives the statistical coefficients for the comparison between predicted and observed pesticide concentrations in Zapote, Chaguana and the whole catchment. Again, the model is not good to predict concentrations in the Zapote river. This weakness was expected as suspended sediments predictions were also not good[33]. Although calibration was done with few data, the pesticide prediction for the Chaguana river showed a better agreement. Figure 4.15 shows the comparison between measured and predicted pesticide concentrations for the entire Chaguana basin. In the figure, the sampling periods are depicted with different symbols. The model showed a good prediction agreement for pesticide concentrations during the rainy season (March 2002). This agreement decreases as less rain falls in the basin. However, there should be more data to get better conclusions.

Table 4.12. Summary of measured and predicted (Pesticide concentrations) statistics for the sampled reaches in Zapote, Chaguana and the whole river system

Statistics		*Zapote*	*Chaguana*	*Basin*
Predicted	**Mean**	0.024 µg/l	0.251 µg/l	0.179 µg/l
	Standard Deviation	0.060 µg/l	0.408 µg/l	0.354 µg/l
	Number of samples	13	28	41
Measured	**Mean**	0.059 µg/l	0.278 µg/l	0.209 µg/l
	Standard Deviation	0.119 µg/l	0.490 µg/l	0.421 µg/l
	Number of samples	13	28	41
Coefficient of Determination (r^2)		0.38	0.78	0.79
Coefficient of Efficiency (E)		0.03	0.60	0.61
Index of Agreement (d)		0.56	0.87	0.87
Mean relative bias (%)		52.43	-0.65	2.387

[33] The model predicts pesticide concentrations as two states: attached and dissolved. The attached part depends on the amount of sediments transported by runoff. If there is poor agreement with the sediment prediction, then the pesticide prediction will also show a poor agreement. The calibration was done with the overall concentration because measured pesticide values were also reported as overall concentrations.

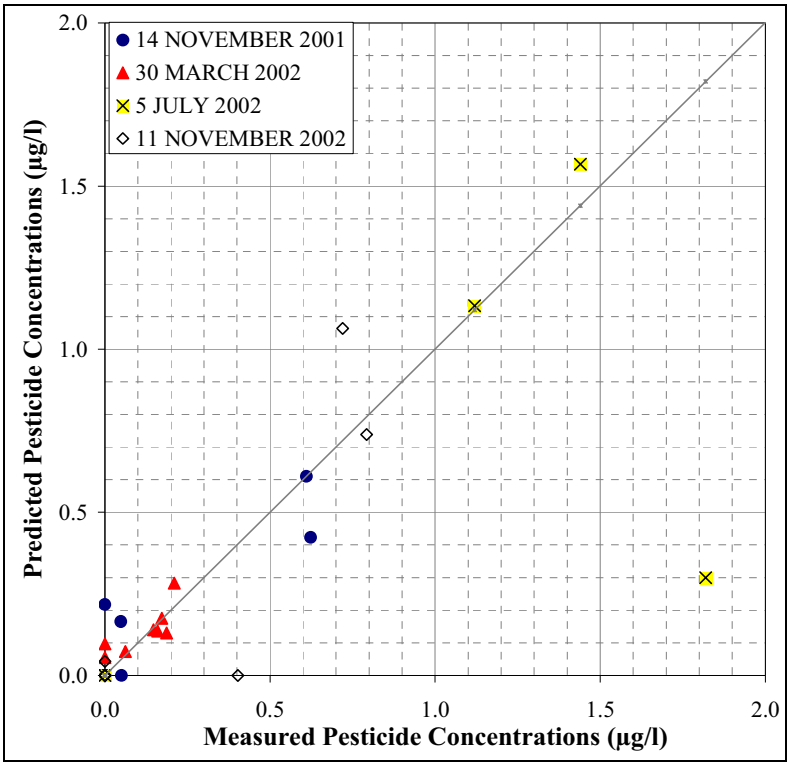

Figure 4.15. Predicted vs. observed pesticide values in all sampled reaches of the basin

4.3.3. SCENARIO EVALUATION

Once the model was calibrated, some scenarios were evaluated to tackle the question proposed at the beginning of the research: what is the impact of fungicide use from the banana sector in the Chaguana basin. This issue was evaluated by running the model in some scenarios by assuming:

- Only propiconazole usage is considered in the evaluation. The model was not calibrated for other pesticide usage in terms of application rates and spraying dates,
- The farms in the evaluated scenarios are using the recommended dose of 100 grams of active ingredient per hectare.
- The application rates and spraying dates are the same for all farms based on the application plan obtained from one typical farm (see chapter 3, Table 3.32)

The AGNPS model was run on three worst-case scenarios: all banana farms in the basin are active (scenario 1), only banana farms in the Chaguana basin are active (scenario 2) and only farms in the Zapote basin are active (scenario 3). The runs were done on the same sampling dates to have an overview of the pesticide predictions based on the planned application

schedules. Figure 4.16 gives the predicted pesticide concentrations for the first scenario (all banana farms are active throughout the year).

