CHAPTER 3. DATA GATHERING PROCESS

Measurements in the Chaguana river, Ecuador

Parts of this chapter were published in

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. Water Science and Technology, 46(9), 37-45.

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.

3.1. INTRODUCTION

The present chapter shows the procedure followed to gather all necessary data to run the evaluated models, mainly AGNPS. When gathering information, data is not necessarily in the correct format for the model execution. Thus, this chapter also presents several processes to convert the collected data. The data structure of each model is shown in the next chapter.

As part of the research objectives, this chapter presents some guidelines for selecting, gathering and processing poor data for environmental evaluation of watersheds in developing countries based on the current experience in Ecuador. The following sections show data processing procedure followed to obtain the most reliable data to be used in the model evaluation phase.

3.2. METHODOLOGY

Before using an environmental model to determine pesticide impacts caused by banana plantations in a watershed, an important step is the collection of representative and reliable data to be used in the model. In Ecuador, like in the majority of developing countries, public and private institutions do not always maintain a good housekeeping of environmental records. In addition, available information is sometimes outdated and not easily accessible. Thus, obtaining information could be difficult to accomplish, and some extra efforts must be done to process the gathered data.

In this Ph.D. thesis, GIS procedures were used in the data evaluation process. The platform selected for GIS processing was ArcView. Most of the GIS generated data was used as input data in the model execution step. However, the majority of data representing the different phenomena are only available as scarcely and irregularly distributed data. Thus, it is necessary to convert those data to a raster format in order to consider spatial variability in the modelling work. Other data were converted to vector format[10] as an intermediate step to aggregate raster data for classification purposes. All data were grouped in two categories:

[10] Vector data in ArcView is represented by shapefiles which is a simple non-topological format for storing attributes and location of data.

- **Primary data** were collected directly from available national databases or measured in the field. These data were evaluated according to the source, year of publication and scale in order to maximise information extraction (Table 3.1). For example, edaphology data was extracted initially from printed maps (only existing soil taxonomy groups in the area); later, these data were complemented with own soil sampling campaigns (Matamoros *et al.* 2002) and lab analysis (texture, soil moisture content).

Table 3.1. Primary data obtained in the Chaguana river basin

Data	*Primary Source*	*Year*	*Scale*	*Data extraction procedure*
Topography	Printed Maps	1970	1:50000	Scan of 4 topographical sheets
Geology	Printed Maps	1970	1:250000	Scan of 1 geological sheet
Edaphology	Printed Maps	1970	1:250000	Scan of 1 edaphological sheet
Land use	Digital Data	1998	1:50000	Visualization in ArcView
Climate	Database	Depends on station	Not applicable	5 georeferenced weather stations
Soil	Field Measurement	2001	Not applicable	30 georeferenced sampling sites
Water Quality	Field Measurement	2002	Not applicable	26 georeferenced sampling sites

- **Secondary data** are more elaborated data that include spatial variability in grid format. For the current research, it was generated from primary data by using kriging interpolation, accepted equations or methodologies applied in similar situations. The majority of information was generated as raster format because the evaluated models are raster-based. The cell size for the generated raster data was 1 hectare. Multiple thematic maps were developed to extract input data for the models (Table 3.2).

Table 3.2. Thematic maps generated from primary data on the Chaguana river basin

Thematic Maps	*Geographic Feature*	*Generation procedure*	*Reference*
Elevations	Point	Digitising from scanned maps	GIS procedure
Digital Elevation Model	Grid	Interpolation procedure	GIS procedure
Slope	Grid	ArcView Avenue statement	GIS procedure
Geologic Units	Polygon	Digitising from scanned maps	GIS procedure
Taxonomic Units	Polygon	Digitising from scanned maps	GIS procedure
Weather Stations	Point	Add as an event theme in ArcView	GIS procedure
Precipitation	Grid	Interpolation procedure	GIS procedure

Thematic Maps	*Geographic Feature*	*Generation procedure*	*Reference*
Runoff Erosivity Factor	Grid	Use of RUSLE equations in Map Calculator	Renard *et al.* (1997)
Sampling Sites	Point	Add as an event theme in ArcView	GIS procedure
Clay – Silt – Sand content	Grid	Interpolation procedure	GIS procedure
USDA Soil Texture	Grid	Boolean Algebra	Benham *et al.* (2001)
Very fine sand content	Grid	Interpolation procedure	GIS procedure
Soil Moisture content	Grid	Interpolation procedure	GIS procedure
Saturation content	Grid	Use of equations in Map Calculator	Saxton *et al.* (1986)
Field capacity content	Grid	Use of equations in Map Calculator	Saxton *et al.* (1986)
Wilting point content	Grid	Use of equations in Map Calculator	Saxton *et al.* (1986)
Bulk density	Grid	Use of equations in Map Calculator	Saxton *et al.* (1986)
Saturated hydraulic conductivity	Grid	Use of equations in Map Calculator	Saxton *et al.* (1986)
Hydraulic soil group	Grid	Boolean Algebra	USDA-NRCS (1986)
Organic matter content	Grid	Interpolation procedure	GIS procedure
Soil albedo	Grid	Use of equation in Map Calculator	Baummer *et al.* (1994)
Soil Map	Polygon	Data aggregation from soil grids by using Taxonomic Map as a mask	GIS procedure

3.3. SELECTION OF THE STUDY AREA

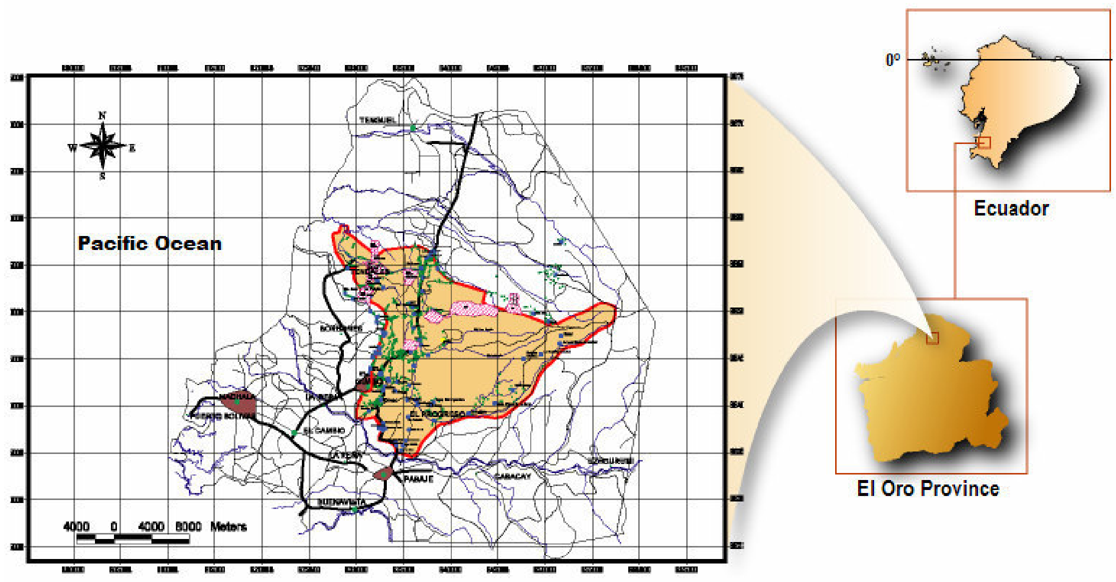
According to the Ecuadorian Central Bank, the Ecuadorian banana activity covers approximately 139000 Ha distributed over 7 coastal provinces. However, the banana activity is not always the only crop affecting a specific site. Thus, an impact assessment of pesticides only coming from banana plantations could be impossible to achieve[11]. For that reason, the current research selected a site in such a way that the main goal could be achieved by minimizing other crop interference. The criteria applied for site selection were:

- As pesticides are transported mainly through the hydrological cycle, a river basin should be the most acceptable site to perform the assessment. However, pesticide drift from adjacent basins could influence the final assessment of the selected river basin.

[11] Other crops sometimes use the same agrichemicals as banana farms do (e.g. glyphosate)

- The basin should be median or low size (less than 50000 Ha) in order to avoid multiple crop activities. The larger the basin, the bigger the chance to find more than one crop activity in the basin.
- At least some farmers should agree to collaborate with research (farm management records, pesticide usage, soil and climate records and accessibility)

After visiting some potential study sites in Ecuador, the Chaguana River Basin was selected. It is located in the El Oro Province in the South-Western part of the country (Figure 3.1). This river basin is approximately 32000 Ha, which could be considered as a hydrological system between a rural catchment and a river basin (Maidment 1996). The Chaguana system does not discharge directly to the sea, but to a bigger watershed called the Pagua River Basin. However, the sea tidal influence reaches up to 6 km upstream the Chaguana basin's outlet.



Source: Centre of Environmental Studies (CEMA) – ESPOL, 2002

Figure 3.1. Location of the Chaguana river basin

3.4. TOPOGRAPHICAL DATA

3.4.1. AVAILABLE CARTOGRAPHICAL INFORMATION

In Ecuador, the Geographical Military Institute (IGM) is the official organism to keep, publish and distribute all topographical information. Most of the available information is from 1970, and it is distributed as printed maps at 1:50000 scale. On those printed maps, elevation contours are displayed every 40 meters with scattered elevation points representing measured bench marks above sea level. Terrains between 0 and 40 meters do not show any elevation contour. Contours on printed maps were drawn based on aero-photogrammetric procedures.

Because topographical information is always used as a reference for other types of information, cartographical data were obtained from those printed maps, with geographical characteristics given in Table 3.3

Table 3.3. Cartographical characteristics used in the geo-referencing data process

Coordinate System / Projection	Universal Transverse Mercator – Zone 17 S
Datum	Provisional South American Datum 1956
Spheroid	International 1924
Horizontal and Vertical Units	Meters
Latitude of Origin	0º
Central Meridian	-81º
False Easting (meters)	500000
False Northing (meters)	10000000

The selected study site, the Chaguana River Basin, is covered by 4 printed topographical sheets: Machala, Tendales, Uzhcurrumi and Ponce Enriquez. The basin is enclosed in a rectangle whose boundary coordinates are shown in Table 3.4. The basin's centroid is located at 641000 E and 9647000 N. The maximum recorded elevation on the basin is 3267 meters and the minimum elevation is 1 meter above sea level. The banana sector in the basin is located below 60 meters and above 4 meters elevation levels.