From the figure, it can be seen that the worst situation is produced during the rainy season when PECs can reach values higher than 5 ppb in the basin outlet. Conversely, during the dry season (mainly July), the river does not show significant values of propiconazole in all river reaches. The higher concentrations are produced because applications of propiconazole are more frequent during the rainy season and more particles attached with pesticide are transported with runoff into the river reaches. In the dry season, the transport of pesticide-soil particles from runoff is decreased because of the lack of rain events[34].

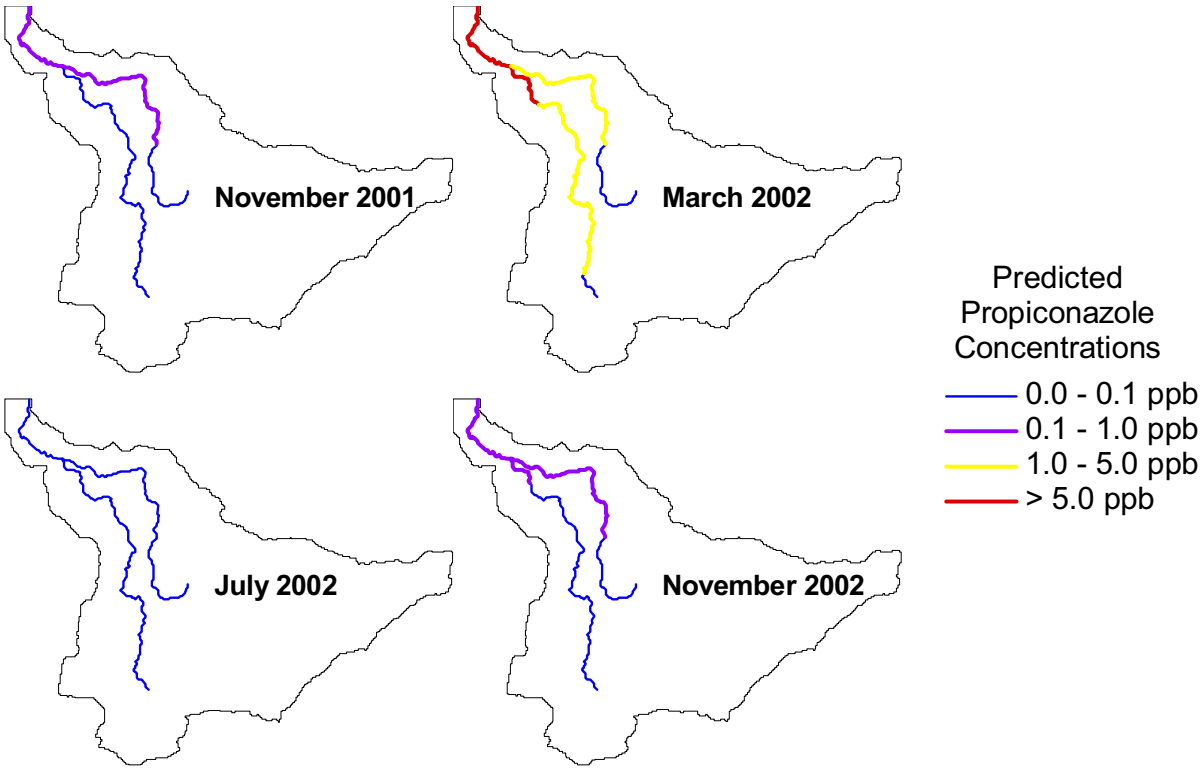

Figure 4.16. Predicted propiconazole concentrations when all banana farms in the basin are spraying with the same application schedule

Based on the analysis of scenarios 2 and 3, the pesticide pollution from Chaguana farms is more significant than that from Zapote farms. As can be seen in figure 4.17, during the rainy season (March event), the pesticide concentration near the outlet is still higher than 5 ppb, even when the Zapote farms are not active during the year.