Table 3.4. River basin boundary locations

Boundary	North	East	South	West
UTM Coordinate	9658919 N	656048 E	9634843 N	625952 E

3.4.2. GENERATION OF DIGITAL ELEVATION MODEL AND OTHER NEEDED TOPOGRAPHICAL CHARACTERISTICS FOR THE STUDY SITE

The Digital Elevation Model (DEM) is raster type data that contains spatially distributed elevation information to allow an automatic delineation of watersheds[12]. DEM for the Chaguana Basin was not available at the time of the study; thus, it was necessary to generate it based on existing topographical maps following the procedures bellow:

[12] Watershed delineation is based on the eight-direction pour point model (Jenson and Dominque 1988)

1. *Digital Conversion of Printed Maps*: Topographical information was converted into digital data by first scanning topographical sheets and then geo-referencing scanned images.
2. *Extraction of Elevation Data*: All elevation contours were digitized into a polyline coverage called *CONTOURS*. Polyline vertexes were then converted into an elevation point coverage by using the EDIT TOOL extension (Tchoukanski 2002). It was necessary to edit the elevation layer to obtain the best terrain representation, which includes manual extrapolation of additional elevation points and addition of some measured terrain levels. Scattered points were also added manually to adjust the shape of the derived river streams to the shape of existing ones[13].
3. *Cell Size Selection*: Before generating a raster image such as a DEM, it is important to define the size of the cell containing elevation data within the raster format. Dealing with watershed delineation, a rule called *thousand-million*[14] is usually applied to obtain the minimum recommended cell size for the watershed assessment (Maidment 1996). For the current watershed assessment, the regional area enclosing the basin is 750×10^6 m^2 (≈ 25 km × 30 km). Therefore, the minimum recommended cell size was 750 m^2 (≈ 27 m × 27 m). However, the cell size was defined as 10000 m^2 (100 m × 100 m) because the majority of agricultural practice management in Ecuador is performed on one-hectare basis. The selected cell size will be used in all generated raster maps.
4. *Data Interpolation*: Finally, the DEM is generated by using an interpolation method with all elevation point data. The selected method was the *Universal Kriging Interpolation Procedure* included in the KRIGING INTERPOLATOR 3.2 extension (Boeringa 2002). The main problem with interpolation is the potential creation of systematic errors in the generated surface. Therefore, the interpolation procedure should be repeated as many times as necessary together with the point layer editing process and river stream generation in order to obtain the best DEM for the study site. Figure 3.2 shows the generated digital elevation model (*FILLEDEM100* raster file) with a 100 m cell size.

[13] Further explanation on stream derivation from DEM will be explained in the hydrological data processing section 3.5.

[14] The rule states that the regional area enclosing the evaluated watershed can be divided down to one million cells without compromising too much effort in the assessment. Further watershed subdivision might not add more precision to the watershed assessment. In addition, the minimum watershed area that could be delineated in that region is obtained by multiplying one thousand times the cell area.

Once the DEM is obtained, other important topographical data for modelling purposes can be obtained by using accepted equations within Avenue statements and macro tools (Map Calculator) in ArcView.

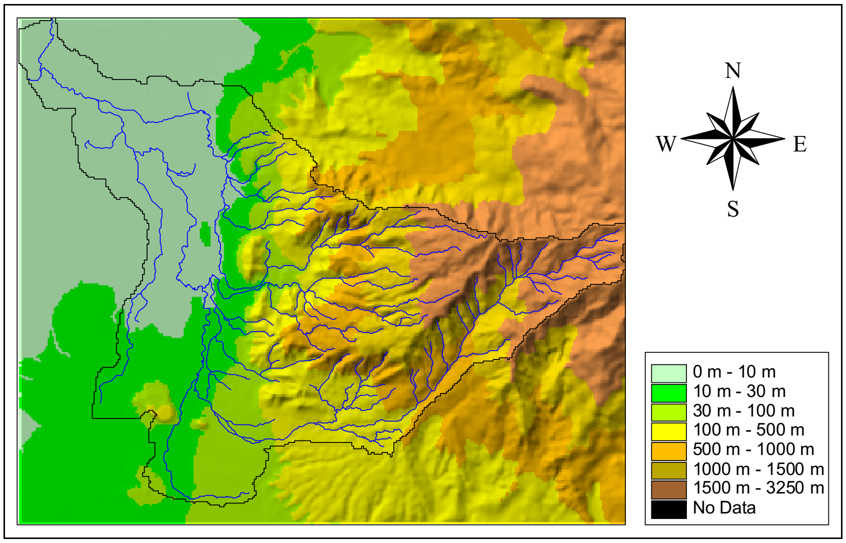


Figure 3.2. Digital Elevation Model with a 100 m cell size

Terrain Slope: The terrain slope angle representation (*SLOPE_PERCENT* raster file) can be generated by using AVENUE statements. The value of each cell is represented as percent rise. Figure 3.3 shows the generated slope raster image. The applied Avenue statement was

SLOPE_PERCENT = FILLEDEM100.Slope(Nil,TRUE) [3.1]

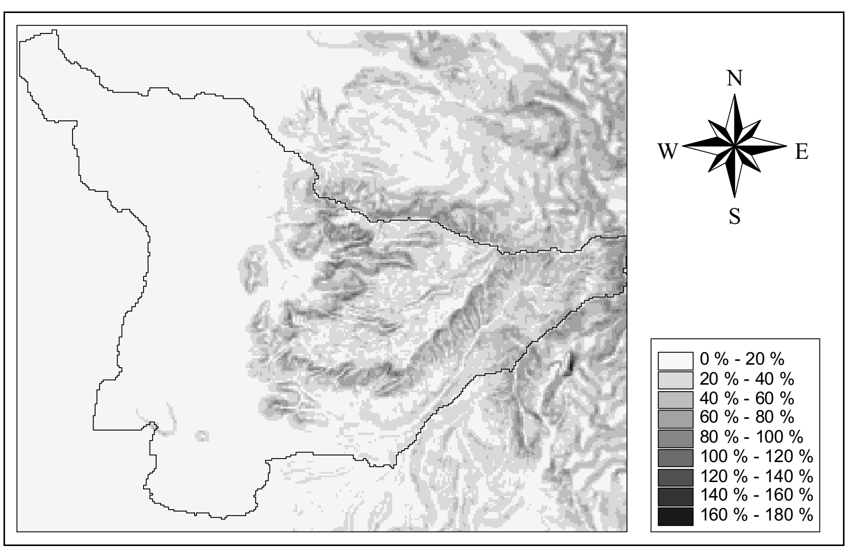


Figure 3.3. Slope percent of the Chaguana river basin

3.5. HYDROLOGICAL DATA

3.5.1. EXISTING FLOW GAUGING STATIONS

When modelling chemicals in rivers, the hydrological data is a very important piece of information to help in quantifying chemical amounts transported by the river flow. In the current research, it was possible to obtain some hydrological information (periodical water flows) in three existing gauging stations from the former National Institute of Water Resources[15]. The locations of the existing stations are shown in Figure 3.4.

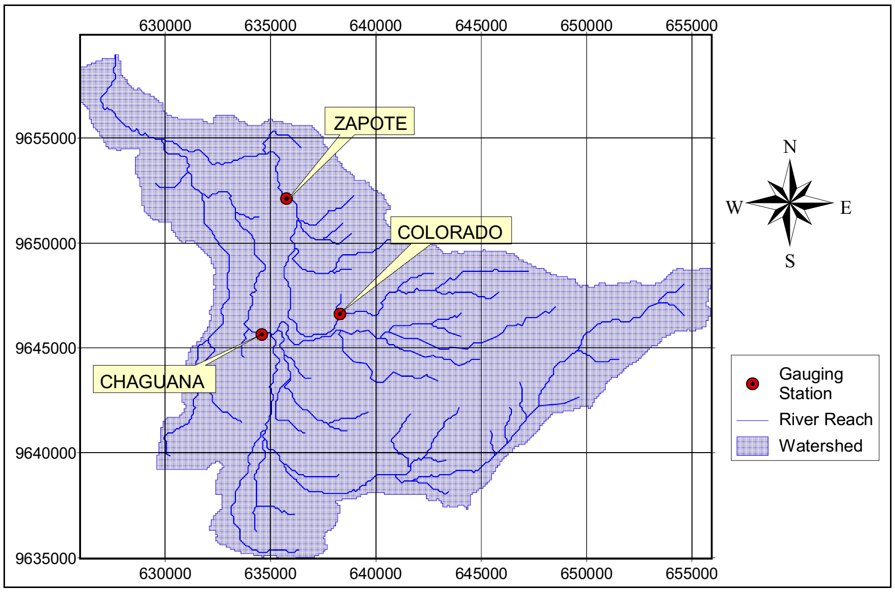


Figure 3.4. Location of existing gauging stations

However, there are some drawbacks related to these records.

- Apparently, the Water Agency is no longer measuring flow data on those stations. The supplied hydrological data only represents a 4-year non-continuous period of measurements (1978 – 1980 and 1982 – 1983).
- The data represent only average monthly flows.
- Some monthly flows are missing at all stations (probably not measured)
- There is no reliable information regarding the measurement method. However, there is a possibility that the flows have been measured by using the existing level marks attached to the bridge's piles at the gauging points (photo 3.1).

[15] INERHI (Spanish acronym) was changed by law in 1994 to the National Council of Water Resources (CNRH)

Photo 3.1. Level marker attached to a bridge's pile

For that reason, the available data was statistically processed to obtain average flow values to be used only in the hydrological calibration of the model. Table 3.5 shows median values of recorded flows within the historical period. The reason of using the median is explained later in the evaluation of meteorological records. From the table, April is the month with the highest flows recorded in the basin. This is strongly related with the rainy season in Ecuador lasting from January to May (see Section 3.6).

Table 3.5. Median flow measurements on existing gauging stations (1979 – 1982)

Gauging Station	***Drainage Area (Ha)***	***Measured monthly flows (m^3/s)***											
		JAN	***FEB***	***MAR***	***APR***	***MAY***	***JUN***	***JUL***	***AUG***	***SEP***	***OCT***	***NOV***	***DEC***
Zapote	5772	0.704	0.491	1.125	1.161	0.756	0.483	0.596	0.451	0.363	0.297	0.690	0.669
Colorado	2422	0.350	0.551	0.731	1.813	0.859	0.478	0.513	0.306	0.252	0.284	0.209	0.268
Chaguana	16199	0.762	1.440	1.646	4.453	1.621	1.121	0.619	0.571	0.346	0.481	3.470	1.173

3.5.2. FIELD FLOW AND CROSS SECTION MEASUREMENTS DONE IN THE RESEARCH

Due to limitations in the flow data, a measurement campaign was planned at some river points to obtain in-situ velocities and cross section areas. These measured data were also used as part of another research whose objective was the hydraulic characterisation of the Chaguana River by using the HEC – RAS model (Vivas 2004). The location of measured river points is shown in Figure 3.5. The monitored points were selected mainly on the basis of accessibility. The campaign was conducted in July 2002.

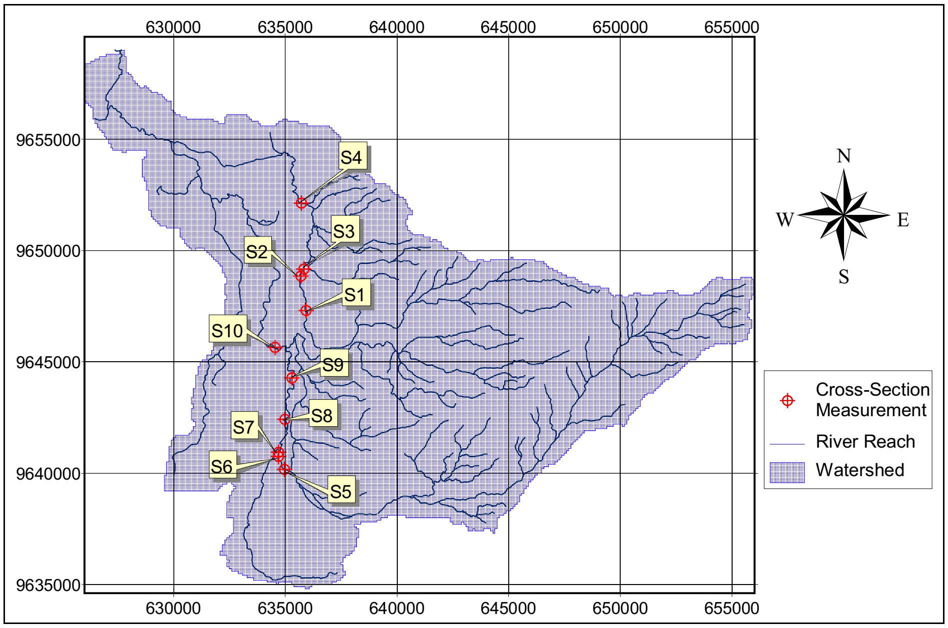


Figure 3.5. Location of measured cross-sections during July 2002

Velocities were measured with a GENERAL OCEANICS Digital Flowmeter, Model No. 2035MKIV. To obtain a representative velocity, the probe was located at one third of the river depth on several places in the same cross section. Then, an average velocity was estimated for the specific cross section. The cross section profiles were obtained by topographic measurements (photo 3.2). Measured data and calculated flows are shown in Table 3.6.

Photo 3.2. Measurement of the cross section S1 in the Zapote river

Table 3.6. Flow estimations at Chaguana river basin during July 2002

River	*Monitored Point*	*Cross Section (m^2)*	*Average Velocity (m/s)*	*Estimated Flow (m^3/s)*
Zapote	S1	1.55	0.320	0.49
	S2	2.03	0.293	0.59
	S3	3.70	0.164	0.61
	S4	3.34	0.226	0.76
Chaguana	S5	2.72	0.222	0.60
	S6	1.23	0.500	0.62
	S7	3.68	0.289	1.07
	S8	1.91	0.651	1.24
	S9	3.03	0.455	1.38
	S10	7.92	0.224	1.77

3.5.3. Watershed Delineation and Other Hydrological Data Needed for Selected Models

Flow data is mainly represented as vector data. However, some hydrological information should be spatially distributed (raster format) for modelling purposes. The majority of those data can be generated from topographical information by using GIS techniques as shown in Figure 3.6.

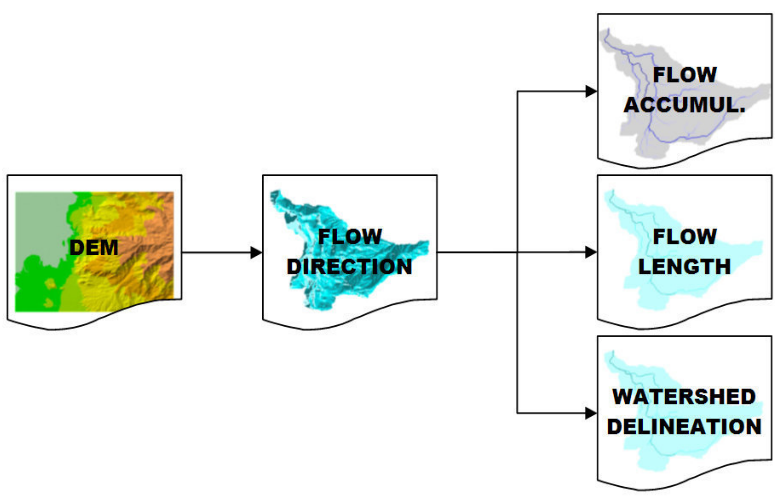


Figure 3.6. Generation of other hydrological data

In this figure, the flow accumulation map generated from the DEM actually shows the cumulative area that drains into a specific cell. By applying colour ramps to the cell values, the map could resemble a river network. However, there are some deviations from the actual river system that can be produced because of:

- **DEM resolution:** The river line in the vector format has actually no real width. The raster cell size is significantly bigger than the visual representation of that line. Thus, the bigger the size of the cell in the DEM, the greater the deviation in the generated river network. In the current research, the cell's size was fixed to 1 hectare as explained in Section 3.4.2.
- **Accurate Topography:** Actual depressions and elevations of the river area may not be well represented by the digital elevation model. Thus, the generated DEM should be adjusted to reflect actual topographical elements at specific sites. There are several methodologies to overcome this problem such as *the burn-in river* method (Maidment 1996). However, such method did not produce good results in the current research. Instead, elevation cell values were changed manually at some areas to force the generated river network to fall within the actual river line.

Figure 3.7 shows a comparison of the same basin spot between the original generated river network and the corrected one by applying the manual adjustments to the Digital Elevation Model. In the figure an extreme case is showed: a conflictive zone where two independent river streams run almost parallel for about 3 km.

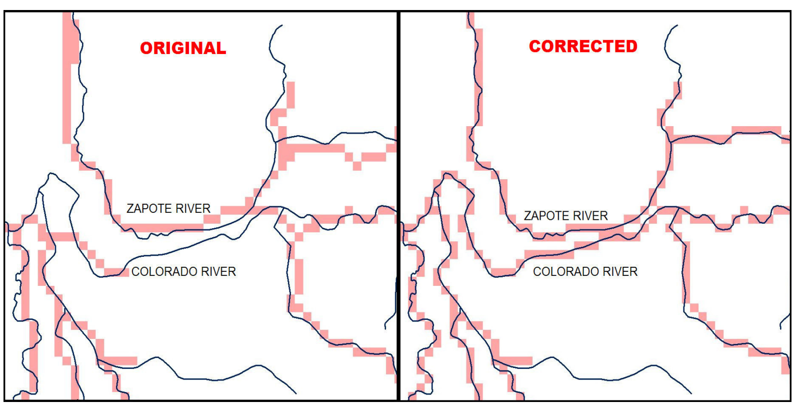


Figure 3.7. Impact of DEM adjustments over river network generation

In the figure, the darker line represents the actual river network while the connected squares represent the generated network. The flow of both rivers (Zapote and Colorado) is from right to left. The original generated network (left) represents the upper part of the Colorado River incorrectly as discharging into the Zapote River. The problem is solved by manual adjustment of the DEM as shown in the corrected river network (right). When modelling non-point sources, this problem must be solved first; otherwise, drainage areas and environmental concentrations could be under- or overestimated.

3.6. METEOROLOGICAL DATA

3.6.1. EXISTING METEOROLOGICAL STATIONS

Meteorological information was gathered among INAMHI[16] weather stations, which are located nearby the Chaguana river basin. However, none of the stations is inside the study site. Historical records on the stations are not continuous, except for one station, and even the recorded period is not the same from one station to another. Table 3.7 shows the nine meteorological stations that potentially could be considered in the river basin and Figure 3.8 depicts the location of the stations related to the watershed. In the table, the distance between each station and the centroid of the basin is also shown. The MACHALA station has the most detailed available information in the area with 27 years of continuous records and measurement of 8 meteorological parameters.

Table 3.7. General description of meteorological stations in the study area

Station	*Distance from Basin's center (Km)*	*Easting (m)*	*Northing (m)*	*Elevation (m)*	*Record Period*	*Type of Data*
Pagua	14.34	636936	9660648	13	1982 – 2000	Monthly precipitation Maximum, minimum and average temperature
Pasaje	16.06	635332	9631874	16	1982 – 1999	Monthly precipitation Maximum, minimum and average temperature
Uzhcurrumi	21.63	657402	9632796	290	1976 – 1999	
Machala - UTM	22.44	621504	9635790	5	1973 – 1999	Monthly precipitation Maximum 24h precipitation Maximum, minimum and average temperature Average relative humidity Sun hours and cloudiness Maximum wind speed and direction
Tenguel	23.25	634815	9669311	10	1965 – 1971 1973 – 1976 1979 – 1980	Monthly precipitation
Tendales	30.89	668850	9633547	750		Monthly precipitation

[16] Instituto Nacional de Meteorología e Hidrología (Ecuadorian National Institute of Meteorology and Hydrology) is the official organization in keeping meteorological records.