[34] The model is mainly a runoff model which predicts pesticide concentrations coming from chemicals attached to eroded particles and chemicals washoff from the crop. Pesticide falling directly into the river by drift or point source is not considered by the model.

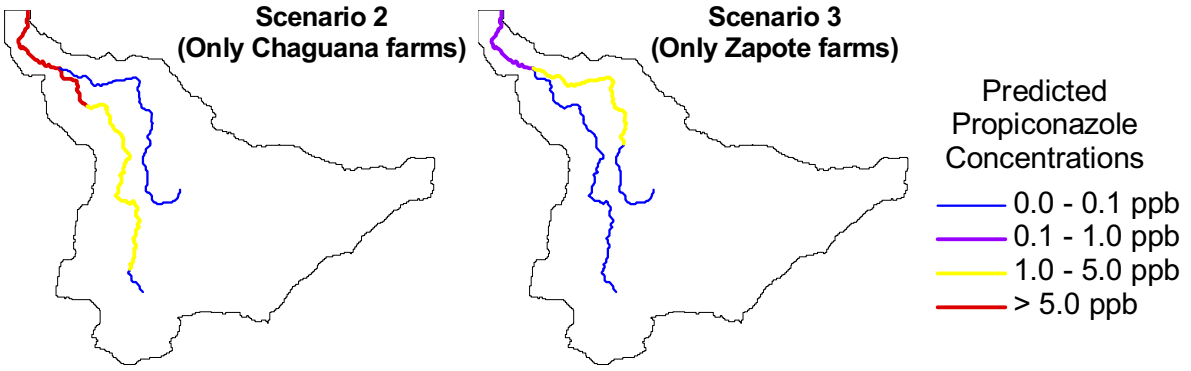

Figure 4.17. Predicted propiconazole concentrations for scenarios 2 and 3 during rainy season (March)

In addition, an analysis of pesticide predictions was done with synthetic daily rainfall data for a typical year to observe how pesticide concentration varies with rain events. The analysis was done only in the basin outlet by considering that all farms in the basin are active and they are using the same propiconazole application schedule (Scenario 1). Figure 4.18 shows the daily predictions for the basin outlet. The application schedule, based on the reference farm, is also shown in the figure by arrows. Ten propiconazole applications (100 g.a.i/Ha) were applied for running the model. It is seen that propiconazole is used more frequently during the rainy season (7 applications), and it is used at least every 10 days between the end of March and beginning of May. These extra applications are causing high predicted propiconazole concentrations during the rainy season.

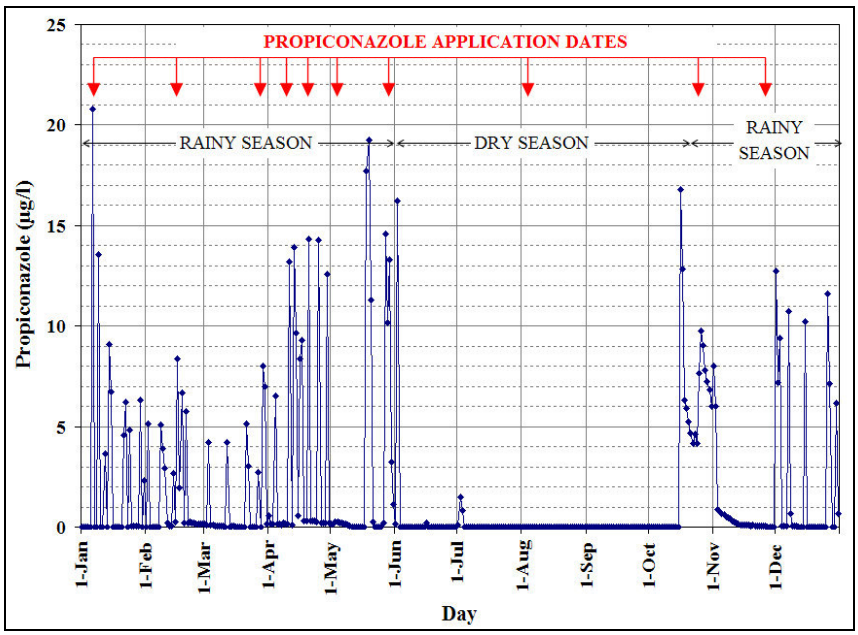

Figure 4.18. Predicted daily propiconazole concentration for the basin outlet after the model calibration.