Station	Distance from Basin's center (Km)	Easting (m)	Northing (m)	Elevation (m)	Record Period	Type of Data
Huertas	46.49	652199	9601784	1530	1971 – 1977, 1980 1982 – 1986, 1988 1990 – 1991	Monthly precipitation
Santa Isabel	47.51	687718	9638279	1550	1965 – 1987	Monthly precipitation Maximum 24h precipitation Average temperature Average relative humidity Sun hours and cloudiness
La Cuca	49.33	603918	9614373	20	1982 – 1989	Monthly precipitation Minimum and average temperature Sun hours and cloudiness Average relative humidity

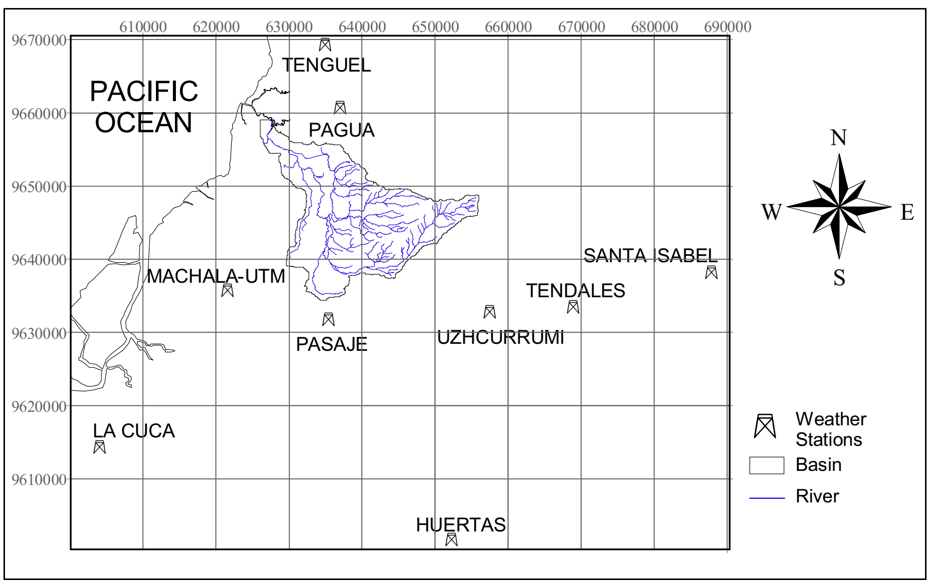


Figure 3.8. Location of available weather stations surrounding the Chaguana basin

3.6.2. MONTHLY PRECIPITATION DATA

The gathered meteorological data were used for the hydrological assessment of the river basin. However, the scope of the analysis is restricted to the average flow condition in the river system within a typical year of precipitation because the only available precipitation data are mean monthly values in all meteorological stations. The existing data is not sufficient at all, though it is the only available data to make the assessment. The actual data show no continuity, and records are not measured during the same periods from one station to another. Thus data selection is a critical step in this basin assessment.

The ideal situation to evaluate a basin's hydrology is to have meteorological stations within the evaluated basin, but that is not always the case. In such cases, some hydrological methods allow the use of data (daily, monthly or yearly data) from nearby stations only, whereas other methods use data from all stations within a certain radius of influence from the basin's centroid. The main drawback of these methods is that the resulting estimated precipitation data is assumed to be the same for the entire basin. In addition, recorded data include both normal and extreme (very high rainfall) events. In order to get a value representative from the whole data set of a station, the median value is used because it is not influenced by very high rain values as the average value is. Finally, the more the stations are geographically distributed, the more the estimated precipitation value is representative for the basin.

For the Chaguana river basin, two stations are nearest to the centroid of the basin (within a radius of 20 km), Pagua and Pasaje (Table 3.7). Their common period of data recording represents 18 years of measurements (1982-1999). Based on these two stations, Table 3.8 shows the mean (median) monthly values estimated for the entire basin. The total annual estimated precipitation is also shown.

Table 3.8. Mean monthly precipitation (mm) estimated from stations located in a radius of 20 km and data recorded between 1997 and 2000.

Station	*Jan*	*Feb*	*Mar*	*Apr*	*May*	*Jun*	*Jul*	*Aug*	*Sep*	*Oct*	*Nov*	*Dec*	*Annual*
Pagua	185.3	230.9	197.5	121.3	29.0	38.9	42.8	30.9	39.7	53.2	34.4	44.7	1048.4
Pasaje	100.9	223.6	211.3	96.6	27.9	30.6	26.6	26.4	26.8	41.4	37.3	53.4	902.8
MEAN	***143.1***	***227.2***	***204.4***	***109.0***	***28.5***	***34.8***	***34.7***	***28.7***	***33.2***	***47.3***	***35.9***	***49.1***	***975.6***

Now considering a radius of influence of 31 km from the centroid of the basin, 4 additional stations could be evaluated to obtain mean monthly precipitation for the entire basin during the same record periods: Machala, Uzhcurrumi, Tenguel and Tendales. By analyzing the record periods, there are three common periods: 1973 – 1976 (Machala, Tendales and Tenguel Stations), 1976 – 1983 (Machala, Tendales and Uzhcurrumi), and 1982 – 1999 (Pagua, Pasaje, Uzchcurrumi and Machala Stations), as shown in tables 3.9, 3.10 and 3.11 respectively.

Table 3.9. Mean monthly precipitation (mm) estimated from stations located in a radius of 31 km and data recorded in 1973-1976.

Station	*Jan*	*Feb*	*Mar*	*Apr*	*May*	*Jun*	*Jul*	*Aug*	*Sep*	*Oct*	*Nov*	*Dec*	*Annual*
Tenguel	110.8	352.7	319.6	102.1	75.9	37.9	43.1	24.9	41.1	54.5	43.7	35.3	1241.6
Machala	74.1	222.2	196.0	37.2	26.2	12.7	9.7	16.1	20.9	21.9	14.3	17.8	669.1
Tendales	72.3	156.0	137.9	74.0	43.6	46.2	17.6	12.8	19.9	10.9	15.9	30.4	637.3
MEAN	***85.8***	***243.6***	***217.8***	***71.1***	***48.6***	***32.3***	***23.5***	***17.9***	***27.3***	***29.1***	***24.6***	***27.8***	***849.3***

Table 3.10. Mean monthly precipitation (mm) estimated from stations located in a radius of 31 km and data recorded in 1976-1983.

Station	*Jan*	*Feb*	*Mar*	*Apr*	*May*	*Jun*	*Jul*	*Aug*	*Sep*	*Oct*	*Nov*	*Dec*	*Annual*
Machala	92.4	155.9	106.6	78.2	29.8	47.5	33.9	39.3	36.9	79.3	44.6	57.5	801.8
Uzhcurrumi	87.3	98.9	149.2	76.9	33.6	17.1	14.0	10.6	14.6	17.8	21.2	52.0	593.0
Tendales	91.0	74.4	154.1	69.0	58.0	39.9	27.2	10.0	11.4	30.9	65.3	64.3	695.2
MEAN	***90.2***	***109.7***	***136.6***	***74.7***	***40.5***	***34.8***	***25.0***	***20.0***	***20.9***	***42.7***	***43.7***	***57.9***	***696.6***

Table 3.11. Mean monthly precipitation (mm) estimated from stations located in a radius of 31 km and data recorded in 1982-1999.

Station	*Jan*	*Feb*	*Mar*	*Apr*	*May*	*Jun*	*Jul*	*Aug*	*Sep*	*Oct*	*Nov*	*Dec*	*Annual*
Pagua	185.3	230.9	197.5	121.3	29.0	38.9	42.8	30.9	39.7	53.2	34.4	44.7	1048.4
Pasaje	100.9	223.6	211.3	96.6	27.9	30.6	26.6	26.4	26.8	41.4	37.3	53.4	902.8
Machala	183.7	253.9	259.1	110.3	37.3	45.9	42.7	38.0	46.9	69.6	53.3	48.0	1188.5
Uzhcurrumi	117.1	117.8	135.4	99.1	34.9	24.6	15.9	11.9	12.6	23.3	26.2	71.5	689.9
MEAN	***146.7***	***206.5***	***200.8***	***106.8***	***32.3***	***35.0***	***32.0***	***26.8***	***31.5***	***46.9***	***37.8***	***54.4***	***957.4***

By considering a radius of 50 km, three additional stations are available: La Cuca, Huertas and Santa Isabel. The two last ones are located around 1500 meters above sea level; the other one is located around 20 meters. Within this influence circumference, three common periods can be analyzed, which represent more than 4 years of records: 1965 – 1971 / 1973 – 1976 / 1980 (Tenguel and Santa Isabel Stations), 1973 – 1976 / 1980 (Machala, Tenguel, Tendales, Huertas and Santa Isabel Stations), and 1982 – 1986 (Pagua, Pasaje, Machala, Uzhcurrumi, Huertas, La Cuca and Santa Isabel Stations). Tables 3.12, 3.13 and 3.14 show the mean monthly precipitation for each evaluated scenario.

Table 3.12.Mean monthly precipitation (mm) estimated from stations located in a radius of 50 km and data recorded in 1965-1971, 1973-1976 and 1980.

Station	*Jan*	*Feb*	*Mar*	*Apr*	*May*	*Jun*	*Jul*	*Aug*	*Sep*	*Oct*	*Nov*	*Dec*	*Annual*
Tenguel	99.0	109.8	103.9	72.6	62.1	31.0	18.6	24.5	41.1	47.9	30.9	29.2	*670.4*
Santa Isabel	48.0	66.3	76.4	58.7	25.0	13.1	4.8	6.1	13.9	23.7	22.2	34.1	*392.2*
MEAN	***73.5***	***88.1***	***90.1***	***65.7***	***43.5***	***22.3***	***11.7***	***15.3***	***27.5***	***35.8***	***26.6***	***31.6***	***531.3***

Table 3.13.Mean monthly precipitation (mm) estimated from stations located in a radius of 50 km and data recorded in 1973-1976 and 1980.

Station	*Jan*	*Feb*	*Mar*	*Apr*	*May*	*Jun*	*Jul*	*Aug*	*Sep*	*Oct*	*Nov*	*Dec*	*Annual*
Machala	91.4	182.9	171.8	53.5	40.4	18.1	14.6	16.2	27.7	27.3	18.6	19.0	*681.4*
Tenguel	94.1	291.0	211.8	107.2	60.9	37.9	35.5	23.6	42.6	54.2	43.7	35.3	*1037.8*
Tendales	74.5	145.4	144.9	68.9	62.0	60.7	24.8	13.0	20.5	17.0	21.3	33.6	*686.6*
Huertas	215.4	381.4	530.1	372.6	246.2	89.6	18.1	17.5	49.8	28.1	34.2	117.8	*2100.7*
Sta. Isabel	46.0	110.3	136.4	66.1	76.0	19.4	5.6	26.7	8.3	28.8	24.4	34.1	*582.2*
MEAN	***104.3***	***222.2***	***239.0***	***133.7***	***97.1***	***45.1***	***19.7***	***19.4***	***29.8***	***31.1***	***28.4***	***48.0***	***1017.7***

Table 3.14. Mean monthly precipitation (mm) estimated from stations located in a radius of 50 km and data recorded in 1982-1986.

Station	*Jan*	*Feb*	*Mar*	*Apr*	*May*	*Jun*	*Jul*	*Aug*	*Sep*	*Oct*	*Nov*	*Dec*	*Annual*
Pagua	208.0	182.0	276.6	136.1	15.2	103.5	180.7	28.7	54.0	54.0	39.8	38.9	1317.4
Pasaje	58.1	139.4	225.9	142.4	25.3	40.1	22.3	18.1	25.3	43.0	62.6	129.2	931.5
Machala	95.4	229.0	268.7	74.8	31.3	66.6	41.0	36.5	34.5	71.8	57.4	76.2	1083.1
Uzhcurr.	68.1	63.8	143.4	97.9	28.9	14.1	15.4	12.9	15.3	34.5	22.3	103.1	619.7
Huertas	394.8	569.2	318.5	444.0	180.3	23.5	59.8	0.0	29.8	139.9	129.5	368.5	2657.8
La Cuca	27.1	152.3	79.3	31.1	8.2	5.4	6.0	4.6	7.9	9.5	0.7	17.2	349.2
Sta. Isabel	42.6	65.1	74.3	88.2	45.1	1.2	2.2	2.0	16.5	23.7	29.3	58.9	449.1
MEAN	***125.6***	***215.8***	***176.8***	***147.2***	***58.8***	***22.2***	***24.9***	***11.2***	***20.8***	***55.9***	***47.8***	***124.8***	***1031.8***

The tables show how the mean annual precipitation for the basin could vary depending on the number of stations included in the analysis and the number of common recorded climate data. Table 3.15 shows how the estimated precipitation for the basin could vary from high to low events depending on the stations involved in the analysis and the evaluated recording period.

Therefore, the right scenario for the so called "normal" event could be selected based on a probability estimation[17] to have the same or higher annual precipitation than the one in the respective scenario (chance of 50% to have higher events).

Another important aspect of the analysis is the influence of extreme events such "El Niño" which is a period where very heavy rains occur in Ecuador. Normally this recurrent event begins at the middle of one year and ends at the middle of next year. Thus, data are normally influenced during two consecutives years by an "El Niño" event from time to time. Reported "El Niño" events were recorded in 1965-1966, 1968-1970, 1972-1973, 1976-1977, 1982-1983, 1986-1988, 1991-1992, 1994-1995, and 1997-1998 (Villacis *et al.* 2001). Therefore, this extreme event has a recurrence interval of 3 to 6 years. The table also includes some El Niño events in their data range periods for the evaluated scenarios.

Table 3.15. Evaluation of precipitation obtained in different scenarios (involved stations and recorded periods).

Scenario	*Evaluated Weather Stations*	*Period of records measured in weather stations*	*Annual Precipitation (mm/year)*	*Probability to have an event greater or equal*	*Classification of the event*
1	Pagua, Pasaje, Machala, Uzhcurrumi, Huertas, La Cuca, and Santa Isabel	1982 – 1986	1031.77	12.50%	High (humid)
2	Machala, Tenguel, Tendales, Huertas, and Santa Isabel	1973 – 1976, 1980	1017.70	25.00%	⇑
3	Pagua and Pasaje	1982 - 1999	975.58	37.50%	
4	Pagua, Pasaje, Machala, and Uzhcurrumi	1982 – 1999	957.39	50.00%	⇐ Mid (normal)
5	Machala, Tenguel, and Tendales	1973 – 1976	849.31	62.50%	⇓
6	Machala, Uzhcurrumi, and Tendales	1976 – 1983	696.64	75.00%	
7	Tenguel and Santa Isabel	1965 – 1971, 1973 – 1976, 1980	531.3	87.50%	Low (dry)

[17] $$P(i) = \frac{i}{n+1} \quad [3.2]$$

where ***i*** is the position in an ascending or descending arrangement of ***n*** values.

Based on the previous analysis, it can be concluded that scenario 4, which represents 18 years of recorded data, could resemble normal precipitation events because there is a 50% chance that the event could be higher or lower. In addition, the weather stations in this scenario are fairly distributed around the basin (Figure 3.8). For that reason, scenario 4 is the best combination of weather stations and recorded period to represent a "normal" precipitation event for the Chaguana river basin.

Map Interpolation Procedures

The monthly precipitation values estimated previously are considered as unique valid values for the entire basin allowing for spatial variability from evaluating weather stations only in an implicit form (the value changes only when another station is selected). In modelling, it is sometimes better to explicitly show spatial variability in parameters such as climatic records. This goal is achieved by interpolating data from meteorological stations. However, in this Ph.D. study, interpolation is restricted to a few methods because there are not so many available data (less than 9 stations). Based on the selected scenario, table 3.11 shows the monthly average precipitation data used for the interpolation process for obtaining a normal monthly precipitation map.

The selection of a suitable interpolation methodology for the available data is critical in order to obtain a non-distorted spatial distribution of any evaluated parameter (Mitas and Mitasova 1999). A wrong interpolation procedure could lead to false distribution patterns in the simulations (Mitasova *et al.* 1996). Based on previous works done elsewhere (Hutchinson and Bichof 1983; Hutchinson 1996; Hutchinson and Corbett 1993), the *SPLINE* interpolation method appears the most recommended for processing the climatic data. To obtain the precipitation raster maps of every month, the *SPLINE* interpolation option in ArcView[18] was applied to the stations selected to produce a normal precipitation event. Therefore, a raster precipitation map was generated for every month in a "normal" year (Figure 3.9). In addition, the total annual precipitation was generated by adding all monthly maps. Table 3.16 shows the average values for each raster map and its spatial standard deviation. From the table and figure, the wettest month in a typical year is February with 205 mm and the driest month is August with 25 mm of average rain in the whole basin.

[18] A weight of zero is selected to produce a basic Thin Plate Smoothing Spline which is the recommended interpolation for climatic events.

Table 3.16. Total monthly average and annual precipitation for Chaguana river basin

	Jan	*Feb*	*Mar*	*Apr*	*May*	*Jun*	*Jul*	*Aug*	*Sep*	*Oct*	*Nov*	*Dec*	*Annual*
P_{mean}	152.9	205.2	178.6	113.4	28.2	33.0	34.1	25.2	30.4	41.9	37.0	50.9	*930.6*
Stand. Dev.	19.6	36.5	40.2	4.7	1.8	5.8	6.1	7.1	8.8	13.0	7.7	4.9	*132.3*

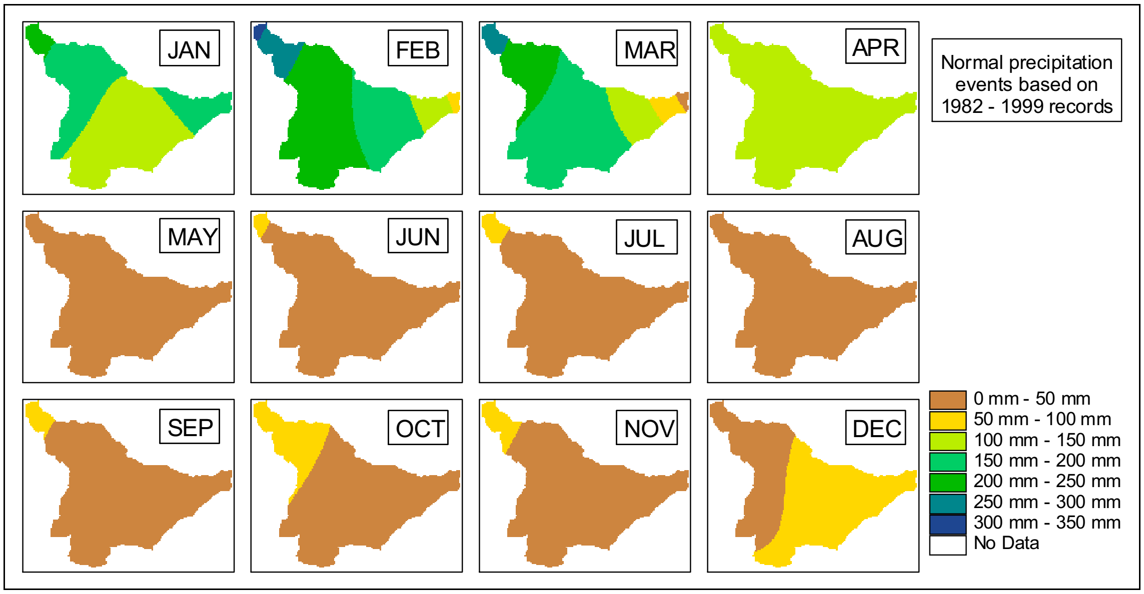


Figure 3.9. Monthly precipitation maps for Chaguana river basin

From figure 3.9, it can be seen that the maps predict more precipitation in the lowlands (west side of the basin) than in the highlands (east side) which might be considered as an uncommon behaviour in a typical precipitation pattern. However, there are two important facts that should be considered when interpreting the interpolated maps:

- The maps do not show the impact of high elevations, because the highest stations (Huertas and Santa Isabel) were not considered in the interpolation procedure. Those stations did not have enough rainfall data to support the interpolation procedure as mentioned before.
- Another important issue is that existing data in the highest stations could be biased as most of their data are influenced by the presence of an El Niño event (extremely high precipitation levels). By reviewing the rain history (interviewing local people) in the areas where the highest stations are, the southern part of Ecuador and the northern part of Peru are very dry (even in the mountain regions).

3.6.3. MAXIMUM 24-HOUR PRECIPITATION

The maximum 24-hour precipitation is a climatic parameter that is useful to obtain other parameters such as the rainfall erosivity factor, peak discharge and so on. In the Chaguana river basin, only three weather stations have records on this parameter: Machala, Pagua and Pasaje. However, these stations do not show the same recording periods for this parameter to produce a representative raster map[19] for a normal event. To overcome this problem, the Gumbel distribution (Benjamin and Cornell 1970) was applied to every monthly value in each station by assuming that these data are recorded continuously through a number N of years.

An EXCEL worksheet was used to evaluate the Gumbel distribution of 24-hour precipitation data in each station. The following equations are used in the calculations.

1. The arithmetic mean 24-hour precipitation (X_{month}) is determined for each month of the year.

$$X_{month} = \frac{\sum_{i=1}^{N} P_{month}(i)}{N} \qquad [3.3]$$

where $P_{month}(i)$ is the 24-hour precipitation for a specific month in year i.

2. The standard deviation (S_{month}) of sample data is determined for each month of the year.

$$S_{month} = \sqrt{\frac{\sum_{i=1}^{N} \left[P_{month}(i) - X_{month}\right]^2}{N-1}} \qquad [3.4]$$

3. Determination of Gumbel distribution fitting parameters: mode (μ) and dispersion (α)

$$\alpha = \frac{\pi}{S_{month}\sqrt{6}} \qquad [3.5]$$

$$\mu = X_{month} - \frac{\gamma}{\alpha} \qquad [3.6]$$

[19] For the interpolation procedure, values used in the stations should represent the same period of measurement.