From the environmental risk point of view, the basin outlet is very important. This last river reach crosses two areas of interest: shrimp farms at both sides of the river and the Tendales town, whose location is shown in Figure 3.13d. The annual predicted propiconazole concentrations can reach up to 20 µg/l during rainy events. Fortunately, these concentrations are still far below the reported geometric mean toxicity level of around 400 µg/l (Jolliet *et al.* 1998). However, there should be some concern in the integrated management of the Chaguana basin because, if the banana sector in the basin is expanded, the expected concentrations should increase causing an increase in environmental problems. An example of these problems is related to the Tendales town. It was observed that this town does not have a good potable water system and people living there use river water for multiple purposes (fishing, laundry, recreation, and so on). Therefore, there would be a potential human health threat during the rainy season as pesticide concentrations would be higher than the maximum allowable limits for human consumption (0.1 µg/l per pesticide – EU limit).

4.3.4. COMPARISON WITH ANOTHER RUNOFF MODEL: SWAT

One of the objectives in the current research was to compare AGNPS results with the outcome of another model. That issue was evaluated in a guided research done at ESPOL as a B.Sc. thesis (Bonini and Guzman 2003) using the SWAT modelling tool. The Chaguana basin was divided in 44 sub-basins, mainly based on the locations of the three existing gauging stations (Colorado, Zapote and Chaguana gauging stations), see figure 4.19.

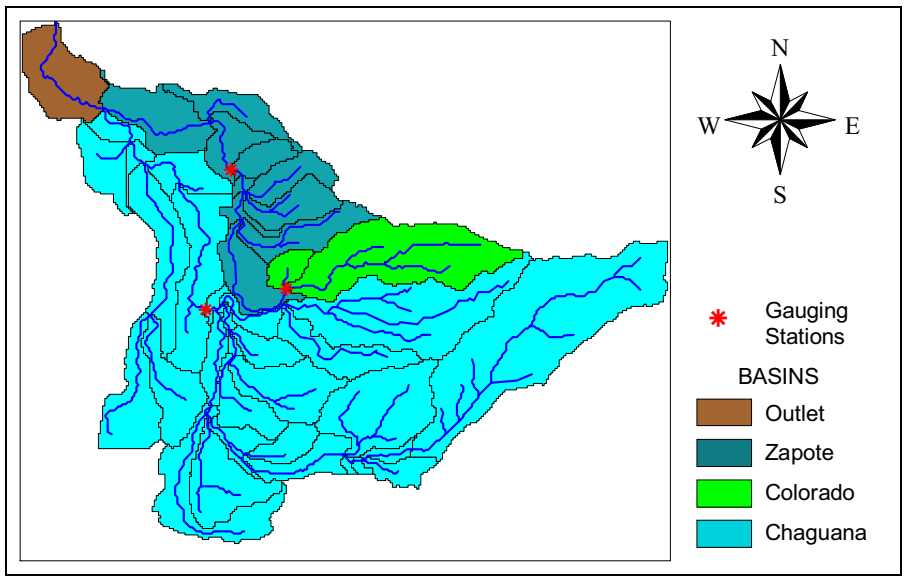

Figure 4.19. SWAT basin division of Chaguana river basin showing the three existing gauging stations.

Although both models require roughly the same amount of data, the data structure of SWAT is quite different from AGNPS. Table 5.10 shows several differences between both models.

Table 4.13. Differences between AGNPS and SWAT model

CHARACTERISTIC	*AGNPS*	*SWAT*
Unit of Analysis	Cell created from Sub-basin division (maximum three cells per sub-basin)	The basic unit of analysis is the sub-basin.
Input Data	All input data, except weather, is entered as one file	Several input files can define data regarding soil, management, chemical, reach, etc.
Climate Data	All climate data is entered as one file per weather station.	Several files are used to input data for climatic parameters, location of weather and gauging stations.
Number of Soil / Landuse types per unit of analysis	Only one type is allowed per AGNPS cell	Many types are allowed depending on the number of Hydrologic Response Units (HRUs) considered per sub-basin.
Information loss due to data aggregation within a cell/basin	Loss varies from 0 up to 50% depending on cell size and the number of data groups within one cell.	Information loss can be decreased significantly by using several HRUs in one sub-basin
ArcView GIS Interaction	The interface is used mainly to input data. The model is run outside GIS.	The ArcView interface is used to input data into the model, to run the model within GIS environment, and to process results
Execution time	For a basin with 192 cells and 78 reaches, the model took around 1 minute to run an annual simulation of daily events	For 44 subbasins, the model took less than 2 minutes to run the same annual event.