4. The probability is then established to have the same or greater rain event after a recurrence interval of precipitation (T_{RETURN})[20] occurs. Then, a variate (ω) value is calculated based on this estimated probability.

$$\text{Probability} = \frac{1}{T_{RETURN}} \quad [3.7]$$

$$\omega = -\ln\left[-\ln\left(1 - Probability\right)\right] \quad [3.8]$$

Table 3.17 shows some calculated values of probability and variate for a recurrence interval of precipitation. The return period values shown in the table are the most commonly used in rain analysis.

Table 3.17. Probability – variate values based on a specific recurrence interval of precipitation

Return Period	Probability of event occurrence	Variate (ω)
2 years	0.50	0.3651
5 years	0.20	1.4994
10 years	0.10	2.2504
25 years	0.04	3.1985

5. Finally, a monthly precipitation for a specific recurrence interval is estimated based on the Gumbel fitting parameters for each month (μ,α) and the corresponding variate calculated for that interval. Based on the probability shown in table 3.17, the 2-year precipitation event could be considered as a "normal" event because there is a 50% chance to have greater or lower events.

$$P(i) = \mu + \frac{\omega(i)}{\alpha} \quad [3.9]$$

Tables 3.18, 3.19 and 3.20 show the calculated values of maximum 24-hour precipitation for each month in Machala, Pasaje and Pagua stations respectively. Those predicted values were interpolated to produce raster maps of maximum 24-hour precipitation for each month in every recurrence interval (2 years, 5 years, 10 years and 25 years) as shown in Figure 3.10.

[20] The return period (T_{RETURN}) is the average interval in years between the occurrence of an event of stated magnitude and an equal or more serious event.

Table 3.18. Estimated maximum 24-hour precipitations for each recurrence interval in Machala station

T_{RETURN}	*Jan*	*Feb*	*Mar*	*Apr*	*May*	*Jun*	*Jul*	*Aug*	*Sep*	*Oct*	*Nov*	*Dec*
2 years	57.1	70.6	60.5	37.0	19.7	14.0	8.7	7.8	8.3	8.5	12.1	25.8
5 years	106.5	114.4	94.3	75.4	41.9	40.1	15.4	16.0	15.8	11.6	29.1	60.3
10 years	139.2	143.5	116.7	100.9	56.6	57.3	19.8	21.5	20.7	13.8	40.3	83.1
25 years	180.5	180.2	145.1	133.0	75.1	79.1	25.4	28.3	27.0	16.4	54.6	111.9

Table 3.19. Estimated maximum 24-hour precipitations for each recurrence interval in Pasaje station

T_{RETURN}	*Jan*	*Feb*	*Mar*	*Apr*	*May*	*Jun*	*Jul*	*Aug*	*Sep*	*Oct*	*Nov*	*Dec*
2 years	38.1	29.8	51.3	42.7	12.8	9.6	6.1	2.8	5.9	8.1	18.9	45.4
5 years	96.1	55.6	85.1	86.0	35.8	19.7	12.3	3.8	12.2	15.8	39.3	59.6
10 years	134.5	72.7	107.5	114.7	51.0	26.3	16.4	4.5	16.4	20.9	52.8	69.0
25 years	183.0	94.3	135.7	150.9	70.2	34.7	21.5	5.4	21.8	27.3	69.9	80.9

Table 3.20. Estimated maximum 24-hour precipitations for each recurrence interval in Pagua station

T_{RETURN}	*Jan*	*Feb*	*Mar*	*Apr*	*May*	*Jun*	*Jul*	*Aug*	*Sep*	*Oct*	*Nov*	*Dec*
2 years	56.0	64.9	55.6	46.7	17.7	15.6	14.9	6.2	11.4	16.8	6.4	21.8
5 years	88.9	102.6	100.9	89.2	40.4	39.6	39.7	10.3	37.6	53.8	8.4	55.1
10 years	110.7	127.6	130.9	117.4	55.5	55.4	56.1	13.0	55.0	78.3	9.7	77.2
25 years	138.2	159.2	168.7	152.9	74.6	75.5	76.8	16.5	76.9	109.3	11.4	105.0

This precipitation data were used in the calculation of the rainfall erosivity factor for each station considered in the analysis, but only the 2-year and the 10-year return period which are used in the selected pesticide models.

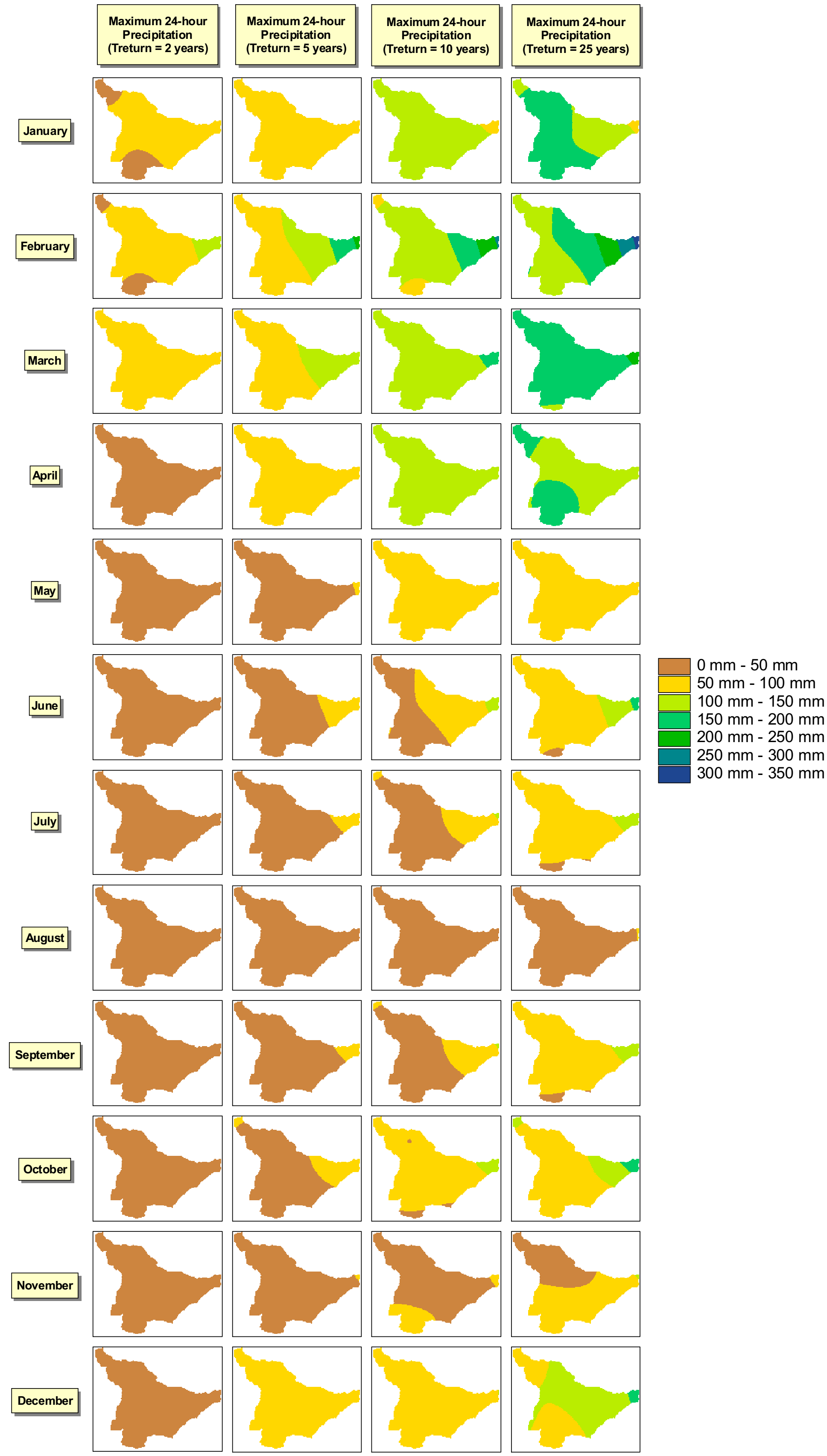


Figure 3.10. Maximum 24-hour precipitation maps for 2-, 5-, 10- and 25-years return periods

3.6.4. AIR AND DEW-POINT TEMPERATURE DATA

Air and dew-point temperature were data required for both pesticide models (SWAT and AGNPS). They can be input as monthly or daily values. Because of the restriction on data availability, only monthly values were generated. Similar to precipitation, air and dew-point temperatures generally show a tendency with topography. In the current assessment, a larger error could be expected because interpolation was done with only three stations (Machala, Pagua and Pasaje). However, this error could be within an acceptable range as explained next. The non-point source pesticide pollution (banana plantations) only affects a part of the catchment area (below 50-m terrain level). Useful stations for temperature evaluation are also located in the same area. Therefore, any error obtained above 50 m level will be irrelevant to the pesticide assessment.

Because the dew-point temperature is not usually monitored in Ecuadorian weather stations, some estimation was performed to obtain that variable. There are several approaches to estimate dew-point temperature such as

- The International Temperature Scale Procedures, established by the International Committee of Weights and Measures (ICWM 1989; Hardy 1998). Although an actual scaling method was proposed in 1990, some equipment could still have scaling methods created in 1968, 1975 or 1976. However, temperature differences between ITSP-90 and ITSP-68 are only in the range of 0.005 and 0.007 °C when the measured temperature is between 20 and 30°C. Therefore, the pesticide assessment will not be affected for the selection of any scale procedure due to this negligible difference in the current assessment context.
- An empirical relationship (Tetens 1930) that uses dry air temperature (T_{AIR}) and relative humidity (RH) to calculate dew-point temperature with ±0.1°C accuracy.
- Some software based on previous formulations such as EZAir (Parks 1998)

In the current research, the following equations based on those methodologies were applied to estimate the dew-point temperature. Those equations are

$$T_{DEW} = -\frac{\left[237.314\ln\left(\frac{E_T}{6.1078}\right)\right]}{\left[\ln\left(\frac{E_T}{6.1078}\right)\right] - 17.26902} \quad [3.10]$$

$$E_T = E_W - 0.63\left(T_{AIR} - T_{WET}\right) \quad [3.11]$$

$$E_W = 6.112\, e^{\frac{(17.67\,Twet)}{(Twet)+243.5}} \quad [3.12]$$

Where,

T_{AIR} Maximum Temperature or Dry Air Temperature

T_{WET} Minimum Temperature or Wet Bulb Temperature

T_{DEW} Dew-point Temperature

E_W Saturation Vapour Pressure over water

E_T Actual Vapour Pressure

An ArcView avenue script called Monthly DewPoint Estimation was developed to draw the corresponding raster map based on maximum (dry) and minimum (wet) temperature raster maps. The dry and wet temperature raster maps were produced with a *SPLINE WITH TENSION* interpolation performed with data from Machala, Pagua and Pasaje Stations. Table 3.21 shows the dry air (T_{AIR}) and wet-bulb (T_{WET}) temperature, as mean measurements, used in the interpolation process. The results obtained for the Chaguana basin are also shown in this table. It can be seen that temporal variation through the year on all stations and the basin estimate is quite small (less than 5ºC). In addition the spatial variation in the basin is also small for all months (less than 1.4ºC). Figure 3.11 shows temperature raster maps for a "normal" year.

Table 3.22. Measured maximum and minimum temperatures used to estimate the dew-point temperature map for the Chaguana basin

Station	*Parameter*	*JAN*	*FEB*	*MAR*	*APR*	*MAY*	*JUN*	*JUL*	*AUG*	*SEP*	*OCT*	*NOV*	*DEC*
Machala	T_{AIR} (°C)	30.5	30.6	31.2	31.2	30.2	28.6	26.9	26.6	26.8	26.0	27.2	29.4
	T_{WET} (°C)	21.9	22.3	22.7	22.4	22.1	21.1	19.9	19.9	20.1	20.5	20.5	21.4
Pagua	T_{AIR} (°C)	30.6	30.6	31.4	31.3	30.2	28.5	27.3	26.4	26.9	26.4	27.7	29.4
	T_{WET} (°C)	21.5	21.1	22.0	22.2	21.9	20.8	20.1	19.9	20.1	20.4	20.6	21.2
Pasaje	T_{AIR} (°C)	31.1	31.9	31.9	31.5	30.1	28.1	27.9	27.3	28.5	28.9	30.1	31.6
	T_{WET} (°C)	21.7	22.2	22.4	22.3	21.5	21.1	20.4	20.1	20.3	20.5	20.7	21.1
Estimated values for the basin	T_{AIR} (°C)	30.6	30.6	31.4	31.3	30.2	28.5	27.2	26.5	26.9	26.3	27.6	29.4
	T_{WET} (°C)	21.6	21.4	22.2	22.2	21.9	20.9	20.1	19.9	20.1	20.4	20.6	21.2
	T_{DEW} (°C)	17.6	17.2	18.2	18.4	18.4	17.4	16.7	16.8	16.9	17.7	17.3	17.6

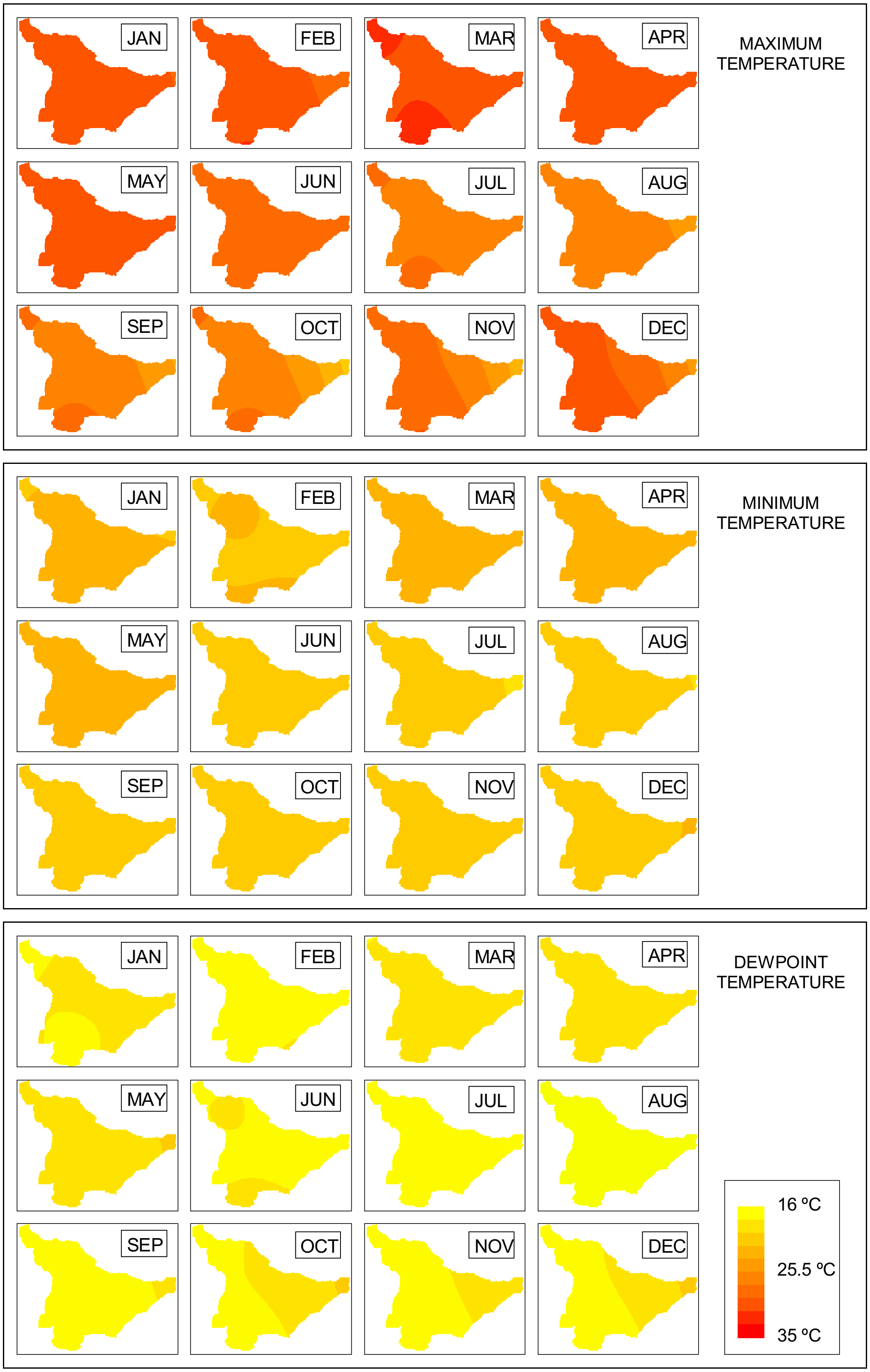


Figure 3.11. Maximum, minimum and dew-point temperature raster maps for a "normal" year

3.6.5. SOLAR RADIATION DATA

Solar radiation data are important to account for pesticide degradation due to photolysis. Only SWAT model requires solar radiation as daily or monthly information. In the present study, solar radiation data is lacking in the existing meteorological stations. Only the Machala Station has records on total effective sunshine hours for every month (Table 3.22). From this data, it can be seen that the study area is mostly covered by clouds during the entire year.

Table 3.22. Average total monthly sunshine hours for Machala station

Month	***Jan***	***Feb***	***Mar***	***Apr***	***May***	***Jun***	***Jul***	***Aug***	***Sep***	***Oct***	***Nov***	***Dec***
Total Sun Hours	65.21	61.54	96.25	93.88	78.43	43.18	40.16	39.62	32.78	21.68	37.64	57.84
Average Hours/day	2.1	2.2	3.1	3.1	2.5	1.4	1.3	1.3	1.1	0.7	1.3	1.9

In order to get solar radiation data for the study area, the following procedure was used

1. ***Gathering information from an existing global on-line database at the NASA Langley Atmospheric Sciences Data Centre:*** This database gives information on solar insolation and other parameters for a 1° grid cell (around 1.2×10^6 Ha) on a 10-year average basis. These data are not necessarily representative for a specific point within the grid cell, and estimated solar data are normally higher than ground measurements (Whitlock *et al.* 2000). In the present study, the Chaguana river basin is located in the grid within 80°E, 3°S, 79°E and 4°S coordinates. Table 3.23 shows the estimated NASA solar radiation on a 10-year average basis for this grid cell.

Table 3.23. Estimated solar radiation from the NASA Langley Atmospheric Sciences Data Center

Month	***Jan***	***Feb***	***Mar***	***Apr***	***May***	***Jun***	***Jul***	***Aug***	***Sep***	***Oct***	***Nov***	***Dec***
Rad ($MJ/m^2.day$)	16.02	16.85	16.96	16.52	16.09	14.54	15.19	16.60	17.68	17.24	18.14	17.28

2. ***Use of the LARS-WG software tool:*** Due to the overestimation mentioned in the previous paragraph, another procedure is needed to get more representative estimates for the study area. There is one software tool that uses sunshine hour measurements as input to get solar radiation estimates. This software was developed at the University of

Bristol, UK and it is called LARS-WG (Racsko *et al.* 1991; Semenov *et al.* 1998; Semenov and Barrow 1997). This tool is basically a stochastic weather generator and specifically can model solar radiation by using empirical distributions of wet and dry weather series. When there are no solar radiation measurements, sunshine hours are converted to global radiation by means of a regression relationship between these two variables (Rietveld 1978).

However, daily values of sunshine hours are needed to accomplish the weather generation. The input file in LARS-WG needs a daily record of precipitation, minimum and maximum temperature and sunshine hours for the entire period of analysis. In the present case study, only total monthly values for the record period are available. In order to overcome this problem, the following assumptions were made:

- Total monthly sunshine hours were converted to average daily sunshine hours for every month of each monitored year. This value was used as a daily value for every day in the current evaluated month.
- Total monthly precipitation was also converted to average daily precipitation. However, this value was only used as daily value for the total days in a month minus one. Because, the maximum precipitation value in every month was also available, this value was used for that day in the current month.
- In the existing records, temperature values represent the average value in the month, so this value was used as a daily value for every day in the current month.

Table 3.24 shows the average daily solar radiation generated from the Machala station weather data for every month. Figure 3.12 shows how the tendency of the solar radiation curve (left axis) follows the one of the sunshine hours' curve (right axis). Despite the good agreement, it is always better to have some solar radiation measurements to validate these estimations.