The calibration processes run for SWAT model followed the same considerations of the AGNPS model calibration by changing the same parameters (curve number, C and P factors). The coefficients of efficiency, determination and agreement were also estimated to look for the goodness of fit between the measured and predicted values (Table 4.14). Figures 4.20, 4.21 and 4.22 give the comparison between predicted and measured values for flow, suspended sediments and pesticide concentrations. The outcome of the SWAT model was shown to be more accurate than the AGNPS model. The efficiency (E) of the three calibration processes was above 0.8 for the entire Chaguana basin. Although predictions for the Zapote were improved, they are still below the ones for the Chaguana river.

Table 4.12. Summary of measured and predicted statistics for the sampled reaches in the whole river system (SWAT runs)

Statistics	*Flow*	*SS*	*Pesticide*
Coefficient of Determination (r^2)	0.91	0.97	0.99
Coefficient of Efficiency (E)	0.82	0.93	0.98
Index of Agreement (d)	0.95	0.98	0.99
Mean relative bias (%)	15.66	27.09	2.87

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.

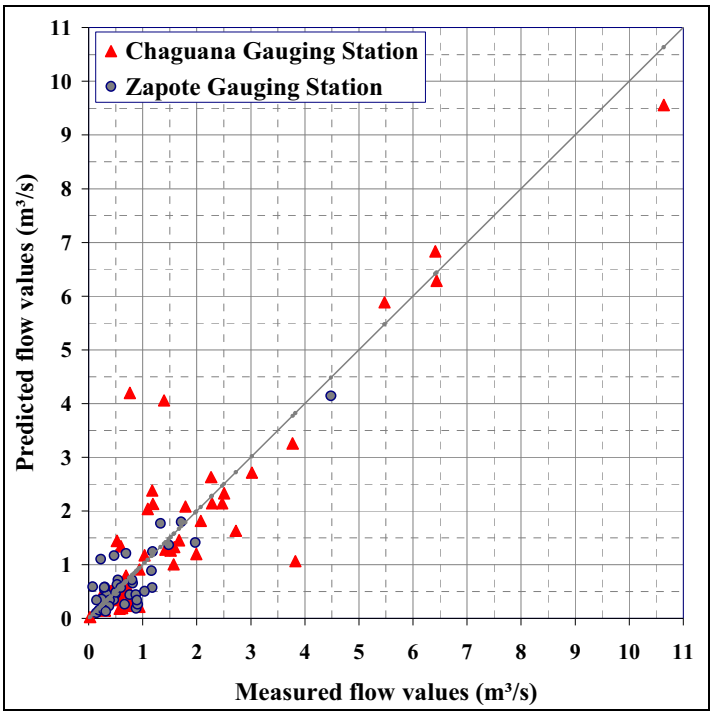

Figure 4.20. Comparison of predicted (SWAT) and measured values for flow in the whole sampled watershed.

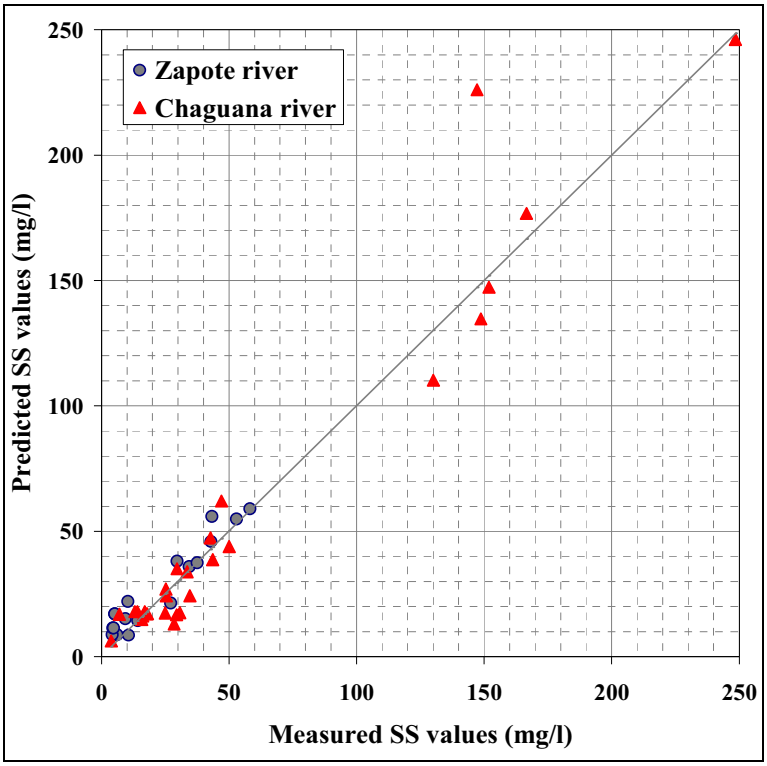

Figure 4.21. Comparison of predicted (SWAT) and measured values for suspended sediment in the whole sampled watershed.

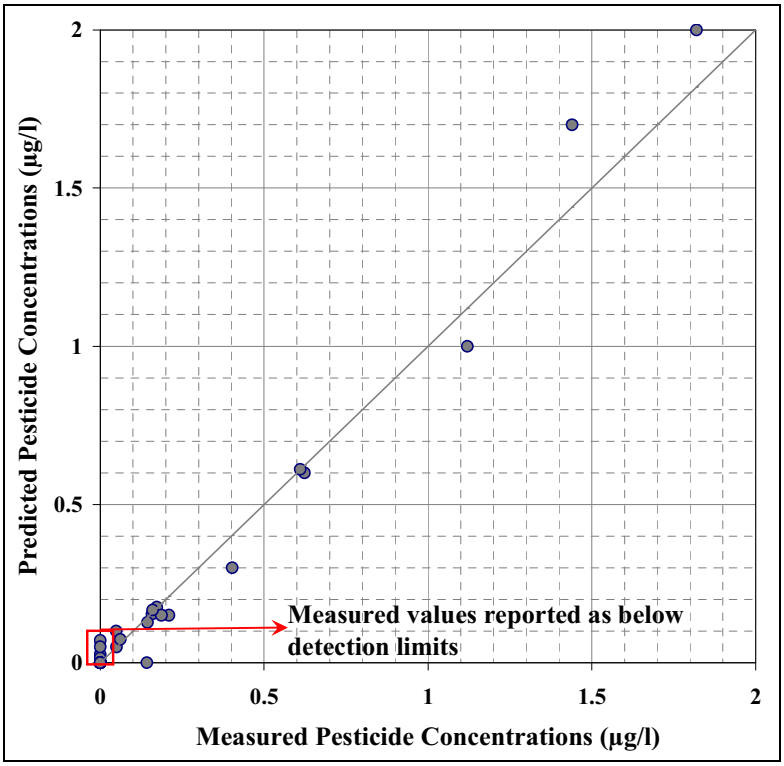

Figure 4.22. Comparison of predicted (SWAT) and measured values for pesticide concentration in the whole sampled watershed.

4.3.5. DISCUSSION

Another important issue observed in the model comparison is the influence of values reported as "Not Detected - ND" or "Traces" over the net efficiency of the comparison between predicted and measured values. There is not enough information regarding the way those ND values should be considered in a model assessment. In European countries, those ND values are assumed to be ⅓ of the corresponding detection limit. In USA, some researchers prefer to assume ½ of the detection limit such as in a soil pollution study conducted in Seattle (Glass 2000). In this study, the half-value criterion was selected because some overlapping problem with the quantified values occurred when the reported detection limits varied on different tests.

In the current research, the ND values were included as "zero" values to ease the calculations (see the marked square at the origin of Figure 4.20). The reason to do that was that the laboratory reported different detection limits for the same pesticide on each sampling campaign, ranging between 0.05 and 0.15 µg/l.

Therefore, it was a simplification to put those reported values as zero. In addition, a sampled reach was characterised to have no pesticide presence when the lab reported no indication at all for the pesticide in the evaluated reach, neither ND value nor "Traces" indication. This type of reach represents around 50% of the sampled reaches in the basin for all sampling campaigns.

As a result, there are two important questions that should be answered in future investigations to assure that a model can predict pesticide concentrations in a basin based on what Ecuadorian labs are reporting now:

1. *Can a zero value predicted by any model be properly compared with an absence of a reported value in the corresponding measured point?*
2. *How significant are the observations below detection threshold compared to the existing Environmental Policy in Ecuador[35] and abroad?*

[35] Right now, Propiconazole is not regulated in the Ecuadorian Environmental Law. However, world banana markets could force regulators and producers to control that fungicide in the near future.