Table 3.24. Estimated average monthly solar radiation by using LARS-WG in the Machala station

Month	*Jan*	*Feb*	*Mar*	*Apr*	*May*	*Jun*	*Jul*	*Aug*	*Sep*	*Oct*	*Nov*	*Dec*
Mean Rad (MJ/m^2.day)	9.7	10.1	12.8	12.3	10.4	7.2	7.2	7.1	7.2	5.8	7.1	8.8

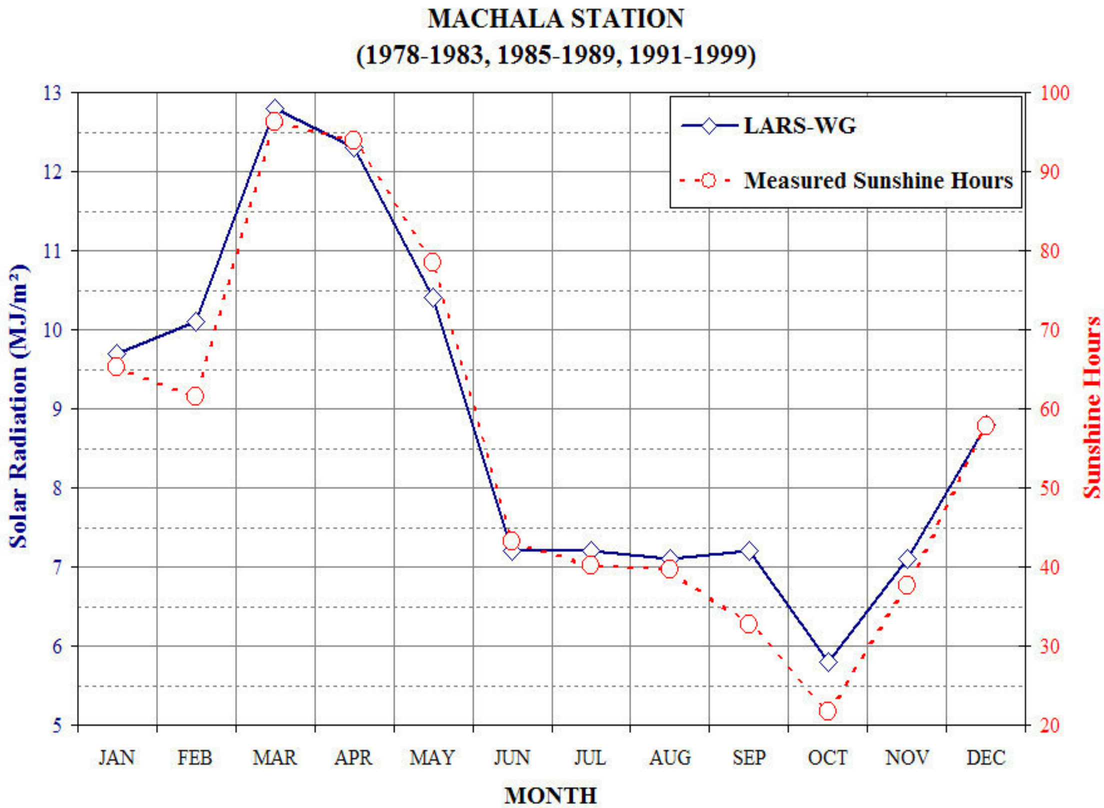


Figure 3.12. Estimated solar radiation and measured sunshine hours in the Machala station

Solar radiation is a parameter that varies not only by the weather conditions but also by the geographic latitude of a specific site. Thus, a solar radiation map would show the spatial variation of this parameter within the map. However, the current study site covers a small geographic spatial variation (less than 0.3° in latitude), which would produce only a slight variation in the parameter values. For that reason, the resulting raster map for the study area will be one with the same value in every raster cell, showing only temporal (monthly) variations and not spatial variations. Values obtained for the Machala station can be accepted as valid solar radiation values for the entire basin.

3.7. SOIL DATA

Another important group of data concerns soil information such as soil texture, permeability, and organic matter. In the Chaguana river basin, the only available information is the edaphology map which shows general information on soil taxonomic units. There is no spatial variability within each taxonomic unit. A sampling campaign was planned to overcome this lack of information.

3.7.1. SOIL SAMPLING CAMPAIGNS

The soil sampling campaign was conducted in March 2001. To plan the campaign, it was necessary to determine the extension of the survey, the parameters to be analysed and the places to be sampled. The following paragraphs will show the methodology applied in the sampling plan.

Sampling Depth and Soil Properties: As previously stated, the available soil data was too general for the selected watershed. Thus, it was necessary to generate more specific soil maps for the study site. In the context of pesticide assessment, the surface soil is more important than deeper soils (Boesten *et al.* 1999; Moorman *et al.* 1999). Thus, the first 50 cm of soil below ground level was evaluated. In addition, soil properties such as soil-water content, bulk density, organic matter and sand-silt-clay content were selected as the main objectives in the soil analysis.

Number of Sampling Sites: The second problem was to answer how many sampling sites are sufficient for the assessment. To solve this problem, the following criteria were considered:

- Both geological and edaphological maps define zones where soils can have very similar properties.
- The locations of banana farms are very important from the pesticide assessment point of view.
- The sampling sites depend on accessibility (existing main roads, consent to enter service roads in the farms). Based on this, the potential surface area to be sampled in the watershed is defined by a variable buffer area not more than 25 meters surrounding the existing and accessible roads.

- Although soil variability is a fact of nature, precision in catchment assessment is not the same as precision in farm assessment. From an agricultural point of view, the largest area representing a soil unit to be sampled is 40 hectares (Jacobsen 1999). From a landscape point of view, around 3 to 6 samples should be taken per landscape unit[21] (Manitoba Government 2000)
- It is necessary to use an adequate statistical technique to determine the minimum number of samples.
- The available budget for the sampling campaign is limited.
- The available time for the whole pesticide assessment is restricted to three years.

Thus, the most appropriate design is the stratified sampling design which has already been used in other types of assessment such as the prediction of organochlorine pesticides in sediments and animal tissues in a river basin (Black *et al.* 2000). For the present study, the sampling size was determined on the basis of statistical analysis (Gilbert 1987). Then, considering the other criteria, the recommended number of stratified sampling sites was fixed to 30 locations.

Considering only the watershed surface area (around 32000 Ha), it apparently seems that very few samples are used to evaluate an important problem in the basin (around one sample per every 1100 Ha). However, the current evaluation seems more appropriate when zoning the sampling points. These can be better seen in Figure 3.13 where sampling points are put over GEOLOGY, LANDUSE, ROAD and EDAPHOLOGY maps. In the road map (Figure 3.13d), it can be seen that a large area is difficult to be sampled due to the lack of roads.

<u>Number of samples per site:</u> Due to soil variability, it was necessary to have more than one soil sample per site. In every sampling site, two soil samples were obtained: a core undisturbed sample and a composite disturbed sample (cross-shaped sampling pattern). A cross-shape sampling pattern represents four sampling points at a certain distance away from the core sample (center of the cross), and separated 90° from each other with respect to the center (Figure 3.14). The radius of the cross-shaped pattern was variable depending on the topography. After collection, the sub-samples were mixed together to form a composite

[21] The landscape unit is a geographical land subdivision based on topography, vegetation, land use, regional context and built infrastructure. According to EPA, it is a designation to identify repeating patterns associated with dominant land uses in an area, and defined by the relative proportions of forest, agriculture, and developed (urban) land cover contained in that area.

sample. The core sample was used to determine the soil bulk density and the composite sample was used to determine the other soil properties.

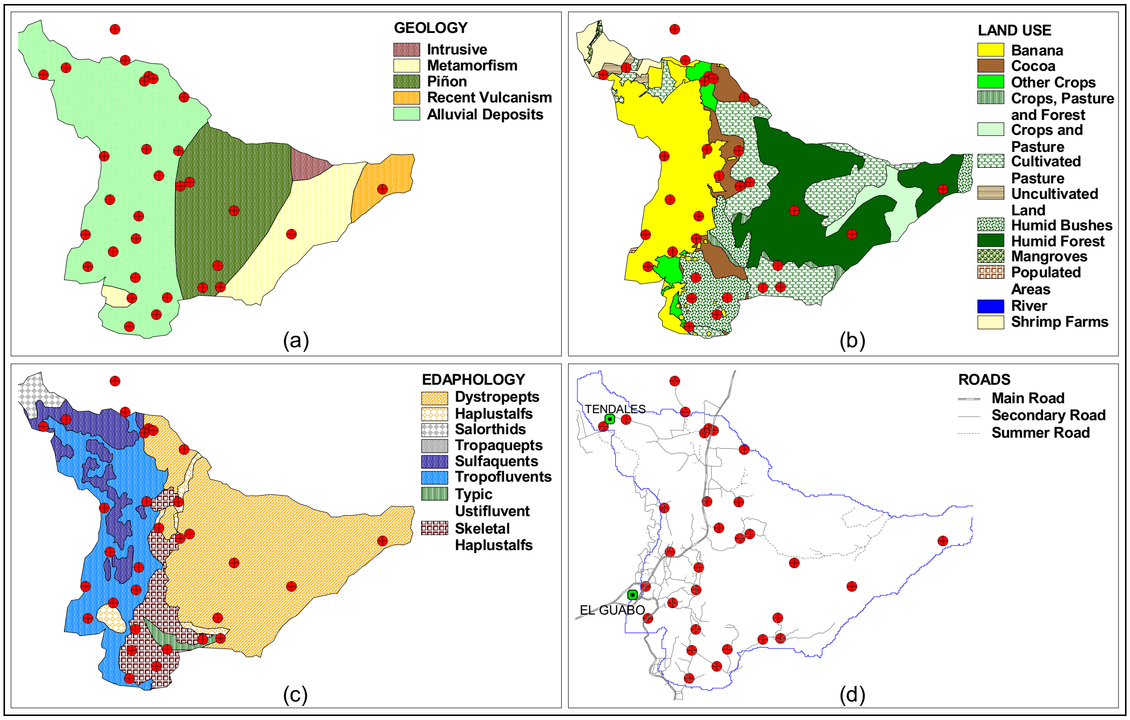


Figure 3.13. Soil sampling point locations compared to (a) Geology Map, (b) Landuse Map, (c) Edaphology Map, and (d) Road Map in the watershed.

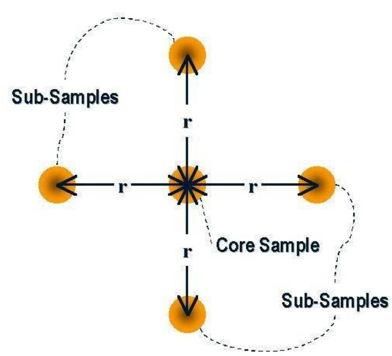


Figure 3.14. Cross-shaped sampling pattern adopted in soil exploration

<u>**Issues in sampling site determination:**</u> As written before, thirty locations were selected as sampling sites. However, during the sampling campaign, two main factors influenced the decision to obtain only 28 samples: