2007

Multisensor Fusion Remote Sensing Technology For Assessing Multitemporal Responses In Ecohydrological Systems

Ammarin Makkeasorn
University of Central Florida

MULTISENSOR FUSION REMOTE SENSING TECHNOLOGY FOR ASSESSING MULTITEMPORAL RESPONSES IN ECOHYDROLOGICAL SYSTEMS

by

AMMARIN MAKKEASORN
B.S. Suranaree University of Technology, 1999
M.S. Texas A&M University - Kingsville, 2002

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Civil and Environmental Engineering
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Summer Term
2007

Major Professor: Ni-Bin Chang

ABSTRACT

Earth ecosystems and environment have been changing rapidly due to the advanced technologies and developments of humans. Impacts caused by human activities and developments are difficult to acquire for evaluations due to the rapid changes. Remote sensing (RS) technology has been implemented for environmental managements. A new and promising trend in remote sensing for environment is widely used to measure and monitor the earth environment and its changes. RS allows large-scaled measurements over a large region within a very short period of time. Continuous and repeatable measurements are the very indispensable features of RS.

Soil moisture is a critical element in the hydrological cycle especially in a semiarid or arid region. Point measurement to comprehend the soil moisture distribution contiguously in a vast watershed is difficult because the soil moisture patterns might greatly vary temporally and spatially. Space-borne radar imaging satellites have been popular because they have the capability to exhibit all weather observations. Yet the estimation methods of soil moisture based on the active or passive satellite imageries remain uncertain. This study aims at presenting a systematic soil moisture estimation method for the Choke Canyon Reservoir Watershed (CCRW), a semiarid watershed with an area of over 14,200 km^2 in south Texas. With the aid of five corner reflectors, the RADARSAT-1 Synthetic Aperture Radar (SAR) imageries of the study area acquired in April and September 2004 were processed by both radiometric and geometric calibrations at first. New soil moisture estimation models derived by genetic

programming (GP) technique were then developed and applied to support the soil moisture distribution analysis. The GP-based nonlinear function derived in the evolutionary process uniquely links a series of crucial topographic and geographic features. Included in this process are slope, aspect, vegetation cover, and soil permeability to compliment the well-calibrated SAR data. Research indicates that the novel application of GP proved useful for generating a highly nonlinear structure in regression regime, which exhibits very strong correlations statistically between the model estimates and the ground truth measurements (volumetric water content) on the basis of the unseen data sets. In an effort to produce the soil moisture distributions over seasons, it eventually leads to characterizing local- to regional-scale soil moisture variability and performing the possible estimation of water storages of the terrestrial hydrosphere.

A new evolutionary computational, supervised classification scheme (Riparian Classification Algorithm, RICAL) was developed and used to identify the change of riparian zones in a semi-arid watershed temporally and spatially. The case study uniquely demonstrates an effort to incorporating both vegetation index and soil moisture estimates based on Landsat 5 TM and RADARSAT-1 imageries while trying to improve the riparian classification in the Choke Canyon Reservoir Watershed (CCRW), South Texas. The CCRW was selected as the study area contributing to the reservoir, which is mostly agricultural and range land in a semi-arid coastal environment. This makes the change detection of riparian buffers significant due to their interception capability of non-point source impacts within the riparian buffer zones and the maintenance of ecosystem integrity region wide. The estimation of soil moisture based on RADARSAT-1 Synthetic Aperture Radar (SAR) satellite imagery as previously developed was used. Eight commonly used vegetation indices were calculated from the reflectance obtained from Landsat 5 TM satellite images. The vegetation indices were individually used to classify

vegetation cover in association with genetic programming algorithm. The soil moisture and vegetation indices were integrated into Landsat TM images based on a pre-pixel channel approach for riparian classification. Two different classification algorithms were used including genetic programming, and a combination of ISODATA and maximum likelihood supervised classification. The white box feature of genetic programming revealed the comparative advantage of all input parameters. The GP algorithm yielded more than 90% accuracy, based on unseen ground data, using vegetation index and Landsat reflectance band 1, 2, 3, and 4. The detection of changes in the buffer zone was proved to be technically feasible with high accuracy. Overall, the development of the RICAL algorithm may lead to the formulation of more effective management strategies for the handling of non-point source pollution control, bird habitat monitoring, and grazing and live stock management in the future.

Soil properties, landscapes, channels, fault lines, erosion/deposition patches, and bedload transport history show geologic and geomorphologic features in a variety of watersheds. In response to these unique watershed characteristics, the hydrology of large-scale watersheds is often very complex. Precipitation, infiltration and percolation, stream flow, plant transpiration, soil moisture changes, and groundwater recharge are intimately related with each other to form water balance dynamics on the surface of these watersheds. Within this chapter, depicted is an optimal site selection technology using a grey integer programming (GIP) model to assimilate remote sensing-based geo-environmental patterns in an uncertain environment with respect to some technical and resources constraints. It enables us to retrieve the hydrological trends and pinpoint the most critical locations for the deployment of monitoring stations in a vast watershed. Geo-environmental information amassed in this study includes soil permeability, surface temperature, soil moisture, precipitation, leaf area index (LAI) and normalized difference

vegetation index (NDVI). With the aid of a remote sensing–based GIP analysis, only five locations out of more than 800 candidate sites were selected by the spatial analysis, and then confirmed by a field investigation. The methodology developed in this remote sensing-based GIP analysis will significantly advance the state-of-the-art technology in optimum arrangement/distribution of water sensor platforms for maximum sensing coverage and information-extraction capacity.

Effective water resources management is a critically important priority across the globe. While water scarcity limits the uses of water in many ways, floods also have caused so many damages and lives. To more efficiently use the limited amount of water or to resourcefully provide adequate time for flood warning, the results have led us to seek advanced techniques for improving streamflow forecasting. The objective of this section of research is to incorporate sea surface temperature (SST), Next Generation Radar (NEXRAD) and meteorological characteristics with historical stream data to forecast the actual streamflow using genetic programming. This study case concerns the forecasting of stream discharge of a complex-terrain, semi-arid watershed. This study elicits microclimatological factors and the resultant stream flow rate in river system given the influence of dynamic basin features such as soil moisture, soil temperature, ambient relative humidity, air temperature, sea surface temperature, and precipitation. Evaluations of the forecasting results are expressed in terms of the percentage error (PE), the root-mean-square error (RMSE), and the square of the Pearson product moment correlation coefficient (r-squared value). The developed models can predict streamflow with very good accuracy with an r-square of 0.84 and PE of 1% for a 30-day prediction.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

Background Information

Remote sensing (RS) is one of the environmental observing systems, which collect information of an object or event from a distance. Remote sensing (RS) technology is a promising engineering marvel that has been implemented in many environmental applications to improve the quality of point measurements at the ground level. There are more than several hundred space-borne satellites, air-borne platform, and ground-based radar stations being used for environmental montoring today. Thousands of terrestrial microsensors and oceanic buoys that form part of the environmental observing systems at different scales are also monitoring the minute changes of the multimedia environment on this planet. A few space-borne satellites have made their way even to the outer space and send back cosmic information so that we can understand more about the outside universe. However, the processing of images collected by those environmental observing systems might involve using some artificial intelligence tools.

There are several methods that allow computers to solve problems without being explicitly programmed such as neural networks, artificial intelligence, self-improving systems, self-organizing systems, for example. Genetic programming (GP) allows computers to create computer programs to solve problems without being programmed explicitly by human (Koza, 1992 and Banzhaf et al., 1998). It is a method to generate computer programs based on the natural selection principle described by Darwin as the nature selection (Darwin, 1859). The GP is a subset of evolutionary algorithms (EA). EAs comprise the genetic programming, evolution

1

strategies, evolutionary programming, and genetic algorithms (Ryan, 2000). Each of them evolves differently, but the basic algorithm is still based on the same evolutionary process as shown in Figure 1-1. Nedjah et al. (2006) presented a collection of theory and case studies of the genetic programming in a series of Studies in Computational Intelligence, volume 13. They discussed many aspects of GP's strength in many applications.

The difference between GP and the other methods is that GP solves problems by creating computer programs via natural selection to best fit the problems that it tries to solve, while the other existing methods utilize specialized structures. For instance, neural networks utilize weight factors in order to adapt its specialized network to the problems. The specific structure such as the neural network constraints itself from solving many problems due to the fact that not all the problems have structure the same way as neural networks. This dissertation is therefore composed of four concatenated independent studies utilizing spatial analysis technology (i.e., remote sensing and geographic information system, GIS) and evolutionary computing modeling technology to explore and improve the fundamental understanding of a few critical components in the area of ecohydrology.
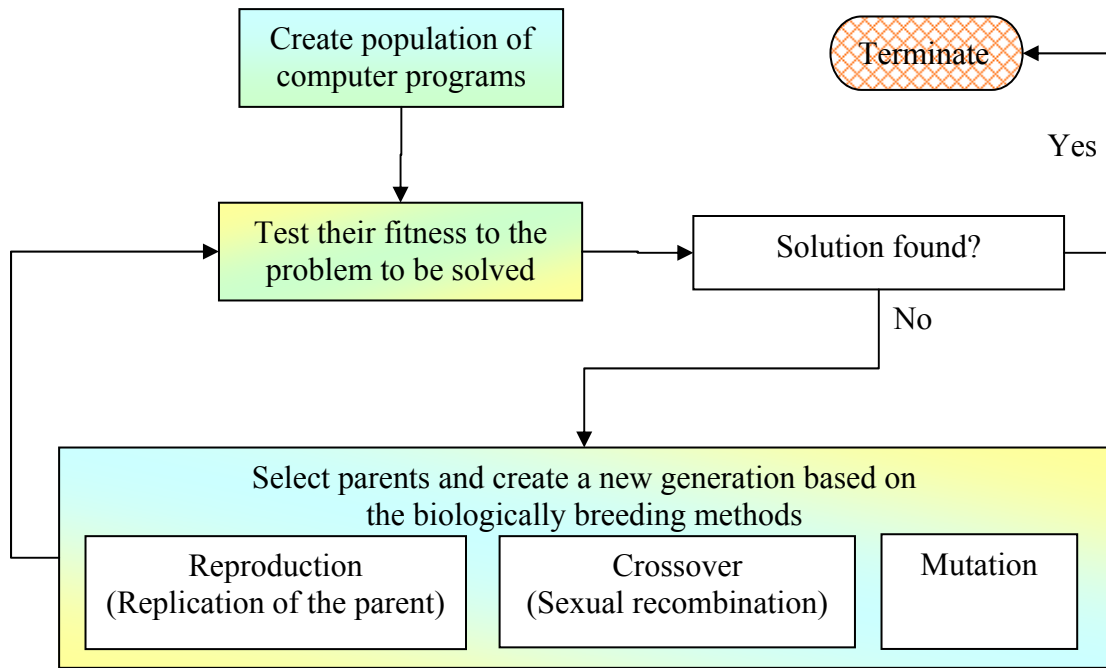
Figure 1-1. Flowchart for the conventional genetic algorithm

Research Objectives and Scope of Work

Land development has been increasingly expanded in the past few decades due to the great increase of the world population (U.S. Census Bureau, 2006). Given the adverse environmental impacts of human activities and the increases in socioeconomic development in the past century, it is important to investigate, monitor and evaluate the consequences in term of environmental impacts from such activities and development. Tracking the changes of ecosystems and land use/land cover would give some ideas to stakeholders toward a truly sustainable development. This study has a main objective to utilize remote sensing resources and available EA techniques to develop new environmental remote sensing and management system for hydrological applications.

A suite of innovative forecasting technologies developed in this dissertation would provide a foundation for 1) more accurate and reliable water resources management strategies in any basin; and 2) an adaptable resource management tool that can be developed for all basins. Further, numerous secondary benefits to society may result from this technology development, including 1) increased warning times for destructive flood/drought events and other natural hazards; 2) reductions in economic hardship resulting from both drought and flood; 3) increased societal awareness of the availability and importance of fresh water; and 4) the underpinnings for achieving regional-scale sustainability of human and cultural landscapes as well as physical landscapes, both natural and managed. Therefore, research objectives are to:

- establish a database of geo-environmental features (such as land cover, land use patterns, soil texture, and others) within a specified river basin by integrated remote sensing, geographical information system, and global positioning system technologies;

- develop a suite of advanced forecasting models that link applied genetic algorithm modeling with an integrated geo-environmental database; and

- apply the modeling technology to a specified river basin to evaluate the validity of these methods in making rapid forecasts of some critical components in the area of ecohydrology.

Dissertation Organization

This dissertation is organized in six (6) chapters. Chapter 2 presents a literature review and discussion regarding the derivation of soil moisture at a watershed scale using space-borne microwave sensor onboard the RADARSAT-1 Synthetic Aperture Radar (SAR) and artificial

4

intelligent approach. This work has been published in the *Water Resources Research* (Makkeasorn et al., 2006).

Chapter 3 presents the use of the derived soil moisture in chapter 2 incorporated with data from Landsat 5 TM satellite to detect the change of riparian forests in a semi-arid watershed -- the Choke Canyon Reservoir Watershed (CCRW), South Texas. A new evolutionary computational, supervised classification scheme (RIparian Classification Algorithm, RICAL) was developed and used to identify the change of riparian zones in a semi-arid watershed temporally and spatially. The case study uniquely demonstrates an effort to incorporating both vegetation index and soil moisture estimates based on Landsat 5 TM and RADARSAT-1 imageries while trying to improve the riparian classification in the CCRW, South Texas. First of all, an estimation of soil moisture based on RADARSAT-1 Synthetic Aperture Radar (SAR) satellite imagery was conducted using genetic programming (GP) approach. Eight commonly used vegetation indices were calculated from the reflectance obtained from Landsat 5 TM satellite images. The vegetation indices were individually used to classify vegetation cover in association with genetic programming algorithm.

Chapter 4 describes an optimal site selection technique using a grey integer programming (GIP) model to discern remote sensing-based geo-environmental patterns in an uncertain environment with respect to some technical and resources constraints. It enables us to retrieve the hydrological trends and pinpoint the most critical locations for the deployment of monitoring stations in a vast watershed. Practical implementation is demonstrated by a case study conducted in Choke Canyon Reservoir watershed (CCRW), which is a semi-arid watershed in south Texas. Geo-environmental information amassed in the chapter includes soil permeability, surface

temperature, soil moisture, precipitation, leaf area index (LAI) and normalized difference vegetation index (NDVI). With the aid of a remote sensing–based GIP analysis, only five locations out of more than 800 candidate sites were selected by the spatial analysis, and then confirmed by a field investigation. The methodology developed in this remote sensing-based GIP analysis will significantly advance the state-of-the-art technology in optimum arrangement/distribution of water sensor platforms for maximum sensing coverage and information-extraction capacity.

Chapter 5 describes the uses of sea surface temperature (SST), Next Generation Radar (NEXRAD), meteorological data collected from the weather station presented in Chapter 4, and the USGS stream data to forecast the actual streamflow using genetic programming. This study case concerns the forecasting of stream discharge of a complex-terrain, semi-arid watershed. This study elicits microclimatological factors and the resultant stream flow rate in river system given the influence of dynamic basin features such as soil moisture, soil temperature, ambient relative humidity, air temperature, sea surface temperature, and spatial precipitation. Three weather stations are deployed as a supplementary data-gathering network in addition to three (3) USGS gage stations in the semi-arid CCRW, South Texas. Standard model consists of 21 independent variables (2 sea surface temperature indices, NEXRAD rainfall intensity, 15 meteorological inputs, 3 discharges) for an output of the total discharge of the river basin. The end results are GP-derived models that can predict streamflow 30 days in advance.

References

Banzhaf, W., Nordin, P., Keller, R.E., and Francone, F.D., 1998. Genetic programming – an introduction: on the automatic evolution of computer programs and its applications. Morgan Kaufmann Publishers, Inc., San Francisco, USA, 470.

Darwin, C., 1859. On the origin of the species. Murray.

Koza, J.R., 1992. Genetic programming: on the programming of computers by means of natural selection. A Bradford Book, The MIT Press, Cambridge, USA, 819.

Makkeasorn, A., Chang, N.B., Beaman, M., Wyatt, C., and Slater, C., 2006. Soil moisture estimation in a semi-arid watershed using RADARSAT-1 satellite imagery and genetic programming. Water Resources Research 42, W09401, doi:10.1029/2005WR004033.

Nedjah, N., Abraham, A., and de Macedo Mourelle, L. (Eds.), 2006. Genetic systems programming. Springer, Netherlands, 230.

Ryan, C., 2000. Automatic re-engineering of software using genetic programming. Kluwer Academic Publisher, Massachusetts, USA.

U.S. Census Bureau, 2006. World population information. Population Division, Retrieved from http://www.census.gov/ipc/www/world.html on in August 2006

CHAPTER 2
SOIL MOISTURE ESTIMATION IN A SEMI-ARID WATERSHED USING RADARSAT-1 SATELLITE IMAGERY AND GENETIC PROGRAMING

Introduction

Soil moisture is one of the fundamental hydrologic parameters in terrestrial hydrology. The ecosystem in semi-arid or arid areas is normally driven by soil moisture in most cases. It has long been recognized that soil moisture in the root zone regulates atmospheric energy exchange at land surface, which plays a key role in flood and drought genesis. Soil moisture also plays a key role in surface-subsurface water exchanges through infiltration and percolation processes. Accurate measurement of soil moisture at the ground level may aid in the estimation of crop yield, plant stress, and watershed runoff. Soil moisture obviously varies in space and time. Multi-temporal spatially-varied soil moisture values are normally required as inputs into the hydrological, meteorological, and ecological models supporting the estimation of rainfall-runoff process, the prediction of meteorological pattern, and the assessment of ecosystem. The surface soil moisture measurement, however, is very difficult to obtain over a large area due to a variety of soil permeability values and associated soil textures. The point measurements can practically be used on a small-scaled area, but it is not possible to acquire such information effectively in

large-scale watersheds. Consistency of measuring in-situ soil moisture is barely obtainable even on a local scale.

Satellite derived remotely sensed images may help promote realization of the variations in intensity of electromagnetic energy reflected or emitted from the Earth's surface [*Lu, 2005*]. Space-borne radar imaging satellites have become a common means of earth observation in the past two decades [*Freeman*, 1992]. The specific imagery produced is determined by the wavelength of the electromagnetic energy that is being sensed, and the physical properties of the matter that determine the reflection and emission of the energy. Passive and active sensors are the two major types of radar remote sensors for soil moisture measurement. Yet the estimation methods of soil moisture based on the satellite imageries remain uncertain [*Salgado et al.*, 2001; *Glenn and Carr*, 2004]. Passive microwave system had been explored the capability of measuring soil moisture remotely [*Owe et al.*, 1988; *Jackson et al.*, 1993]. Later on the active microwave systems were developed and used for earth observations. Synthetic Aperture Radar (SAR), one of the active remote sensing schemes, has shown its capability of measuring soil moisture in the work of *Ulaby* [1974], *Olmsted* [1993], *Dubois et al.* [1995], *Moran et al.* [2000], *Njoku et al.* [2000], *Salgado et al.* [2001], *Baghdadi et al.* [2002], *Wilson et al.* [2003], and *Glenn and Carr* [2004]. RADARSAT-1 is a space-borne Synthetic Aperture Radar (SAR) satellite equipped with an active microwave sensor. The active microwave sensor provides all-weather data imaging capabilities for data acquisition because it does not rely on any external microwave source [*ASF*, 1999]. The space-borne SAR can provide the hydrographical features, such as soil moisture, flood zone, and snow cover area [*Shi and Dozier, 1995 and 1997*], on a regional scale due to its large footprint. It is well suited to large-scaled, hydrological applications.

Due to the sensitivity of backscattered microwave energy to dielectric constant, the SAR has the potential for measuring water content in the surface soil indirectly [*Ulaby*, 1974; *Dubois et al.*, 1995]. When using the space-borne SAR to remotely detect water content in the surface soil, the time constraint is almost negligible. The RADARSAT-1 is able to capture surface soil moisture over a large area in a matter of seconds, if the study area is within its swath. However, not only does the dielectric constant affect the SAR, but also many other factors as well. As reported in the work of *Dubois et al.* [1995], *Moran et al.* [2000], *Salgado et al.* [2001], and *Baghdadi et al.* [2002], the radar backscatter responds to the surface roughness and vegetation cover too. *Ulaby* [1974] and *Olmsted* [1993] also mentioned that the radar backscatter responds to surface slope as well. The aspect is the horizontal direction of slope. While the forward and backward slopes reflect backscatter toward and away from the incoming direction of the radar signals, the aspect of slope also affects the backscatter likewise. Depending on the direction of the incoming radar signal, the aspect could return the signal back to its incoming direction, or the signal might be reflected away from its source.

Estimation of soil moisture based on SAR measurement (i.e., ERS or RADARSAT-1) was made possible via developing linear regression models [*Freeman*, 1992; *Dubois et al.*, 1995; *Moran et al.*, 2000; *Moeremans and Dautrebande*, 2000; *Salgado et al.*, 2001; *Glenn and Carr* 2003; *Nolan*, 2003] and nonlinear regression model [*Zribi and Dechambre*, 2002] in a single land use/land cover from several hundreds m^2 to several km^2 based on traditional statistical regression theory. Studies using neural network models and inversion approaches to retrieve soil moisture based on passive microwave remotely sensed data can be found elsewhere [*Narayanan and Hirsave*, 2001; *Frate et al.*, 2003; *Wigneron et al.*, 2003]. Up to this point, there is an exceptional difficulty to derive a highly complex model in dealing with multiple land use/land

cover environment simultaneously within a vast watershed while collecting ground data might be extremely time consuming and difficult. The further development that differentiates this study to the others is to use evolutionary computation approach for fulfilling soil moisture estimation that uniquely links the SAR imagery with topographical and geographical features, such as slope, aspect, vegetation cover, and soil permeability without touching surface roughness, a parameter that is hard to have generic measurement across different land use patterns in a vast watershed. Genetic Programming (GP), one of the evolutionary computing techniques, is the next best advancement to create best-selective nonlinear regression models in terms of multiple independent variables when dealing with multiple land use/land cover situation. The soil moisture measurement in the Choke Canyon Reservoir Watershed (CCRW), a semi-arid watershed in south Texas, is of interest in this study since it consists of various types of land use patterns, such as row crops, pasture, evergreen forest, and range within an area of 14,200 km^2. Figure 2-2 and Table 1 jointly present land use/land cover of the CCRW. The evolutionary computation using the GP as a means is thus proposed in this study to estimate surface soil moisture using space-borne SAR along with relevant topographic and geographic features. In particular, the aspect in conjunction with the RADARSAT-1 SAR data, soil permeability, vegetation cover, and slope itself are incorporated into the set of independent variables, and they are collectively used to derive a representative soil moisture model in the case study. Both root-mean-square error (RMSE) and the square of the Pearson product moment correlation coefficient (R-square) are used to verify the effectiveness of model development.

National Land Cover Data 1992
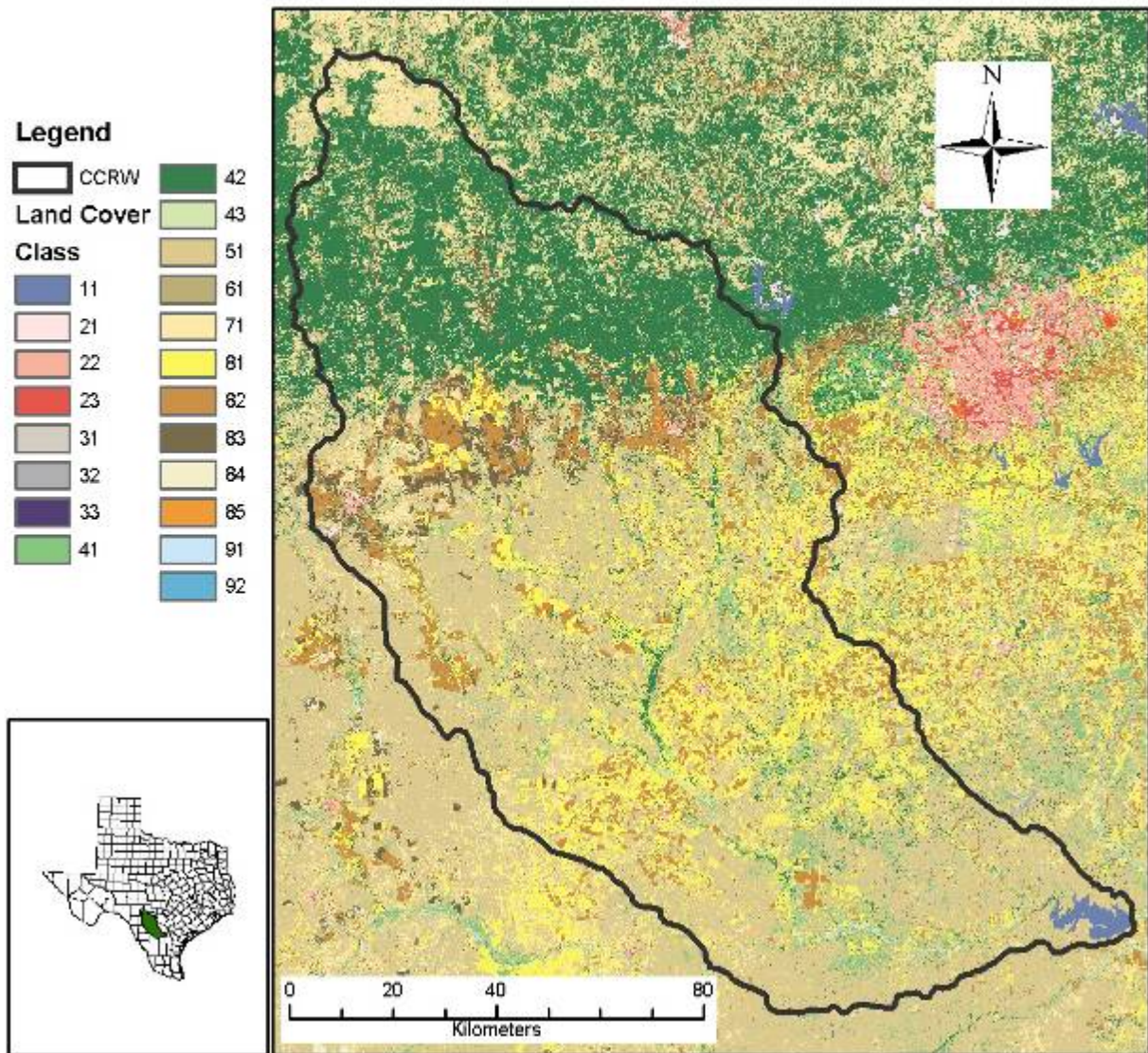
Figure 2-2. A description of the land covers in CCRW (see also Table 2-1).

Background

The Choke Canyon Reservoir Watershed (CCRW) is composed of several land use/land cover types. Farming and livestock husbandry are major land use patterns in the past few decades. The farmland is often graded and plowed, and irrigation may change the soil moisture in some seasons periodically. The livestock in south Texas is naturally fed on grass in open areas and ranches. Mixed land uses in this area introduce complexity of soil moisture distribution. Figure 2-2 is the National Land Cover Data (NLCD) showing the land cover in the watershed (Distributed Active Archive Center, U.S. Geological Survey EROS Data Center, available at http://landcover.usgs.gov). Table 2-1 complements the description of the NLCD image. It shows the land use patterns in this area mainly include these from evergreen forest in the upstream area to cropland and ranges in the middle stream areas, and down to shrubland in the lower stream of the watershed. Landscape in south Texas, however, is intimately tied with the geological structure. Figure 2-3 shows the geographic environments and geological features of the CCRW. The CCRW encompasses 14,200 km^2 out of the 43,300 km^2 Nueces River Basin. Elevations in the CCRW range from 42 m above sea level near the dam to 740 m at the Edwards Plateau near the divide of the watershed upstream. To the north, topography strongly influences the hydrology of the watershed. In the upper portion of the watershed, the steep slopes and arid terrain of the Balcones Escarpment rise into the Edwards Plateau. These hills, cliffs, exposed rock, and clay soil, while acting as sinks at the beginning of a precipitation, cause rapid runoff during large storm events resulting in flashflood. As the streams cross the Edwards Aquifer Recharge zone, they lose a significant portion of their flow through faults and solution cavities (Karst topography). Downstream of the Balcones fault zone, the landscape tends to flatten as the water

13

flows south and east into the South Texas Brush Country where slopes range from 0 to 10%. Placement of USGS stream gages above and below the fault zone helps to quantify the water losses in the fault zone and to provide early warning information of any potential flooding in the downstream areas (see Figure 2-3). Right above the Choke Canyon Reservoir (CCR) there are two USGS stream gages measuring the total inflow of the streams that flow into the reservoir. According to the historical flow measurements recorded in decades, the hydrological pattern of this watershed comprises two seasons: wet and dry seasons (see Figure 2-4). The stream data are available at http://waterdata.usgs.gov/tx/nwis/rt. The upper portion of the CCRW is not included as part of the study area because of its unique geological structure of bedrock. There are exposed rocks and gravels in some areas, while the others are covered barely by a very thin layer of soil, if any. Therefore this area is not deemed valuable for soil moisture study.

Table 2-1. Classification system used for National Land Cover Data.

Value	Class
11	Open water
12	Perennial ice/Snow
21	Low intensity residential
22	High intensity residential
23	Commercial/Industrial/Transportation
31	Bare rock/Sand/Clay
32	Quarries/Strip Mines/Gravel Pits
33	Transitional
41	Deciduous Forest
42	Evergreen Forest
43	Mixed Forest
51	Shrubland
61	Orchards/Vineyards/Other
71	Grasslands/Herbaceous
81	Pasture/Hay
82	Row Crops
83	Small Grains
84	Fallow
85	Urban/Recreational Grasses
91	Woody Wetlands
92	Emergent Herbaceous Wetlands
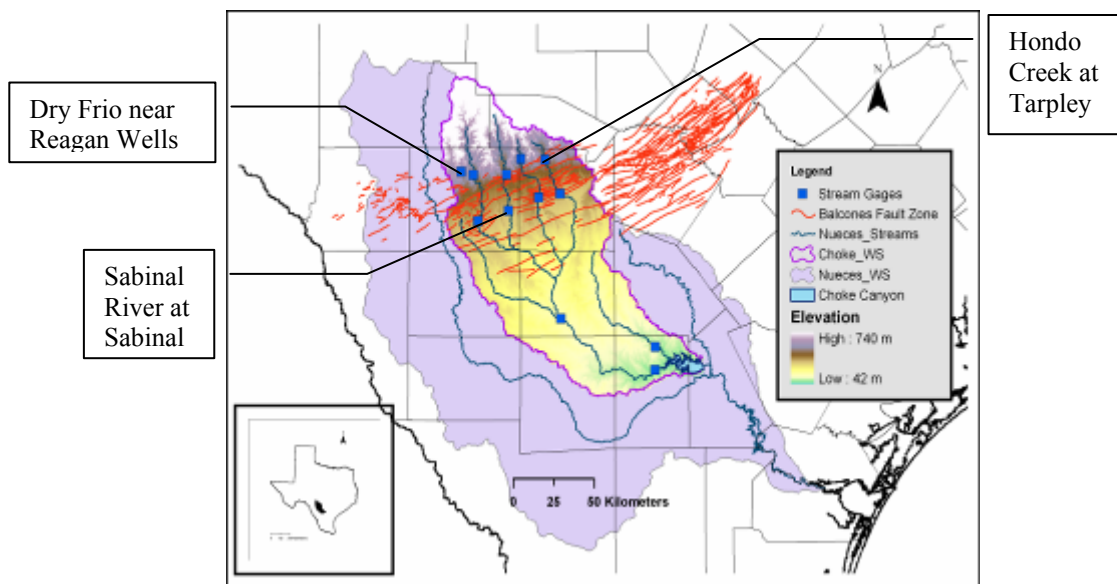
Figure 2-3. Geographic and geological features of CCRW.

Fault lines, where water recharges to the underground water aquifer, are shown in red. The Texas Hill Country comprises hills and valleys located above the Balcones zone. The differences of slope above and below the fault zone are obvious. Streams flow southward to the east, and are merged together before flowing into the Choke Canyon Reservoir. The USGS gage stations are located above and below the recharge zone. One gage is located at the middle of the watershed where all streams are merged. The other two gage stations are located at the downstream immediately before the streams flow into the CCRW.
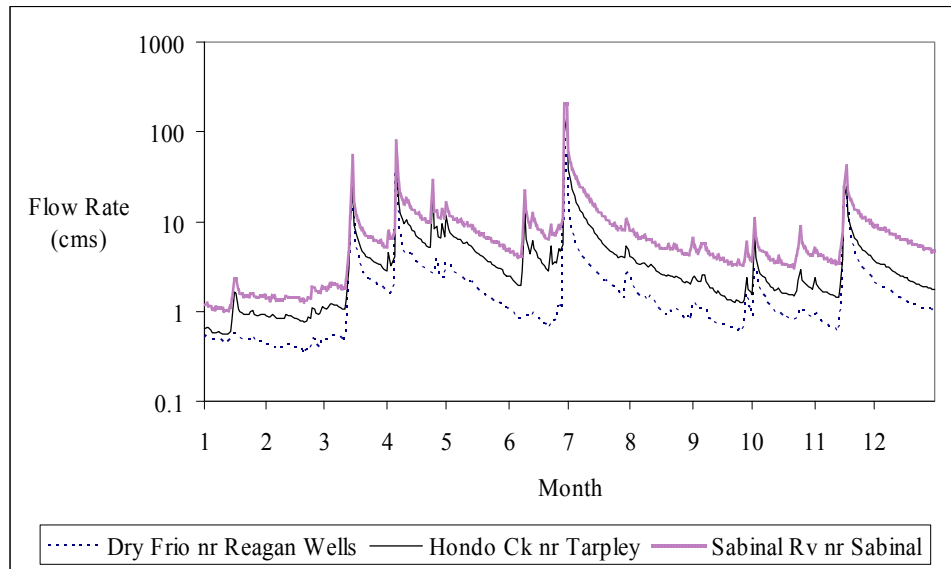
Figure 2-4. Historical streamflow of Dry Frio River, Hondo Creek, and Sabinal River.
A log-plot of mean streamflows measured at the USGS stations that are located at Dry Frio River near Reagan Wells, Hondo Creek at Tarpley, and Sabinal River at Sabinal. The high flow rates occurred in April 2004 (wet month) throughout the time frame of the SAR data acquisition on April 19, 2004. In the September 2004 the flow rates were very low nearly at the base flow, which was considered as a dry month.

Methodology

Field data collection

Modeling the soil moisture in this study requires emphasizing the efforts of data synthesis of Synthetic Aperture Radar (SAR) imagery, slope, aspect, soil permeability, and Normalized Difference Vegetation Index (NDVI). Modeling outputs based on genetic programming technique are supposed to compare against intensive ground truth samples in the same region. Yet there was an exceptional difficulty to acquire the ground truth data at the resolution of RADARSAT-1 SAR data in the vast study area. Two sampling campaigns for ground truth measurements were made in April and September 2004. They were carried within 24 hours before and after the SAR data acquisition in order to capture the synchronous soil moisture patterns. At least four types of land cover, including grassland, shrubland, row crop and

16

deciduous forest, were included in both April and September campaigns in 2004. Some evergreen forested land upstream was also selected to enhance the credibility of ground truthing (see Table 2-1 and Figure 2-2). It looks like there are only 7 fields sampled in the ground truth data acquisitions in Figure 2-5 and Figure 2-6. In fact, 434 and 63 surface soil moisture measurement points were collected for building up the ground truth database in April and September 2004, respectively. Each measurement point was chosen at least 50 m away from any road or building nearby to increase the data integrity. This could avoid struggling with some misleading results in the end by using strayed backscatter of SAR imagery influenced by the construction work in comparison to the ground truth data points. The distance between any two measurement points is at least 13 m apart to ensure that there is only one ground truth measurement point that is associated with one pixel of SAR imagery. We navigated to each measurement point with a handheld Global Positioning System (GPS) unit with a capability of reading location of submeter accuracy [Trimble Navigation Ltd., 2004]. The GPS unit used in this study was a Trimble handheld GPS model GEO XT. To reduce the uncertainty, each ground truth data would comprise 3 measurements within a vicinity of 2 m in radius, and then, we took average of the 3 associated measurements at each measurement point. In addition, the target areas for the ground truth measurements must be chosen in the proximity of the ground control points (i.e., corner reflectors) in order to minimize the horizontal error of the ground truth points relative to the SAR geometrically corrected.
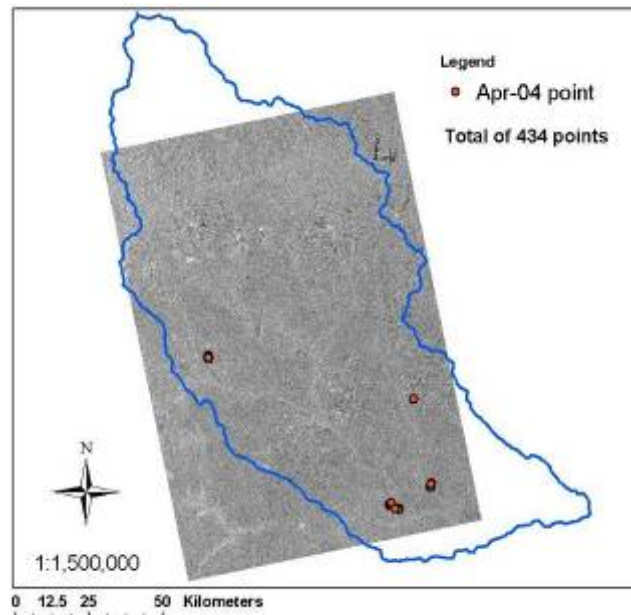
Figure 2-5. Soil moisture sampling sites visited in April 2004.
Four hundred and thirty-four (434) ground truth measurement points were collected within 24 hours
before and after the SAR data acquisition on April 19, 2004. The measurements were done on flat bare
soil, high-density mesquite trees, deciduous forest, and grassland.
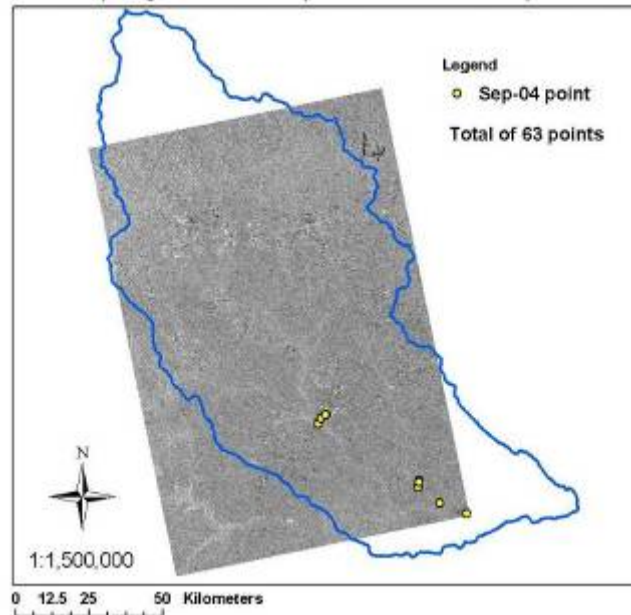
Figure 2-6. Soil moisture sampling sites visited in September 2004.
Sixty-three (63) ground truth measurement points were collected within 8 hours before the SAR data
acquired on September 12, 2004. The ground truths were done on evergreen forest, raw crop, brush land,
high-density mesquite trees, and grassland.

All ground truth measurements of soil moisture in this study were collected within the top 5 cm of soil by using The FieldScout™ TDR 300 soil moisture meter [see also Le He´garat-Mascle et al., 2003; Wilson et al., 2003; Spectrum Technologies, Inc., 2004]. The TDR method has been popular for it may provide measurements of in situ soil moisture content with good accuracy in the work of Topp et al. [1980], Roth et al. [1992], and Walker et al. [2001]. The TDR 300 sensor rods used in our measurements were 12 cm in length. We measured the soil moisture content on the top 5 cm of soil surface by inserting the probe at an angle of 25° from the flat ground. Prior to use, the TDR probe was calibrated against gravimetric measurement method within a range between 10 and 50% moisture (converted the gravimetric to the volumetric moisture content). An average value of three gravimetric measurements was used to calibrate each TDR measurement.

Genetic programming

The well-known approach invented by Koza et al. [2003, p. 3] has given statements about the main point of genetic programming as ''. . .high-return human-competitive machine intelligence.'' It generally approaches the solution by evolving over a series of generations of regression model using the evolutionary search based on the Darwinian principle of natural selection (from J. R. Koza, http://genetic-programming.org, last updated on 16 September 2004). The principle of Evolutionary Computation (EC) is rooted from Genetic Algorithms (GA) first developed by Holland [1975], Evolution Strategies (ES) developed by Rechenberg and Schwefel [from Back et al., 1997], and Evolutionary Programming (EP) developed by Fogel et al. [1966]. All three of them were eventually combined into one entity called ''Evolutionary Computation''

[Gagne and Parizeau, 2004]. Under the EC framework, the Genetic Programming is generally considered as an extension of GA.

The GP is the heuristic iterative search technique that obtains the best solution in a given decision on the basis of an algorithm that mimics the evolution of genetic life forms [see also Cramer, 1985; Heywood and Zincir-Heywood, 2002; Song et al., 2003]. It starts with solving a problem by creating massive amount of random functions in a population pool. This population of functions is progressively evolved over a series of generations. The search for the best result in the evolutionary process involves applying the Darwinian principle of nature selection (survival of the fittest) including crossover, mutation, duplication, and deletion. Regression models generated from the GP are free from any particular model structure [Chang and Chen, 2000]. It could be the best solver for searching highly nonlinear spaces for global optima via adaptive strategies. In recent years, the GP has been proved useful for solving highly nonlinear environmental problems [Chang and Chen, 2000].

The Linear Genetic Programming (LGP) expresses instructions as a line-by-line instruction. Execution of the program is a mimic of calculating multiple calculations in a normal calculator as simple line-by-line processing steps [Heywood and Zincir-Heywood, 2002; Song et al., 2003]. In this study we use the GP software called Discipulus®, which is developed by Francone [1998]. The codes are defined in terms of functions and terminal sets that modify the contents of internal memory and program counter. Discipulus® uses LGP algorithm to produce multiple lists of instructions representing models with the best fit to its training and calibrating data. While the training and the calibrating data are used as the basis genotype to build models, another independent data set is used to validate the generated models. The validating data are

untouched by Discipulus® during the process of modeling development. The validating data are used only to test the fitness of the surviving models.

Integrated framework

Figure 2-7 summarizes all the work flows of this analytical framework. The Alaska Satellite Facility (ASF) handled the image transcriptions and the level-0 processing, including radiometric and geometric calibrations, and geocoding. The data, thereafter, were transferred to Texas for the level-1 processing, including georeference, translation, and data extraction. The translation was done only when the georeferencing process did not reduce the horizontal error less than RMSE of 12.5 m, which is the SAR pixels' size. One scene of the CCRW image is a composition of many SAR images called frames. The frames were captured within approximately 15 s to compose a complete image of the watershed. Each frame was processed with the same algorithms to maintain consistency throughout all data. Since a complete image of the CCRWis composed of many frames, mosaicking was performed to combine frames together. However, the mosaicking procedure must not be done before the data extraction because the SAR data could be altered because of the image resample in the mosaicking process.

While the soil permeability map can be created from STATSGO database, the NDVI data, derived from the AVHRR sensor, may address the seasonal changes of plants' productivities. Slope and aspect data can be easily derived from a Digital Elevation Model (DEM). After all the input data (SAR, slope, aspect, soil permeability, and NDVI) were calibrated and imported into the GIS framework, the data are extracted from each layer and tabulated for uses in later GP analyses. Once regression models were developed from different

21

attempts using GP as a means, the best model can then be chosen for mapping soil moisture on a watershed scale.

Since most of the ground data in this study fall in the frame 72, it was chosen as our reference frame for the mosaic. The cutline method was employed for the mosaic to maintain most of the frame image, which covered more than 70% of the study area. To generate images of soil moisture, all the input data were imported into the ArcGIS framework to process raster calculations. At the end of the raster calculations, an image of soil moisture is created smoothly according to generated GP model.
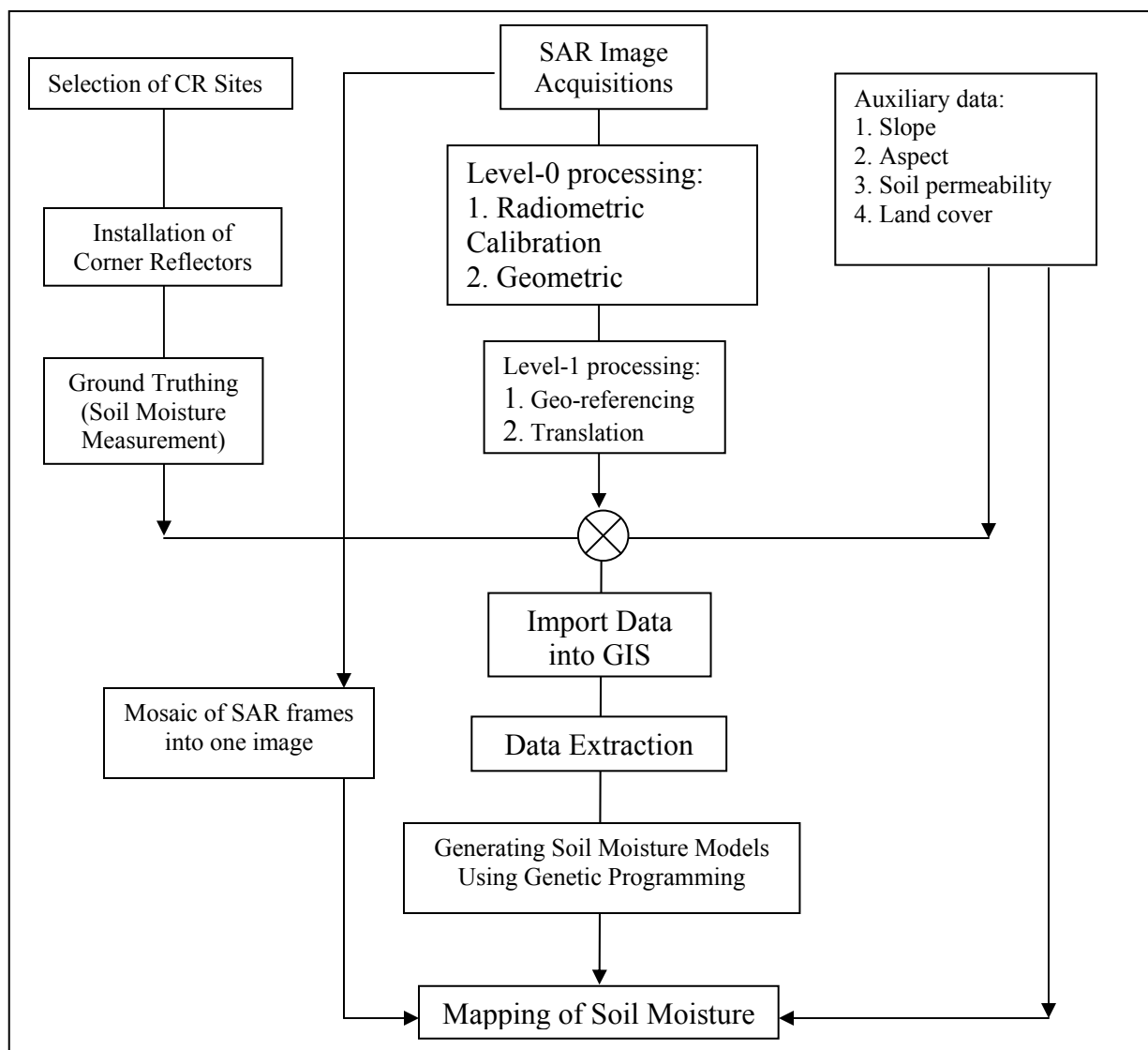
Figure 2-7. Work flow chart of the RADARSAT-1 soil moisture study.

Data Synthesis and Processing

RADARSAT-1 SAR data

This analysis counts on RADARSAT-1 SAR imageries acquired in April and September 2004. The standard beam mode in ascending orbits of RADARSAT-1 was selected for this study. The standard full-resolution imagery covers approximately 100 km x 100 km with the pixel size of 12.5 m (i.e., 25 m resolution). This implies any feature that is smaller than 12.5 m cannot be differentiated by RADARSAT-1 directly. The electromagnetic pulse used in RADARSAT-1 SAR is in the C-band frequency (5.3 GHz; 5.66 cm wavelength) [ASF, 1999]. Speckle reduction caused by the surface terrain was not performed on the SAR data because of the flatness of the study region (i.e., the lower CCRW) [Zribi et al., 2005]. Furthermore, the speckle is considered as a property of the backscatter; thus there would not be a calibration problem because the pixel backscatter measurement is repeatable [Freeman, 1992]. Minimum number of image processing is our target in order to minimize the alteration of backscatter measurements as much as we practically can. Only normalized radiometric correction to compensate for speckle due to the inherent radar image distortion was carried out on the basis of the ASF SAR Processing Algorithm [Olmsted, 1993]. The spatial resolution of the processed data can be kept at its original resolution.

To ensure the accuracy of the data, radiometric and geometric corrections were deemed necessary to all data with the aid of corner reflectors [Freeman, 1992; ASF, 2002]. Radiometric calibration is required to assure the correct interpretation and information of the signal. Geometric (spatial) calibration is required to assure the correct dimensions and position, and adjust for any distortion of the SAR imagery. The corner reflector has been widely used for

calibration of SAR data from the early age of the technology [Sarabandi et al., 1992; Sarabandi, 1994]. Before the installation of the five corner reflectors, we used the Satellite Tool Kit[®] (STK) (available at http://www.stk.com/) to determine the correct orientations for pointing our corner reflectors to SAR acquisition pathway, and then the Two Line Element (TLE) was used to determine the look direction of each corner reflector after finding out its GPS coordinate, see http://www.celestrak.com. In general, the two known backscatter measurements used by ASF constantly to perform the SAR calibrations include the Amazon rain forest in Brazil and site-specific corner reflectors installed in Alaska [ASF, 2002]. These midlatitude corner reflectors provide additional references for both radiometric and geometric calibrations in this application [Lu, 2005; Small et al., 1997; Williams, 2004]. To remove the center-bias phenomena and the background noise, the SAR data were processed from pixel intensity to backscatter coefficient, σ_0 (sigma-naught). For ASF's purpose, σ_0 is defined as

$$\sigma_0 = 10 \cdot \log\left\{ a2 \cdot \left[d^2 - \left(a1 \cdot n(r) \right) \right] + a3 \right\} \qquad (2\text{-}1)$$

where d is pixel intensity (0–255), a1 is noise scaling, a2 is linear conversion, a3 is offset, and n(r) is noise as a function of range. The coefficients are found in the Radiometric Data Record (part of the CEOS leader file) [Olmsted, 1993]. The σ_0 is expressed in decibel (dB). The σ_0 was, afterward, converted to a digital number (DN) to be used for deriving the soil moisture model as the following:

$$DN = \left(\sigma_0 * 10 \right) + 255 \qquad (2\text{-}2)$$

Thus the SAR imagery acquired in April 2004 was geometrically corrected using ground control points, including corner reflectors and even some more references, such as street

intersections, SPOT satellite imageries, and Digital Ortho Quarter Quads (DOQQs) optical images. The RMSE after the georeferencing became less than 8 m. The other SAR imagery acquired in September 2004 was then rectified on the basis of the April 2004 SAR data. By overlaying the two SAR data together, the spatial error between the two acquisitions can be minimized further. This technique is normally used in multitemporal study to detect changes in time. The RMSE eventually reaches a level of less than 2 m.

Slope and aspect data

Both slope and aspect may directly influence the return signals to the SAR sensor [van Zyl et al., 1993; Jeremy, 2002; Baghdadi et al., 2002; Le He´garat-Mascle et al., 2002]. The required DEM data in this study can be downloaded from Texas Natural Resources Information System at http://www.tnris.state.tx.us/DigitalData/DEMs /dems.htm. Once the DEM is obtained and imported into the ArcGIS workspace, we used the ArcGIS 8.3 Slope function and Aspect function to derive the slope and aspect data, respectively. The Slope and Aspect functions are located under the Surface Analysis submenu in the Spatial Analyst Toolbar [ESRI, 2004a, 2004b]. Once the Slope function is open, the program would require an input surface, which is the DEM data. We specify the output measurement to be in percent slope by selecting the option percent. The output cell size in both slope and aspect analyses is maintained at 30 m as same as the pixel size of the DEM in order to minimize the discrepancies in data synthesis and processing.

On the basis of such rationale, Figure 2-8 summarizes the image of slope in the CCRW. The image shows percent slopes with a range from 0 to 131.5%. As stated before, aspect data are required to represent the direction of the slopes because the slope data derived in GIS platform

only represent the magnitude of slopes, not the direction of the slopes. For deriving the aspect data, the DEM data, which were just used to derive the slope, are the only input required in the GIS software. The aspect is measured clockwise in 0° due north, 90° due east, 180° due south, 270° due west, and 360° due north again [see ESRI, 2004b]. Figure 2-9 summarizes the image of the aspect data with appropriate color indicator. The majority of the aspect values indicate the direction of slopes toward south (green) and southeast (cyan) directions that bear the similar natural directions of the stream system in the watershed.
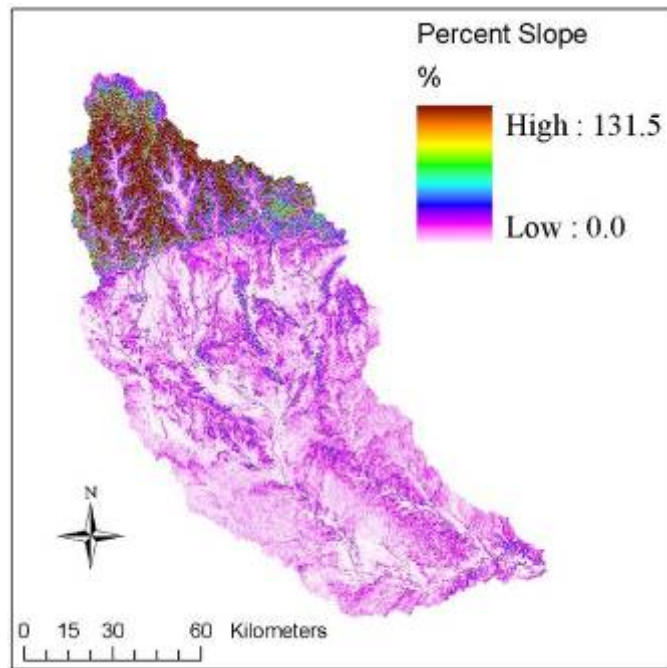
Figure 2-8. The map of slope in CCRW.
The slope map is showing the variation of slopes in the CCRW. Most area below the Balcones zone is very flat and is used mostly for agriculture and livestock. The upper area is the Edwards Plateau that comprises hills.
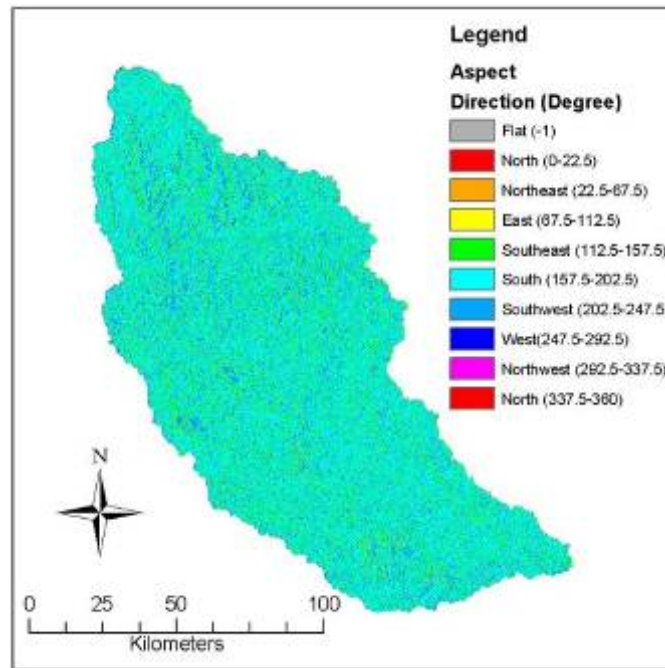
Figure 2-9. The map of aspect of slope in CCRW.
The aspect map shows the major directions of slopes are facing toward the South and the Southeast. It is the actual direction that the streams flow toward the Gulf of Mexico.

Soil permeability data

In the field of Geotechnical Engineering soils can be classified into groups and subgroups on the basis of their engineering behavior. Many general characteristics of soils can be used to express their description, but the grain size is a common use in many classification systems [Das, 1999]. STATSGO is the State Soil Geographic Data Base developed by the United States Department of Agriculture–Natural Resource Conservation Service (USDA-NRCS). STATSGO was created by generalizing soil survey maps, county general soil maps, state general soil maps, and state major land resource area maps. More information can be found at Earth System Science Center [2004]. The purpose of using the soil map is to incorporate the soil permeability into the model. The soil permeability refers to the ability of water and air to move through saturated soil. The permeability of soil is influenced by many factors, such as size and shape of the soil

particles, degree of saturation, and void ratio. For a given soil, permeability is inversely proportional to soil density. A map of soil permeability made up of 31 soil types in the CCRW is presented in Figure 2-10. The development of soil moisture model would benefit from incorporating soil permeability, measured in inch/hr, along with some other geoenvironmental variables and SAR.
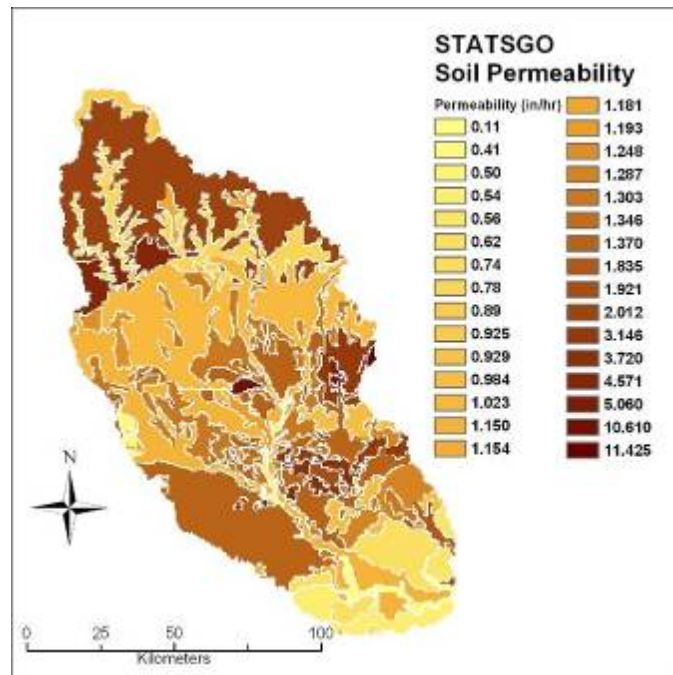
Figure 2-10. STATSGO soil classifications in CCRW.
Thirty-one classes of STATSCO soil classifications were found in the watershed. The average permeability (inch/hr) of the soils was included in the derivation of the soil moisture model.

Normalized difference vegetation index (NDVI) data

The NDVI data represent the density of plant growth of the vegetation that covers the land [D'Souza et al., 1993]. It is a measurement of the density of green vegetation on the ground. Not only does the NDVI represent the greenness, but also it can roughly measure the features of land surface as well. A total range of index values can correspond to leaves, trees, shrub, grassland, forests, bare soils, exposed rocks, sand, or snow [Weier and Herring, 1999]. All of

those land surface features also give out different roughness [Gupta et al., 2002], and they consequently affect the radar backscatter. In addition, the density of plant growth could represent the depletion of soil moisture via transpiration. Thus the NDVI data may be used to estimate soil moisture in combination with SAR data.

The NDVI can be derived from a multispectral sensor's data, such as Advanced Very High Resolution Radiometer (AVHRR) or LandSat. The NDVI values range from -1 to 1. Cloud cover is the common obstacle for this type of sensor. Masking out the cloud can be done by composing an image from multiple images acquired from the sequenced data. Weier and Herring [1999] approximated the corresponding values of NDVI to many land features. Very low values of NDVI (0.1 and below) correspond to barren areas of rock, sand, or snow. Moderate values (0.2–0.3) represent shrub and grassland, while high values (0.6–0.8) indicate temperate and tropical rain forests. Spatial variation of plant density and plant species would not be very phenomenal in this semiarid river basin. While the difference of soil moisture could vary within a few meters, the vegetation density could be relatively the same within a few acres. Considering the estimates of soil moisture for the whole watershed with an area of 14,200 km^2, 1-km pixel size of AVHRR-derived NDVI may still address 14,200 samples of land cover pattern in totality. The AVHRR is therefore chosen to derive NDVI in this study because of its shorter satellite repeat cycle. In applications, the NDVI is converted to digital number by adding 1 to the NDVI and then multiplying it by 100 to generate the digital number (DN) for use in the model. The DN value of NDVI, varied from 0 to 200, is thus used for developing the soil moisture model. The AVHRR NDVI data are provided by the USGS EROS Data Center (EDC) DAAC [Distributed Active Archive Center, 2005]. The data sets generated by USGS actually contain weekly and biweekly NDVI composite products for public uses. The AVHRR NDVI data used in the study

were acquired in May and September 2004 to surrogate the changes of vegetation density during the wet and dry seasons (see Figure 2-11a and 10b). It should also be noted that SAR may penetrate the sparse vegetation cover because of having larger wavelength. Thus the water content in vegetation would not significantly affect SAR backscattering signals in this semiarid vegetation area.

(a) May 2004 (wet month)   (b) September 2004 (dry month)
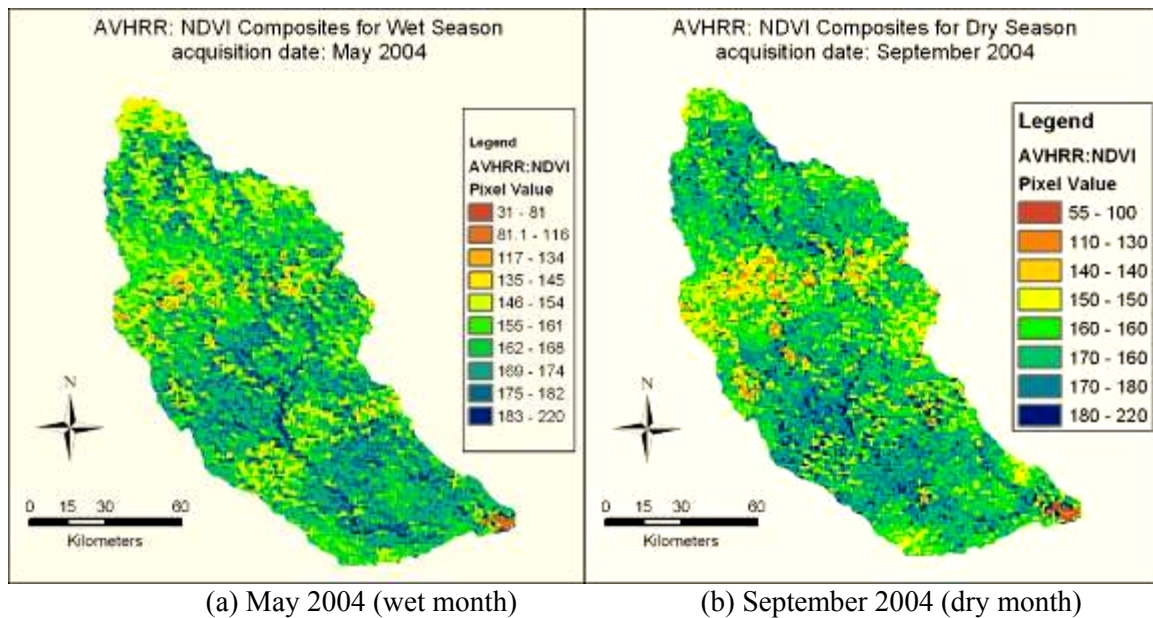
Figure 2-11. Maps of NDVI in CCRW in acquired in April and September 2004.
Maps of AVHRR NDVI images show high values of greenness along the river corridor (shown in dark blue). The plowed land and farms are likely presented in orange and yellow. Texas brush land, ranches and grassland are shown in dark and light greens. Red area in the maps indicates the water body of the Choke Canyon Reservoir.

Model development for soil moisture estimation

Overall, the backscatter might be more strongly influenced by roughness than soil moisture. Roughness is the variations in the height of a physical surface. It is a relative measurement in term of scale. This is because grain size is an order of magnitude or smaller than the roughness element (clods, clumps and row structures) that drive the ''roughness'' response in the backscatter signal on the order of mm to cm. While it is true that soil texture or grain size

plays a role in how cloddy or rough a soil will be after tillage, it is the tillage that determines the size of clods in agricultural area. Yet it is impossible to keep track of the changes of roughness in the dynamic system all by the in situ measurement in the watershed with an area of over 14,200 km^2, whereas this type of measurement can be done easily in a small study area. Therefore integrative use of the soil permeability addressing the feature of soil texture and the NDVI values implying the inherent density of plant species are used to collectively reflect some sort of surface roughness in the GP model development. This would also reduce the possible correlation among exogenous variables in the model from a statistical sense. The model should also be valid to some other watersheds where the soil permeability may fall into the same range, between 0.5 in/hr and 3.7 in/hr, as the soil permeability values found in this watershed.

To increase the model credibility two thirds of both data sets collected from the two ground truth campaigns were combined together for model calibration and one third of them were randomly picked out for model validation. Only the samples with the soil moisture less than 50% are adopted in ground truth campaigns because the TDR probe was normally calibrated in the range of 0 to 50%. This is of concern because some soil contains high content of salinity which induces error in probe readings abnormally and the gravels in the soil did affect the probe readings as well.

Data extraction, which is required for model development, is a process of retrieving the value of the pixel that lies underneath the ground truth measurement point. First, all the data have to be imported into the same coordinate system in GIS workspace. The ground truth database may provide a set of accurate locations using the sub-meter GPS. Then we may extract those associated values of NDVI, slope, aspect, soil permeability, and backscatter coefficient (σ_0) from

individual map layer based on the GPS measurements, respectively. Even though there are discrepancies of cell sizes in different input databases when preparing for regression analysis, with such a data extraction strategy there is no need to resample data spatially in order not to disturb the data integrity. Most importantly, SAR data are not averaged for use in regression. Thus the required data are eventually extracted one layer at a time using GRIDSPOT add-in script developed by Rathert [2003]. The data are then exported into Microsoft® Excel and are randomly shuffled. The shuffled input data (i.e., two thirds of the ground truthing data points) are then fed into the GP model for calibration. Once the model can be properly calibrated, validation may be performed by using the rest of data points (i.e., one third of the ground truthing data points). The outputs using the calibrated soil moisture model can be compared against the unseen ground truth soil moisture samples (i.e., the validating data set). The estimation of soil moisture, as a consequence, is expected to be a function of the SAR data, the surface slope, the aspect of slope, the soil permeability, and the NDVI, as expressed below:

$$VWC = fn\left(V0, V1, V2, V3, V4\right) \qquad (2\text{-}3)$$

where VWC is the percent volumetric water content in soil, V0 is the SAR backscatter coefficient that is converted to DN value (0–255), V1 is the slope value in percent, V2 is the aspect value in degree, V3 is the STATSGO soil permeability (in/hr), and V4 is the NDVI that is converted to digital number (0–200).

Results and Discussion

A GP-derived soil moisture model was proved useful to accommodate the soil moisture estimates of the CCRW. With the aid of the GP algorithm, the soil moisture model is shown in equation (4). The model is presented in forms of compound functions because of the high complexity of its nature.

GP model:

$$\text{Soil Moisture (\% volumetric)} = \sqrt{2 \cdot \left(\left| A1 * V0 \right| + V4 \right)} + 4 \cdot (A3) - 4 \cdot (V3) - \frac{V1}{0.924} \qquad (2\text{-}4)$$

$$A1 = \frac{\left| Cos(A2) \right|}{V3} - \left| Sin(A5) \right| + 2 \cdot (A8)^2$$

$$A2 = \frac{\left| Sin(A5) \right| + V4}{0.924}$$

$$A3 = A4 - \left| Sin(A5) \right|$$

$$A4 = 2 \cdot \left[Sin\left(2 \cdot (A9)^2 \right) \right]^2$$

$$A5 = \left\{ \frac{Sin\left[Cos\left(\sqrt{A6} + V4 \right) \right]}{V3} \right\} + V4$$

$$A6 = 2 \cdot (A7) + 4 \cdot (A8) + 4 \cdot (V4)$$

$$A7 = \left[\frac{1}{0.16 \cdot (A4)} \right]^2$$

$$A8 = Sin\left(2 \cdot (A9)^2 \right)$$

$$A9 = 2 \cdot (V4) + 2.03$$

$$V0 = \text{SAR backscattering coefficient (0-255)}$$

V1 = slope (%)

V3 = soil permeability (in/hr)

V4 = NDVI (0 – 200)

Note that independent variable V2 is not included in this best selected GP model. V0 is SAR backscatter coefficient (digital number: 0–255); V1 is slope value in percent (%); V3 is soil permeability (in/hr); and V4 is NDVI (digital number: 0–200).

The effect of slope in the estimates of soil moisture in the GP-derived model is only valid where the slope is below 2% in this study. The first reason is that the model was calibrated with the slope values less than 2%. There was no ground truth in the steep slope area where the soil moisture measurement was taken. Therefore the model cannot estimate soil moisture where the slope value is out of range. Second, the majority of the CCRW are flat and thus the terrain correction algorithm, which is normally used to compensate for foreshortening and shadowing phenomenon, was not applied to the SAR data. Because of this reason the slope and aspect information was purposely included into the GP model to compensate for the lack of terrain correction.

In the GP model, ''frequency of use'' would be the only way to quantitatively delineate the relative importance of input factors (i.e., exogenous variables) being included in the regression models. ''Appearing frequency'' in the evolutionary computing process provided by Discipulus® was employed to evaluate the relative importance of those exogenous variables. After GP had generated millions of evolutionary models, each input factor was counted as how many times the input factor was used in the models in a way that contributes to the fitness of the models. A value of 100% (i.e., frequency of use) indicates that the input variable is used in 100%

of the generated models. Table 2-2 summarizes the statistics. It shows SAR, soil permeability, and NDVI are mostly important in all scenarios. Slope and aspect, however, are relatively not as important as the other parameters in the models for estimating soil moisture since slope and aspect data were used only 70% and 50%, respectively. Higher ''appearing frequency'' of the NDVI data in the model selection process indicates that vegetation greenness is equally important predictor of soil moisture as backscatter since vegetation greenness is physically related to soil moisture via transpiration. The best model chosen out of many millions of generated models does not include the aspect data eventually. To validate the GP models, the calculated soil moisture values were compared against the measured soil moisture values in the unseen sub–dataset pair-wise. Figure 2-12 reflects a summary of a comparison between the measured soil moisture and the estimated soil moisture based on the April 2004 data set. The model presents the value of R-square of 0.72 and the values of corresponding RMSE of 3.4%. On the other hand, Figure 2-13 demonstrates the same comparison using the value of R-square of 0.69 and the value of RMSE of 2.3% based on the September 2004 data set. Table 2-3 summarizes the statistical evaluation of the models based on r-square value between the predicted values and the unseen observed data.

Table 2-2. Relative importance analysis for input factors in the regression models

Input	Frequency of Use (%)
Backscattering Coefficient	100
Slope	70
Aspect	50
Soil Permeability	100
NDVI	100

Table 2-3. Statistical evaluation of regression models (Unseen data are used.)

Model	R-square		RMS Error	
	APR Data	SEP Data	APR Data	SEP Data
GP Model	0.72	0.69	3.4	2.3
Multiple linear regression Model	0.36	0.34	8.9	10.1
Nonlinear regrssion model # 1	0.27	0.36	11.5	9.4
Nonlinear regression model # 2	0.30	0.33	11.8	8.8

Research findings also indicate that the ground truth measurements with larger value of soil moisture are likely to generate disturbance in model development (see Figure 2-12 and Figure 2-13) because the TDR probe is normally calibrated for soil moisture levels that are below 50%. Also, the more the soil moisture measurements spreading out of the normal range, the higher the chance that it may inherently bear with more measurement errors. Further, the different salinity content in soils area wide might also affect the measurement accuracy. Yet soil salinity measurement using remote sensing, such as the Scanning Low Frequency Microwave Radiometer (SLFMR) [Le Vine et al., 1994, 1998], in such a vast watershed is out of current research capability in remote sensing community.
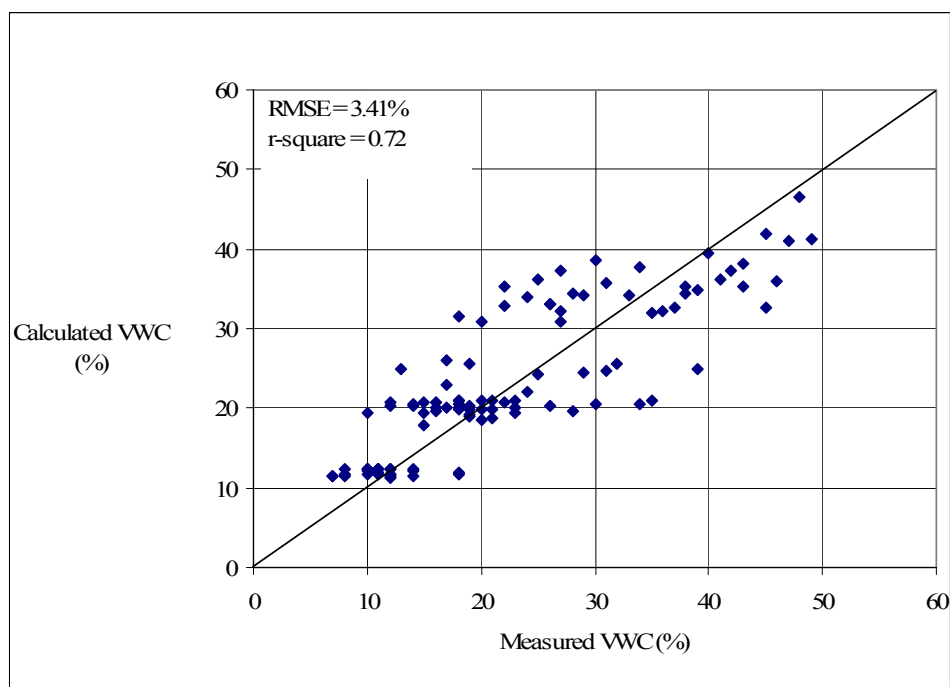
Figure 2-12. Plot of measured vs. calculated soil moisture of the unseen samples collected in April 2004. The unseen samples used in the calculations are independent from the samples used to calibrate the model.
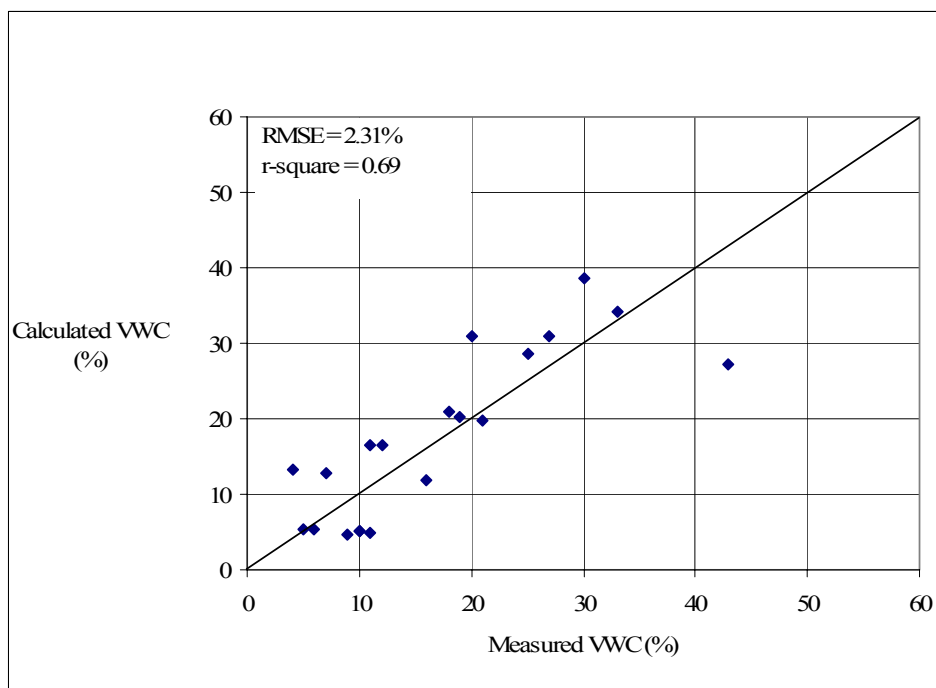
Figure 2-13. Plot of measured vs. calculated soil moisture of the unseen samples collected in September 2004.
The unseen samples used in the calculations are independent from the samples used to calibrate the model.

With the aid of soil moisture model derived by the GP technique, Figure 2-14 and Figure 2-15 present the soil moisture estimations watershed wide. The grid cell resolution of Figure 2-14 and Figure 2-15is kept at 25 m. Two maps of soil moisture were eventually generated on the basis of the same soil moisture model derived in equation (4). Figure 2-14 shows the map of soil moisture in April 2004 that has a mean soil moisture value of 16.3% by volume. Figure 2-15 shows the map of soil moisture in September 2004 that has a mean soil moisture value of 15.5% by volume. High values of soil moisture are present along the river corridors in both maps.
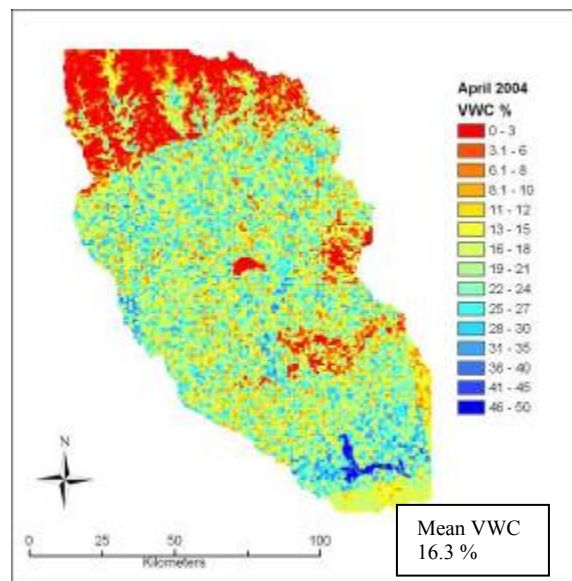
Figure 2-14. The soil moisture map derived from RADARSAT-1 SAR data acquired in April 2004 The average volumetric soil moisture is 16.3 %. The low estimates of soil moisture at the top area due to the high percent slopes. The patch of low soil moisture at the center in the map occurs from the large value of soil permeability. The high values of soil moisture at the bottom of the map occur from the very small values of soil permeability.
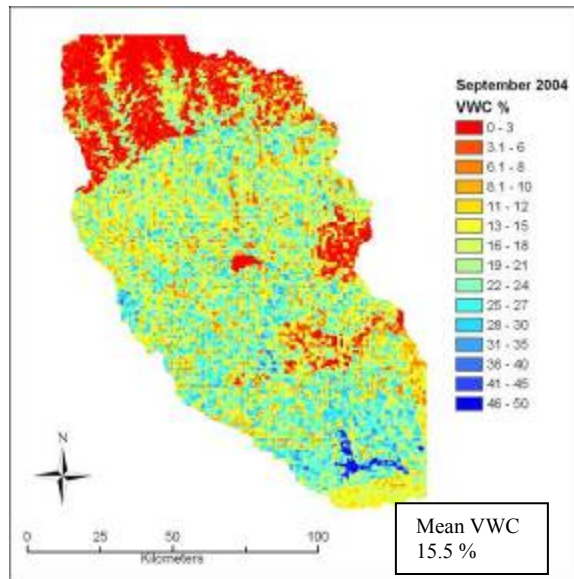
Figure 2-15. The soil moisture map derived from RADARSAT-1 data acquired in September 2004
The average volumetric soil moisture is 15.5 %. The extreme values of percent slope and soil
permeability affect the estimation of soil moisture as similar as Figure 2-14.

The soil permeability shows a significant effect in the estimates of soil moisture in the

model. According to equation (4), the soil permeability ''V3'' is a subtrahend and a denominator

in different part of expressions in the model simultaneously. Obviously the soil moisture

decreases as the soil permeability increases. This is a reason for Figure 2-14 and Figure 2-15 to

show a patch of very low soil moisture at the center of the maps due to the very high value of

soil permeability as evidenced in Figure 2-10. Similarly, the patches of high soil moisture areas

at the bottom of Figure 2-14 and Figure 2-15 are due to the very low soil permeability as

evidenced in Figure 2-10 too. This phenomena follows the physical sense that the soil with high

permeability allows water and air to move more freely, thus it retains less water content.

According to equation (4), the slope parameter is a subtrahend. It confirms that the increases of

slope reduce the water content in the soil. Obviously the steep slope (large value of slope) would

generate more runoff, and this drains the water from the soil more. This explains why there is a

presence of very low soil moisture patch at the top of Figure 2-14 and Figure 2-15 because of the high values of slope as evidenced in Figure 2-8.

The effects of the NDVI to the soil moisture can be analyzed by two scenarios. First, high value of the NDVI refers to high density of plants' leaves or high greenness in canopy level. This could imply the abundance of water available in the soil that the plants can use for their photosynthesis. It could also imply that the evapotranspiration is supportive for carboxylation where the NDVI is high. The second scenario could be the opposite in a way that high density of plants' leaves causes high transpiration rate, and consequently the soil moisture should be low because of the water depletion. However, the second scenario has a flaw that if the water available in the soil is low, how could the plants maintain high productivities (indicated by the density of green leaves)? Thus the first scenario has the higher probability of being true. According to the model (equation (4)), the NDVI parameter only adds its value into the model since it is never being used as a subtrahend. It may be concluded that the NDVI implies the abundance of water content in the soil rather than the depletion of the soil moisture.

The question ''why use radar images for soil moisture estimation when NDVI is so readily available?'' can be answered on the basis of such findings now. On the basis of the model in equation (4), it indicates that vegetation greenness is an equally important predictor of soil moisture as backscatter. Even though the AVHRR-derived NDVI was used to support the modeling analysis with 1-km spatial resolution, spatial variation of plant density and plant species would not be phenomenal in this semiarid river basin and the modeling outputs as shown in Figure 2-14 and Figure 2-15 can still maintain the resolution of 25 m. Although order of magnitude difference in speckle over a few pixel domains will negate site specific results, the

model was not derived by using averages of several pixels of SAR measurements in model development. Data extraction was made for the SAR measurement and its corresponding value of ground truth by referring to the GPS record having submeter accuracy. Hence speckle noise across pixels is not a main concern here.

It is very difficult to determine the soil roughness because of several factors that influence the roughness over such a vast watershed. The roughness of the same type of soil can be totally different depending on its use. In situ measurement of the soil surface roughness in such a vast watershed is impossible. The surface roughness, a factor that could be more significant than soil moisture in determining backscatter coefficient, can be collectively addressed by the NDVI and the soil permeability in the model. In our study area, the NDVI values do not have drastic change over time. This is true at least for three types of land covers, including pasture, shrub, and riparian buffer. Only crop row may have a relatively big change, as evidenced by Figure 2-16. This is a supportive finding for the modeling work since we have tried to use the NDVI to address part of the surface roughness effect.
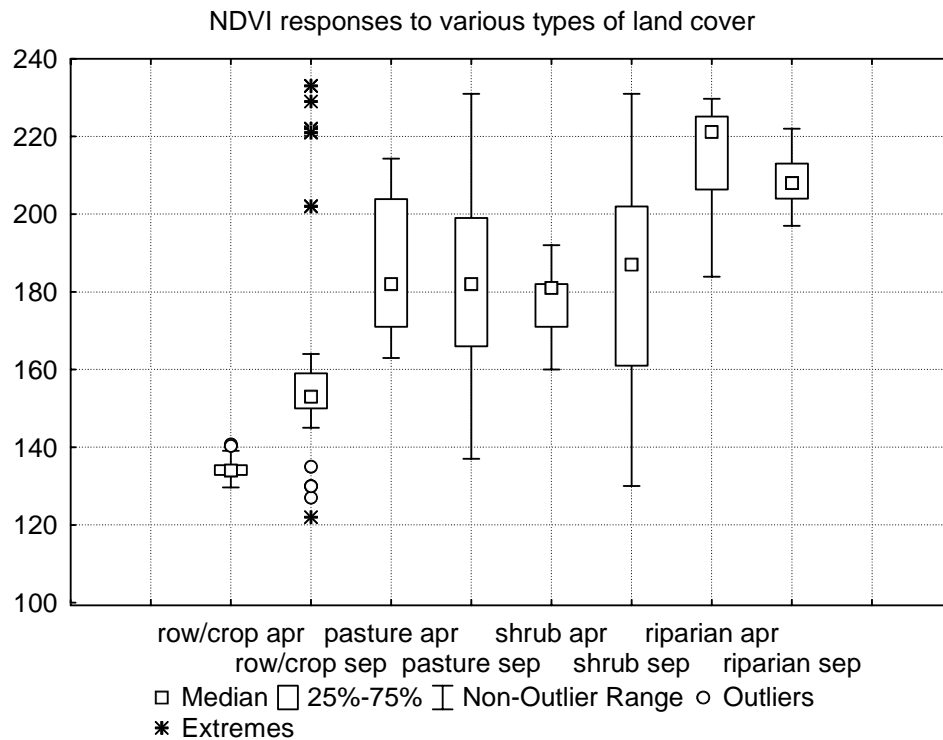
Figure 2-16. The box plots present NDVI values corresponded to various types of land cover over a time period from April to September 2004. NDVI of row/crop increased. NDVI of pasture insignificantly changed over the time period. The shrub area responded to a wider range of NDVI values in September than in April. The NDVI of the riparian area dropped over the time period.

For the purpose of comparisons, this study also developed some multiple linear and nonlinear regressions using stepwise approach with all related input variables. As shown in equations (5), (6), and (7), a multiple linear regression and two multiple nonlinear regressions were created in order to compare the outputs against those from the GP model. In the process of developing the two nonlinear regression models, a conventional power law type fit is calibrated including eleven coefficients and the five independent variables. As an alternative another nonlinear regression is created with a higher degree of complexity by fusing many nonlinear forms that best fit individual independent variable. This process mimics the law of nature selection, which is similar to the selection algorithm which is utilized by the genetic

programming, by manually selecting the regression type that best fits each independent variable to the ground data pair-wise. The selected forms of regression are fused together to form the second nonlinear regression model as shown in equation (7). Fourteen coefficients and five independent variables are included.

Linear regression:

$$\text{Soil Moisture (\% volumetric)} = -1.016 + 0.084(V0) - 4.021(V1) - 0.002(V2) \quad (2\text{-}5)$$
$$- 9.126(V3) + 0.138(V4)$$

Nonlinear regression 1:

$$\text{Soil Moisture (\% volumetric)} = -18.299 + 63.98 \cdot (V0)^{-0.516} + 89.62 \cdot (V1)^{-0.025} \quad (2\text{-}6)$$
$$+ 27.79 \cdot (V2)^{-0.602} + 155.33 \cdot (V3)^{-0.041} - 603.92 \cdot (V4)^{-0.196}$$

Nonlinear regression 2:

$$\text{Soil Moisture (\% volumetric)} = 10.890 + \frac{29063844.9}{772415.4 + V0} - 15.309 \cdot (V1)^{0.24} \quad (2\text{-}7)$$
$$+ \frac{1.063}{0.036 \cdot (V2)^{-0.613}} - 57.064(V3) + 36.608 \cdot (V3)^{2}$$
$$- 2.738 \cdot (V3)^{3} - 0.961 \cdot (V3)^{4} + \frac{727313.0}{3738763.4 + V4}$$

where V0 is the SAR backscatter coefficient that is converted to DN value (0–255), V1 is the slope value in percent, V2 is the aspect value in degree, V3 is the STATSGO soil permeability (in/hr), and V4 is the NDVI that is converted to digital number (0–200).

The correlation between the measured soil moisture and the calculated soil moisture in the linear regression model and the two nonlinear regression models are weak. The highest R-square value was only 0.36 at its best with the RMS error of 8.9%. Table 2-3 concludes the evaluation of these models on the basis of the same criteria. Figure 2-17 presents box plots of the observed and predicted soil moisture. The GP model shows the best result that most resembles the observed data with somewhat similar interquartile range. On the other hand, the linear and nonlinear regression models produce results with approximately a half interquartile range of the observed data. In regard to the estimations, the linear regression model tends to overestimate soil moisture, while the nonlinear regression models tend to underestimate soil moisture. By comparing equations (4), (5), (6), and (7) in totality, the nonlinear structure of the GP model generated is viewed solid than the other counterparts. As opposed to the conventional nonlinear regression models, GP increases the chance of developing successful nonlinear functions because of its unbounded complexity. Yet the attempt to mimic the law of nature selection by manually selecting a best fit regression type takes a lot of time and efforts. While GP seeks the best forms of regressions through millions of selections, it is impossible for a human to do the equivalent task. It can be concluded that the GP-derived model in this study is much better than its counterparts no matter if they are either multiple linear or nonlinear regression models.
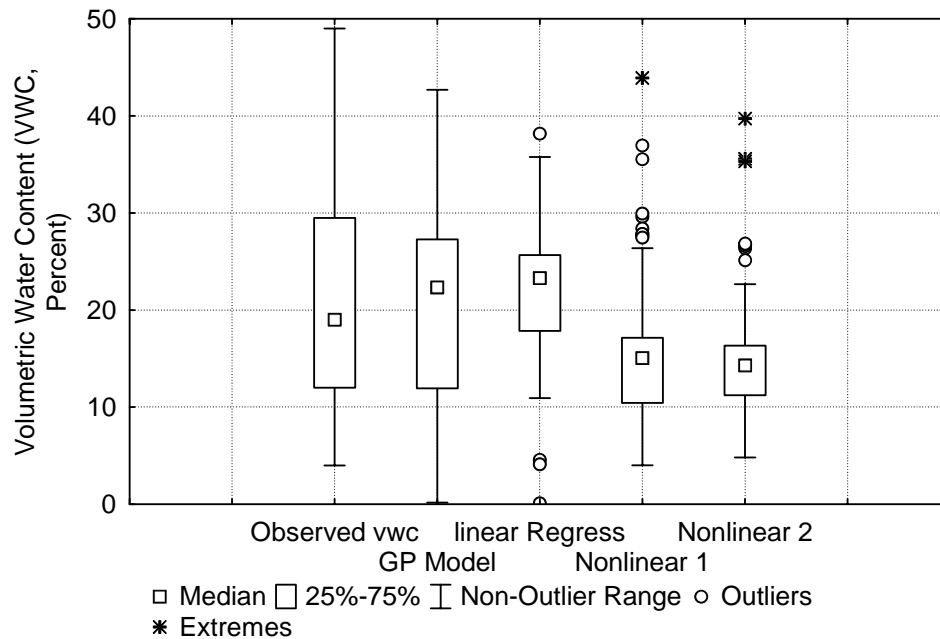
Figure 2-17. Box plots of observed and calculated soil moisture generated from different models. The result of GP model represents very similarly to the observed data. Its interquartile range is almost as wide as the interquartile range of the observed data, while the other models only result in a half of the interquartile range. The linear regression model results the over estimation, and the two nonlinear models result the under estimations.

Conslusions

This study presents a systematic data synthesis and analyses that lay down the foundation for the multitemporal soil moisture estimation in the study area. It uniquely demonstrates the use of remote sensing of hydrologic fluxes, states, and parameters, including combined active microwave and optical observations, to improve the understanding of the soil moisture variability in the terrestrial hydrosphere. The GP-derived soil moisture model is proved useful to identify the correlations between soil moisture measurements, SAR backscatter coefficient, geographical and topographical features at a watershed scale. The slope and the aspect (direction of slope) were particularly included in the development of the models to enhance the formulation credibility but are proved insignificant in this terrain because of the flatness. Yet the NDVI and

soil permeability data may significantly influence the estimates of soil moisture. The GP model exhibits a credible record supported statistically by R-square value of 0.72 and RMSE of 3.4 based on the April 2004 data set, and R-square value of 0.69 and RMSE of 2.3 based on the September 2004 data set. When comparing to the multiple linear and nonlinear regression models, the GP model provides an acceptable agreement with observed measurements under conditions in which slope is less than 2% on average in the lower portion of the CCRW.

Such a case study in Texas promotes the scientific justification of new measurements, involving satellites and artificial intelligence algorithms that potentially support several key scientific regimes:

i. the application of new technologies for remote sensing hydrologic quantities for terrestrial hydrologic interpretation;

ii. completion of studies on appropriate spatial and temporal sampling scales of new synergy of optical and microwave sensors for satisfying specific scientific objectives; and

iii. enhancement the information on flood and drought prediction systems indirectly, which is deemed ecologically important for the basin management authority, especially in the semiarid coastal region, south Texas.

Comparing the work to others [Moran et al., 2000; Salgado et al., 2001; Glenn and Carr, 2003], this GP-derived soil moisture model provides a nonlinear functional form that enhances hydrologic model capability and performance through modern data collection, assimilation, and analysis techniques to incorporate remotely sensed observations, which may include efforts to resolve spatial scale discrepancies between in situ and satellite observations. The quantification

of soil moisture will be used to estimate water availability of the watershed in general seasons in order to assist hydrologists, engineers, and stakeholders in managing water resources in this semiarid watershed. Such development also serves the scientific basis in the future for observing and modeling large scale terrestrial water-storage dynamics with emphases on how these processes are affected by the heterogeneity of soil, vegetation, precipitation, and topography and even their interaction with various biogeochemical cycles.

References

Alaska Satellite Facility, 1999. RADARSAT-1 Standard Beam SAR Images. report. Geophysical Institute, University of Alaska Fairbanks, Fairbanks, Alaska, USA.

Alaska Satellite Facility, 2002a. ASF Inrterferometric SAR Processor (AISP) Calibration Report. Version 4.0. Alaska Satellite Facility, Geophysical Institute, University of Alaska Fairbanks, AK, USA.

Alaska Satellite Facility, 2002b. Terrain Correction. tutorial. Advanced Product Development (APD), Geophysical Institute, University of Alaska - Fairbanks, AK, USA. URL: http://www.asf.alaska.edu/apd/software/tutorial/terrain_correction.html, last update FEBRUARY 2002.

Back, T., Hammel, U., and Schwefel, H.-P., 1997. Evolutionary computation: Comments on the history and current state. IEEE Transactions on Evolutionary Computation 1 (1):3-17.Fogel, L.J., A. J. Owens, and M. J. Walsh (1966), Artificial Intelligence through Simulated Evolution. John Wiley & Sons, New York.

Baghdadi, N., King, C., Chanzy, A., and Wigneron, J.P., 2002. An empirical calibration of the integral equation model based on SAR data, soil moisture and surface roughness measurement over bare soils. International Journal of Remote Sensing 23 (20), 4325 – 4340.

Chang, N.B. and Chen, W.C., 2000. Prediction of PCDDs/PCDFs emissions from municipal incinerators by genetic programming and neural network modeling. Waste Management & Research 18, 341 – 351.

Cramer, N.L., 1985. A representation for the adaptive generation of simple sequential programs. Proceeding of the International Conference on Genetic Algorithms and Their Application, 183-187.

D'Souza G., Belward, A.S., and Malingreau, J.P., 1993. Advances in the Use of NOAA AVHRR Data for Land Applications, Remote Sensing Vol. 5. Kluwer Academic Publishers, London, UK.

Das, B.M., 1999. Soil Classification. Fundamentals of Geotechnical Engineering, Brooks/Cole, Pacific Grove, CA, USA, 35 – 44.

Del Frate, F., Ferrazzoli, P., and Schiavon, G., 2003. Retrieving soil moisture and agricultural variables by microwave radiometry using neural networks. Remote Sensing of Environment 84, 174 – 183.

Dubois, P.C., van Zyl, J., and Engman, T., 1995. Measuring soil moisture with imaging radars. IEEE Trans. Geosci. Remote Sensing 33 (4), 915 – 926.

Earth System Science Center (ESSC), 2004. STATSGO, Soil Information for Environmental Modeling and Ecosystem Management. Pennsylvania State University, PA, USA. URL:// http://www.essc.psu.edu/soil_info/index.cgi?index.html last update 22 June 2004.

ESRI, 2004a. Spatial Analyst: Deriving slope. in ArcGIS 8.3 Help File, Redlands, CA, USA.

ESRI, 2004b. Spatial Analyst: Aspect. in ArcGIS 8.3 Help File, Redlands, CA, USA.

Fogel, L.J., Owens, A.J., and Walsh, M.J., 1966. Artificial Intelligence through Simulated Evolution. John Wiley & Sons, New York.

Francone, F.D., 1998. Discipulus™ Software Owner's Manual, version 3.0 DRAFT. Machine Learning Technologies, Inc., Littleton, CO, USA.

Freeman, A., 1992. SAR calibration: an overview. IEEE Trans. on Geoscience and Remote Sensing 30 (6), 1107 – 1121.

Gagne, C. and Parizeau, M., 2004. Genericity in evolutionary computation software tools: principles and case-Study. Technical report RT-LVSN-2004-01. Laboratoire de Vision et Systemes Numerique (LVSN), Universite Laval, Quebec, Canada. 28 October 2004.

Glenn, N.F. and Carr, J.R., 2003. The use of geostatistics in relating soil moisture to RADARSAT-1 SAR data obtained over the Great Basin, Nevada, USA. Computers and Geosciences 29, 577 – 586.

Glenn, N.F. and Carr, J.R., 2004. The effects of soil moisture on synthetic aperture radar delineation of geomorphic surfaces in the Great Basin, Nevada, USA. Journal of Arid Environment 56 (4), 643 – 657.

Gupta, R.K., Prasad, T.S., and Vijayan, D., 2002. Estimation of roughness length and sensible heat flux from WiFS and NOAA AVHRR data. Adv. Space Res 29 (1), 33 – 38.

Heywood, M.I., and Zincir-Heywood, A.N., 2002. Dynamic page based linear genetic programming. IEEE Trans. on Systems, Man and Cybernetics – Part B, Cybernetics 32 (3), 380 – 388.

Holland, J.M., 1975. Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor, MI, USA.

Jackson, T.J., Le Vine, D.M., Griffis, A.J., Goodrich, D.C., Schmugge, T.J, Swift, C.T., and O'Neill, P.E., 1993. Soil moisture and rainfall estimation over a semiarid environment with the ESTAR microwave radiometer. IEEE Transactions on Geoscience and Remote Sensing 31 (4), 836 - 841.

Koza, J.R., 1992. Genetic Programming: On the Programming of Computers by Means of Natural Selection. MIT Press, Cambridge, MA, USA.

Koza, J.R., 2004. URL: http:genetic-programming.org, last update 16 September 2004.

Le Hégarat-Mascle, S., Zribi, M., Marticorena, B., Bergametti, G., Kardous, M., Callot, Y., Chazette, P., and Rajot, J-L., 2003. Use of ERS/SAR measurements for soil geometric and aerodynamic roughness estimation in semi-arid and arid areas. Geoscience and Remote Sensing Symposium, 2003. IGARSS '03. Proceedings, IEEE International 4, 2272 – 2274.

Le Hégarat-Mascle, S., Zribi, M., Alem, F., Weisse, A., and Loumagne, C., 2002. Soil moisture estimation from ERS/SAR data: toward an operational methodology. IEEE Trans. on Geoscience and Remote Sensing 40 (12), 2647 – 2658.

Le Vine, D.M., Griffis, A.J., Swift, C.T., and Jackson, T.J., 1994. ESTAR: A synthetic aperture microwave radiometer for remote sensing application. Proc. IEEE 82, 1787 – 1801.

Le Vine, D., Kao, M., Garvine, R., and Sanders, T., 1998. Remote sensing of ocean salinity: results from the Delaware coastal current experiment. J. Atmos. Oceanic Tech. 15, 1478 – 1484.

Lu, Z., 2005. Satellite radar remote sensing: applications to the study of Earth sciences and natural resources. Technical Seminars, USGS/EROS, Sioux Falls, SD. URL:http://www.asf.alaska.edu/educational/seminarpdf/2005/satellite_radar_remote sensing.pdf

Moeremans, B. and Dautrebande, S., 2000. Soil moisture evaluation by means of multi-temporal ERS SAR PRI images and interferometric coherence. Journal of Hydrology 234, 162 – 169.

Moran, M.S., Hymer, D.C., Qi, J., and Sano, E.D., 2000. Soil moisture evaluation using multi-temporal synthetic aperture radar (SAR) in semiarid rangeland. Agricultural and Forest Meteorology 105, 69 – 80.

Narayanan, R.M. and Hirsave, P.P., 2001. Soil moisture estimation models using SIR-C SAR data: a case study in New Hampshire, USA. Remote Sensing of Environment 75, 385 – 396.

National Land Cover Data (NLCD), 2003. USGS EROS Data Center (EDC) DAAC. Retrieved in October 2003 from http://landcover.usgs.gov.

Njoku, E.G., Wilson, W.J., Yueh, S.H., and Rahmat-Samii, Y., 2000. A large-antenna microwave radiometer-scatterometer concept for ocean salinity and soil moisture sensing. IEEE Trans. Geosci. Remote Sensing 38 (6), 2645 – 2655.

Nolan, M., 2003. DInSAR measurement of soil moisture. IEEE Transactions in Geoscience and Remote Sensing 41 (12), 2802 – 2813.

Olmsted, C., 1993. Alaska Satellite Facility Scientific SAR User's Guide: ASF-SD-003. Geophysical Institute, University of Alaska Fairbanks, Alaska, USA, 53.

Owe, M., Chang, A.T.C., and Golus, R.E., 1988. Estimating surface soil moisture from satellite microwave measurements and a satellite derived vegetation index. Remote Sensing of Environment 24, 331 – 345.

Rathert, D., 2003. GRIDSPOT. in ESRI Support Center, GIS Programmer/Analyst. Resource Data Inc., www.resdat.com. E-mail: drathert@resdat.com.

Roth, C.H., Malicki, M.A., and Plagge, R., 1992. Empirical evaluation of the relationship between soil dielectric constant and volumetric water content as the basis for calibrating soil moisture measurements by TDR. Journal of Soil Science 43, 1 – 13.

Salgado, H. A., Génova, L., Brisco, B., and Bernier, M., 2001. Surface soil moisture estimation in Argentina using RADARSAT-1 imagery. Canadian Journal of Remote Sensing 27 (6), 685 – 690.

Sarabandi, K., 1994. Calibration of a polarimetric synthetic aperture radar using a known distributed target. IEEE Trans. on Geoscience and Remote Sensing 32 (3), 575 – 582.

Sarabandi, K., Pierce, L.E., and Ulaby, F.T., 1992. Calibration of a polarimetric imaging SAR. IEEE Trans. on Geoscience and Remote Sensing 30 (3), 540 – 549.

Shi, J., and Dozier, J., 1995. Inferring snow wetness using C-band data from SIR-C's polarimetric synthetic aperture radar. IEEE Trans. Geosci. Remote Sensing 33 (4), 905 – 914.

Shi, J., and Dozier, J., 1997. Mapping seasonal snow with SIR-C/X-SAR in mountainous areas. Remote Sensing of Environment 59 (2), 294 – 307.

Small, D., Holecz, F., Meier, E., Nuesch, D., and Barmettler, A., 1997. Geometric and radiometric calibration of RADARSAT images. Proc. of Geomatics in the Era of RADARSAT, Ottawa, Canada.

Song, D., Heywood, M.I., and Zincir-Heywood, A.N., 2003. A linear genetic programming approach to intrusion detection, in E. Cantú-Paz et al. (Eds.): GECCO 2003, LNCS 2724, Springer-Verlag Berlin Heidelberg, 2325 – 2336.

Spectrum Technologies, Inc., 2004. User's Manual Catalog # 6430FS: Field Scout™ TDR 300 Soil Moisture Meter. Plainfield, IL, USA. URL: http://www.specmeters.com/pdf/6430FS.pdf

Topp, G.C., Davis, J.L., and Annan, A.P., 1980. Electromagnetic determination of soil water content: Measurements in coaxial transmission lines. Water Resources Research 16 (3), 574 – 582.

Trimble Navigation Limited, 2004. GEO XT: datasheet and specifications. 749 North Mary Ave., Sunnyvale, CA, USA. URL: http://www.trimble.com/geoxt.shtml, retrieved in September 2004.

Tromso Satellite Station (TSS), 1996. TSS-JOR-1071/RNJ-96. Marketing Department, Tromso Satellite Station, Norway.

Ulaby, F., 1974. Radar measurement of soil moisture content. IEEE Transactions on Antennas and Propagation AP-22 (2), 257 – 265.

USGS EROS Data Center (EDC) DAAC, 2005. AVHRR Normalized Difference Vegetation Index (NDVI) Composites. USGS EROS Data Center (EDC) DAAC. URL: http://edc.usgs.gov//products/landcover/ndvi.html, last update 16 Feb 2005.

Van Zyl, J.J., Chapman, B.D., Dubois, P., and Shi, J., 1993. The effect of topography on SAR calibration. IEEE Trans. on Geoscience and Remote Sensing 31 (5), 1036 – 1043.

Walker, J.P., Willgoose, G.R., and Kalma, J.D., 2001. One-dimensional soil moisture profile retrieval by assimilation of near-surface measurements: A simplified soil moisture model and field application. Journal of Hydrometeorology, 2 (4), 356 – 373.

Weier, J. and Herring, D., 1999. Measuring vegetation (NDVI & EVI). Earth Observatory, NASA, USA. Retrieved from http://eobglossary.gsfc.nasa.gov//Library/MeasuringVegetation /index.html

Wigneron, J.P., Calvet, J.C., Pellarinb, T., Van de Griend, A.A., Berger, M., and Ferrazzoli, P., 2003. Retrieving near-surface soil moisture from microwave radiometric observations: current status and future plans. Remote Sensing of Environment 85, 489 – 506.

Williams, J., 2004. Introduction to uncalibrated data, calibration/validation. Report. Alaska Satellite Facility (ASF), Geophysical Institute, University of Alaska Fairbanks, AK, USA.

Wilson, D.J., Western, A.W., Grayson, R.B., Berg, A.A., Lear, M.S., Rodell, M., Famiglietti, J.S., Woods, R.A., and McMahon, T.A., 2003. Spatial distribution of soil moisture over 6 and 30 cm depth, Mahurangi river catchment, New Zealand. Journal of Hydrology 276, 254 – 274.

Yeh, G.T., Cheng, H.P., Cheng, J.R., Lin, H.C., and Martin, W.D., 1998. A Numerical Model Simulating Water Flow and Contaminant and Sediment Transport in WAterSHed Systems of 1-D Stream-River Network, 2-D Overland Regime, and 3-D Subsurface Media (WASH123D: Version 1.0). US Army Corps of Engineers, Waterways Experiment Station, Technical Report CHL-98-19 prepared for the US Environmental Protection Agency.

Zribi, M. and Dechambre, M., 2002. A new empirical model to retrieve soil moisture and roughness from C-band radar data. Remote Sensing of Environment 84, 42 – 52.

Zribi, M., Baghdadi, N., Holah, N., and Fafin, O., 2005. New methodology for soil surface moisture estimation and its application to ENVISAT-ASAR multi-incidence data inversion. Remote Sensing of Environment 96, 485 – 496.

CHAPTER 3
DATA MINING ANALYSIS FOR CHANGE DETECTION OF RIPARIAN BUFFER ZONE WITH MULTIPLE SATELLITE REMOTE SENSING IMAGERIES IN A SEMI-ARID WATERSHED

Introduction

The usefulness of wetland and riparian systems for ecosystems are great considerations for their prevention and rehabilitation (EPA, 2005). The ecosystems that provide many beneficial services to humans, animals, and natural processes are threatened with their losses and degradation. Riparian zones are the interfaces between waterways and uplands. Riparian buffers are the vegetated areas in the riparian zones that functions as buffers between the uplands and the waterways. They are widely recognized as an important part of the aquatic and floodplain ecosystems (EPA, 2005). The riparian buffers provide benefits such as erosion resistance, riverbank stability, wildlife habitats and corridors, and water quality protection (Osborne and Kovacic, 1993; Cooper et al., 1995; Palone and Todd, 1997; Abbey, 1998; Chapman and Ribic, 2002). The buffers can catch and filter out sediment and debris from surface runoff which can carry off nutrients with them (EPA, 2005). With the help of the riparian buffers, the nutrients can settle out and be absorbed by vegetation in the buffers. The buffer also stabilizes river bank by root systems which hold the soil together. Trees and shrubs reduce the velocity of runoff in steep slope landscape, thus increasing the amount of sediment settling on the land (Abbey, 1988). A vast diversity of wildlife and vegetation is also found in the riparian ecosystems. Riparian forests appear to support a particularly rich and abundant small mammal community (Chapman and Ribic, 2002). Continuous stripes of riparian forests also serve as wildlife travel corridors. Aquatic habitats can benefit from the riparian buffers characteristics such as the improvement

quality of water, provision of living and hiding habitats by root systems, and shading effect of trees along the river banks which help to maintain cooler water temperature in the summer. The cooler water holds more dissolved oxygen than the warmer water, providing more oxygen to aquatic life. The riparian buffer is categorized as one of structural BMPs to prevent and reduce the intrusion of nutrients and pollutant transport into water systems (EPA, 2005). Netusil (2006) evaluated the economic valuation of riparian corridors to signify the measurable values of riparian buffers.

Awareness of the great loss of wetland and riparian systems has been recognized. Laws such as the Clean Water Act were enacted to protect, restore, and recreate wetland and riparian systems (EPA, 2005). Planning and actions for protection, restoration, and creation of wetland and riparian systems require understanding of function, location, biophysical condition, geography and geomorphology of the riparian zones. Mapping riparian zones provides information required for the management of wetland and riparian systems. Monitoring for the changes of the wetland and riparian systems offers information to be used for evaluation, protection, and restoration (Lee and Marsh, 1995; Narumalani et al., 1997; Hattermann et al., 2006).

Remote Sensing (RS) technology has been implemented for Earth observation since the 1970's. Remote sensing for land cover classification has become more popular than ground survey mapping because of many advantages. The RS can observe the locations that are difficult or impossible to ground access such as the deep forest of Amazon or the middle of a large desert. The cost of utilizing RS is also reducing and the availability of the RS to public is increasing. New space-borne sensors continuously monitor the Earth from the atmospheric to the ground

surface and shallow water. RS provides constant monitoring spatially and temporally. The spatial resolution can be acquired ranging from meters to kilometers. Day- and night-time observations are feasible in all weather conditions. Multiple electromagnetic spectrals collected digitally can be manipulated for mixed-spectral analysis (e.g., vegetation indices), which provides physical properties that offer meaningful measurements. As a consequence, space-borne RS has been used to aid in monitoring, mapping, delineation, and analysis of riparian zones (Lee and Marsh (1995), Muller (1997), Narumalani et al. (1997), and Mertes (2002). Remote sensing and geographic information systems (GIS) together implement a leap in mapping riparian zones for computational hydrologic models, such as the integration of AnnAGNPS and the REMM (Yuan et al., 2004) or the model SWIM (Hattermann et al., 2006). The models are used to simulate the hydrological processes, nutrient transport, erosion, and vegetation dynamics at the river basin or watershed scales. The integration of riparian buffers' information and the hydrologic models allows stakeholders to analyze the impacts of wetlands and riparian buffers on water, nutrient, and particulate transports that are relevant for the implementation of BMPs in the context of watershed management planning.

Land cover is a mixture of vegetation, soil, shadow, or man-made objects. The spectral response of each object on the land is different depending on the reflectance of the object. Vegetation indices (VIs) quantify the vigor of vegetated land cover (Campbell, 2002). Spectral-mixed analysis, such as vegetation indices, has been developed to interpret the remote sensing data of land cover by a more meaningful way. Vegetation indices derived can be used to differentiate green vegetation from the red or brown vegetation or the brightness and color of the background soil. Once vegetation index can be calculated based on the remote sensing data, stakeholders would be able to use the vegetation index to analyze the biomass in riparian zones.

However, the numerous vegetation indices developed in the past three decades create a question about the index screening and selection process. Thirty five (35) vegetation indices have been reviewed by Bannari et al. (1995). Derek and Brunke (2001), however, evaluated ten (10) popular vegetation indices for their capacities to derive forest biophysical information. There is a renewed interest of using an appropriate vegetation index to improve land cover classification, particularly for the riparian buffers. A total of forty-one vegetation indices are summarized in Table 3-1.

Classification is the essential tool in mapping regions of interests. Data classification is the process of grouping the same type of data together. In the aspect of digital image data classification, the classification is to group similar pixels into classes that represent the information of similar interest. However, there are many classification methods available, however, which complicates such clustering analysis. It is a subjective matter as to which classification method may provide the best performance for a specific task.

This study includes a review of several common classification schemes, including Iterative Self-Organizing Data Analysis Techniques (ISODATA), K-Means Algorithm, Fuzzy C-Mean, Decision Tree, and Genetic Programming (GP) algorithm (Jensen, 1996; Campbell, 2002). Then, the study thoroughly summarizes all the vegetation indices developed over the past four decades. Eight vegetation indices were selected and tested for riparian classification using genetic programming. Secondly, one conventional classification approach (ISODATA combined with maximum likelihood) was selected and compared with a white box expert system (GP) to discern the GP application potential. The GP with comparative advantage was eventually used to perform the classification for the purpose of mapping the riparian buffers along the stream

corridor in a semi-arid watershed of South Texas. The hypothesis in this investigation is that soil moisture along riparian buffer is influential in shaping evolution of riparian buffers. Hence, the case study uniquely demonstrates an effort to incorporating both vegetation index and soil moisture estimates based on Landsat 5 TM and RADARSAT-1 imageries trying to improve the riparian classification. Three-dimensional graph analysis was employed to collectively summarize many possible combinations of scenario in terms of their comparative advantage. It strives to find the most suitable vegetation index in association with each specific classification scheme. Such approach may lead to the identification of the best classification approach and the importance of input data that are most suitable for temporal and spatial riparian classification and change detection.

Background

Descriptions of study area

The Choke Canyon Reservoir Watershed (CCRW) is comprised of several land use/land cover types. Farming and livestock husbandry have been the major land use pattern in the past few decades. The farmland is often graded and plowed, and irrigation may change the soil moisture in some seasons periodically. The livestock in south Texas is naturally fed on grass in open areas and ranches. Mixed land uses in this area introduce complexity in plant species. On the other hand, landscape in south Texas is intimately tied with the geological structure. CCRW encompasses 14,200 km^2 out of the 43,300 km^2 Nueces River Basin. Major landscapes are flat as the water flows due southeast toward the Gulf of Mexico where slopes range from 0-10%. Figure 3-1 shows the CCRW and its features. Streams in this watershed normally have the width in the range of 5 to 30 meters. Floodplains are not very wide in the upper area of the watershed, but

very large floodplains can be found in the center and lower areas of the CCRW. Agricultural activities, such as cattle farming and crops, are found adjacent to the streams along the river banks. Figure 3-2 shows an ephemeral stream surrounded by crops with a 1-m-resolution aerial photograph of the Site α (see Figure 3-1). The ground truth photographs of riparian Site β and Site δ taken in April 2006 are showed in Figure 3-3 and Figure 3-4. The stream at Site β is very shallow and wide, while the stream at Site δ is deep and narrow. Agricultural activities in the areas are the major causes of the non-point source pollution of this watershed. A data mining analysis for change detection of riparian buffers is needed to assist in the surface water pollution prevention and management.
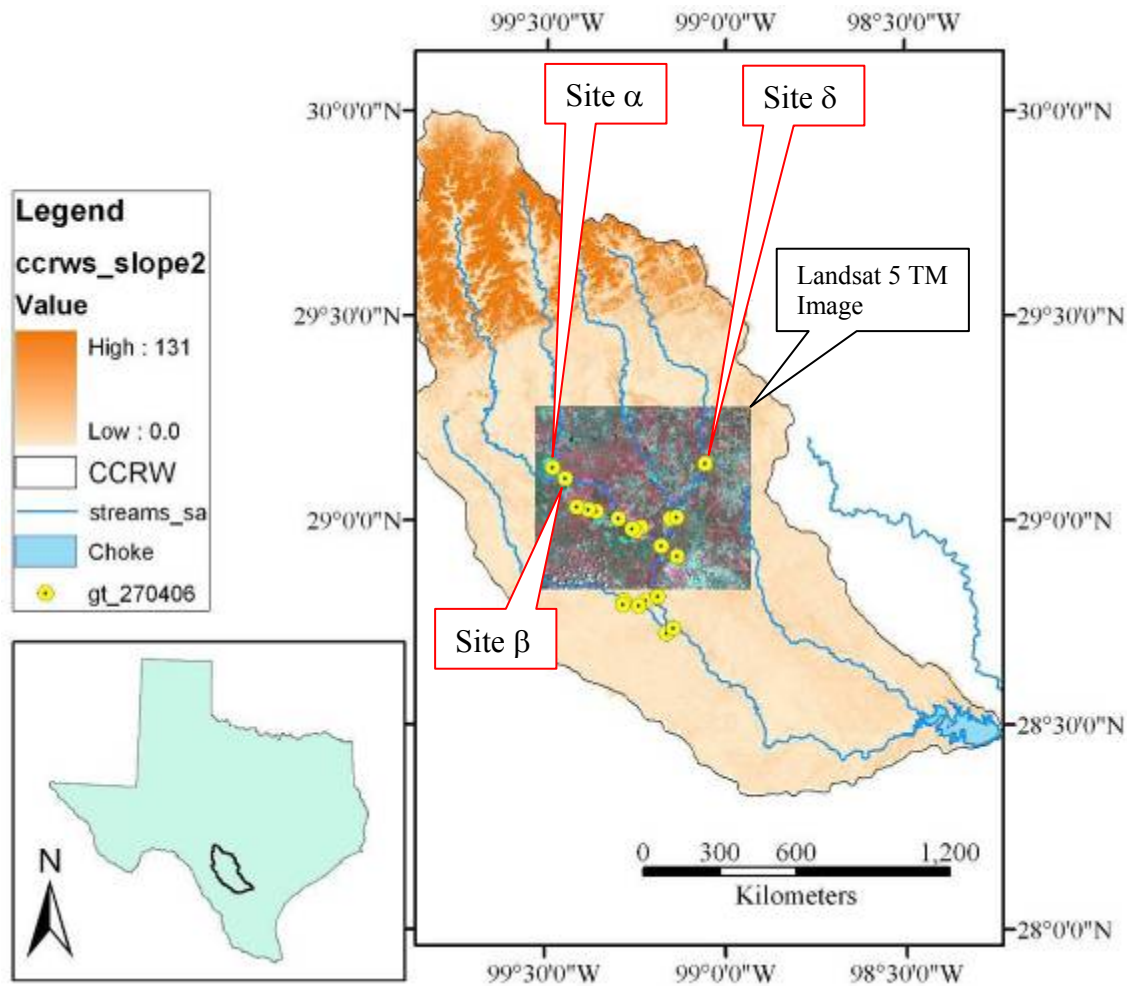
Figure 3-1. The location of the CCRW in south Texas.
CCRW is located at the central-south Texas (bottom left corner). The Choke Canyon Reservoir is located at the southeastern of the watershed. Landsat 5 TM imagery used in this study focuses at the center of the watershed. Ground information was investigated on locations as indicated by the yellow dots on the map. Twenty-two sampling sites, assigned in yellow dots, were investigated for the existence of riparian buffers.
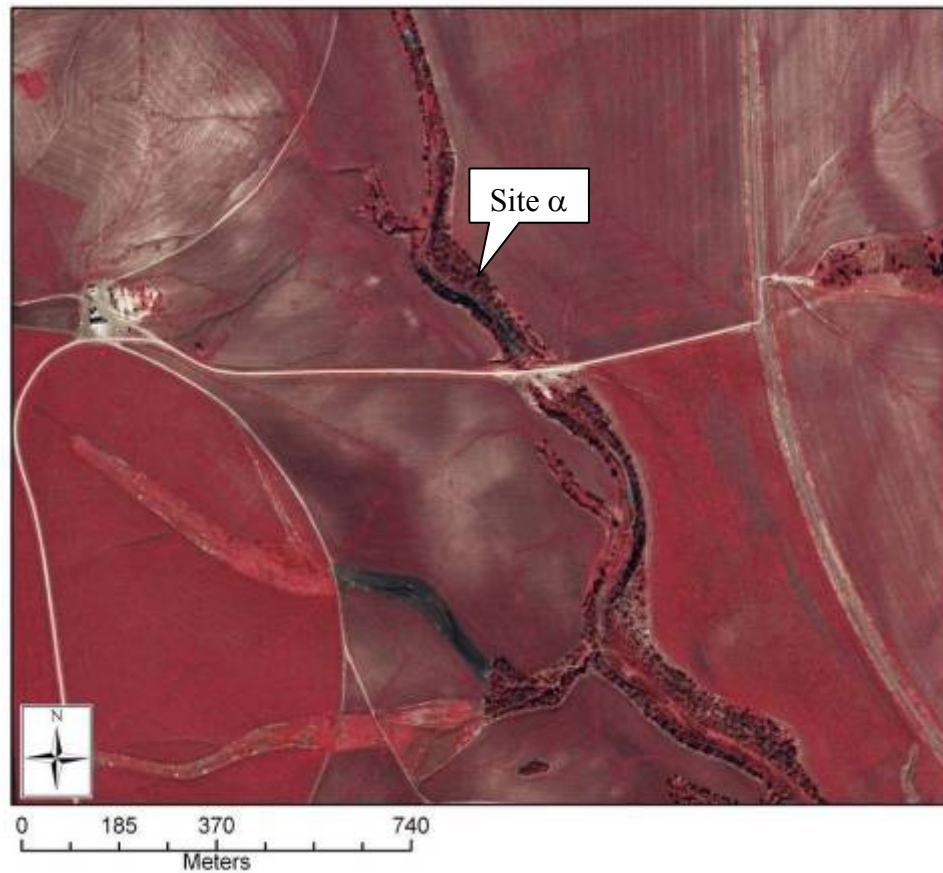
Figure 3-2. An aerial photograph of the site α.
A very high spatial resolution (1-m resolution) aerial photograph reveals the details of the riparian buffer at Site α. The buffers are very narrow along both sides of the stream. The stream is surrounded by agricultural activities, which contribute pollution to the stream via runoffs. (Source: Texas Commission on Environmental Quality (TCEQ) at http://www.tceq.state.tx.us/gis/doqviewer.html).

Figure 3-3. Ground image of site α.
The image was taken at the Site α marked in Figure 3-2. The riparian buffer is found along a small creek adjacent to an agricultural area. This type of riparian area is commonly found throughout the watershed.

Figure 3-4. Ground images of site βand site δ.
Very steep banks can be found throughout the stream systems in the CCRW. Excessive erosions are normal. Water level could rise up from the bottom-dried stage to many feet high.

Field data collection

In 2004, two campaigns for ground data collections of soil moisture measurements were carried on in April and September as the efforts of data synthesis of RADRASAT-1 Synthetic

Aperture Radar (SAR) imagery, slope, aspect, soil permeability, and land cover. Seven (7) field sampling sites were investigated with a collection of four hundreds and thirty four (434) soil moisture sampling points in April and sixty three (63) sampling points in September. The seven (7) sampling sites covered a broad range of different land covers including row crops, shrubland, pasture/hay, grasslands/herbaceous, deciduous forest, and riparian zone. The distance between any two ground truthing points were selected at least 13 meters apart to ensure that there is only one ground truth measurement point that is associated with one pixel of SAR imagery. Each measurement point was also selected at least 50 meters away from any road or building to avoid the interferences from any reflected radar signal. We navigated to the sampling points with a handheld Global Positioning System (GPS) unit that is capable of reading location with sub-meter accuracy (www.trimble.com/geoxt.shtml). To reduce the uncertainty, each sampling point comprised three measurements within a radius of two (2) meters, and then we took the average value of the three (3) readings. To minimize the uncertainty even further, most of the ground truth measurements were particularly selected in the proximity of the ground control points (i.e., corner reflector) in order to minimize the horizontal error relative to the SAR geometrically corrected. The outputs of soil moisture content based on GP technique were intensively compared against the extensive ground truth data in the same region. The GP-based soil moisture model was derived in the form of high complexity compound functions for generating wet and dry seasons of soil moisture maps. The GP-derived soil moisture model was proved useful to accommodate the soil moisture estimates of the CCRW with r-square 0.72 against unseen dataset for verification purpose.

Twenty-two (22) sampling sites were visited in April 2006 to investigate the existence of riparian buffers. It is assumed that the conditions of riparian vegetation in 2004 are not

significantly different than the riparian vegetation of the same area in 2006 within the same season. USGS (2007) reports air and water temperatures revealing that the climate pattern in the year 2004 and 2006 were very similar (see Figure 3-5). It provides a clue that the evolution pattern of riparian forests within 2004 and 2006 were not significantly different. Therefore, the ground investigation carried out in 2006 be supportive for the data mining analysis based on the remote sensing imageries acquired in 2004. The yellow dots in Figure 3-1 represent the investigated sites of targeted riparian forests. The sampling sites were located throughout the river systems in the central CCRW area where a footprint of Landsat 5 TM scene covers. A handheld Global Positioning System (GPS) unit with 3-m accuracy was used in the navigation and ground truth mapping.
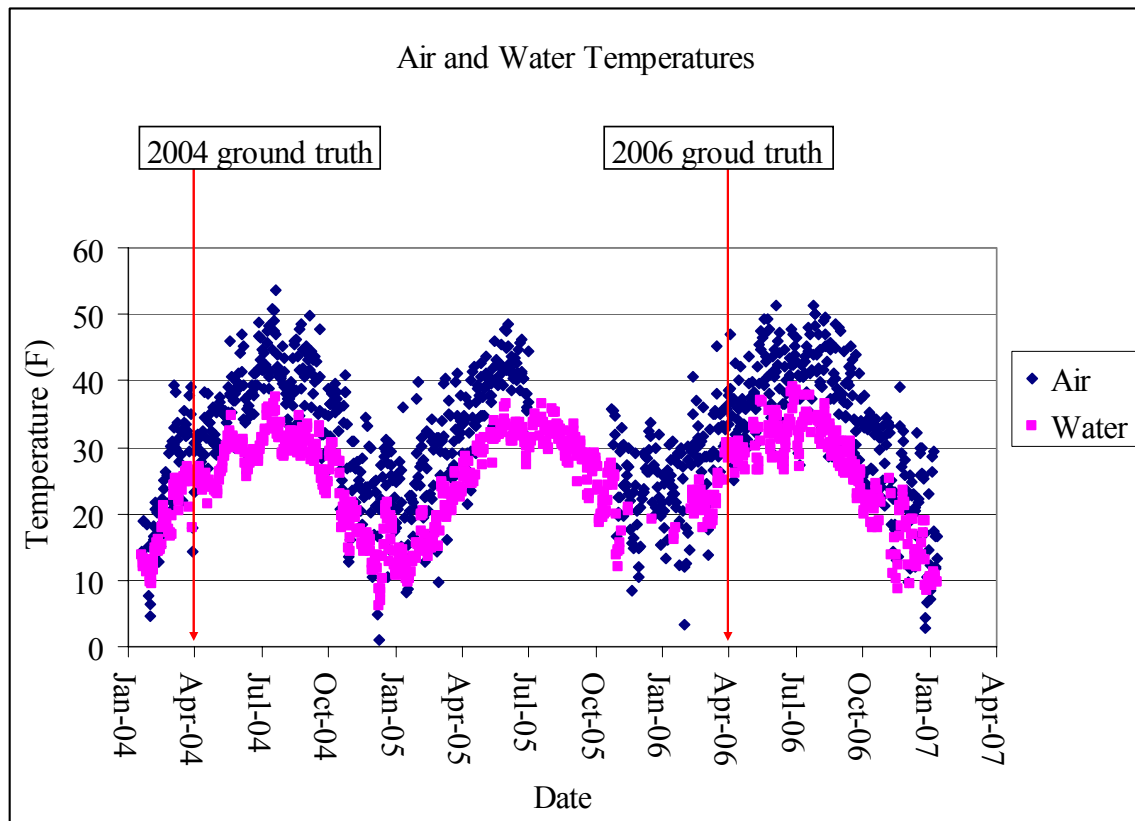
Figure 3-5. Air and water temperatures at the USGS gage 08067118 acquired between 2004 and 2006. Air and water temperature data reveal that the climate in the year 2004 and 2006 were very similar. These data support the assumption that the riparian forests of the two years were not significantly different. Therefore, the ground investigation carried on in 2006 could be acceptable for the analysis of the remote sensing imageries acquired in 2004. Data acquired from USGS gage station 08067118 from http://waterdata.usgs.gov

Riparian buffers in the study area vary in size and composition from location to location. Vegetation found in the buffers is mostly composed of trees, shrubs, and grass. Steep slopes are normally found along the river banks. Various species of plants compose a riparian buffer. Tree species found in most riparian area in the study area are Texas Ash (*Fraxinus Texensis*), Catclaw Acacia (*Acacia Greggii Gray*), Pecan (*Carya Illinoiensis*), and Indigobush (*Amorpha Fruticosa*) (see Figure 3-6). Patches of herbaceous plants and grass are found on most riverbanks. Indigobush are normally found on relatively dry riverbanks where the diurnal effects of wet and dry seasons are present. They are commonly found along ephemeral streams and dried land (see

Figure 3-6 b and c).   Texas Ash or Pecan trees are generally found on riverbanks where the streams have water all year round. They are commonly found at perennial streams (see Figure 3-6 a and d). The heights of plants are between 50 to 70 feet for Texas Ash and Pecan and between 15 to 30 feet for Indigobush and Catclaw Acacia. The Texas Ash and Pecan generally have footprints (brunch span) between 30 to 50 feet. The Indigobush and Catclaw Acacia have footprints between 10 to 30 feet. The size of riparian buffers varies from less than 30 feet to more than 300 feet wide, while in some areas a riparian buffer does not exist at all. The density of plants in a riparian varies from a row of trees to many rows of trees. Agricultural activities and ranching activities are common practices in the study area. They are always found adjacent to the riparian buffers (see Figure 3-3). In most area the slope of the stream banks are very steep. Erosion can be found wherever bare soil is present along the riverbanks. Many riverbanks are geographically canyon-like. However, there are many natural ramps that allow animals to access the streams (see Figure 3-4). Animal footprints found at many accesses cause damages to herbaceous plants and grass on riverbanks.
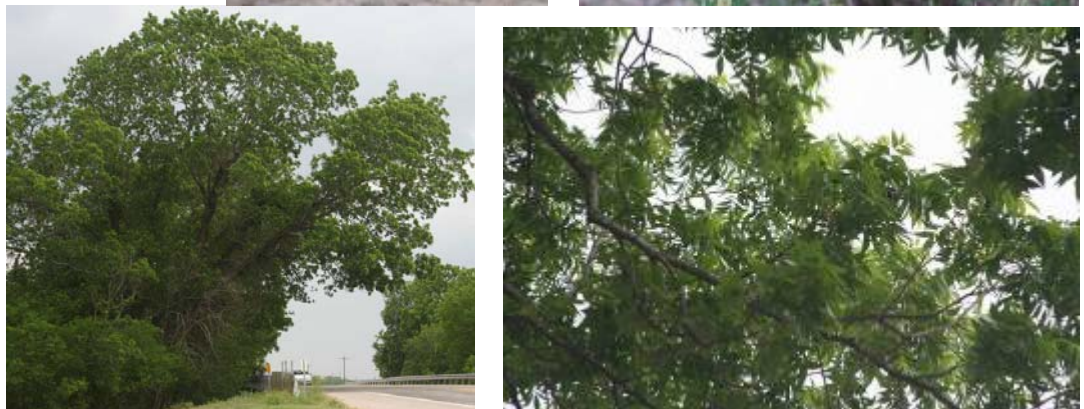
Figure 3-6. The common plant species in the watershed
Texas ash (A), Indigobush (B), Catclaw Acacia (C), and Pecan (D). The Texas ash and the Pecan are usually found along the parenial rivers in the watershed, while the Indigobush and Catclaw Acacia are normally found along the ephemeral rivers.

Remote sensing

At the beginning of the 21st century there are many remote sensing instruments available for public. A non-comprehensive list of seventeen (17) RS instruments is concluded by Mertes (2002). The remote sensing instruments have been utilized to observe the earth in many aspects from the atmosphere down to the ocean including 24 measurements (King, 2004). Many new instruments are now recording data across the electromagnetic spectrum. The advantages that remote sensing may offer include the repeatability of observation temporally and spatially, the large-scaled coverage, the ability to observe very difficult terrains, and the ability to process the data digitally. These advantages make the remote sensing very cost-effective in many applications, especially when a large area must be observed. The planet's landforms, vegetation, and resources can be revealed on regional, continental, and even global scales by the use of aerial and satellite images. The change detection of riparian zones in the CCRW can be made possible through using the advantages of remote sensing technology at the core.

Many satellites providing various types of remote sensors are currently observing the earth. They include Advanced Land Observing Satellite (ALOS), Advanced Earth Observing Satellite-II (ADEOS-II), Earth Observing System PM1 (Aqua), Tropical Rainfall Measuring Mission (TRMM), Land Satellite (LandSat), Satellite Probatoire d' Observation de la Terre (SPOT), European Remote Sensing Satellite (ERS), RADASAT-1 and -2 (JAXA, 2003). Transient phenomena, such as temporal changes of landscapes or seasonal vegetation vigor can be observed by comparing remotely sensed images acquired at different times (USGS, 2006). Because of the advantages of the remote sensing technology, such as the large scaled observation, repeated availability, and all-weather capability (i.e., in a certain type of sensors),

many researchers employed aerial and satellite images for the study of riparian ecosystems (Narumalani et al, 1997; Goetz et al., 2003).

Regarding to be the most popular space-borne sensor used for Earth's surface observations for many decades is the Landsat satellites. The Landsat program was established in the U.S. to photograph Earth's land surface in 1960s (NASA, 2006a). The program has a series of Earth-observing satellites since Landsat 1 launched in 1972 until Landsat 7 operating at the present. The Landsate satellites have taken multispectral digital photographs of Earth's surface including the lands and surrounding coastal regions for more than 30 years. The continuous observations enable people to study the dynamic changes of the Earth due to natural causes and human actions. We chose the Landsat images for the derivation of vegetation indices in this study due to the many advantages including the availability, the consistency, the accuracy, spatial and temporal resolutions, and the ability of manipulate of the data. Landsat data are widely used in many landcover studies, i.e. (Muller, 1997; Narumalani et al., 1997; Scott et al., 2000; Nagler et al., 2001; Apan et al., 2002; Congalton et al., 2002; Zhao et al., 2003; Gutiérrez et al., 2004; Miltner et al., 2004; Shupe and Marsh, 2004; Ingram et al., 2005; Muukkonen and Heiskanen, 2005; Sivanpillai et al., 2005)

Vegetation indices

Vegetation indices have been developed to qualitatively and quantitively assess vegetation covers using spectral measurements (Bannari et al., 1995). The first earth resources satellite, a.k.a. Landsat-1, launched in 1972 was a remarkable effort of mankind to use electromagnetic spectral response for evaluating vegetation covers. The uses of the red and near-infrared spectral bands of the sensors on board satellites are very well suited for assessing

vegetation covers (Weier and Herring, 2006). The green vegetation strongly absorbs red light (Landsat band 3) by the photosynthetic pigments, such as chlorophyll a. In contrast, the near-infrared wavelengths are half reflected by and half passing through the leaf tissues regardless of their color (USDA-ARS, 2006).

Thirty five (35) vegetation indices were reviewed by Bannari et al. in 1995. Most of the vegetation indices use the red and the near-infrared bands, while some other vegetation indices incorporate additional parameters to compensate for atmospheric and/or background corrections. In 2001, Peddle and Brunke did another review on ten (10) most commonly used vegetation indices in the forestry applications. Each vegetation index has strength and weakness though. For instance, the Ratio Vegetation Index (RVI) does not perform well when the vegetation cover is less than 50%, but it is the best at the dense vegetation cover (Jackson et al., 1983). In contrast, the Normalized Difference Vegetation Index (NDVI) was first proposed by Rouse (1973) to reduce the effect of sensor degradation by normalizing the spectral bands. It is generally agreed that the NDVI is sensitive to low dense vegetation such as semi-arid areas (Tucker and Miller, 1977; Kerr et al., 1989; Nicholson et al., 1990). Many other vegetation indices were developed to correct for effects of confounding factors, such as background effects (soil brightness and soil color), atmospheric effects (absorption and scattering), and the effects of the sensors' responses and calibration. Soil Adjusted Vegetation Index (SAVI) is an example of a vegetation index developed by Huete (1988) to describe the effects of soil brightness in the background. Examples of vegetation indices that are insensitive to atmospheric effects include Global Environment Monitoring Index (GEMI) (Pinty and Verstraete, 1992) and Enhanced Vegetation Index (EVI) (Huete et al., 1999). The EVI is regularly produced from the MODIS program (Huete et al., 1999). It was developed to improve the sensitivity in high biomass regions while reducing the

71

atmospheric effects. Although vegetation indices are normally classified into two categories by Lautenschlager and Perry (1981), Huete (1989), Baret and Guyot (1991), and Bannari et al. (1995), this chapter classified the vegetation indices into three groups (see Table 3-1).

Selecting the right vegetation index might greatly affect the accuracy of change detection of riparian forests. Eight (8) vegetation indices across three different categories, including SR, NDVI, SAVI, SAVI-2, RDVI, GEMI, MSR, and NLI. NDVI, were deemed promising for this watershed study because they are more suitable than other vegetation indices for sparse vegetated areas in this semi-arid environment. SAVI and SAVI-2 are particularly appealing because they are corrected by the disturbance of soil background, which significantly influences the sensors in the sparse vegetated areas, such as CCRW. The atmospheric-corrected vegetation indices are also included in the screening process even though the atmospheric correction of the Landsat data has been also performed prior to calculating of vegetation indices. An extensive work in selecting the most suitable vegetation index would be necessary to search for a vegetation index that offers the most accurate result.

Table 3-1. Summary of most vegetation indices from 1972 to 1996, adopted from Bannari et al. (1995) and Peddle and Brunke (2001).

Index	Formula	Empirically derived	Mathematically and physically derived	Soil Adjusted	Atmospheric Adjusted	Author and Year
Ratio Vegetation Index (RVI)	$\dfrac{NIR}{R}$	✓				Pearson and Miller, 1972
Vegetation Index Number (VIN)	$\dfrac{R}{NIR}$	✓				Pearson and Miller, 1972
Transformed Vegetation Index (TVI)	$\sqrt{(NDVI + 0.5)}$	✓				Rouse et al., 1974
Green Vegetation Index (GVI)	(-0.283MSS4 – 0.660MSS5 + 0.577MSS6 +0.388MSS7)	✓		✓		Kauth and Thomas, 1976
Soil Brightness Index (SBI)	(0.332MSS4 + 0.603MSS5 + 0.675MSS6 +0.262MSS7)	✓		✓		Kauth and Thomas, 1976
Yellow Vegetation Index (YVI)	(-0.899MSS4 + 0.428MSS5 + 0.076MSS6 - 0.041MSS7)	✓			✓	Kauth and Thomas, 1976
Non Such Index (NSI)	(-0.016MSS4 + 0.131MSS5 - 0.425MSS6 +0.882MSS7)	✓			✓	Kauth and Thomas, 1976
Soil Background Line (SBL)	MSS7 – 2.4MSS5	✓		✓		Richardson and Wiegand, 1977
Differenced Vegetation Index (DVI)	2.4MSS7 – MSS5	✓				Richardson and Wiegand, 1977
Misra Soil Brightness Index (MSBI)	(0.406MSS4 + 0.600MSS5 + 0.645MSS6 + 0.243MSS7)	✓		✓		Misra et al., 1977
Misra Green Vegetation Index (MGVI)	(-0.386MSS4 - 0.530MSS5 + 0.535MSS6 + 0.532MSS7)	✓		✓		Misra et al., 1977

Index	Formula	Empirically derived	Mathematically and physically derived	Soil Adjusted	Atmospheric Adjusted	Author and Year
Misra Yellow Vegetation Index (MYVI)	(0.723MSS4 - 0.597MSS5 + 0.206MSS6 - 0.278MSS7)	✓			✓	Misra et al., 1977
Misra Non Such Index (MNSI)	(0.404MSS4 - 0.039MSS5 - 0.505MSS6 + 0.762MSS7)	✓			✓	Misra et al., 1977
Perpendicular Vegetation Index 1 (PVI-1)	$\sqrt{\left(\rho_{soil} - \rho_{veg}\right)^2_R + \left(\rho_{soil} - \rho_{veg}\right)^2_{NIR}}$	✓		✓		Richardson and Wiegand, 1977
Ashburn Vegetation Index (AVI)	(2.0MSS7 – MSS5)	✓				Ashburn, 1978
Greenness Above Bare Soil (GRABS)	(GVI – 0.09178SBI + 5.58959)	✓		✓		Hay et al., 1979
Multi-Temporal Vegetation Index (MTVI)	NDVI(date 2) – NDVI(date 1)	✓				Yazdani et al., 1981
Greenness Vegetation and Soil Brightness (GVSB)	$\dfrac{GVI}{SBI}$	✓		✓		Badhwar, 1981
Adjusted Soil Brightness Index (ASBI)	(2.0 YVI)	✓		✓	✓	Jackson et al., 1983
Adjusted Green Vegetation Index (AGVI)	$GVI - (1 + 0.018GVI)YVI - \dfrac{NSI}{2}$	✓		✓	✓	Jackson et al., 1983
Transformed Vegetation Index (TVI)	$\dfrac{\left(NDVI + 0.5\right)}{\left\|NDVI + 0.5\right\|}\sqrt{\left\|NDVI + 0.5\right\|}$	✓				Perry and Lautenschlager, 1984
Differenced Vegetation Index (DVI)	NIR – R; near-infrared(NIR) and visible red(R)	✓				Clevers, 1986

Index	Formula	Empirically derived	Mathematically and physically derived	Soil Adjusted	Atmospheric Adjusted	Author and Year
Normalized Difference Greenness Index (NDGI)	$\dfrac{G-R}{G+R}$; visible green(G)	✓				Chamard et al., 1991
Redness Index (RI)	$\dfrac{R-G}{R+G}$	✓		✓		Escadafal and Huete, 1991
Normalized Difference Index (NDI)	$\dfrac{NIR-MIR}{NIR+MIR}$; mid-infrared(MIR)	✓				McNairn and Protz, 1993
Normalized Difference Vegetation Index (NDVI)	$\dfrac{NIR-R}{NIR+R}$		✓			Rouse et al., 1974
Perpendicular Vegetation Index 2 (PVI-2)	$\dfrac{NIR-aR-b}{\sqrt{a^2+1}}$ $NIR=aR+b$		✓	✓		Jackson et al., 1980
Soil Adjusted Vegetation Index (SAVI)	$\dfrac{NIR-R}{NIR+R+L}\cdot(1+L)$ $0 < L < 1$ with typical value $L = 0.5$		✓	✓		Huete, 1988
Transformed SAVI (TSAVI)	$\dfrac{\left[a(NIR-aR-b)\right]}{(R+aNIR-ab)}$		✓	✓		Baret et al., 1989
Weighted Difference Vegetation Index (WDVI)	$NIR-\dfrac{NIR_{soil}}{R_{soil}}\cdot R$		✓	✓		Clevers, 1989
Transformed SAVI (TSAVI)	$\dfrac{\left[a(NIR-aR-b)\right]}{\left[R+aNIR-ab+X(1+a^2)\right]}$		✓	✓		Baret and Guyot, 1991

Index	Formula	Empirically derived	Mathematically and physically derived	Soil Adjusted	Atmospheric Adjusted	Author and Year
Atmospherically Resistant Vegetation Index (ARVI)	$\dfrac{NIR - RB}{NIR + RB}; RB = R - \gamma(B - R)$ $\gamma = \dfrac{\rho_{red}}{\rho_{blue} - \rho_{red}}$		✓		✓	Kaufman and Tanré, 1992
Global Environment Monitoring Index (GEMI)	$\dfrac{\eta(1 - 0.25\eta) - (\rho_r - 0.125)}{1 - \rho_r}$ when $\eta = \dfrac{2(NIR^2 - R^2) + 1.5NIR + 0.5R}{NIR + R + 0.5}$		✓		✓	Pinty an Verstraete, 1992
Transformed Soil Atmospherically Resistant Vegetation Index (TSARVI)	$\dfrac{[a_{rb}(NIR - a_{rb}RB - b_{rb})]}{[RB + a_{rb}NIR - a_{rb}b_{rb} + X(1 + a_{rb}^2)]}$ When $NIR = a_{rb}RB + b_{rb}$		✓	✓	✓	Bannari et al., 1994
SAVI-1	$\dfrac{NIR - R}{NIR + R + L} \cdot (1 + L)$ where $L = 1 - 2.12 \cdot NDVI \cdot WDVI$		✓	✓		Qi et al., 1994
SAVI-2	$\dfrac{2NIR + 1 - \sqrt{(2NIR + 1)^2 - 8(NIR - R)}}{2}$		✓	✓		Qi et al., 1994
Non-Linear Index (NLI)	$\dfrac{NIR^2 - R}{NIR^2 + R}$		✓			Goel and Qin, 1994
Angular Vegetation Index (AVI)	$\tan^{-1}\left\{\dfrac{\lambda_3 - \lambda_2}{\lambda_2}[NIR - R]^{-1}\right\} + \tan^{-1}\left\{\dfrac{\lambda_2 - \lambda_1}{\lambda_2}[G - R]^{-1}\right\}$		✓		✓	Plummer et al., 1994

Index	Formula	Empirically derived	Mathematically and physically derived	Soil Adjusted	Atmospheric Adjusted	Author and Year
Re-normalized Difference Vegetation Index (RDVI)	$\dfrac{NIR - R}{\sqrt{NIR + R}}$		✓			Roujean and Breon, 1995
Modified Simple Ratio (MSR)	$\dfrac{\dfrac{NIR}{R} - 1}{\sqrt{\dfrac{NIR}{R}} + 1}$		✓			Chen, 1996
Enhanced Vegetation Index (EVI)	$EVI = 2 \cdot \dfrac{(NIR - R)}{(L + NIR + C_1 R + C_2 B)}$; B = blue band		✓	✓	✓	Huete and Liu, 1996

where

      NIR = near infrared
      MIR = mis infrared
      R = electromagnetic wavelength that correspond to visible red
      MSS4 = Landsat MSS band 4
      MSS5 = Landsat MSS band 5
      MSS6 = Landsat MSS band 6
      MSS7 = Landsat MSS band 7

Classification method

Remotely sensed data can be converted to useful information by classification. Many techniques are available for classification and interpretation of the remotely sensed data. Multi-dimensional data, such as the combination of Landsat TM, vegetation indices, and soil moisture, can be processed for riparian buffers change detection using a variety of algorithms. They include hard classification (supervised and unsupervised approaches), fuzzy classification (used in conjunction with supervised and unsupervised algorithms), and hybrid classification approaches (involving the used of ancillary data). An integration of supervised and hybrid classification approach is used in this study. The ground data are divided into two sets for training and verification, respectively. The training dataset is used to perform the supervised classification, while the verification dataset is used for accuracy assessment. The ancillary data, including vegetation indices and soil moisture, may be incorporated into the Landsat TM dataset using a per-pixel logical channel classification (Jensen, 1996).

Tools being used in expert systems, such as decision tree, artificial neural network (ANN), and genetic algorithm (GA), are broadly employed for classification by practitioners and scientists in this field. A number of studies used ANNs to classify remotely sensed satellite data (Shupe and Marsh, 2004; Ingram et al., 2005; Muukkonen and Heiskanen, 2005; Sangle and George, 2005; Filippi and Jensen, 2006). Hepner et al. (1990) suggested that the ANN techniques have a potential advantage of minimizing training sets, and thus the less ground truth data are required. However, Shupe and Marsh (2004) showed in their work that maximum likelihood method gave higher accuracy for classifying vegetation than ANN method did, based on the Landsat TM and SAR data.   Filippi and Jensen (2006) extensively used supervised,

unsupervised, and fuzzy learning vector quantization (FLVQ) ANNs to classify coastal wetland environment based on the hyperspectral AVIRIS image. The FLVQ was found dramatically more computational efficient than the supervised and unsupervised ANNs that were trained with back propagation multilayer perceptron (MLP). The FLVQ required only 3.6% of the computational time used by the MLP ANN with 400 neurons. On the other hand, Sangle and George (2005) showed in their work that fuzzy neural networks did not give good accuracies when compared to other types of ANNs. In addition, pure back propagation neural networks were found to be weak when dealing with multiple inputs from a large pool of variables (Ritchie et al., 2003). The GP is a self-adaptive, evolutionary computational method. It deploys a heuristic search through all dimensions for all possible programs. The GP can reveal detailed information with regards to constraints and structures of the outcomes. The advantage of the GP implements the identification of effectiveness of datasets and constraints. The GP algorithm can optimize inputs using the evolutionary features of GP to evolve the architecture of ANNs (Koza and Rice, 1991). Genetic Programming Neural Network (GPNN) is an optimized neural network architecture that was first proposed by Koza and Rice (1991). It was proved more powerful than stepwise logistic regression (SLR) or classification and regression trees (CART) in the work of Ritchie et al. (2006). The drawback of GPNN is the expenses of computational time and complexity. A larger pool of inputs and evolutionary time are required to improve the results in GPNN.

There are many dimensions of inputs used for the classification. The algorithm that is most suitable for the task must be able to perform data mining analysis in order to sort out and select datasets that mostly impact the outcome. The ANN and the GA hold the advantage over the cluster analysis and decision tree because they are able to perform data mining analysis. The

79

ANN is regarded as a black box due to the inherent complexity for analyzing the weights and bias terms. In contrast, the GP is somewhat a white box expert system allowing for evaluation of its outcomes. Hence, GP is chosen in this study to perform the classification analysis. Yet a combination of ISODATA (unsupervised approach) and maximum likelihood (supervised classification) was sequentially performed and used to compare to the GP outputs.

Methodology

Riparian classification algorithm (RICAL)

The RICAL algorithm is proposed in this chapter. The algorithm includes a per-pixel logical channel classification using multi-spectral imagery (Landsat 5 TM) incorporating with ancillary datasets including vegetation index and soil moisture. The ancillary datasets are used to decrease the chance of misclassification (Jensen, 1996). The Landsat 5 TM data include band 1, 2, 3, and 4 which corresponds to visible blue, visible green, visible red, and near infrared. Each band is post-processed to convert the digital number (DN) to reflectance according to the revised algorithm in the work of Chander and Markham (2003). Vegetation indices are calculated from the reflectance and added into the Landsat image as band 5. Soil moisture map is derived from RADARSAT-1 SAR imagery and ancillary data including STATSGO soil permeability data, DEM (slope and aspect), and National Land Cover Data (NLCD) (Makkeasorn et al., 2006). The soil moisture estimation model is created at the watershed scale in a companion study using genetic programming approach, which can be used directly in this study. The soil moisture data is incorporated into the Landsat image as band 6. Consequently, the final images have a total of 6 layers including Landsat band 1, 2, 3, and 4, a vegetation index as layer 5, and soil moisture data as layer 6 for change detection analysis.

The riparian forests were investigated in April 2004. Geographic locations of the investigated riparian forests were recorded with a handheld GPS unit with 3-meter accuracy. The locations of riparian forests and non-riparian forests are imported into GIS framework. The collected information was then imported into GIS framework. The ground data of riparian forests and non-riparian areas are superimposed on the final map layer. Pixel values of all the layers of the image are linked to the data points. Each data point contain multi-dimensional data including Landsat band 1, 2, 3, and 4, a vegetation index, a soil moisture content, and a ground data ('1' as riparian forest, '0' as not riparian area). The data points were then fed into GP. The GP do the evolutionary computation to classify a pixel for riparian forest identification. To signify the philosophy a schematic of RICAL algorithm is presented in Figure 3-7.
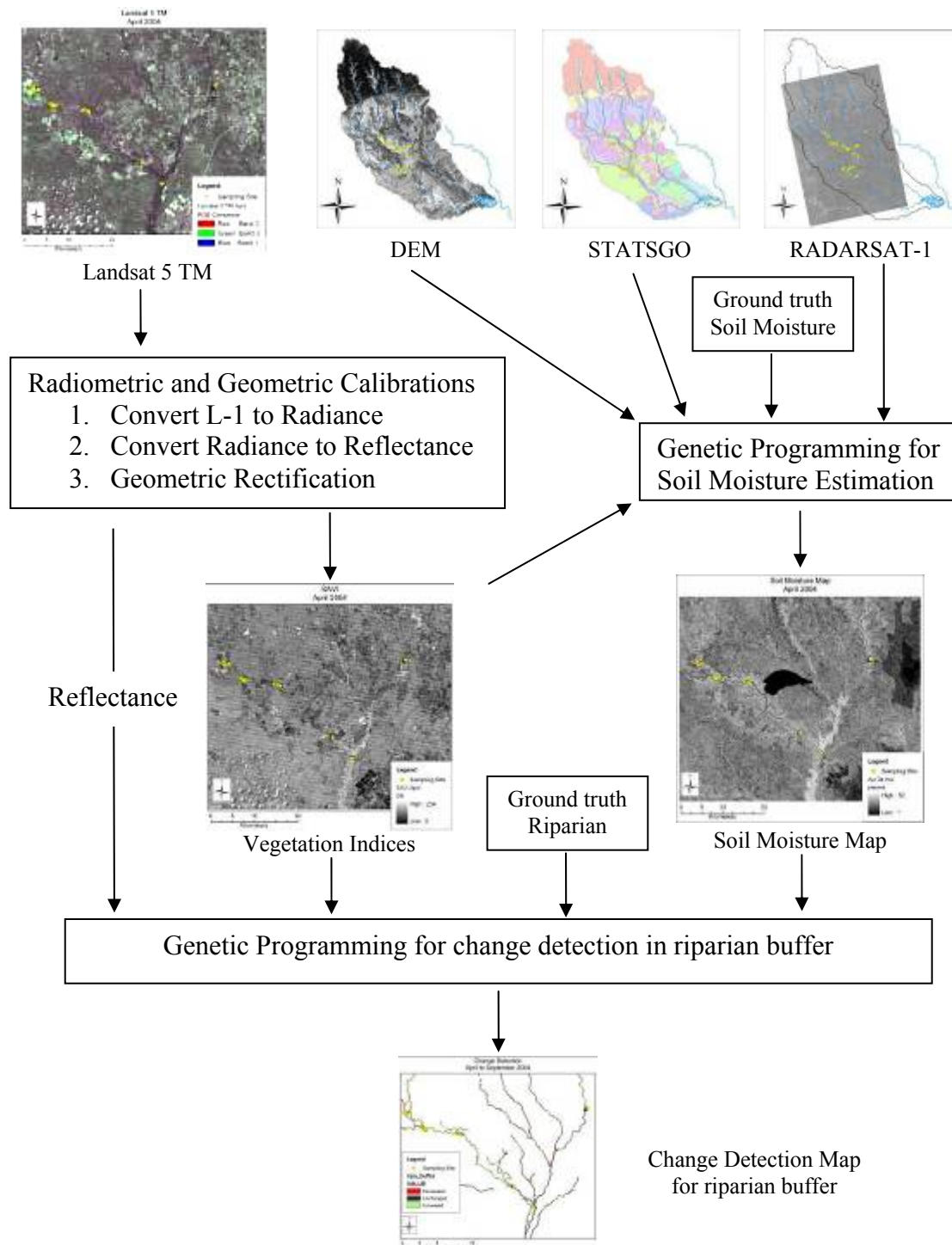
Figure 3-7. Schematic of the riparian change detection.
Preparations of input data for GP classification begin with a Landsat 5 TM image. Radiometric calibrations ensure the consistency between scenes. Surface's reflectance is calculated for the derivation of vegetation indices. Data of each layer including band 1, 2, 3, 4, and the vegetation indices are extracted for the preparation of GP's input datasets.

Preparation of vegetation indices

Two Landsat 5 TM images from April 7, 2004 and September 30, 2004 are used in the analysis. The Landsat 5 TM image acquired on April 7, 2004 was geometrical rectified to 1-m resolution images using obvious landmarks such as intersections and local runways. The second-order polynomial method was used to transform the Landsat 5 TM image with the RMS error of 11.3 meters, essentially smaller than a pixel of Landsat 5 TM image. The September 2004 Landsat 5 TM image was rectified based on the April Landsat image using the image-to-image method. The RMS error was obtained at 6.8 meters. The nearest-neighbor resample was used in the rectification process. Radiometric calibrations are applied to the normalized data from multiple scenes, sensors, and platforms into a common scale. The digital number (DN) data are converted to at-sensor spectral radiance (L_λ) using Eq.(3-1).

(3-1). A reduction of the variability between data acquired at different time is achieved through a conversion from radiance (L_λ) to reflectance (ρ) using Eq.(3-2). Such conversion may normalize for different solar irradiances at different time of acquisitions.

The sensors on board Landsat satellites measure radiance (the amount of light coming from a surface passing through the atmosphere). The calibrated Landsat data can be converted to radiance values using the following equation from Chander and Markham (2003):

where
$$L_\lambda = \left(\frac{LMAX_\lambda - LMIN_\lambda}{Q_{cal\max}} \right) \cdot Q_{cal} + LMIN_\lambda \qquad (3\text{-}1)$$

L_λ pectral radiance at the sensor's aperture in $\frac{W}{\left(m^2 \cdot sr \cdot \mu m\right)}$;

Q_{cal} quantized calibrated pixel value in DNs;

$Q_{cal\,min}$ minimum quantized calibrated pixel value (DN = 0) corresponding to

$LMIN_\lambda$;

$Q_{cal\,max}$ maximum quantized calibrated pixel value (DN = 255) corresponding to

$LMAX_\lambda$;

$LMIN_\lambda$ spectral radiance that is scaled to $Q_{cal\,min}$ in $\dfrac{W}{\left(m^2 \cdot sr \cdot \mu m\right)}$;

$LMAX_\lambda$ spectral radiance that is scaled to $Q_{cal\,max}$ in $\dfrac{W}{\left(m^2 \cdot sr \cdot \mu m\right)}$;

The post-calibration values for the above equation can be obtained in the header file of the Landsat data (Chander and Markham, 2003), see also Appendix D.

The surface's reflectance is a characteristic of the surface that is independent from the intensity of incoming light. The reflectance can be calculated as the radiance divided by the incoming light, a.k.a. irradiance. The reflectance is practically the most valuable value that is very difficult to obtain. USGS-EROS is responsible for the radiometric calibration of Landsat satellites, and they have revised algorithms to convert the Landsat data to top-of-atmosphere reflectance. In order to compare the two Landsat scenes that were obtained at different time, the reduction of scenes' variability must be minimized. By converting the spectral radiance to reflectance, the differences of solar irradiance due to the time difference between data acquisitions can be removed. The combined surface and atmospheric reflectance of the earth is calculated as follows (Chander and Markham, 2003).

$$\rho_P = \frac{\Pi \cdot L_\lambda \cdot d^2}{ESUN_\lambda \cdot \cos\theta_s} \tag{3-2}$$

where

ρ_P unitless planetary reflectance;

L_λ spectral radiance at the sensor's aperture;

d   earth-sun distance in astronomical units;

$ESUN_\lambda$ mean solar exoatmospheric irradiances;

θ_s solar zenith angle in degrees.

The values of the parameters are referred in Chander and Markhem (2003), see also Appendix D. Figure 3-8 shows the two geometric and radiometric calibrated images of Landsat 5 TM of the river system in CCRW. The April 2004 image shows green fields scattered around the river corridors, while the September image shows a sign of bare soil in brownish red color in the same area. Sampling sites are presented as the yellow dots on the maps. A total of 22 locations were investigated for the presences of riparian vegetation.
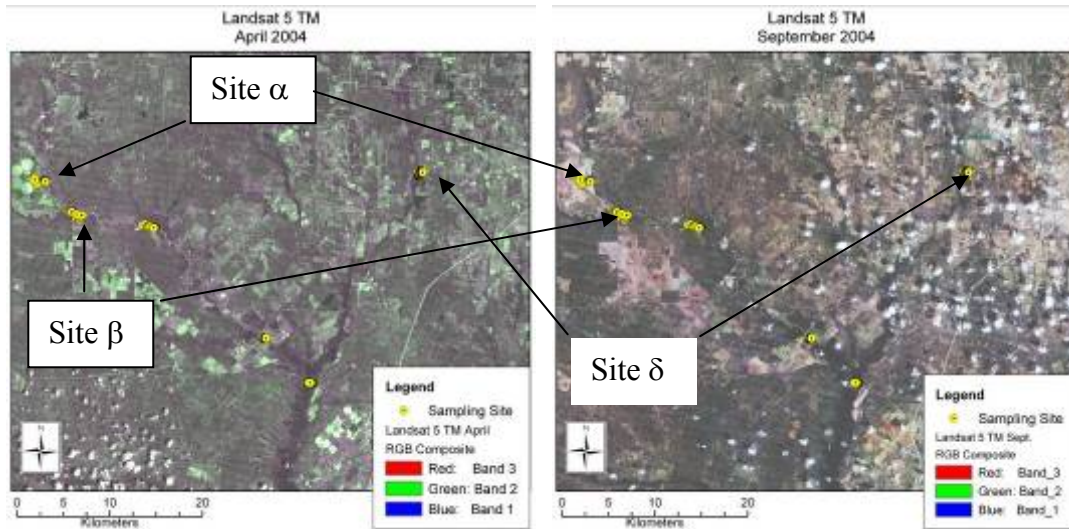
Figure 3-8. Landsat 5 TM satellite captured the central area of CCRW's river system in April and September 2004.
The images are composed of red:green:blue composites in accordance to the Landsat 5 TM band 3:2:1, respectively. Some of the green fields changed to brownish red color due to the fact that the vegetation cover had been harvested, thus the bare soil was exposed.

Vegetation indices were calculated from the calibrated reflectance obtained earlier. The values of vegetation indices were extracted according to the ground truth database. Two sets of ground truth data are riparian forest and non-riparian area. The vegetation indices datasets were prepared as input datasets for GP analysis by combining them with the Landsat 5 TM band 1, 2, and 3, SAVI, and soil moisture estimates. Figure 3-7 shows the schematic of the input data preparation. The SAVI maps of the CCRW are shown in Figure 3-9. The two images are geometrically and radiometrically corrected. The histogram of the September image was adjusted to match the histogram of the April image for accurate comparison. The average SAVI value of the September image is smaller than the average SAVI value of the April image, which indicates that the overall area in April was greener than the same area in September. A large-scaled image of infrared aerial photograph (1-m resolution) is also shown in comparison with the Landsat 5 TM image of the same area in Figure 3-10 (see also Figure 3-3). We cannot observe a riparian

buffer that has a footprint smaller than 30 m due to the limitation of spatial resolution of the

Landsat 5 TM image (30 m).
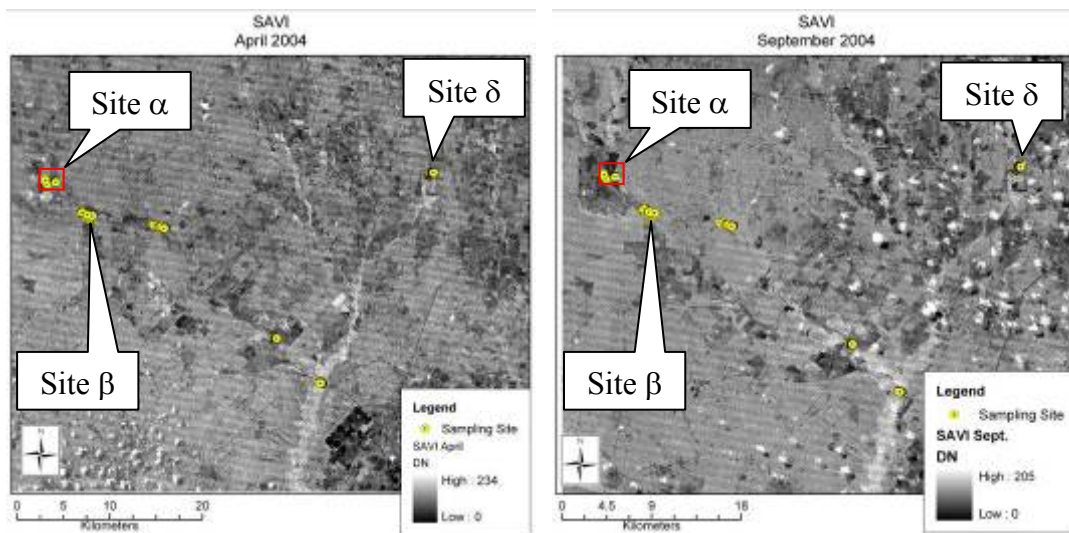
Figure 3-9. SAVI is calculated from the reflectance values of the Landsat 5 TM images.
The histogram of the September image is adjusted to match the histogram of the April image for accurate
comparison. The two images reveal the changes of the vegetation covers of the CCRW. The vegetation
covers decreased from April to September 2004. Ground truth locations along the rivers are presented in
the yellow dots.
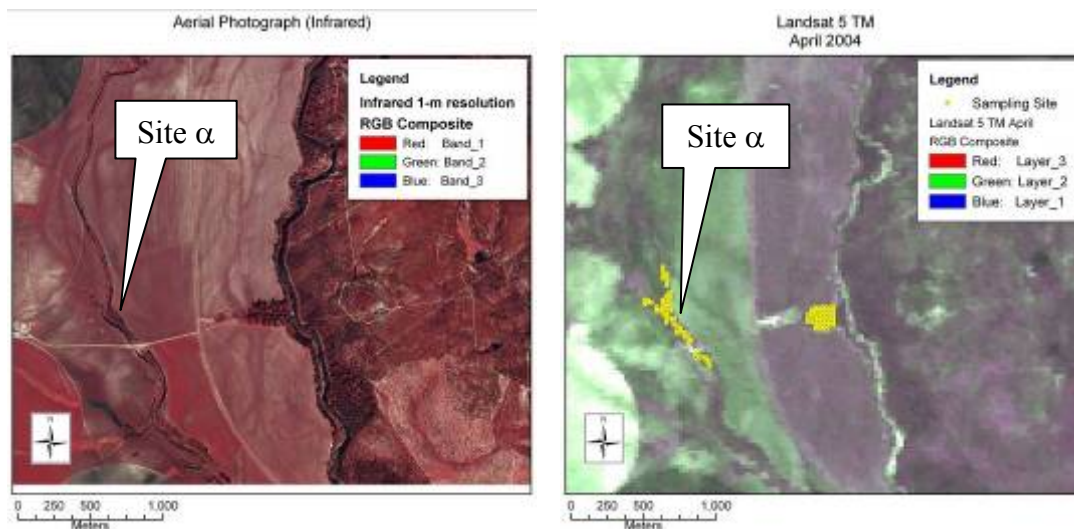
Figure 3-10. The infrared aerial photograph (left) is compared to the Landsat 5 TM image.
The 30-m spatial resolution of the Landsat image could not capture the details of the riparian zones. The
arrow points to the sampling location showed in Figure 3-3.

Eight vegetation indices were used to classify riparian forest using GP approach. This procedure was used to find the most suitable vegetation index for riparian classification. The simple ratio (SR) provided the highest accuracy, but not significantly higher than other vegetation indices. An average value of the accuracy is 67.1 percent. Figure 3-11 summarizes the classification accuracy based on vegetation indices only. Since the outcomes are similar across all vegetation indices, it is preferable that more widely used vegetation indices would be the best fit in the semi-arid watershed. SAVI therefore was selected and used in many fields of studies and applications. SAVI, SAVI-2, and other soil adjusted indices were also found favorable ones due to the correction for soil effects. MSR, RDVI, and NLI are similar to the NDVI except that they are normalized for non-linear effect. The final candidates that may be used in the subsequent full-scale analysis are NDVI, SAVI, SR, and MSR. NDVI is more suitable for sparse vegetation cover, while SR is good for dense vegetation cover. Yet the SAVI offers correction for background effects, and MSR provides a more linear relationship while retains the fundamental basis of SR that reduces the sensitivity to noise.

Accuracy Assessment for VI selection
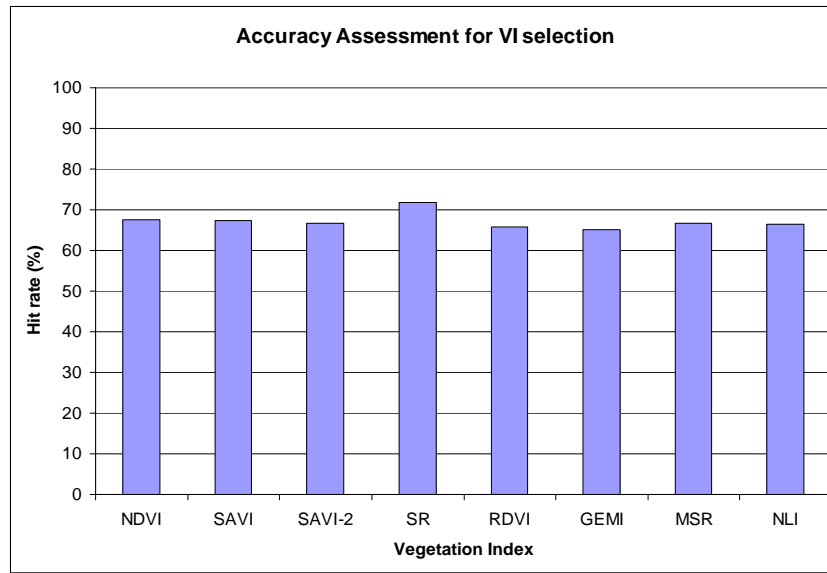
Figure 3-11. Classification accuracy across vegetation indices based on GP approach
is showing that all of the vegetation indices are equally good. SR seems to offer better accuracy (71%)
while the rest of the VIs yielded an average accuracy of 67%. NDVI, SAVI, SR, and MSR are chosen for
the following analysis later in this chapter.

Preparation of soil moisture maps

Soil moisture information is included as an ancillary dataset to help separate the riparian

forest from any other vegetation covers with similar spectral responses. A per-pixel logical

channel classification method is used to add the soil moisture data as an additional layer in the

Landsat spectral metric. Two soil moisture maps have been created based on RADARSAT-1

SAR images and other ancillary data (Makkeasorn et al., 2006). Estimates of soil moisture are

performed based on SAR technology (Zribi et al., 2005). This analysis counts on RADARSAT-1

SAR imageries acquired in the same time period of April and September 2004. The

RADARSAT-1 standard full-resolution SAR imagery covers approximately 100 km x 100 km

with the pixel size of 12.5 m. The electromagnetic pulse used in RADARSAT-1 SAR is in the C-

band frequency (5.3 GHz; 5.66 cm wavelength) (ASF, 1999). To ensure the accuracy of the data,

radiometric and geometric corrections were performed to all data by Alaska Satellite Facility (ASF) based on the SAR Processing Algorithm Document by Olmsted (1993) with the aid of corner reflectors (Freeman, 1992). Before the installation of the five (5) corner reflectors, we used the Satellite Tool Kit® (STK) provided by Analytical Graphics, Inc. (www.agi.com) to support the determination of the correct orientations for pointing our corner reflectors to SAR acquisition pathway. Then the Two Line Element (TLE) was used to determine the look direction of each corner reflector after finding out its GPS coordinate.

In general, the two known backscatter measurements used by ASF constantly to perform the SAR calibrations include the Amazon rain forest in Brazil and site-specific corner reflectors installed in Alaska. With the additional corner reflectors recently installed in south Texas, ASF can ensure the accuracy of the radiometric calibrations. These mid-latitude corner reflectors provide additional references for both radiometric and geometric calibrations in this application (Small et al., 1997). To remove the center-bias phenomena and the background noise, the SAR data were processed from pixel intensity to "sigma-naught", backscatter coefficient. For ASF's purpose, sigma-naught (σ_0) is defined as:

$$\sigma_0 = 10 \cdot \log \left\{ a2 \cdot \left[d^2 - \left(a1 \cdot n(r) \right) \right] + a3 \right\} \tag{3-3}$$

where, d is pixel intensity (0 – 255), a1 is noise scaling, a2 is linear conversion, a3 is offset, and n(r) is noise as a function of range. The coefficients are found in the Radiometric Data Record (part of the CEOS leader file) (Olmsted, 1993). The σ_0 is expressed in decibel (dB). The σ_0 was, afterwards, converted to a digital number (DN) as the following:

$$DN = (\sigma_0 * 10) + 255 \qquad\qquad (3\text{-}4)$$

Thus, the SAR imagery acquired in April 2004 was geometrically corrected using ground control points. The RMSE after the geo-referencing became less than 8 m. The other SAR imagery acquired in September 2004 was then image-to-image rectified on the April 2004 image. By overlaying the two SAR data together, the spatial error between the two acquisitions can be minimized further. This technique is normally used in multi-temporal study to detect changes over time. The RMSE of the spatially horizontal error eventually reaches a level of less than 2 m based on the image-to-image geo-referencing.

The soil moisture maps were derived using the advantages of the evolutionary process based on GP algorithm (Makkeasorn et al., 2006). A series of crucial topographic and geographic features included in this process are: slope, aspect, vegetation cover, and soil permeability to compliment the well calibrated SAR data (see Figure 3-12). Research indicates that the GP proved useful for generating a highly non-linear structure that exhibits correlations between the model estimated soil moisture and the ground truth measurements. These soil moisture maps were included in the mapping of riparian buffers based on the idea that the hydrologic regimes of the riparian zones are subjected to shallow vadose zones. Thus, the riparian zones should present high values of soil moisture content.
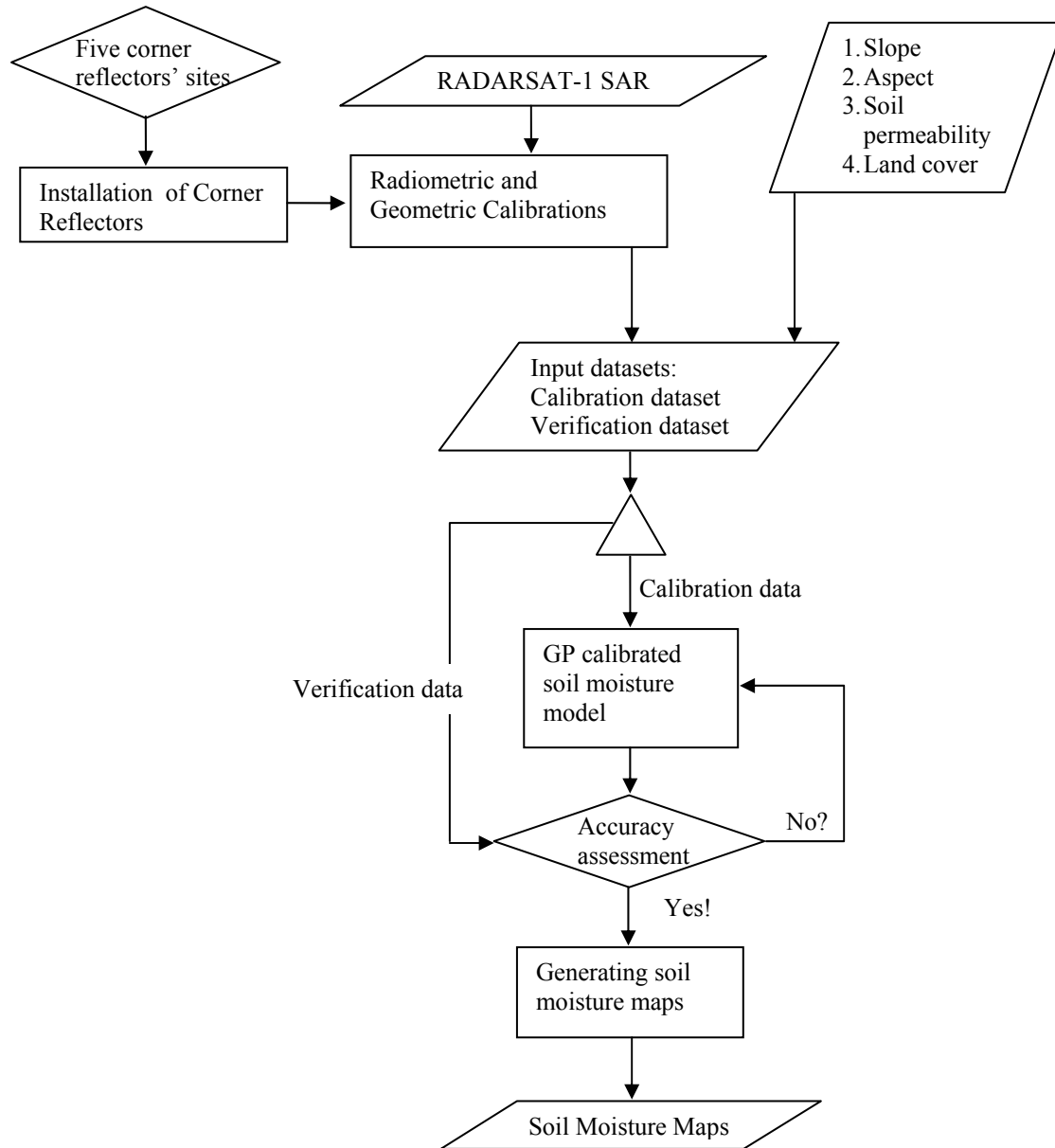
Figure 3-12. Schematic of generation of soil moisture maps using GP-based evolutionary process. Input data used to generate soil moisture map include RADARSAT-1 SAR, DEM-derived slope and aspect, STATSGO-based soil permeability, and NLCD land cover data.

Classification using genetic programming

Reflectance values of the Landsat 5 TM were calculated following through the revised radiometric calibration in the study of Chander and Markham (2003). The reflectance values at the sample sites were exported as tabular format to MS® Excel. The Landsat 5 TM band 1, 2, 3, and 4 were used in conjunction with the vegetation indices and the soil moisture data. Two classifying values of the ground truth data are riparian forest assigned as "1", and non-riparian area assigned as "0". The whole dataset was divided into two parts equally in number of data points. The first set of data is used by GP to calibrate evolutionary computational programs. Millions of symbolic regression programs (a.k.a. nonlinear regression models) can be generated using mathematical operations in GP algorithm including addition, subtraction, multiplying, dividing, and trigonometry. All models generated by GP may be measured up against the unseen data sets as an independent cross measure. Accuracy assessment can be performed by counting the 'hit rate'. 'Hit' means the models correctly assign the same classified pixel as the ground truth data; otherwise the models misclassify the pixel. Higher hit rates indicate better models. Since the second dataset is 'unseen' by the GP it means that GP never knows that the second dataset exists during the evolutionary processes. Therefore, this method of accuracy assessment is believed to be free from bias. GP chooses the best 30 models out of more than twenty millions (20,000,000) models in the program. The best model is manually selected based on the experiences of the author.

Another measurement for model evaluation is the percent of use. Within the selected 30 best models, the GP counts the use of each input parameters, which are the Landsat band 1, 2, 3, 4, SAVI, and soil moisture, as an impact index to reflect the importance of each input parameter. If an input parameter is used in all the 30 selected models, the input parameter may be expressed

as 100% usage. Based on the idea of the evolutionary processes, the unimportant variable would be eliminated during the evolutionary processes. This is the advantage of GP's white box characteristic. If a black box algorithm were used, one could hardly find out what input parameter is important and how important it is.

Classification using a combined unsupervised and supervised classification

The classification processes were performed using Erdas Imagine 9.0 software suit. The Landsat TM data, the corresponding vegetation indices, and the corresponding soil moisture were combined together using a per-pixel logical layer. They include Landsat TM band 1-4, a layer of vegetation index, and a layer of soil moisture data. Two images collected in April and September 2004 data served as the basis for temporal analysis in the context of change detection.

Firstly, unsupervised classification based on ISODATA algorithm was performed to produce a 100-class image. Convergence threshold is set to 0.990. Secondly, supervised classification was manually performed to consolidate the 100 classes into 2 classes, which are riparian forest and non-riparian area. Details on how to use the functions in Erdas Imagine 9.0 software package can be found in Erdas Imagine Tour Guides (Leica Geosystems, 2005). Accuracy assessment was performed by comparing the classified pixels to the ground truth data. Seventy (70) pixels were selected for the accuracy assessment.

Results and Discussion

A summary of the accuracy assessments is shown in Figure 3-13. When the sole classification is the vegetation index, an accuracy of 67.1% was obtained across all the vegetation indices. When using the vegetation indices along with the other inputs, GP can

classify riparian forests at more than 95% accuracy. SAVI provided the highest accuracy compared to all the other vegetation indices. The result confirmed that SAVI is more suitable for sparse vegetation areas such as this semi-arid watershed. This conclusion is in agreement with Kerr et al. (1989) and Nicholson et al. (1990)
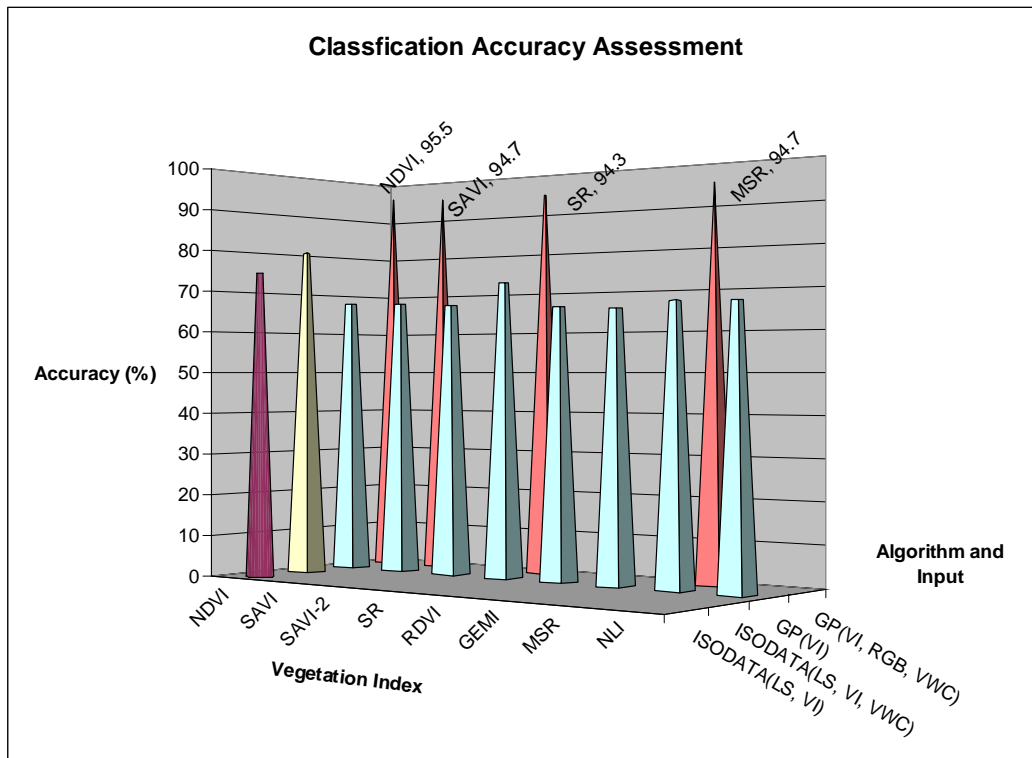
Figure 3-13. A summary of accuracy assessment for all classification algorithms and input parameters is showing that GP-based classification with Landsat band 1, 2, 3, soil moisture, and vegetation index offers the highest accuracy for classifying riparian buffers in a semi-arid watershed in south Texas.

The results also suggested that the inclusion of soil moisture data did not significantly improve the classification given the fact that soil moisture was not often used by GP during the evolutionary processes. Even though GP considers soil moisture as a less important parameter compared to the other data, it does not mean that soil moisture must be taken out of the consideration. In contrast, soil moisture is physically a very important factor of the changes of land cover. This is a fault of the GP-based model. The performance-based summary of all input

parameters in GP analysis is shown in Table 3-2. This analysis showed that soil moisture data were used only 13% and 20% of the time during the evolutionary process for April and September datasets, respectively. This conclusion could be explained that vegetation index had already been used to generate the soil moisture data prior to classification process (Makkeasorn et al., 2006). Thus, the soil moisture data became partially redundant information when another vegetation index was reused during the classification process. The GP detected the redundancy so that it gave relatively less importance to the soil moisture data in the evolutionary process. A similar explanation is also applied to the less frequently use of Landsat 5 TM band 3 (ρ_3). Band 3 was already used for deriving vegetation indices, thus it also became partially redundant information. Since the vegetation indices measure the greenness of vegetation covers, the Landsat 5 band 2 (visible green) automatically became redundant information. The frequently use of band 1 (which corresponds to the visible blue) can be explained by inspecting the Landsat images on a larger scale. It became apparent that most of the riparian vegetation created shadows due to the sun angle and the thick leaves. The shadows in the background of the riparian zones are highly sensitive to the Landsat band 1. Therefore, GP picked up Landsat band 1 more frequently.

Table 3-2. Frequency of input parameters used in the GP classification process.

INPUT	FREQUENCY	
	April Data	September Data
Band 1 (visible blue)	70%	87%
Band 2 (visible green)	83%	100%
Band 3 (visible red)	100%	80%
Band 4 (near infrared)	100%	47%
SAVI	90%	53%
VWC	13%	20%
Accuracy	94.7%	90.3%

Overall, the two RICALs derived from this study were eventually presented in the form of a compound function in Eq.(3-5) and Eq.(3-6) associated with April and September 2004 data, respectively. Although six input parameters were used in the classification processes including Landsat 5 TM band 1-4, a vegetation index, and the soil moisture, the evolutionary processes of the GP primarily selected four of them including band 1, 2, 3, and the vegetation index. It can be concluded that the soil moisture data and band 4 (near-infrared) were discarded eventually.

From ground data investigation, the riparian forests corresponded directly to band 2 (ρ_2) and vegetation indices. In Eq.(3-5) ρ_2 is the numerator that subtracts values from the equation. The value of term A_1 decreases when the value of ρ_2 increases. SAVI is the denominator in Eq.(3-5). Therefore, if the values of ρ_2 and SAVI are increasing, the value in the square bracket would approach zero, which would make the value of Eq.(3-5) greater than 0.5. The pixel value would be classified as riparian forest when the calculated pixel value is greater than 0.5 because the threshold was set at 0.5. In contrast, if the values of ρ_2 and SAVI are small, the term in the square bracket may become a large negative number, thus the pixel value, calculated by Eq.(3-5), became less than 0.5. The pixel would be classified as non-riparian area.

$$RICAL(Apr04) = -\left[\frac{(0.1272730976 \cdot A_1) + 1.3985419472 \cdot \rho_3 - 7.4382195025}{SAVI}\right] + 0.9765300750$$

(3-5)

where

$A_1 \quad = (0.0587134747 \cdot \rho_1) - (4.2407217411 \cdot \rho_2) + 2 \cdot Sin(1.5499701499 - \rho_2)$
$\quad\quad + \rho_3 + 0.9765300732$

$\rho_1 \quad$ = Landsat 5 TM: reflectance band 1

ρ_2 = Landsat 5 TM: reflectance band 2

ρ_3 = Landsat 5 TM: reflectance band 3

ρ_4 = Landsat 5 TM: reflectance band 4

SAVI   = Soil Adjusted Vegetation Index that is derived from the reflectance values of Landsat 5

          TM data

Eq.(3-6) is another RICAL derived from the September dataset. If the pixel value is less than 0.5, the pixel is classified as non-riparian area; otherwise the pixel is assigned as riparian forest. SAVI is the nominator and the ρ_2 is a denominator in term B_3 of Eq.(3-6). If the values of SAVI and ρ_2 increase or decrease together corresponding to the riparian forest, then the term B_3 would increase or decrease, respectively. However, the term cosine B_3 would only have a range of [-1, 1], allowing the value of term B_2 to be in a range of [-5.264598321, -2.418568659]. The values of B_2 within such a range allow the cosine B_2 term to be in a range of [-1, 0.6347]. Low values of ρ_3 (Landsat 5 TM band 3) correspond to the riparian forest. If the value of ρ_3 is zero, then the square bracket term in B_1 could be in the range of [-0.45, 0.7]. Thus, the value of B_1 would be in a range of [0.7, 1.85]. The cosine B_1 would then be in a range of [-0.276, 0.76]. In summary, the value of Eq.(3-6) depends on ρ_2 and ρ_3. In the riparian forests, the value of ρ_2 is larger than the value of ρ_3. In contrast, the non-riparian area would have a small ρ_2 and a large ρ_3.

$$RICAL\ (Sep\,04) = -0.5786151885\ \cdot \left[2 \cdot Cos\left(B_1\right) + \rho_2 - \rho_3\right]$$

(3-6)

where

$$B_1 \quad = \left[\frac{Cos\left(B_2\right) - \rho_3}{-1.4270856380}\right] + 1.1500550234$$

$$B_2 \quad = \left[1.4723014831 \cdot Cos(B_3) \right] - 3.8415834903$$

$$B_3 \quad = \frac{\left[(2.6970122281 \cdot SAVI) + (2 \cdot \rho_1) \right]}{\rho_2}$$

ρ_1 = Landsat 5 TM: reflectance band 1

ρ_2 = Landsat 5 TM: reflectance band 2

ρ_3 = Landsat 5 TM: reflectance band 3

ρ_4 = Landsat 5 TM: reflectance band 4

SAVI   = Soil Adjusted Vegetation Index that is derived from the reflectance values of Landsat 5 TM data

Table 3-2 summarizes the frequency of each parameter picked up during the evolutions of the GP-based riparian classification models. The frequency of each parameter represents how often such a parameter would be used in the formulated models. Each evolutionary process in GP may generate at least twenty million (20,000,000) riparian classification models using predefined operators indicative of the reliability of the assessment. The best model was manually selected out of the final thirty (30) candidate models ultimately based on its accuracy. The criterion of selection was applied by comparing value of classified pixels to the pixels in the unseen dataset. The unseen dataset was used solely for verification purposes. Since the unseen ground data was not used during the calibration process, it was confirmed that the bias can be significantly reduced. The analysis revealed that the accuracy was at 94.7% for the April dataset and 90.3% for September 2004 datasets (see Table 3-2). Table 3-2 also shows that ρ_1, ρ_2, ρ_3, ρ_4 and SAVI were mostly used over 70% of the time during the evolutionary computation processes. On the

other hand, the soil moisture data were only used 13% of the time in the April dataset and 20% of the time in the September dataset, respectively. These results led to conclude that the soil moisture information does not significantly improve the riparian classification.

In the final stage of this study, a combination of GIS, remote sensing, and GP-based classification algorithms provide change detection maps of riparian buffers (see Figure 3-14). The mapping of riparian buffer zone reveals seasonal change, allowing effective water resource management at the watershed scale. An evaluation of the changes in the detection of riparian buffers reveals that the riparian buffers are spatially different. The riparian buffers on the Frio River increased during the study time period while the riparian buffers on the other rivers were generally unchanged (see Figure 3-14).

The magnified images of the sampling Site α are shown in Figure 3-15 in April and September 2004, respectively. It is clearly seen that the agricultural lands were densely covered with vegetation in April 2004, while the soil was exposed on the same lands in September 2004 (see Figure 3-15 a and b). The riparian forest zone on the ephemeral creek decreased from April to September in concert with the agricultural vegetation because the creek was fed by the runoff water from the nearby irrigation system (see Figure 3-16). In contrast, the riparian forest of the perennial river increased over the time period. It can be explained that during the growing season significant amounts of groundwater were drafted and used for the agricultural crops causing the groundwater level to decrease in April. Consequently, the riparian forest was starved for water. After the harvesting season, the irrigation had stopped or slowed down, thus the level of groundwater rose up, replenishing the riparian forest in the perennial river in September (see Figure 3-16).
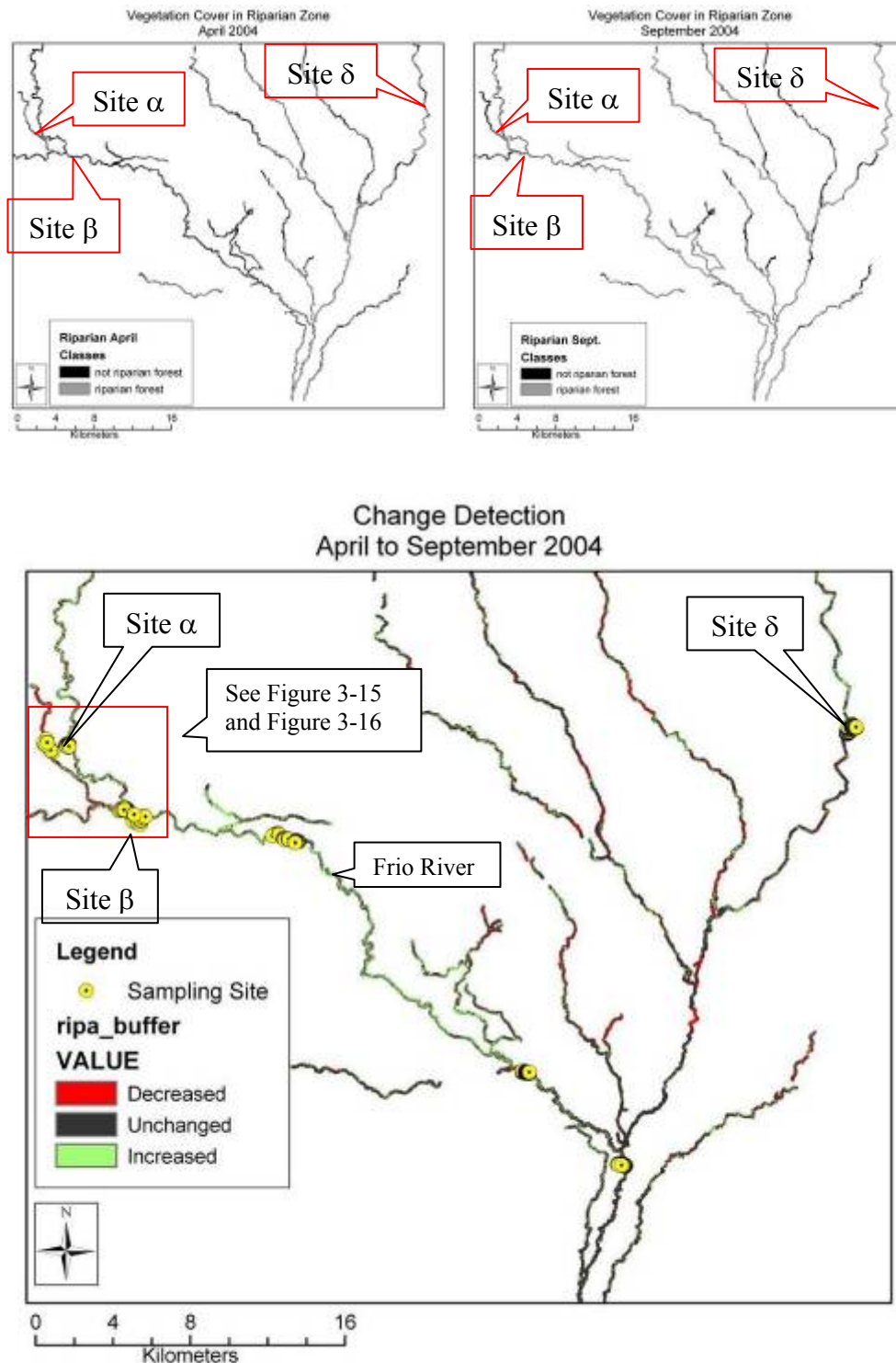
Figure 3-14. Change detection map of riparian buffers in CCRW
is presenting some increased, decreasied, and unchanged riparian zones. The green color indicates the increased vegetation cover in the riparian areas, while the red color indicates the decreased vegetation cover in the areas.
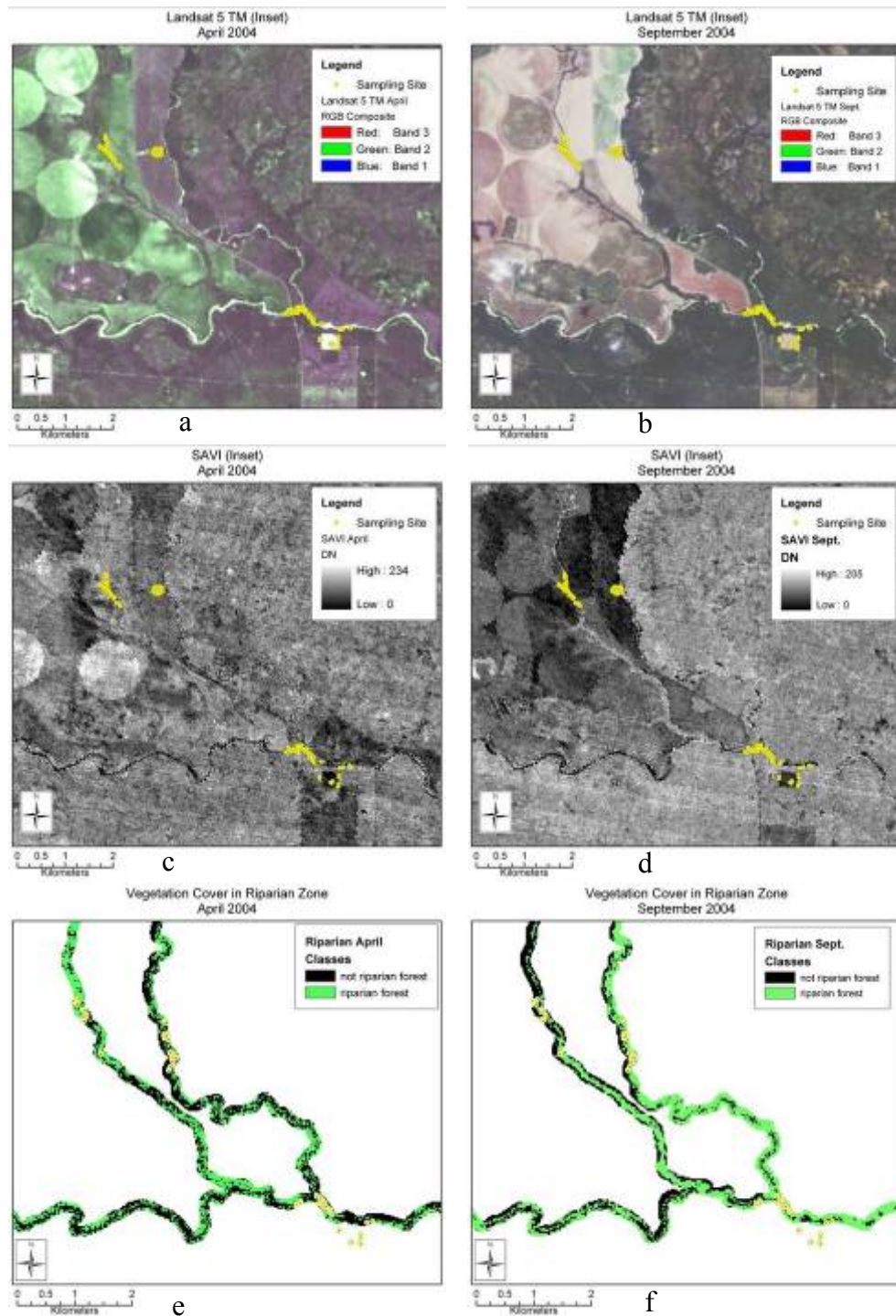
Figure 3-15. Landsat 5 TM, SAVI, and riparian classification images of April and September data are showing the changes of the vegetation cover. These images are the large-scaled area of the inset in Figure 3-14. The (a) and (b) are Landsat images, (c) and (d) are VI imageries, and (e) and (f) are the riparian forest classification imageries.
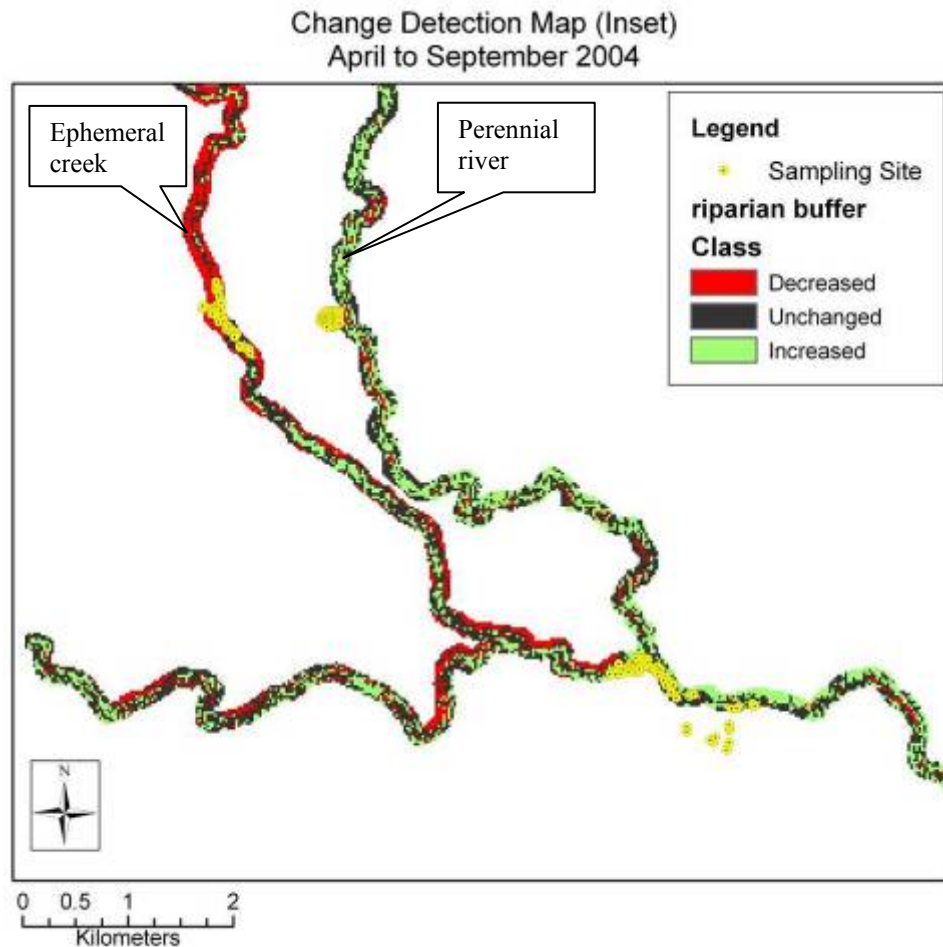
Figure 3-16. The zoomed-in map reveals the changes of the riparian zones between April and September 2004.
While a riparian zone was subjected to decreasing vegetation cover, the other riparian zone gained vegetation cover. The difference of change may be influenced by the practices of land management surrounding the area.

Conclusions

This chapter develops a unique RICAL algorithm that is proved useful in this analysis. The RICAL outputs showed that soil moisture information had minimal influence for improving the classification of riparian buffer zones. This is not really agreeable knowing that soil moisture, in fact, is physically the most important factor. It is believed that the redundancy of vegetation index used in the development of the soil moisture model and the vegetation index used in the riparian classification causes the rejection of the soil moisture parameter during the evolutionary

process. The GP was fault in this regard, however, it does a very good estimation in classifying riparian buffers. The white box characteristic of the GP offers the opportunity to analyze highly nonlinear structure subject to evolutionary numerical rules and constraints. The GP also allows us to evaluate the importance of each input parameter as well. It shows very high accuracy as compared to the unseen ground truth data.

All of the vegetation indices yielded similar results with very little differences. Even though we hoped that the soil-adjusted vegetation index would provide better accuracy due to the fact that the soil background would be adjusted in the semi-arid watershed. This finding is in agreement with many other studies that SAVI performs better with sparse vegetation cover and semi-arid areas. If the value of L in SAVI were measured from the sampling areas and used, it could have been found that the SAVI could provide even better results. Although the ground truth measurement of the soil's reflectance was not considered, which involves using field spectroradiometer measurements, the data mining effort could greatly improve the accuracy. Based on the data mining feature embedded in the GP it can be concluded that the Landsat 5 TM band 1 (corresponding to visible blue spectral) greatly helps to improve the classification accuracy. It is believed that the shadows and background effects of riparian forests make the band 1 useful for detecting riparian buffers. The change detection of riparian buffers can temporally and spatially be made and used in many applications for water resources management at the watershed scale.

References

Abbey, D.G.A., 1988. Tangipahoe crossing: restoring and preserving one of Louisiana's designated scenic rivers. Erosion control: stay in tune. Proceeding Conference. XIX International. Erosion Control Association. New Orleans, LA, 107-124.

Alaska Satellite Facility (ASF), 1999. RADARSAT-1 Standard Beam SAR Images, report. Geophysical Institute, University of Alaska Fairbanks, Fairbanks, Alaska, USA.

Apan, A.A., Raine, S.R., and Paterson, M.S., 2002. Mapping and analysis of changes in the riparian landscape structure of the Lockyer Valley catchment, Queensland, Australia. Landscape and Urban Planning 59, 43-57.

Bannari, A., Morin, D., Bonn F., and Huete, A.R., 1995. A review of vegetation indices. Remote Sensing Reviews 13, 95-120.

Baret, F. and Guyot, G., 1991. Potentials and limits of vegetation indices for LAI and APAR assessment. Remote Sensing of Environment 35, 161-173.

Campbell, J.B., 2002. Introduction to Remote Sensing, 3rd edition. The Guilford Press, New York, USA, 621. ISBN 1572306408.

Chander, G. and Markham, B., 2003. Revised Landsat-5 TM radiometric calibration procedures and postcalibration dynamic ranges. IEEE Transactions on Geoscience and Remote Sensing 41 (3), 2674-2677.

Chapman, E.W. and Ribic, C.A., 2002. The impact of buffer strips and stream-side grazing on small mammals in southwestern Wisconsin. Agriculture, Ecosystems and Environment 88, 49-59.

Congalton, R.G., Birch, K., Jones, R., and Schriever, J., 2002. Evaluating remotely sensed techniques for mapping riparian vegetation. Computers and Electronics in Agriculture 37, 113-126.

Cooper, A.B., Smith, C.M., and Smith, M.J., 1995. Effects of riparian set-aside on soil characteristics in an agricultural landscape: Implications for nutrient transport and retention. Agriculture, Ecosystems and Environment 55, 61-67.

Derek, R.P. and Brunke, S.P., 2001. A comparison of spectral mixture analysis and ten vegetation indices for estimating Boreal Forest biophysical info. Canadian Journal of Remote Sensing 27 (6), 627-635.

Filippi, A.M. and Jensen, J.R., 2006. Fuzzy learning vector quantization for hyperspectral coastal vegetation classification. Remote Sensing of Environment 100, 512-530.

Freeman, A., 1992. SAR calibration: an overview. IEEE Transactions on Geoscience and Remote Sensing, 30(6), 1107-1121.

Goetz, S.J., Wright, R.K., Smith, A.J., Zinecker, E., and Schaub, E., 2003. IKONOS imagery for resource management: Tree cover, impervious surfaces, and riparian buffer analyses in the mid-Atlantic region. Remote Sensing of Environment 88, 195-208.

Gutiérrez, M., Johnson, E., and Mickus, K., 2004. Watershed assessment along a segment of the Rio Conchos in Northern Mexico using satellite images. Journal of Arid Environments 56, 395-412.

Hattermann, F.F., Krysanova, V., Habeck, A., and Bronstert, A., 2006. Integrating wetlands and riparian zones in river basin modeling. Ecological Modelling. doi:10.1016/j.ecolmodel .2005.06.012

Hepner, G.F., Logan, T., Ritter, N., and Bryant, N., 1990. Artificial neural network classification using a minimal training set: Comparison to conventional supervised classification. Photogrammetric Engineering and Remote Sensing 56, 469-473.

Huete, A.R., Justice, C., and van Leeuwen, W., 1999. MODIS vegetation index (MOD 13): algorithm theoretical basis document, version 3. University of Arizona, Tucson, AZ, 129.

Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment 25, 295-309.

Huete, A.R., 1989. Soil influences in remotely sensed vegetation-canopy spectra. Theory and Applications of Optical Remote Sensing. Wiley, Washington, USA, 107-141.

Ingram, J.C., Dawson, T.P., and Whittaker, R.J., 2005. Mapping tropical forest structure in southeastern Madagascar using remote sensing and artificial neural networks. Remote Sensing of Environment 94, 491-507.

Jackson, R.D., 1983. Spectral indices in n-space. Remote Sensing of Environment 13, 409-421.

Japan Aerospace eXploration Agency (JAXA), 2003. Earth Observation: Satellites and Spacecraft. Http://www.jaza.jp/missions/projects/sat/eos/index_e.html, retrieved on December 5, 2005.

Jensen, J.R., 1996. Introductory Digital Image Processing – A Remote Sensing Perspective, 2nd edition. Prentice Hall, New Jersey, 316.

Kerr, Y.H., Imbernon, J., Dedieu, G., Hautecoeur, O., Lagouarde, J., and Seguin, B., 1989. NOAA AVHRR and its uses for rainfall and evapotranspiration monitoring. International Journal of Remote Sensing, 10, 847-854.

King, M., 2004. 24 EOS Measurements. Goddard Space Flight Center. URL: http://eospso.gsfc.nasa.gov/eos_homepage/for_scientists/24measurements.php. Last updated March 14, 2000.

Koza, J.R. and Rice, J.P., 1991. Genetic generation of both the weights and architecture for a neural network, Volume 2. IEEE Press, 397-404.

Lautenschlager, L.F. and Perry, C.R., 1981. Comparison of vegetation indices based on satellite-acquired spectral data. American Statistical Association. Proceedings of the Section on Survey Research Methods, Michigan, USA, 77-82.

Lee, C.T., and Marsh, S.E., 1995. The use of archival Landsat MSS and ancillary data in a GIS environment to map historical change in an urban riparian habitat. Photogrammetric Engineering and Remote Sensing 61 (8), 999-1008.

Leica Geosystems, 2005. Erdas Imagine Tour Guides. Geospatial Imagine, LLC, Norcross, Georgia.

Makkeasorn, A., Chang, N.-B., Beaman, M., Wyatt, C., and Slater, C., 2006. Soil moisture estimation in a semi-arid watershed using RADARSAT-1 satellite imagery and genetic programming. Water Resources Research 42, W09401, doi:10.1029/2005WR004033.

Mertes, L.A.K., 2002. Remote sensing of riverine landscapes. Freshwater Biology 47:799-816.Muller E. 1997. Mapping riparian vegetation along rivers: old concepts and new methods. Aquatic Botany 58, 411-437.

Miltner, R.J., White, D., and Yoder, C., 2004. The biotic integrity of streams in urban and suburbanizing landscapes. Landscape and Urban Planning 69, 87-100.

Muller, E., 1997. Mapping riparian vegetation along rivers: old concepts and new methods. Aquatic Botany 58, 411-437.

Muukkonen, P. and Heiskanen, J., 2005. Estimating biomass for boreal forests using ASTER satellite data combined with standwise forest inventory data, Remote Sensing of Environment 99, 434-447.

Nagler, P.L., Glenn, E.P., and Huete, A.R., 2001. Assessment of spectral vegetation indices for riparian vegetation in the Colorado River delta, Mexico. Journal of Arid Environments 49, 91-110.

Narumalani, S., Zhou, Y., and Jensen, J.R., 1997. Application of remote sensing and geographic information systems to the delineation and analysis of riparian buffer zones. Aquatic Botany 58, 393–409.

NASA, 2006a. The Landsat Program. http://landsat.gsfc.nasa.gov. last update January 8, 2007.

Nicholson, S.E., Davenport, M.L., and Malo, A.D., 1990. A comparison of the vegetation response to rainfall in the Sahel and East Africa, using NDVI from NOAA AVHRR. Climate Change 17, 209-214.

Olmsted, C., 1993. Alaska Satellite Facility Scientific SAR User's Guide. ASF-SD-003, Geophysical Institute, University of Alaska Fairbanks, Alaska, USA, 53.

Osborne, L.L. and Kovacic, D.A., 1993. Riparian vegetated buffer strips in water-quality restoration and stream management. Freshwater Biology 29 (2), 243–258.

Palone, R.S. and Todd, A.H., editors, 1997. Chesapeake Bay riparian handbook: a guide for establishing and maintaining riparian forest buffers. USDA Forest Service, NA-TP-02-97, Radnor, PA.

Peddle, D.R. and Brunke, S.P., 2001. A comparison of spectral mixture analysis and ten vegetation indices for estimating Boreal Forest biophysical info. Canadian Journal of Remote Sensing 27 (6), 627-635.

Pinty, B. and Verstraete, M.M., 1992. GEMI: A non-linear index to monitor global vegetation from satellites. Vegetatio, 101, 15-20, as cited in Peddle DR, Brunke SP. 2001. A comparison of spectral mixture analysis and ten vegetation indices for estimating Boreal Forest biophysical info. Canadian Journal of Remote Sensing 27 (6), 627-635.

Rouse, J.W., Haas, R.H., Schell, J.A., and Deering, D.W., 1973. Monitoring vegetation systems in the Great Plains with ERTS. Third ERTS Symposium, NASA SP-351 I, 309-317.

Ritchie, M.D., Motsinge,r A.A., Bush, W.S., Coffey, C.S., and Moore, J.H., 2007. Genetic programming neural networks: a powerful bioinformatics tool for human genetics. Applied Soft Computing 7, 471-479.

Ritchie, M.D., White, B.E., Parker, J.S., Hahn, L.W., and Moore, J.H., 2003. Optimization of neural network architecture using genetic programming improves detection of gene—gene interactions in studies of human diseases. BMC Bioinformatics 4, 28p. doi:10.1186/1471-2105-4-28

Sangle, P.S. and George, S.M., 2005. Alternate neural network models as supervised classifiers for satellite data. Journal of Environmental Informatics 6 (2), 80-92.

Scott, R.L., Shuttleworth, W.J., Goodrich, D.C., Maddock, (III) T., 2000. The water use of two dominant vegetation communities in a semiarid riparian ecosystem. Agricultural and Forest Meteorology 105, 241-256.

Shupe, S.M. and Marsh, S.E., 2004. Cover- and density-based vegetation classifications of the Sonoran Desert using Landsat TM and ERS-1 SAR imagery. Remote Sensing of Environment 93, 131-149.

Sivanpillai, R., Smith, C.T., Srinivasan, R., Messina, M.G., and Ben Wu X., 2005. Estimating regional forest cover in East Texas using Enhanced Thematic Mapper (ETM+) data. Forest Ecology and Management 218, 342-352.

Small, D., Holecz, F., Meier, E., Nuesch, D., and Barmettler, A., 1997. Geometric and Radiometric Calibration of RADASAT Images. Proceeding of Geomatics in the Era of RADARSAT, Ottawa, Canada.

Tucker, C.J. and Miller, L.D., 1977. Soil Spectra Contributions to Grass Canopy Spectral Reflectance. Photogrammetric Engineering and Remote Sensing 43 (6), 721-726.

U.S. EPA, 2005. National Management Measures to Protect and Restore Wetlands and Riparian Areas for the Abatement of Nonpoint Source Pollution. Office of Water, Washington, D.C.

USDA-ARS, 2006. How a vegetation index works. U.S. Water Conservation Laboratory, Remote Sensing Research Program. Retrieved from http://www.uswcl.ars.ag.gov /epd/remsen/vi/VIWorks.htm in November 2006.

USGS, 2006. How Images Are Categorized. Arial Photographs and Satellite Images. online edition, U.S. Department of Interior – U.S. Geological Survey, Reston, VA, USA. http://erg.usgs.gov/isb/pubs/booklets/aerial/aerial.html. Retrieved in July 2006.

USGS, 2007. USGS surface-water daily data for Texas. National Water Information System: Web Interface. Http://waterdata.usgs.gov. Retrieved in January 2007.

Weier, J. and Herring, D., 2006. Measuring vegetation (NDVI & EVI). Earth Observatory, NASA. http://earthobservatory.nasa.gov/Library/MeasuringVegetation/printall.php.

Yuan, Y., Bingner, R.L., and Rebich, R.A., 2003. Evaluation of AnnAGNPS nitrogen loading in an agricultural watershed. Journal of American Water Resources Association 39 (2), 457-466.

Yuan, Y., Bingner, R.L., Williams, R.G., Lowrence, R.R., Bosch, D.D., and Sheridan, J.M., 2004. Integration of annagnps and remm for watershed riparian buffer system assessment. Abstract. Proceedings American Water Resources Association Summer Specialty Conference "Riparian Ecosystems and Buffers: Multi-scale Structure, Function, and Management". Olympic Valley, California, June 28-30 2004.

Zhao, S., Fang, J., Ji, W., and Tang, Z., 2003. Lake restoration from impoldering: impact of land conversion on riparian landscape in Honghu Lake area, Central Yangtze. Agriculture Ecosystems and Environment 95, 111-118.

Zribi, M., Baghdadi, N., Holah, N., and Fafin, O., 2005. New methodology for soil surface moisture estimation and its application to ENVISAT-ASAR multi-incidence data inversion. Remote Sensing of Environment 96, 485-496.

CHAPTER 4 OPTIMAL SITE SELECTION OF HYDROLOGICAL MONITORING STATIONS USING SATELLITE REMOTE SENSING AND GREY INTEGER PROGRAMMING

Introduction

The availability of adequate fresh water is a fundamental requirement for the sustainability of human and terrestrial landscapes. Hydrologic cycle significantly influences ways of livings in most societies such as urbanizations, ecological conservation, agricultural development, and industrialization. The hydrologic cycle is the process of water circulation on the Earth beginning as the precipitation. Once precipitation reach the ground, some of it may fill depressions (ponds, lakes, any storages), other part may penetrate the ground by infiltration to replenish soil moisture and then percolate into groundwater reservoir. Overflow in those depressions may flow on the Earth's surface as runoff or into defined channels such as streams and reservoirs. Yet, another part of the precipitation may be intercepted by vegetation and structural objects and would eventually return to the atmosphere by evaporation and transpiration. Consequently, hydrological monitoring networks that must monitor at the watershed scale should be deployed at strategically critical locations to advance the state-of-the-art environmental monitoring and sensing capability. There is an acute need to develop optimum arrangement/distribution of water sensor platforms for maximum sensing coverage and information-extraction capacity. Such optimally distributed monitoring networks may significantly advance our understanding of the integration and coupling of Earth surface and subsurface flow processes as mediated by the presence and flux of fresh water.

Each type of environmental monitoring system has unique monitoring requirements that are subjected to different constraints. Yet monitoring network assessment has been performed using many differing approaches. In general, they can be differentiated into two categories, including statistically-driven method and simulation-driven method in the optimization context for optimal site selection. For example, a statistical measure of information content was used as a means for assessing the effectiveness of a particular monitoring network configuration in Canada (Pickett and Whiting, 1981) and a study on interpolation techniques was undertaken in the Nerthelands resulting in an evaluation of errors involved in spatial analysis (Egmond and Onderdelinden, 1981). One statistically-driven method was proposed by Harmancioglu et al. (1999) using entropy, as a measure of uncertainty, to study the amount of information between sampling sites. If, over time, two sampling sites are producing the same information then one could be discontinued with a minimal loss of information. Kriging, which is a geostatistical and probabilistic approach, has been used for network optimization for decades. This method was often applied by the fields of soil science, ground water, and mining industry. However, extensive applications were also performed for surface water-quality, precipitation networks, and precipitation chemistry (Hughes and Lettenmaier, 1981; Eynon, 1988; Venkatram 1988; Jager et al., 1990; Ben-Jemaa et al., 1995; Christensen et al., 1997). In particular, Modak and Lohani (1985a, b, and c) aggregated the Air Quality Index (AQI) in support of a Minimum Spanning Tree (MST) optimization analysis for a network rationalization in Taipei, Taiwan. Behar and McElroy (1987) applied air quality simulation model and population exposure information to produce representative patterns and then employed the "Figure of Merit (FOT) and "Sphere of Influence (SOI)" to determine the minimum number of sites required. Lo et al. (1996) employed an integrated approach by using a steady state water quality simulation model to delineate the

biochemical oxygen demand (BOD) and dissolved oxygen (DO) curves providing a basis for the subsequent Kriging analysis for selecting the optimal locations of water quality monitoring sites in Keelung River, Taiwan. Odem (2004) developed an integrated statistics and optimization model to design the water quality monitoring network basin wide. The watershed characteristics and water-quality data were first analyzed using principal components analysis (PCA), cluster analysis (CA), and discriminant analysis (DA). The multivariate methods were used to group similar sampling sites in terms of water-quality, watershed vegetation, watershed geology, and watershed geomorphology information. Then the costs for physical access, laboratory testing and overhead were also gathered for each site and used to formulate an objective function that was minimized by the simulated annealing (SA) algorithm for monitoring network costs with respect to the constraint of a user-specified number of sampling sites in the final network. In the field of hydrology, SA was used to optimize a rainfall gaging network with the objective of providing estimates of mean precipitation while minimizing the estimation variance (Pardo-Iguzquiza 1998). Dixon, et al. (1999) also used SA to optimize a suite of river sampling sites on the Logan River and the Albert River in Queensland, Australia.

On the other hand, several integrated simulation and optimization approaches were developed to identify the optimal locations in response to ground water, surface water, and air quality monitoring requirements. Well-known cases include those using an integrated MT3D ground water transport model and integer programming for ground water quality monitoring network design (Hudak et al., 1995), using QUAL2E simulation model and multiobjective programming for surface water quality monitoring network relocation and design (Harmancioglu and Alpaslan,1992; Ning and Chang, 2002; Ning and Chang, 2005), using the Industrial Source Complex (ISC) and Empirical Kinetic Modeling Analysis (EKMA) and deterministic

compromise programming or genetic algorithms-based grey compromises programming for air quality monitoring network design (Chang and Tseng, 2001; Tseng and Chang, 2001; Chang et al., 2006).

The optimal site selection of hydrological monitoring stations in vast watersheds has to be performed in response to many geologic and geomorphologic features, such as soil properties, landscapes, channels, fault lines, erosion/deposition patches, and bedload transport history, however. Within vast watersheds, remote sensing imagery-grid may help reflect the status of precipitation, infiltration and percolation, stream flow, plant transpiration, soil moisture changes, and groundwater recharge that are all intimately related with each other to form water balance dynamics on the surface of these watersheds. This paper describes an optimal site selection strategy using an integrated approach of satellite remote sensing and grey integer programming (GIP) model with respect to some technical and resources constraints in an uncertain environment. GIP enables us to retrieve the hydrological trends and pinpoint the most critical locations for the deployment of monitoring stations in a vast watershed given that many parameters of concern can only be expressed by a set of interval numbers. Geo-environmental information amassed in this analysis for demonstration may include but are not limited to soil permeability, surface temperature, soil moisture, precipitation, leaf area index (LAI) and normalized difference vegetation index (NDVI). Inexact geo-environmental information derived from the satellite remote sensing imagery-grid and expressed by an upper and lower bounds is therefore defined as an integral part of the GIP optimization model. Hence, the purpose of selecting monitoring locations is to locate subset areas that represent the best characteristic of the whole area of interest to meet the cost-effectiveness criteria. A set of interval values reflecting the corresponding inexact geo-environmental information derived from the satellite remote

113

sensing imagery-grid are therefore defined as an integral part of the GIP optimization model. Such synergistic integration weaves a suite of interrelated visionary screening metrics for final site selection throughout the optimization context.

As compared to the traditional approaches, the inclusion of interval numbers (i.e., grey numbers) in a linear-programming based optimization model can augment the capabilities of both statistically-driven and simulation-driven methods. Overall, the use of GIP approach may exhibit a comparative advantage when handling a variety of data streams and information flows collected in all scales and from all fields in relation to statistical output, simulation outcome, and monitoring data for decision making in an uncertain environment. Such an optimization scheme does not have to perform any parameter-based sensitivity analysis for the inclusion of an interval expression of the optimal solution itself can address the systematic uncertainty collectively across all parameters with sensitivity concerns. The following sections introduce the study area in Texas, remote sensing image processing and ground truthing, formulation of the GIP model, and final site screening and selection within the targeted watershed for demonstration.

Study Area

Geographic setting

Choke Canyon Reservoir, which is the largest reservoir in South Texas, was completed in 1982 and filled for the first time in 1987. The reservoir is operated by the City of Corpus Christi. On July 11, 2002, Choke Canyon Reservoir filled to 100% capacity based on inflows is from the Frio River and San Miguel Creek. When full, the lake level is at 220.5 feet above sea level and stores 695,271 acre-feet of water. The normal release is 33 acre-feet of water per day. The crest gates are only used when excessive water, such as during flood events, needs to be released (City

of Corpus Christi, 2003). The Choke Canyon Reservoir Watershed (CCRW), as part of the Nueces River Basin, is composed of several land use/land cover types. Figure 4-1 shows the geographic environments of the CCRW. Farming and livestock husbandry are major land use patterns in the past few decades. The Choke Canyon Reservoir Watershed (CCRW) encompasses 14,200 km2 out of the 43,300 km2 Nueces River Basin. Elevations in the CCRW range from 42 m above sea level near the dam to 740 m at the Edwards Plateau near the divide of the watershed upstream. The farmlands which are mostly located in the middle stream area of in the CCRW is often graded and plowed, and irrigation may change the soil moisture in some seasons periodically. The livestock in south Texas is naturally fed on grass in open areas and ranches mostly located in lower stream area of the CCRW. Mixed land uses in this area introduce complexity of soil moisture distribution.

The availability of clean water to sustain life and human societies in the CCRW is perhaps the most recurrent constraint due to the long-term drought and flood impact and it will remain so for the foreseeable future. In the U.S. nowhere is this more evident than in the arid west and semiarid southwest, where rapid population growth and limited water resources converge to reach near-crisis level during periods of drought. Located in one of the fastest-growing areas in the U.S., the Nueces River Basin provides the opportunity to observe climate and human-induced land-surface changes affecting water availability, water quality, and water use. These attributes reflect the changing relationship between people and water across the "South Texas coastal bend" and make the CCRW a microcosm of contemporary water resource issues and an excellent site to pursue interdisciplinary and integrated hydrologic science, such as hydrological observatory study. Therefore, hydrological monitoring stations or even hydrological observatory must be introduced into the watershed to obtain additional information for

improving the watershed management. These sensors may include but are not limited to precipitation sensors, soil moisture sensors, soil temperature sensors, relative humidity sensors, wind sensors, nutrient sensors, and carbon sensors. They will provide monitoring infrastructure for advanced studies of the hydrology in the context of sustainability.

Geologic features

Landscape in south Texas, however, is intimately tied with the geological structure. To the north, topography strongly influences the hydrology of the watershed. In the upper portion of the watershed, the steep slopes and arid terrain of the Balcones Escarpment rise into the Edwards Plateau. These hills, cliffs, exposed rock, and clay soil, while act as sinks at the onset of a precipitation event, cause rapid runoff during large storm events resulting in flashflood. As the streams cross the Edwards Aquifer Recharge zone, they lose a significant portion of their flow through faults and Karst topography. Downstream of the Balcones fault zone, the landscape tends to flatten as the water flows south and east into the South Texas Brush Country where slopes range from 0-10%. Placement of USGS stream gauges above and below the fault zone helps to quantify the water losses in the fault zone and to provide early warning information of any potential flooding in the downstream areas (see Figure 4-1). Right above the Choke Canyon Reservoir there are two USGS stream gauges measuring the total inflow of the streams that flow into the reservoir. Figure 4-2 is the National Land Cover Data (NLCD) showing the land cover in the watershed (USGS, 2001). It shows the land use patterns in this area include from evergreen forest in the upstream area to cropland and ranges in the middle stream areas, and down to shrub land in the lower portion of the watershed. Table 4-1 complements the description of the NLCD image.

116
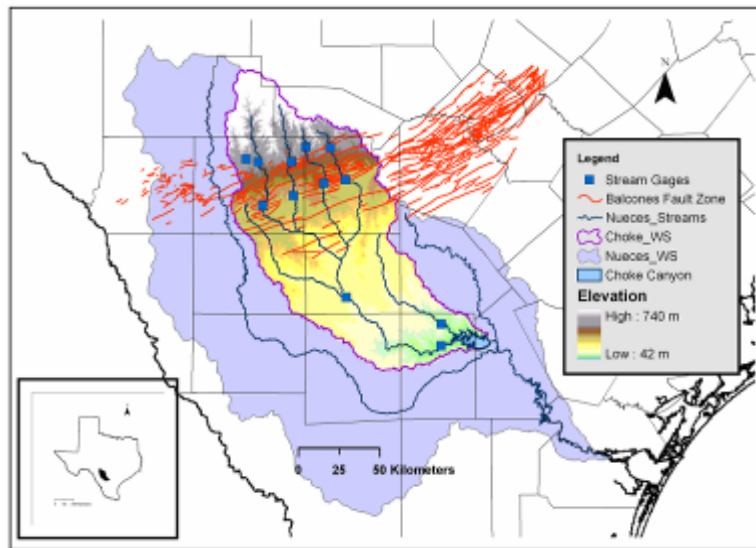
Figure 4-1. Geological features of the Choke Canyon Reservoir
are composed of fault lines along the Edwards Plateau. Below the fault zone the landscape is almost
literally flat with the slope less than 5 percent. This Figure is adopted from Makkeasorn et al (2006).
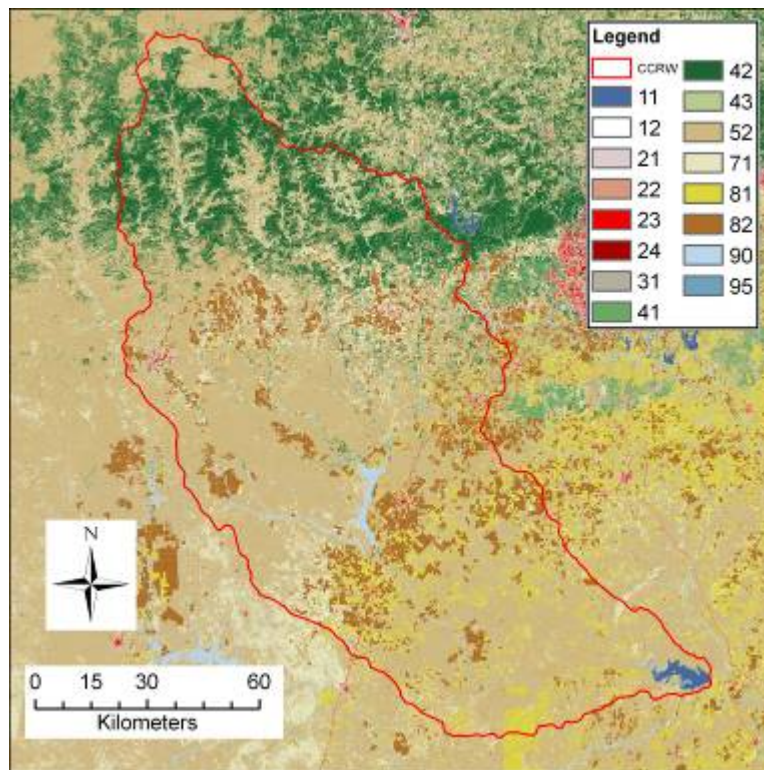
Figure 4-2. Choke Canyon Reservoir Watershed is located in south Texas.
A description of land covers in CCWR is explained in

Table 4-1. Classification system used for National Land Cover Data 2001 (Homer et al., not dated).

Value	Class
11	Open water
12	Perennial Ice/Snow
21	Developed, Open Space
22	Developed, Low Intensity
23	Developed, Medium Intensity
24	Developed, High Intensity
31	Barren Land (Rock/Sand/Clay)
41	Deciduous Forest
42	Evergreen Forest
43	Mixed Forest
52	Shrub/Scrub
71	Grasslands/Herbaceous
81	Pasture/Hay
82	Cultivated Crops
90	Woody Wetlands
95	Emergent Herbaceous Wetlands

According to the historical flow measurements recorded in many decades (USGS, 2005), the hydrological pattern of this watershed comprises two seasons -- wet and dry seasons (see Figure 4-3). The upper portion of the CCRW is not included as part of the study area due to its unique geological structure of bedrock. There are exposed rocks and gravels in some areas, while the others are covered barely by a very thin layer of soil, if any. Water availability has been monitored along the streams in the watershed. However, more information is needed in order to better understand the characteristic of the watershed, and thus to improve the management of water in the Choke Canyon Reservoir.
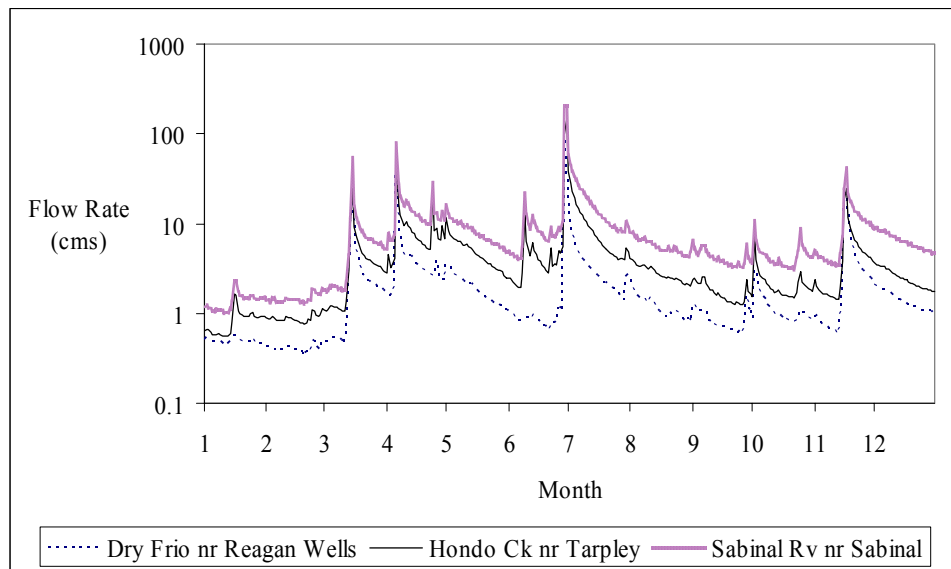
Figure 4-3. A log-plot of mean streamflows measured at the USGS stations that are located at Dry Frio River near Reagan Wells, Hondo Creek at Tarpley, and Sabinal River at Sabinal.

Methodology

Selection criteria and planning scenarios

The most critical step in developing a monitoring program is the clear definition and specification of information needs and monitoring objectives (Timmerman et al., 1997). Information needs and monitoring objectives should be specified to such extent that design criteria for the various elements of the information system can be derived (Timmerman et al., 1997). The siting process of monitoring stations can be seen as a sequence of related screening activities, starting with the definition of spatial information needs and ending with the use of the information product for optimization analysis in an uncertain environment with respect to essential planning scenarios. To build up a set of candidate sites, the study area of interest which is located in the middle and lower portions of the CCRW is divided into five hundred and sixty three cells using 4-km grid. Hydrological, ecological, and geological parameters of each cell (i.e., 4-km by 4-km) were established based on seven criteria. They include surface soil

119

permeability, leaf area index (LAI), normalized difference vegetation index (NDVI), precipitation, fraction-of-photosynthetically active radiation (FPAR), land surface temperature (LST), and surface soil moisture. These parameters that vary to some extent within each cell or grid are prepared to indicate the most sensitive ranges associated with each cell that well correspond to the hydrologic cycle in an uncertain environment. This explains the needs to use a number of interval numbers reflecting the maximum and minimum values of concerned parameters in each cell to delineate the possible change of values associated with them. Aggregating all the varying features of cells for site screening and selection would require employing a special optimization model, the GIP model, which is able to handle multiple data streams with interval numbers defined for all relevant parameters.

In principle, the precipitation would indicate the frequency and extent of the water getting into the watershed. The surface soil permeability would correspond to the process of infiltration and percolation affecting surface and subsurface flows. The evaporation can be realized by using surface temperature and FPAR as surrogate indices. The evapotranspiration would respond to LAI, FPAR, and NDVI collectively. The LST, LAI, and NDVI would correspond to the loss of water through surface soil and vadose zone. To have a more accurate index to address the soil moisture content directly, however, this analysis also counts on RADARSAT-1 SAR imageries acquired in April and September 2004. The standard beam mode in ascending orbits of RADARSAT-1 was selected for this study. These analytical procedures acquiring all the information flows aforementioned have to be done collectively using satellite remote sensing images prior to the formulation of the GIP Model. The target site in a semi-arid watershed, the CCRW, is expected to provide the best realization of hydrological information collectively across all criteria. Planning scenarios can be set up based on differing set of weighting factors

with respect to all seven criteria in decision analysis for site selection. All seven criteria should be jointly considered in decision analysis for drawing up the hydrological monitoring program. Finally, the selective criteria for all grid cells involved in the GIS layers can be defined for site screening and selection as follows:

- The higher the soil permeability, the higher the possibility to install a monitoring station.

- The higher the NDVI and LAI values, the higher the possibility to install a monitoring station.

- The higher the precipitation, the higher the possibility to install a monitoring station.

- The higher the evaporation potential (high value of surface temperature and low value of FPAR), the higher the possibility to install a monitoring station.

- The higher the soil moisture, the higher the possibility to install a monitoring station.

- The closer the riparian buffer zone, the higher the possibility to install a monitoring station.

The planning scenario would involve using the seven criteria for assessment. For simplicity, equal weighting factors associated with all criteria involved are applied for all decision analyses in this study. Overall, the analysis procedure can be summarized by a flowchart as shown in Figure 4-4. It emphasizes the organization of the information flows before feeding them into the GIP model for site screening and selection. Detail discussion of the interface and interaction between these two stages will be shown in the following subsections.

Satellite image processing

MODerate resolution Imaging Spectroradiometer (MODIS) is the instrument onboard Terra and Aqua satellites. Through NASA's Earth Science Enterprise's Program called Earth Observing System (EOS), Terra and Aqua are used to monitor the entire Earth's surface and near-ground atmosphere every 1 to 2 days. Terra orbits around the Earth from north to south passing the equator in the morning, while Aqua travels from south to north across the equator in the afternoon (http://modis.gsfc.nasa.gov/about/). The designed orbit allows Terra and Aqua to work in tandem to observe the same area on Earth in the morning and afternoon. The fundamental purpose of Terra and Aqua is to observe the global changes and processes of the Earth's land surface, oceans, and the lower atmosphere with a fine temporal resolution to capture many Earth's rapid changes. The MODIS frequent flyover complements other space-borne sensors such as Landsat, which has a finer spatial resolution but less frequent return period of 16 days. The 36-band spectral resolution of MODIS (http://modis.gsfc.nasa.gov/about/design.php) also extends the capability over the 6-band spectral resolution of Advanced Very High Resolution Radiometer (AVHRR) (http://noaasis.noaa.gov/NOAASIS/ml/avhrr.html) by offering greater spectral resolution that allows more detailed analyses.
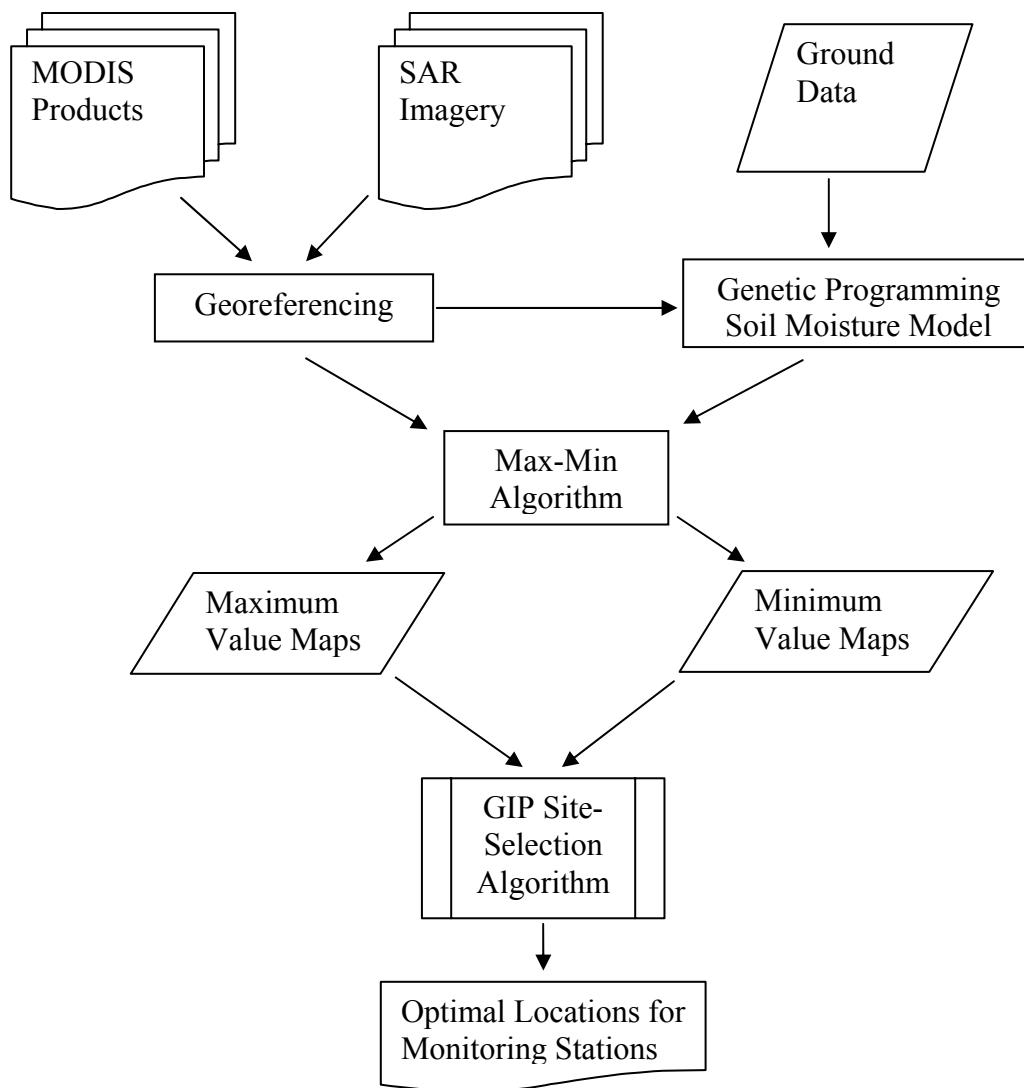
Figure 4-4. Schematic of the site-selection algorithm.
Workflow for the development of site-selection algorithm involves using multiple remote sensing platforms and Genetic Programming Soil Moisture Model.

MODIS instruments are designed to observe electromagnetic spectral from 620nm to 14,385 nm in 36 bands. Three principal spatial resolutions are 250 m, 500 m, and 1000 m at nadir. There are more than thirty five MODIS products available through LP DAAC (2006) for public access. Some products have been post-processed and validated, while the others are in the process of validation. The MODIS Adaptive Processing System (MODAPS) generates MODIS products as gridded output in the Integerized Sinusoidal (ISIN) projection to enhance the

efficient organization and hierarchical management of the data. The MODIS Reprojection Tool (MRT) is MODIS imaging software developed by the South Dakota School of Mines & Technology through LP DAAC to handling the MODIS level-3 data products. The MRT will enable users to perform geographic transformation to the users' preferred coordinate system/cartographic projection, and export the data to other formats other than the HDF-EOS in-house format. The MRT is available for use by all registered users at LP DAAC (2006).

Three MODIS products used in this study are MOD11A2, MOD13A2, and MOD15A2. First, the MOD11A2 contains level 2 Land Surface Temperature (LST) retrieved from Terra and Aqua under clear-sky conditions in day and night time. Hydrologic processes are very sensitive to the surface temperature. The temperature also strongly influences many human activities such as agricultural practices, flood and drought control, or land developments. More details about the MOD11A2 product can be found in Wan (1999). Second, the MOD13A2 is the MODIS Vegetation Index level 2 product. The vegetation indices in the MOD13A2 include NDVI and the enhanced vegetation index (EVI). The NDVI index of MODIS is considered as the extension of NDVI data relative to the AVHRR-derived NDVI. While the AVHRR-NDVI global datasets were retrieved from 1981-1999, the MODIS-NDVI continues to collect the global NDVI providing a long term data collection. The EVI improves VI sensitivity of high biomass regions and reduces the atmosphere influences. Both of the two VIs can extract canopy biophysical parameters effectively (Huete et al., 1999). Yet only NDVI index of MODIS was used in this study. Third, the MOD15A2 is the level-2 MODIS leaf area index (LAI) and fraction of photosynthetically active radiation (FPAR) absorbed by vegetation product. The LAI indicates the number of equivalent layers of leaves vegetation that are present relative to a unit area. FPAR

is the measurement of fraction of available radiation in the specific photosynthetically active wavelengths of the spectrum 0.4 – 0.7 μm that vegetation absorbs (Knyazikhin et al., 1999).

September 2005, January 2006, and April 2006 MODIS images were chosen to represent a year of changes. Each image was reprojected to UTM 14 NAD83 projection using the MRT imaging software as aforementioned. An image of maximum value and an image of minimum value of each MODIS dataset were created by an algorithm to extract the maximum and minimum values of each pixel from the two images of each dataset. The maximum-value image and the minimum-value image represent the maximum and minimum values of data that occurred in September and April 2006, respectively. These datasets represent the diurnal effect to some extent (see Figure 3). The maximum and minimum values of each pixel were thus used as the upper and lower bounds to form the corresponding interval numbers for use in the subsequent GIP analysis. Overall, the post-processed MODIS products provide us with the maximum and minimum datasets of LST, NDVI, LAI, and FPAR at 1km pixel size. The maximum and minimum images were then all imported into the GIS framework in support of GIP modeling analysis for site screening and selection. ESRI® ArcMap™ 9.0 was used for handling all the image processing work. All of the MODIS products used in this study have the native resolution of 1km. However, our calculating grid on the GIS platform are 4 km by 4 km. Zonal Statistics function in the ESRI® ArcMap™ 9.0 was used to extract the highest maximum value and the lowest minimum value of the pixel in each 4-km grid. The highest maximum and the lowest minimum values were adopted for the GIP modeling. Figure 4-5 presents the integrated workflow of the MODIS data preparations collectedly.
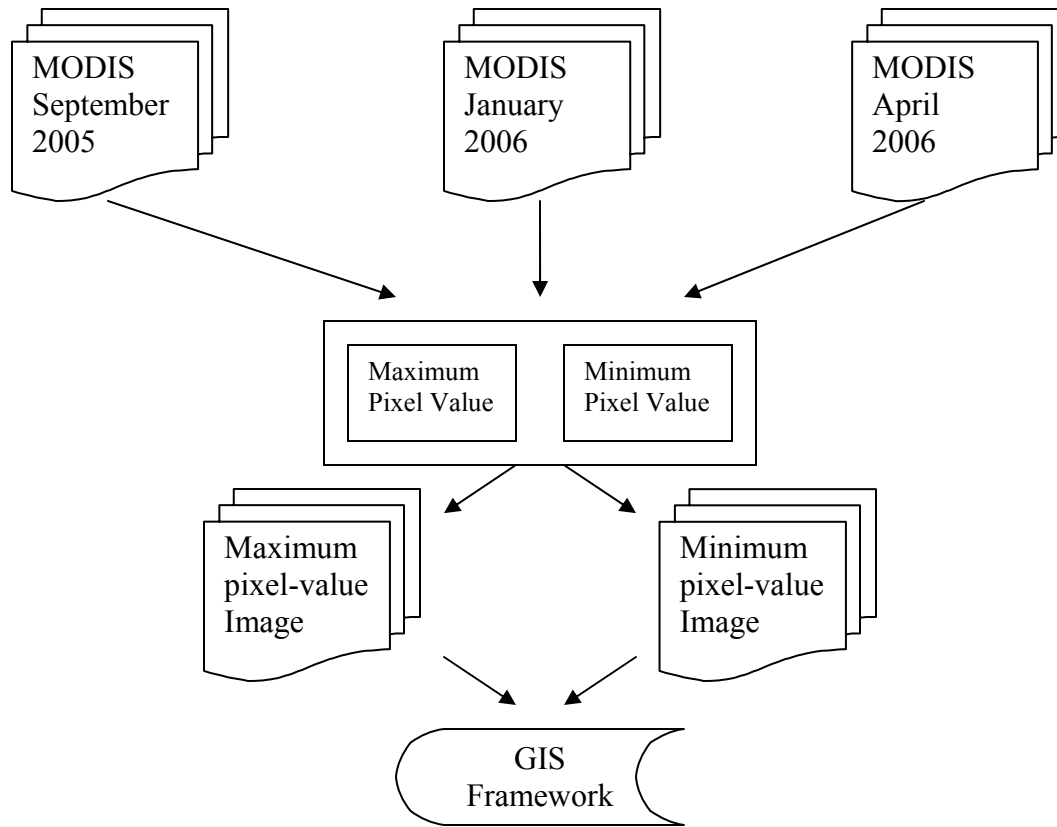
Figure 4-5. Min-max extraction algorithm.
Flowchart shows the post-processing procedures of MODIS data to extract the one-year representative of maximum and minimum pixel values.

Formulation of grey integer programming model

In a multi-criteria decision-making process, uncertainties and risks may arise from both the imprecision of parameter values and the vagueness among inexact decision goals and constraints. While the former can be expressed as grey information (Huang and Moore, 1993), the latter may be related to fuzzy decision analysis (Zadeh, 1965; Bellman and Zadeh, 1970). Applications of fuzzy sets have since flourished, varying from solving the inherent problem of uncertainty in natural resources assessment, to accommodating vagueness and complexities when modelling the environmental systems for decision support. There is also increasing number

126

of applications in environmental risk analyses based on the grey systems theory (Chang and Wang, 1995). In a grey system, all sub-systems are divided into three categories, i.e. white, grey, and black system. The white systems show clear messages completely while the black systems show only unknown information, and the grey systems show partially known and unknown (Wu and Chang, Jan. 2002; Huang and Huang, 1994). TO deal with such grey message in association with differing planning scenario, the parameter value can be defined as a closed interval number with upper and lower limits, $[a^-, a^+]$ accordingly, while the imprecision of its intrinsic characteristics remains. As time progresses and the grey message turns out to be clearer, the range of interval might be narrowed down. The integration between fuzzy set theory and grey system theory can also be seen in the literature (Huang and Huang, 1994). The spectrum of applications rapidly covers various problems of systems analysis that include the use of grey linear programming (GLP), grey fuzzy linear programming (GFLP), grey fuzzy dynamic programming (GFDP), and grey integer programming (GIP) approaches (Huang, et al., 1992, 1993, 1994, 1995a 1995b). Grey multi-objective programming was also applied for solving solid waste and watershed management issues (Chang, et al., 1996a, 1996b, and 1997). The characterization of grey compromise programming and its stability analysis have been fully discussed by Chang, et al. (1999). Grey input-output analysis was developed in the early 2000s (Wu and Chang, 2002). These prior efforts provide this study with a firm tool to apply directly for processing a suite of satellite remote sensing imageries with grey implications, which is the GIP.

The site selection performed by using the GIP model requires having seven matrices (i.e. criterion) in the constraint set while considering all grid cells as possible candidate sites in the screening process. The solution procedure of GIP would follow the algorithm developed by

127

Huang and Moore (1993), Chang, et al. (1996a, 1996b), and Chang et al. (1997), in which the GIP model needs to be solved by two submodels in series or in parallel. Hence, to process a suite of grid values with interval expressions across all cells, the GIP model can be formulated as below in which $I_{i,j}$ are the decision variables (the binary integer variables) in this study; it is equal to 1 when cell i, j th is selected by the screening process; otherwise zero. Subscripts i and j are the cell identity of row i and column j. The objective function is formulated as below.

$$Max \quad F^{\pm} = \sum_{\substack{i=1 \\ \{i,j\} \in Z1}}^{n} \sum_{j=1}^{m} (w_1 a_{i,j}^+ + w_2 b_{i,j}^+ + w_3 c_{i,j}^+ + w_4 d_{ij}^+ + w_5 e_{ij}^+ + w_6 f_{ij}^+) I_{ij}^+$$

$$+ \sum_{\substack{i=1 \\ \{i,j\} \in Z1}}^{n} \sum_{j=1}^{m} (w_1 a_{i,j}^- + w_2 b_{i,j}^- + w_3 c_{i,j}^- + w_4 d_{ij}^- + w_5 e_{ij}^- + w_6 f_{ij}^-) I_{ij}^-$$

(4-1)

where w_1, w_2, w_3, w_4, w_5, and w_6 are the weighting factors

$a_{i,j}^{\pm}$ is soil permeability in the grid I_{ij} which is expressed as $\left[a_{i,j}^-, a_{i,j}^+ \right]$

$b_{i,j}^{\pm}$ is NDVI values in the grid I_{ij}, which is expressed as $\left[b_{i,j}^-, b_{i,j}^+ \right]$

$c_{i,j}^{\pm}$ is precipitation rates in the grid I_{ij}, which is expressed as $\left[c_{i,j}^-, c_{i,j}^+ \right]$

$d_{i,j}^{\pm}$ is evaporation rates in the grid I_{ij}, which is expressed as $\left[d_{i,j}^-, d_{i,j}^+ \right]$

$e_{i,j}^{\pm}$ is soil moisture in the grid I_{ij}, which is expressed as $\left[e_{i,j}^-, e_{i,j}^+ \right]$

$f_{i,j}^{\pm}$ is the distance away from riparian buffers in the grid I_{ij}, which is expressed as $\left[f_{i,j}^-, f_{i,j}^+ \right]$

Z1 is the set of all cells inside the area of interest.

Subject to:

1) Decision consistency constraints: This constraint ensures the decision of the upper bound model and the lower bound in the two submodels should be consistently logical in response to the GIP algorithm (Chang et al. 1997). This would allow a grid to be selected by one submodel with the upper bound and deleted by the other submodel with the lower bound according to the flexibility in GIP (Huang et al. 1995). In site screening and selection, such mathematical formulation ensures that bigger variation over the upper and lower bounds of a particular parameter may end up having the cell to be included as a candidate site due to its upper bound even though its lower bound is very small.

$$I_{ij}^- \leq I_{ij}^+ \qquad \text{for all } I_{ij}^\pm \tag{4-2}$$

2) Budget constraint: This constraint specifies the total number of stations allowed to be installed in the study area. Because of the inclusion of the first constraint, it is not necessary to ensure the satisfaction of this constraint with respect to the lower bound $I_{i,j}^-$ in this case. Otherwise, the constraint would become redundant in the optimization context.

$$\sum_{i=1}^{n}\sum_{j=1}^{m} I_{i,j}^+ \leq N \quad \text{for all } I_{ij}^\pm \in Z1 \tag{4-3}$$

where n and m are the upper bound of the subscripts I and j; N is the total number of monitoring sites in the system.

3) Neighboring cells constraints: This constraint prevents any neighborhood cells from being selected in the study area simultaneously in each individual screening scenario. Again, it is not necessary to ensure the satisfaction of this constraint with respect to the lower bound $I_{i,j}^-$ in this case because of the inclusion of equation (2). Otherwise, the constraint would become redundant in the optimization context. This implies that, with the inclusion of equation (2) in the site

129

screening procedure, the value of $I_{i,j}^-$ should be limited to zero once its corresponding upper

bound $I_{i,j}^+$ is set to zero.

$$I_{i,j}^+ + I_{i-1,j}^+ + I_{i+1,j}^+ + I_{i-1,j-1}^+ + I_{i,j-1}^+ + I_{i+1,j-1}^+ + I_{i-1,j+1}^+ + I_{i,j+1}^+ + I_{i+1,j+1}^+ \leq 1 \quad \text{for all } I_{ij}^\pm \in Z1$$

(4-4)

4) Boundary constraint: This constraint specifies the value of 1 in all boundary cells to close the

algorithm. This technical setting is solely prepared in association with equation(4-4) in order to

make all the constraints formulated in response to equation (4-4) effective.

$$I_{ij}^\pm = 1 \quad \text{for all } I_{ij}^\pm \in Z2 \tag{4-5}$$

where Z2 is the set of boundary cells located outside of the area of interest neighboring to all the

cells in Z1 in the watershed.

Data Synthesis and Analysis

Ground data

Soil moisture is a good indicator to addressing the health of agriculture, hydrology and

meteorology on the Earth's surface (Salgado et al. 2001; Baghdadi et al. 2002; Guswa et al.

2002). The moisture content of soil is the primary indicator for the planning and management of

crops and farms, including the time of seeding, fertilizer, pesticide application, irrigation and

harvest opportunities (Salgado et al. 2001). Surface and subsurface runoff depend heavily on the

moisture content (Wilson et al. 2003). Surface soil moisture estimation has been extensively

developed over the past two decades. Many studies have investigated spatial soil moisture

measurement at 0-30 centimeter depth (Grayson et al. 1997; Western and Grayson 1998, as cited

130

in Wilson et al. 2003). The spatial distribution of surface soil moisture content is limited because of the point measurements over the field scale. One of the most well-known methods of soil moisture measurement in field measurement is time-domain reflectometry (TDR). It is reliable and portable method for collecting non-destructive soil moisture values in many studies (Topp et al. 1980; Roth et al. 1992; Grayson et al. 1999; Walker et al. 2001; Western et al. 2001, as cited in Wilson et al. 2003). By measuring soil moisture at many points within an area, the spatial distribution of soil moisture can be obtained. This method is not practical for soil moisture measuring over the basin-wide area. It is time consuming and time-dependent causing difficulty in evaluation of soil moisture of spatial patterns. Because of the large time-span between the first and the last soil moisture measurements in the field, the spatial patterns of soil moisture within a basin-wide field are already altered. Advances in remote sensing have been studied for their potential to measuring surface soil moisture content. Synthetic Aperture Radar (SAR) is a technology of active microwave remote sensing that can map soil moisture over large areas at high resolution (tens of meters) (Lacoul et al. (n.d.); Olmsted 1993; Romshoo et al. 1999; Kwok et al. 2000; Moran et al. 2000; Salgado et al. 2001; Baghdadi et al. 2002; Wilson et al. 2003). In order to use the SAR technology in soil moisture monitoring applications, knowledge of hydrological characteristics of the given area must be well studied and defined (Baghdadi et al. 2002). Ground observation of soil moisture content must be obtained by monitoring several points of surface soil moisture within a large area against SAR sensor data to seek out correlations between them (Moran et al. 2000; Salgado et al. 2001; Wilson et al. 2003).

For surface soil moisture sampling, a total of 109 sites, 56 sites and 53 sites were investigated on August 1st, 7th, and 14th, 2003, respectively. The Field Scout™ TDR 300 Soil Moisture Meter (Spectrum Technologies, Inc., not dated) was used to measure the volumetric

moisture content (VMC) at the top 5 cm of surface soil. It has the accuracy of ±3.0% volumetric water content in the measuring range of zero percent to saturation (typically 50% VMC). The VMC is the ratio of the volume of water to the total soil volume. At saturation, the VMC will equal to the percentage of pore space of the totally dry soil. The TDR 300 sensor rods used in our measurements are 12 cm in length. The total length of the TDR 300 when it is ready to use is approximately 110 cm. The way we inserted the probe into the target soil was to angle the TDR 300 probe at about 25° to the ground by holding the probe's handles right above the knees (about 50 cm above the ground). This effort verified the RADARSAT-1 SAR measurements of soil moisture in a companion study that can be adopted directly by this study (Makkeasorn et al. 2006).

Besides, to find out the surface soil permeability, ninety-two sites of ground data for soil texture were investigated on June 13, 2003 in the CCRW. At least 32 groups of State Soil Geographic Databases (STATSGO) soil were found mainly in the watershed. The STATSGO was used to address permeability of soil across the study area. Forest and hills are mainly on the upper part of the watershed. Agriculture, Texas shrub lands and grasslands were found around the middle of the watershed. Shrub lands and grasslands were also found around the lower part of the watershed. Figure 4-6 shows the ground data sites presented as red dots fro the verification of soil texture, which can be used later on for soil permeability study.
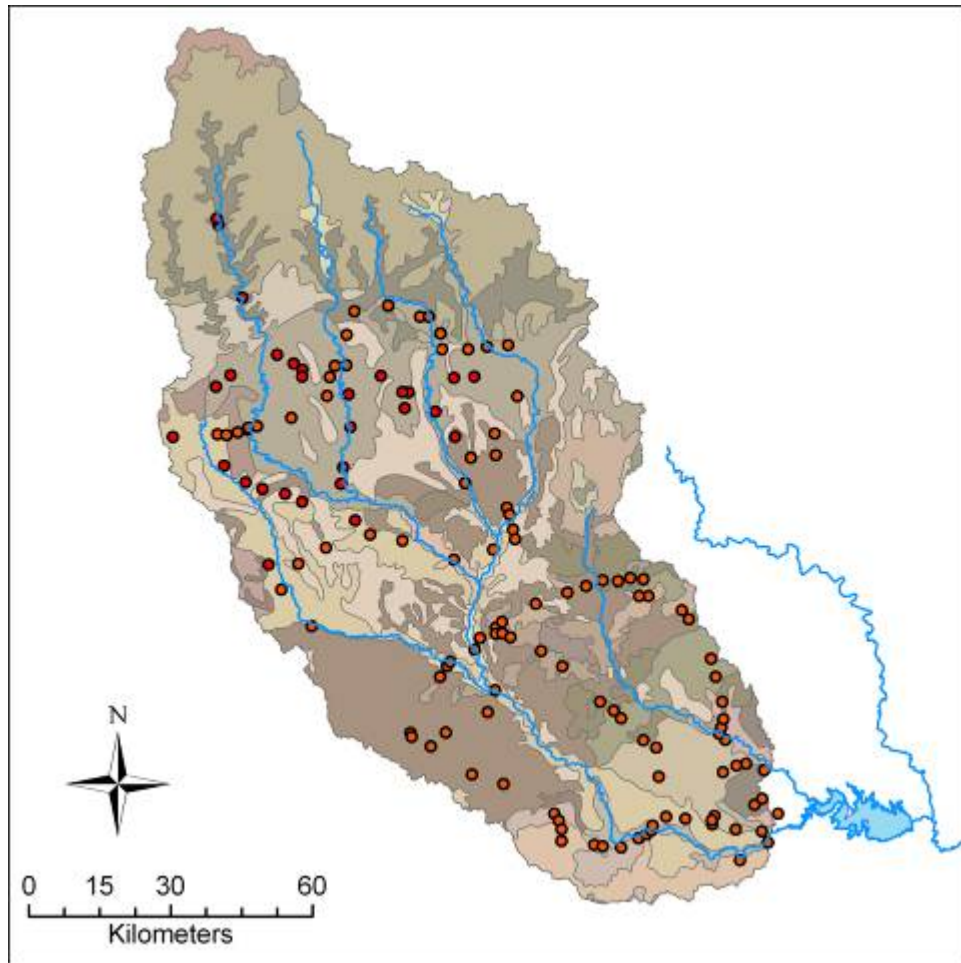
Figure 4-6. Soil sampling sites
Ninety-two sites of investigation for soil textures and land cover are shown in red dots.

The matrix of surface soil permeability

The soils may change over a long period of time (spatially dependent); however, the time scale is very large (thousands of years) so that the change is considered relatively insignificant to this study. With the clear soil texture information, verified by our field efforts, the surface soil permeability can be obtained from a modified State Soil Geographic Databases (STATSGO) and the multi-layer soil characteristics datasets from the conterminous United States (CONUS-SOIL). The estimated permeability rate is related to the saturated hydraulic conductivity of the soil (ESSC, 2002). Ranges of permeability values can be acquired in these databases so that the

variations of the soil permeability can be identified. The algorithm in GIP may therefore discern the optimal locations with respect to the smallest minimum and the largest maximum permeability values. The permeability values of soil groups can also be obtained from the Earth System Science Center, Penn State University for comparison (ESSC, 2002). They determined the mean permeability for each of 11 standard layers for each map unit of each state using datasets from STATSGO Comp and Layer tables. The standard layers were introduced because of the wide variation in number, thickness, and depth between top and bottom of soil layer in the STATSGO datasets.

The smallest minimum and the largest maximum permeability in each soil group can also be identified from the STATSGO's Comp table. By joining the soil ID in the STATSGO map unit with the permeability table in the GIS framework, a map unit with permeability feature was obtained in our study. The featured map unit was then converted to a raster image of soil permeability embedded in each pixel using "*Zonal Statistics* → *Convert to Raster*" feature designed in ESRI® ArcMap™ 8.3. Average permeability in each cell was calculated using the Zonal Statistics in ESRI® ArcMap™ 8.3 software. Figure 4-7 shows the largest maximum and the smallest minimum permeability values spatially averaged in each cell. The darker blue color indicates the higher value of permeability identified. The permeability ranges from 0 to 500 mm/hr. Even though it is impractical to consider the zero value of the soil permeability over such a large area, we did not rule them out in the calculations.
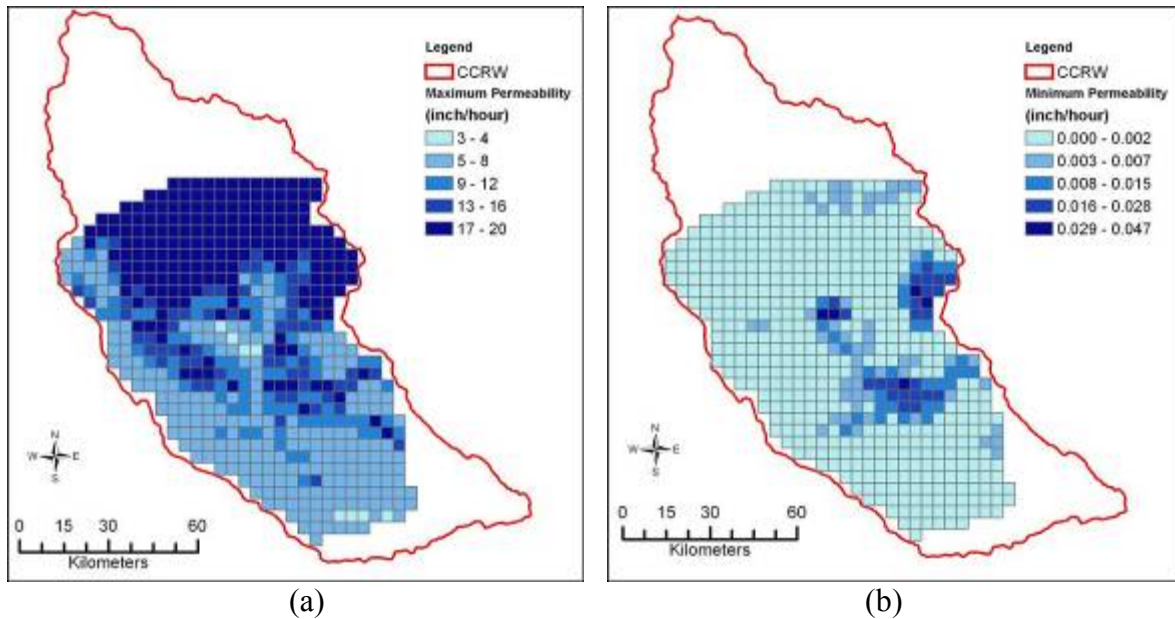
Figure 4-7. The largest maximum and the smallest minimum permeability maps are shown above in (a) and (b), respectively.
The largest maximum permeability ranges from 3 to 20 inches per hour (75 – 500 mm/hr). The smallest minimum permeability ranges from 0.0 to 0.047 inch per hour (0 – 1.2 mm/hr).

The matrix of NDVI

The summary of MODIS Vegetation Index (MOD13) is reported by Huete et al. (1999). In the Global Land Biosphere Data and Resource website (1999) the NDVI is defined as the proportion of photosynthetically absorbed radiation, which is calculated from atmospherically corrected reflectance from the visible red (VR) and near-infrared (NIR) bands. For the MODIS NDVI data, the visible red is the reflectance in the wavelengths of 0.58-0.68 μm and the near-infrared (near IR) is the reflectance in the reflective infrared wavelengths of 0.725 to 1.1 μm. The NDVI can be calculated by the following equation.

$$\text{NDVI} = \frac{NIR - VR}{NIR + VR}$$

(4-6)

The rationale behind this equation is that the visible red is just a part of the spectrum where chlorophyll causes considerable absorption of incoming radiation (green color for the

135

human eyes), and the near IR is in a spectral region where spongy mesophyll leaf structure causes significant reflectance (Tucker 1979, Jackson et al. 1983; Tucker et al. 1991, as cited by Global Land Biosphere Data and Resources, 1999). The ratioing properties of the NDVI offer advantages due to the cancellation of a large proportion of signal variations attributed to noise, atmospheric conditions that change with sun angles, topography, clouds and shadow. The atmosphere reduces the contrast of reflectance between the red and NIR. The net result is a drop in the NDVI signal and an underestimation of the amount of vegetation at the surface. The MODIS NDVI data have been corrected to compensate for the atmospheric effects, the sun angles, the canopy background contamination, the NDVI saturation, and the canopy structural effects, as explained in details in Huete et al. (1999). The MODIS NDVI data used in this study were acquired in September 2005, January and April 2006. The pixel size of the imagery is 1 km. They were spatially averaged (arithmetic mean) into the 4-km grid to fir in our GIS grid system.

The NDVI may be used in the model to imply the degree of evapotranspiration of the hydrologic cycle of the watershed. It's known that the moisture budgets can be estimated with long-term evapotranspiration data over extended areas using satellite remotely sensed data and limited surface meteorological properties (Song et al, 2000). They suggested that the limited resolution of the NDVI data resulted in deviations of the model from the observations in the fields of which surface vegetative conditions were notably different than the surrounding fields. Their conclusion informs us that the evapotranspiration data over distances of tens kilometers can be modeled without significant bias. Our targeted watershed is consistent in topography, which means there are not many dramatic changes of land covers over a large, contiguous area. Therefore, the NDVI used in this study for the selection of monitoring locations should not introduce significant bias into the GIP model.

The maps of maximum and minimum NDVI values are presented in Figure 4-8. These figures show the spatially average values of the NDVI within each 4-km grid. In the NDVI Figures the darker shades indicate higher NDVI values. The NDVI images are in good agreement with the land cover characteristic (see Figure 4-2), where the presences of evergreen forest can be clearly identified at the top and the central portion along the Frio river of the CCRW. In summary, the largest maximum NDVI value was 0.51 and the smallest maximum NDVI value was 0.18, while the largest minimum NDVI value was 0.42 and the smallest minimum NDVI value was 0.05.

(a)   (b)
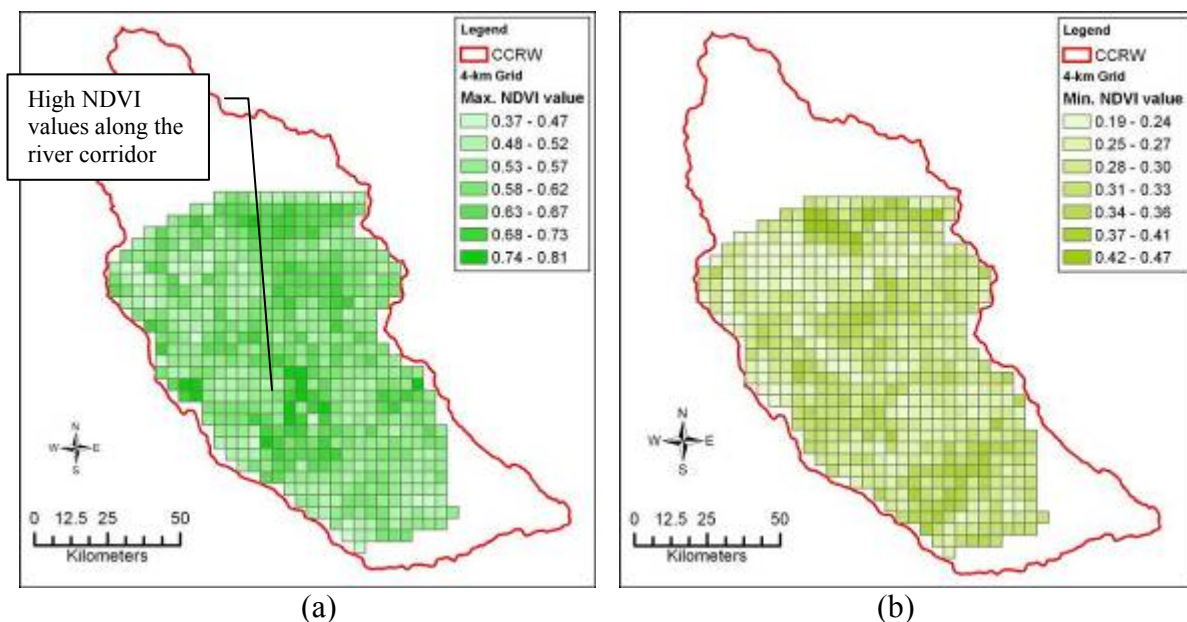
Figure 4-8. The largest maximum and the smallest minimum NDVI maps
of the CCRW between 2005 and 2006 are shown above in (a) and (b), respectively. High NDVI values
are identified along the river corridor.

The matrix of leaf area index and fraction of photosynthetically active radiation

The MOD15A2 is the level-2 Terra and Aqua MODISs' Leaf Area Index (LAI) and Fraction of Photosynthetically Active Radiation absorbed by vegetation (FPAR). The imagery has its native resolution of 1 km. LAI is defined as the one-side leaf area per unit ground area, which is a very important structural property of vegetation canopy. The fraction of available radiation in the photosynthetically active wavelengths, between 400 to 700 nm, that a canopy absorbs is measured as the FPAR (Knyazikhin et al., 1999). The energy (e.g. heat) and mass (e.g. carbon dioxide) are normally absorbed by the vegetation. The LAI and FPAR are the biophysical indices that describe the relations between the canopy and the exchanges of mass and energy. Both LAI and FPAR may reflect some of the ecological and hydrological parameters such as photosynthesis, evapotranspiration, and process of carbon transfer (Myneni et al., 2003). The MOD15A2 data are derived from MODIS surface reflectance (MOD09), the land covers (MOD12), and surface characteristics of land such as vegetation covers. Seven spectral bands of data were used for applications, including 470 nm, 555 nm, 648 nm, 858 nm, 1240 nm, and 2130 nm (Knyazikhin et al., 1999). Details on the derivation techniques and algorithms can be found in the Algorithm Theoretical Basis Document (ATBD) (Knyazikhin et al., 1999).

The MOD15A2 datasets used in this study were acquired in September 2005, January and April 2006. The datasets were downloaded from the LP DAAC (2006). The three data images were reprojected onto the UTM 14N NAD83 map projection using the MRT. The MAX-MIN algorithm was used to create a maximum pixel-value image and a minimum pixel-value image in this study to identify the interval for GIP modeling (see Figure 4-5). The valid range of the LAI and FPAR are in between 0 and 100 normally as shown in Figure 4-9 and Figure 4-10. Figure 4-

9 summarizes all the maximum LAI image and minimum LAI images. Figure 4-10 summarizes the FPAR images on the other hand. High FPAR pixel values were often present along the rivers where the probability of the presence of riparian vegetation was dramatically high, as it can clearly be seen in Figure 4-2. The FPAR images are in good agreement with the LAI images though, which is a reasonable outcome in ecohydrology. Both LAI and FPAR are considered as the surrogate indices of evapotranspiration and heat exchange rate in the aspect of ecohydrology.

(a)   (b)
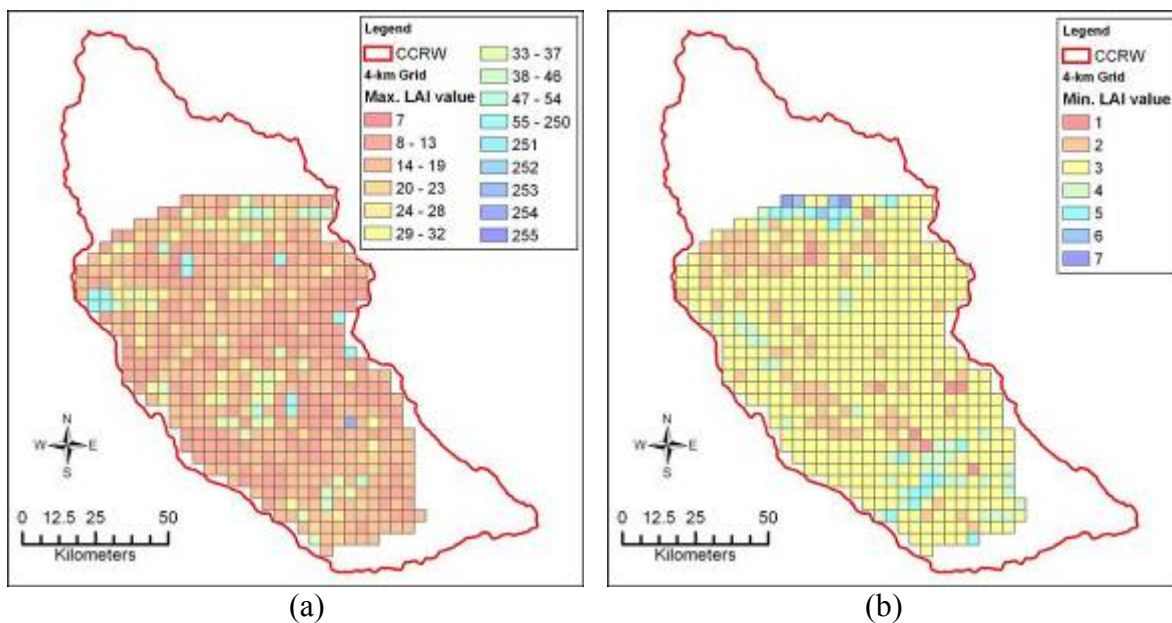
Figure 4-9. The largest maximum and the smallest minimum LAI maps
are shown above in (a) and (b), respectively. The largest maximum LAI values range from 7 to 55, while
the smallest minimum LAI values range from 1 to 7.
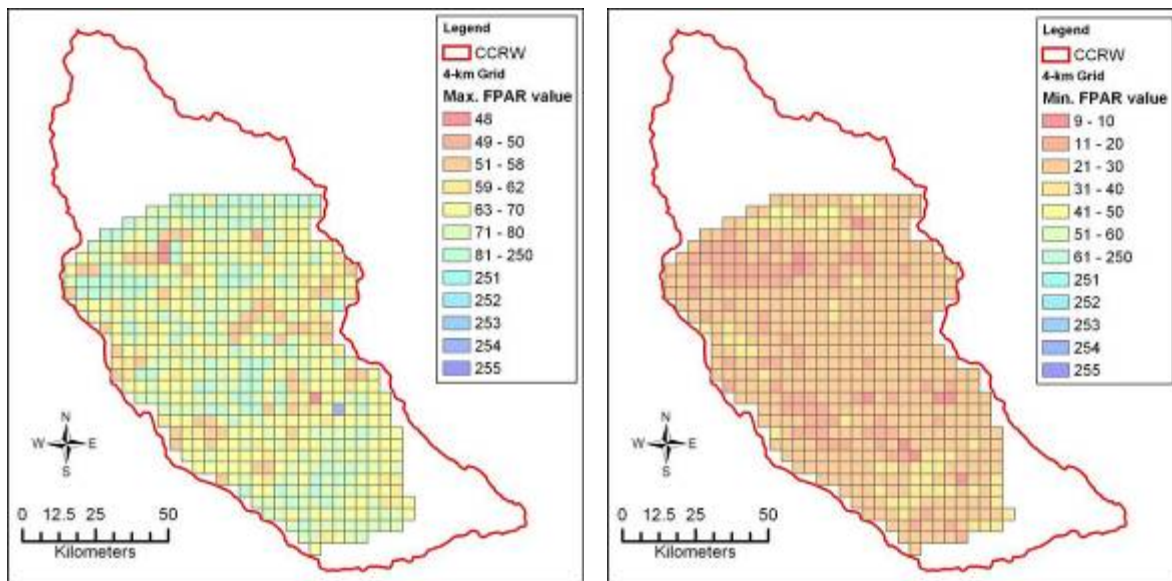
| (a) | (b) |

Figure 4-10. The largest maximum and the smallest minimum FPAR maps
are shown above in (a) and (b), respectively. The largest maximum PFAR values range between 48 and
80. The smallest minimum FPAR values range between 9 and 60.

Table 4-2. Non-terrestrial value index and fill value index are assigned by the MOD15 product team
(http://edcdaac.usgs.gov/modis/mod15a2v4.asp).

LAI and FPAR Value Index	Description
249	Landcover assigned as unclassified/not able to determine
250	Landcover assigned as urban/built-up
251	Landcover assigned as permanent wetlands/inundated marshland
252	Landcover assigned as perennial snow, ice
253	Landcover assigned as barren, sparse vegetation (rock, tundra, desert)
254	Landcover assigned as perennial salt or inland fresh water

The matrix of land-surface temperature (LST)

The per-pixel temperature of land surface in Kelvin is the MODIS product (MOD11). The MODIS data inputs used to derive the MOD11 product include geolocation, radiance, cloud masking, atmospheric temperature, water vapor, snow, land cover, and view angle information (Wan, 1999). MODIS observations in bands 20, 22, 23, 29, 31, 32, and 33 are used to generate the MOD11A2 products. The LST datasets used in this study were retrieved from MODIS data in September 2005, January and April 2006 over the CCRW of Texas. The MOD11 data are provided by the USGS and are available for download from the LP DAAC (2006).

The maximum and minimum LST images were derived using the Max-Min algorithm. The highest maximum temperature was 325 Kelvin and the lowest minimum temperature was 283 Kelvin. High LST values were found mostly in the agricultural areas where baresoil may expose directly to the sunlight. In contrast, the low values of LST were observed mostly along the forest such as the riparian vegetation along the rivers and the Edwards Balcony area above the groundwater recharge zone. The variations of low LST ranges are smaller than the variations of the high LST ranges. The largest variations of LST were noted within mainly the agricultural areas via comparing Figure 4-11 (a) and (b).
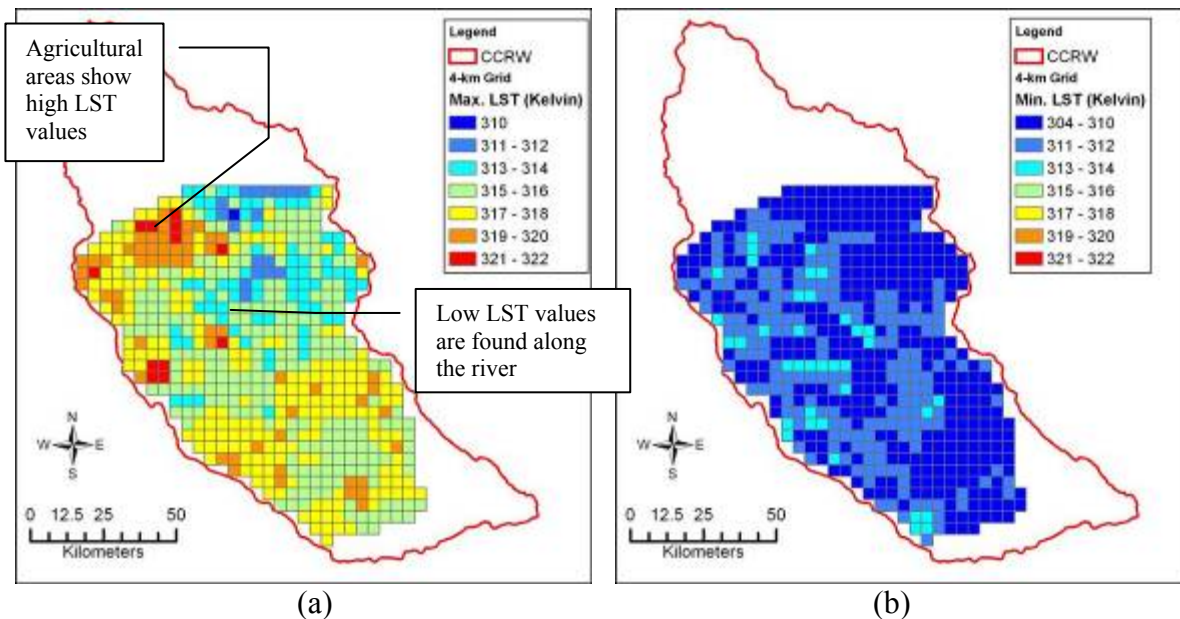
Figure 4-11. The largest maximum and the smallest minimum LST maps
are shown above in (a) and (b), respectively. High values of LST are found mainly in the agricultural
areas, while the low values of LST are found mostly along the rivers as identified above.

The matrix of average total monthly precipitation

Precipitation directly influences the hydrological characteristics of the watershed. Total monthly precipitation data are obtained from PRISM Group (2006). The Oregon Climate Service (OCS) PRISM Group provides access to the highest-quality spatial climate data sets currently available as a public service. The datasets were developed through the PRISM Group funded partly by the USDA Natural Resources Conservation Service, USDA Forest Service (USFS), NOAA Office of Global Programs, National Climate Data Center (NCDC), National Aeronautics and Space Administration (NASA), National Center for Atmospheric Research (NCAR), and the Oregon Climate Service at Oregon State University. Three Total Monthly Precipitation datasets were retrieved including May 2005, September 2005, and January 2006 datasets. The April 2006 dataset was not available at the time of data retrieval, therefore, the May2005 data was used instead. Another reason that May 2005 dataset is chosen over the April 2005 dataset is because

the historical precipitation data collected from 1971 to 2000 indicate that the precipitation occurrences in San Antonio, Texas in the month of May is the greatest (NCDC, 2006). Hence, the September 2005 and January 2006 data were used together with the May 2005 data to maintain the temporal coverage of yearly data and to keep the consistency of the temporal window with the MODIS data. There are about 1,300 data points of precipitation in Texas obtained from NEXRAD grid. Using the concept of interval expression similar to the other practices we had in this study provides the minimum and maximum values associated with each of the 4-km2 cell. The maximum and minimum total monthly precipitation maps are presented in Figure 4-12(a) and (b), respectively. The largest maximum total monthly rain was found 177 mm around the center of the CCRW where the largest riparian forest of the CCRW exists (compare Figure 4-2 and Figure 4-12).

(a)   (b)
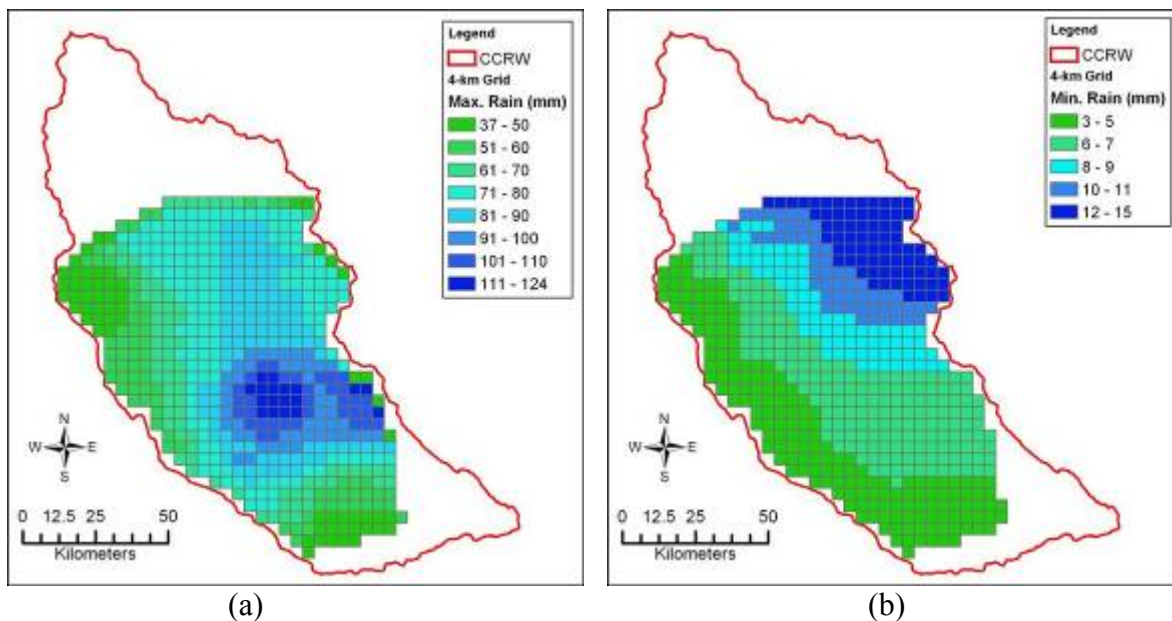
Figure 4-12. The largest maximum and the smallest minimum total monthly precipitation maps are shown above in (a) and (b), respectively. More precipitations are identified around the main riparian forest at center of the CCRW (see also Figure 4-2).

The matrix of soil moisture

Two RADARSAT-1 SAR imageries were acquired in April and September 2004. The estimation of the volumetric water content (VMC) of surface soil based on these two acquisition is derived from the RADASAT-1 SAR imageries and other relevant data, including NDVI, slope, aspect of slope, and soil permeability (Ulaby, 1974; Olmsted, 1993; Dubois et al., 1995; Moran et al., 2000; Baghdadi et al., 2002; Wilson et al., 2003; and Glenn and Carr, 2004). Groudtruth data were used to calibrate and validate the soil moisture estimation model, which was created via using Genetic Programming (GP) technique. The GP-based nonlinear function seeks the best solution through a numerical evolutionary process. The result shows a highly nonlinear structure in regression regime that exhibits strong correlations statistically between the model and the ground data. Two soil moisture maps were cited for this analysis from a companion study (Makkeasorn et al., 2006). The surface soil moisture derived based on SAR and used in this study can be directly drawn from a companion study though (Makkeasorn et al., 2006).

The estimated soil moisture images are shown in Figure 4-13. The value of the pixels is in percent of volumetric water content (% vwc). The Figures show high values of water content are identified around the Frio River at the center of the CCRW. Note that the upper area of the CCRW above the fault zone is not included in the soil moisture model due to the effects of steep slope that cause false interpretations based on RADARSAT-1 SAR data. Also, the lack of data coverage at the bottom of the CCRW did not permit the full coverage of soil moisture estimation in the CCRW.

(a)   (b)
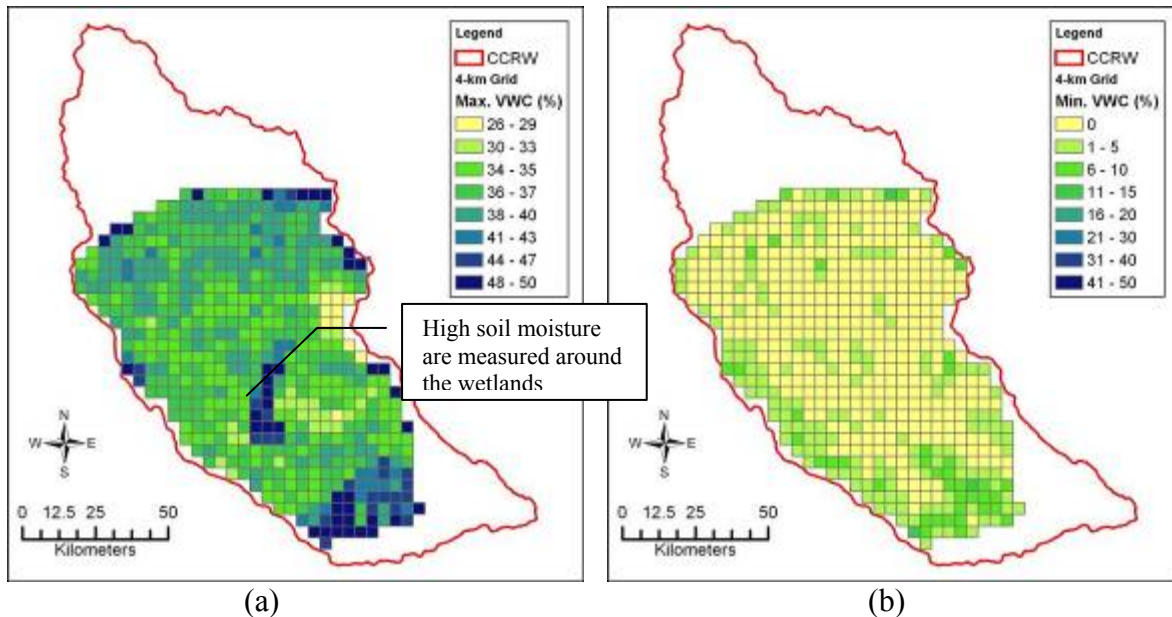
Figure 4-13. The largest maximum and the smallest minimum soil moisture maps
acquired in 2004 are shown above in (a) and (b), respectively. High values of soil moisture are identified
around the main wetland in the center and the lower areas of the watershed (see also Figure 4-2).

Results and Discussion

The study area is divided into five hundred and sixty three (563) cells. Each 2-km x 2-km

grid cell contains between 0 and 4 km of land area in the watershed. Study foci discussed range

from extremely small (leaf surfaces) to extremely to very large (watershed soil property).

Parameters or coefficients of concerns are applied to each cell to address the cells' status,

quality, or characteristics in terms of soil permeability, precipitation, soil moisture, LAI, FPAR,

NDVI, and LST. All of the coefficients are rescaled from zero to one to make them comparable.

The coefficient matrices in the designated GIP optimization model therefore aid in the final

identification of the most suitable locations for hydrological monitoring stations. By running the

GIP model in a software package of Linear, INteractive, and Discrete Optimizer (LINDO) (i.e.,

version 6.1) (Lindo Systems Inc., 2006), seven location preferences are placed in rank order from

one to seven, as shown in Figure 4-14 in which the number in the box associated with each red

145

cell indicates the corresponding rank. Even though we only require three candidates to stand out of the screening process, seven locations shown in red cells are presented because some targets may not be accessible or usable in the real world environment. The locations with higher priority are very well distributed over the watershed. They cover the study region roughly from top to bottom and left to right. Most of the selected locations are located at those spots nearby the rivers, which is considerably sensitive to all hydrologic events.
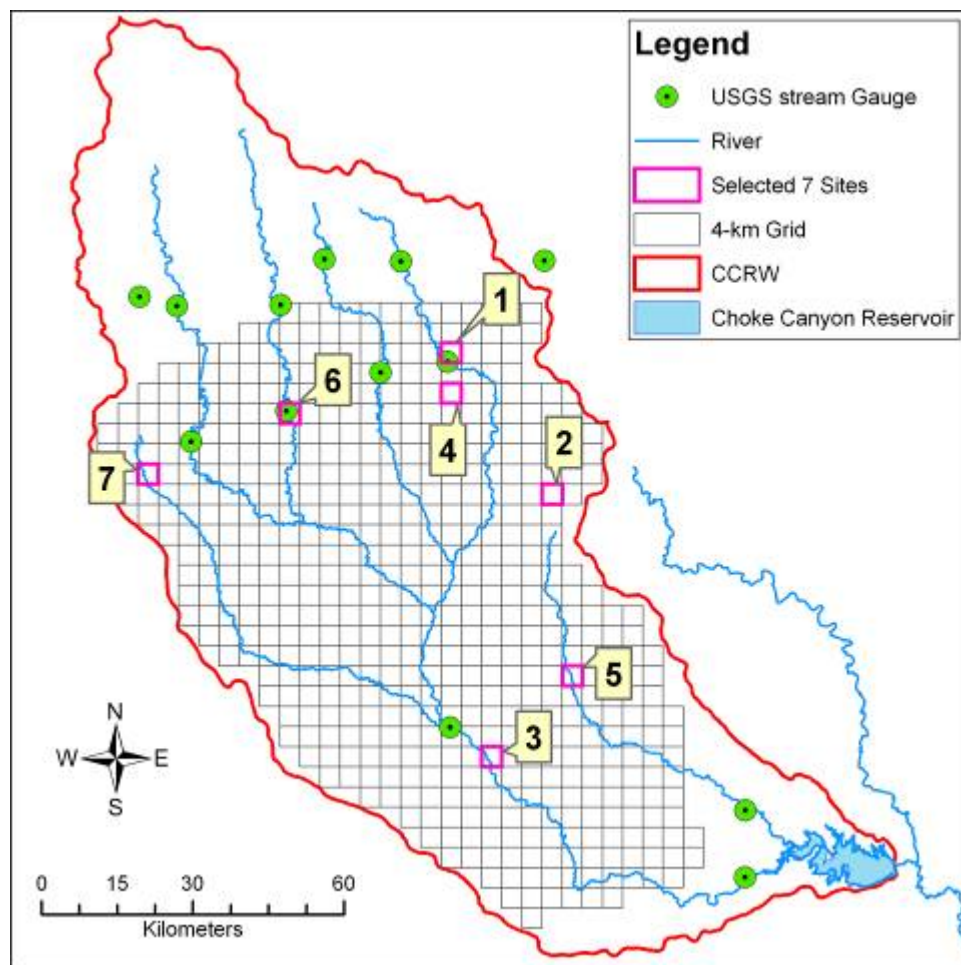
Figure 4-14. Resulting locations selected from the grey integer programming
Seven (7) cells are chosen by the GIP Site-Selection Algorithm. Two of which (#1 and #6) are the cells that contain the USGS stream gages below the fault zone. Cell# 2, 5, and 7 are selected where the USGS stream gage is not present.

Two of the selected cells (#1 and #6) have the USGS stream gage located below the recharge zone. Yet the selected cells #5 and #7 are located on rivers that do not have a USGS stream gage. Cell #2 is also placed in the gap between USGS stream gages. Cell #3 is located below the USGS stream gage where the four rivers converge into one. The pattern of such selections complements the existing USGS stations. Overall five of the seven selections were chosen in the vicinity of the rivers. This is partially due to that the abundance of vegetation existing along the river banks increases the values of VWC, NDVI, LAI and FPAR. Since the governing equation (objective function) in the GIP model was designed to select the cells that have high values of VWC, NDVI, and LAI, the majority of cells along the rivers therefore fall into the category of suitable candidates in the screening process. In contrast, some of the cells at a distance away from the river are competitive due to having the relatively higher LST and low FPAR. Trade-off occurs in site selection for the values of LST and FPAR would normally decrease when the value of LAI increases in the same grid. The decreasing Leaf Area Index means less radiation that is adsorbed by the vegetation, resulting in more radiation that travels to the ground and increase the value of LST. Thus, the probability of the cells located at a distance away from any river might not be low due to the fact that while the presence of maximum LST value would be outweighed by the presence of minimum FPAR. Besides, Figure 4-7(a) shows that about 50% of the grid, mainly in the top half, has the maximum permeability value of 500 mm per hour. On the contrary, Figure 4-7(b) indicates that the majority of the grid has the minimum permeability of less than 1.2 mm per hour. The greater the variability of the highest maximum and the lowest minimum permeability values of the grid, the higher the odds for them to be selected as the final hydrologic monitoring stations.

Many of the 7 cell coefficients are inter-dependent in the hydrological cycle inherently. It may lead to some sensitivity in the optimization process. One of the merits via using the GIP is to avoid running the single-parameter based sensitivity analysis for all relevant parameters of concern in the context of linear programming optimization. The use of interval expression in relation to the GIP model can actually perform a once-through sensitivity analysis that propagates all parameter uncertainties throughout the optimization process within the algorithm and end up having the optimal solution expressed with interval expressions automatically. Overall, GIP model provides an all-inclusive site-screening mechanism based on all factors with all possible variations. The GIP seeks the optimal locations based on the uncertain parameters collectively (i.e. permeability, LST, NDVI, and LAI) and the stochastic processes (i.e., precipitation, soil moisture, and FPAR). The heterogeneity of uncertainties that would defect the conventional, statistical approach may not have negative impact when using GIP for its interval expression may accommodate various sources of uncertainty. This promising approach outweighs all the others in the linear programming regimes and this is the reason why we have to count on such tool in this study.

The hydrology of such an agricultural watershed is very complex. In the United States, many of these types of agricultural watersheds have included intensive crop production systems for at least 50 years and some for much longer. More recently, livestock production systems have become increasingly centralized in concentrated animal feeding operations (CAFO's) and intensive ranch operations. The public has become concerned about their impact on water quality in the river system caused by a combination of pesticides, fertilizers and animal wastes. Besides, the riparian buffers are generally composed of trees, shrubs, herbaceous, and grass that acts as a buffer between land and water body. The mixed vegetations of the riparian buffers provide

benefits such as erosion resistance, riverbank stability, wildlife habitats, and water quality protection. The inclusion of riparian buffers as one of the criteria could lead to develop a sound hydrological monitoring network. This type of information in the CCRW can be drawn from another companion study directly (Makkeasorn and Chang, 2006). Hence, ecohydrology that turns out to be of importance in this regard is an environmental problem-solving discipline of applied science based upon the assumption that sustainable development of water resources is dependent on the ability to maintain ecological processes of water and nutrient circulation and energy flows (Zalewski et al. 1997; Zalewski 2000). All the water fluxes mentioned in the hydrological cycle would deeply affect these ecohydrological components as well. This may affect the site selection of hydrological monitoring stations too.

Conclusions

The deployment of the hydrological monitoring stations can be enhanced by implementing the integrated satellite remote sensing and GIP modeling algorithm developed in this study. It leads to an unprecedented integration of a suite of space-borne satellite remote sensing sensor data, meteorological data, and topography map to evaluate the optimal locations for siting hydrological monitoring stations in the CCRW, south Texas. The GIP modeling technique was used to screen possible candidate locations in the watershed that are significant in providing relative representative hydrologic parameters. The GIP model is proved useful in selecting reasonable candidate locations that are well distributed over the watershed. Updates of datasets such as evaporation or evapotranspiration, if available, would further enhance the outcome in the future. Nevertheless, with the aids of the extensive data collections, reliability, advancement, and consistency of various data streams, integration of the space-borne remote

sensing technologies, and the unique GIP optimization analysis, this analysis indeed advances the Earth-sciences studies in terms of developing optimum arrangement/distribution of water sensor platforms for maximum sensing coverage and information-extraction capacity.

References

Baghdadi, N., King, C., Chanzy, A., and Wigneron, J.P., 2002. An empirical calibration of the integral equation model based on SAR data, soil moisture and surface roughness measurement over bare soils. International Journal of Remote Sensing 23 (20), 4325 – 4340.

Bellman, R. and Zadeh, L.A., 1970. Decision-making in fuzzy environment. Management Science 17, 141 – 164.

Ben-Jemaa, F., Marino, M.A., and Loaiciga, H.A., 1995. Sampling design for contaminant distribution in lake sediments. Journal of Water Resources Planning and Management 121 (1), 71 – 79.

Chang, N.B. and Tseng, C.C., 2001. Assessing relocation strategy of urban air quality monitoring network by compromise programming. Environment International 26, 524 – 541.

Chang, N.B., Chen, Y.L., and Wang, S.F., 1997. A fuzzy interval multi-objective mixed integer programming approach for the optimal planning of metropolitan solid waste management system. Fuzzy Sets and Systems 89 (1), 35 – 60.

Chang, N.B., Ning, S.K., and Chen, J.C., 2006. Multi-criteria relocation strategy of offsite radioactive monitoring network for a nuclear power plant. Environmental Management 38 (2), 197 – 217.

Chang, N.B., Wen, C.G., Chen, Y.L., and Yong, Y. C., 1996b. Optimal planning of the reservoir watershed by grey fuzzy multi-objective programming (II): application. Water Research 30 (10), 2335 – 2340.

Chang, N.B., Wen, C.G., Chen, Y.L., and Yong, Y.C., 1996a. Optimal planning of the reservoir watershed by grey fuzzy multi-objective programming (I): theory. Water Research 30 (10), 2329 – 2334 .

Chang, N.B., Yeh, S.C., and Wu, G.C., 1999. Stability analysis of grey compromise programming and its applications. International Journal of Systems Science 30 (6), 571 – 589.

Christensen, E.R., Phoomiphakdeephan, W., and Razak, A., 1997. Water quality in Milwaukee, Wisconsin versus intake crib location. Journal of Environmental Engineering, ASCE 123 (5), 492 – 498.

City of Corpus Christi, 2003. Lake of Corpus Christi & Choke Canyon Reservoir. Retrieved from http://cctexas.com/?fuseaction=main.view&page=1020 accessed in October 2003.

Deng, J.L., 1989a. Introduction to grey system theory. Journal of Grey System 1 (1), 1 – 24.

Deng, J.L., 1989b. Properties of multivariable grey model GM(1,N). Journal of Grey System 1 (1), 25 – 41.

Dixon, W., Smyth, G. K., and Chiswell, B., 1999. Optimized selection of river sampling sites. Water Research 33 (4), 971 – 978.

Dubois, P.C., van Zyl, J., and Engman, T., 1995. Measuring soil moisture with imaging radars, IEEE Trans. Geosci. Remote Sensing 33 (4), 915 – 926.

Earth System Science Center (ESSC), 2002. Soil Information for Environmental Modeling and Ecosystem Management, Penn State University. Retrieved from http://www.essc.psu.edu/soil_info/index.cgi?index.html accessed in September 2003.

Egmond, N.D.V. and Onderdelinden, D., 1981. Objective analysis of air pollution monitoring network data: spatial interpolation and network density. Atmospheric Environment 15 (6), 1035 – 1045.

EROS Data Center, 2003. Band Designations. Retrieved from http://edc.usgs.gov/products/ satellite/band.html in November, 2003. Sioux Falls, SD.

Eynon, B.P., 1988. Statistical analysis of precipitation chemistry measurements over the Eastern United States. Part II: Kriging analysis of regional patterns and trends. Journal of Applied Meteorology 27 (12), 1334 – 1343.

Glenn, N.F. and Carr, J.R., 2003. The use of geostatistics in relating soil moisture to RADASAT-1 SAR data obtained over the Great Bason, Nevada, USA. Computers and Geosciences 29, 577 – 586.

Glenn, N.F. and Carr, J.R., 2004. The effects of soil moisture on synthetic aperture radar delineation of geomorphic surfaces in the Great Basin, Nevada, USA. Journal of Arid Environment 56 (4), 643 – 657.

Global Land Biosphere Data and Resources, 1999. Normalized Difference Vegetation Index. Retrieved in November 2003 from http://daac.gsfc.nasa.gov/ CAMPAIGN_DOCS/LAND_BIO/ndvi.html

Grayson, R.B., Western, A.W., Chiew, F.H.S., and Blöschl, G., 1997. Preferred states in spatial soil moisture patterns, local and non local controls. Water Resources Research 33 (12), 2897 – 2908.

Grayson, R.B., Western, A.W., Wilson, D.J., Young, R.I., McMahon, T.A., Woods, R.A., Duncan, M., and Bloeschl, G., 1999. Measurement and Interpretation of Soil Moisture for Hydrological Applications, Proceedings of Water 99 Joint Congress 25th Hydrology and Water Resources Symposium and Second International Conference on Water Resources and Environment Research, vol. 1. Institution of Engineers Australia, Brisbane, 5 – 9.

Guswa, D.J., Celia, A., and Rodriguez-Iturbe, I., 2002. Models of soil moisture dynamics in ecohydrology: comparative study. Water Resources Research 38 (9), 1166.

Harmancioglu, N.B. and Alpaslan, N., 1992. Water quality monitoring network design: a problem of multi-objective decision marking. Water Resources Bulletin 28 (1), 179 – 192.

Harmancioglu, N.B., Fistikoglu, O., Ozkul, S.D., Singh, V.P., and Alpaslan, M.N., 1999. Water Quality Monitoring Network Design. Kluwer Academic Publishers, Dordrecht, Netherlands.

Homer, C., Huang, C., Yang, L., Wylie, B., and Coan, M., not dated. Development of a 2001 National Landcover Database for the United States, report. SAIC Corporation, USGS/EROS Data Center, Sioux Falls, SD 57198, USA.

Huang, G. H., Baetz, B. W., and Patry, G. G., 1994. Grey dynamic programming for solid waste management planning under uncertainty. Civil Engineering Systems 11, 43 – 73.

Huang, G.H. and Moore, R.D., 1993. Grey linear programming, its solving approach, and its application to water pollution control. International Journal of Systems Sciences 24 (1), 159 – 172.

Huang, G.H., Baetz, B.W., and Patry, G.G., 1992. A grey lnear programming approach for municipal solid waste management planning under uncertainty. Civil Engineering Systems, 9, 319 – 335.

Huang, G.H., Baetz, B.W., and Patry, G.G., 1993. A grey fuzzy linear programming approach for municipal solid waste management planning under uncertainty. Civil Engineering Systems 10, 123 – 146.

Huang, G.H., Baetz, B.W., and Patry, G.G., 1995. Grey integer programming: an application to waste management planning under uncertainty. European Journal of Operational Research 83 (3), 594 – 620.

Huang, Y.P. and Huang, C.C., 1994. The integration and application of fuzzy and grey modeling methods. Fuzzy Sets and Systems 78, 107 – 119.

Hudak, P.F., Loaiciga, H.A., and Marino, M.A., 1995. Regional-scale ground water quality monitoring via integer programming. Journal of Hydrology 164, 153 – 170.

Huete, A., Justice, C., and Leeuwen, W.V., 1999. MODIS vegetation index (MOD 13): algorithm theoretical basis document, version 3. University of Arizona, Tucson, AZ, 129.

Hughes, J.P. and Lettenmaier, D.P., 1981. Data requirements for Kriging - estimation and network design. Water Resources Research 17 (6), 1641 – 1650.

Jager, H.I., Sale, M.J., and Schmoyer, R.L., 1990. Co-kriging to Assess Regional Stream Quality in the Southern Blue Ridge Province. Water Resources Research 26 (7), 1401 – 1412.

Knyazikhin, Y., Glassy, J., Privette, J.L., Tian, Y., Lotsch, A., Zhang, Y., Wang, Y., Morisette, J.T., Votava, P., Myneni, R.B., Nemani, R.R., and Running, S.W., 1999. MODIS Leaf Area Index (LAI) and Fraction of Photosynthetically Active Radiation Absorbed by Vegetation (FPAR) Product (MOD15) Algorithm Theoretical Basis Document, Retrieved from http://eospso.gsfc.nasa.gov/atbd/modistables.html.

Kwok, R. and Cunningham, G.F., 2000. RADASAT Geophysical Processor System Data User's Handbook. NASA, JPL. JPL D-19149.

Lacoul, M., Honda, K., Yokoyama, and R., Saito, G., not dated. Monitoring and Assessing Rice Crop with Multi-temporal RADASAT Fine beam mode data in Pathumthani and Ayuatthaya Province of Thailand." Asian Center for Research on Remote Sensing (ACRoRS), School of Advanced Technonlogies, Asian Institute of Technology, Pathumthani, Thailand.

Land Processes Distributed Active Archive Center (LP DAAC), 2006. MODIS Data Products, USGS – NASA Distributed Active Archive Center, URL: http://LPDAAC.usgs.gov/modis/dataproducts.asp, last update: February 16, 2006.

Lettenmaier, D.P., Anderson, D.E., and Brenner, R.N., 1984. Consolidation of a stream quality monitoring network. Water Resources Bulletin 20 (4), 473 – 481.

LINDO Systems Inc., 2006. URL:http://lindo.com.

Lo, S.L., Kuo, J.T, and Wang, S.M., 1996. Water quality monitoring network design of Keelung River, northern Taiwan. Water Science and Technology 34 (12), 49 – 57.

Lozar, R.C., Ehlschlaegar, C.R., and Cox, J., 2003. A Geographic Information Systems (GIS) and Imagery Approach to Historical Urban Growth Trends around Military Installations. US Army Corps of Engineers. Engineer Research and Development Center, Washington, DC.

Makkeasorn A. and Chang, N.B., 2006. Use of satellite remote sensing for change detection of riparian buffer zone in Choke Canyon Reservoir Watershed, Texas. Water Resources Research, in review.

Makkeasorn, A., Chang, N.B., Beaman, M., Wyatt, C., and Slater, C., 2006. Soil moisture estimation in a semi-arid watershed using RADARSAT-1 satellite imagery and genetic programming. Water Resources Research 42, 1 – 15.

Modak, P.M. and Lohani, B.N., 1985a. Optimization of ambient air quality monitoring networks: part I. Environmental Monitoring and Assessment 5, 1 – 19.

Modak, P.M. and Lohani, B.N., 1985b. Optimization of ambient air quality monitoring networks: part II. Environmental Monitoring and Assessment 5, 21 – 38.

Modak, P.M. and Lohani, B.N., 1985c. Optimization of ambient air quality monitoring networks: Part III. Environmental Monitoring and Assessment 5, 39 – 53.

Moran, M.S., Hymer, D.C., Qi, J., and Sano, E.E., 2000. Soil moisture evaluation using multi-temporal synthetic aperture radar (SAR) in semiarid rangeland. Agriculture and Forest Meteorology 105, 69 – 80.

Myneni, R., Knyazikhin, Y., Glassy, J., Votava, P., and Shabanov, N., 2003. User's Guide: FPAR, LAI (ESDT: MOD15A2) 8-day Composite NASA MODIS Land Algrithm, Boston University, USA, 17.

National Climate Data Center (NCDC), 2004. NOAA Satellite and Information: National Environmental Satellite, Data, and Information Service. http://www.ncdc.noaa.gov /oa/ncdc.html accessed in Oct. 2004.

National Climate Data Center (NCDC), 2006. Normals, Means, and Extremes, San Antonio, TX (SAT), report, NCDC Ashville, NC.

Ning, S.K. and Chang, N.B., 2002. Multi-objective, decision-based assessment of a water quality monitoring network in a river system. Journal of Environmental Monitoring 4, 121-126.

Ning, S.K. and Chang, N.B., 2005. Screening and sequencing analysis for the relocation of water quality monitoring network by stochastic compromise programming. Journal of American Water Resources Association 41 (5), 1039 – 1052.

Njoku, E.G., Wilson, W.J., Yueh, S.H., and Rahmat-Samii, Y. 2000. A large-antenna microwave radiometer-scatterometer concept for ocean salinity and soil moisture sensing. IEEE Trans. Geosci. Remote Sens. 38 (6), 2645–2655.

Odem, K.R., 2004. Optimizing a Sampling Network using Multivariate Statistics in a Simulated Annealing Algorithm. Ph.D. Dissertation, University of Tennessee, Knoxville, TN.

Olmsted, C., 1993. Alaska SAR Facility Scientific SAR User's Guide. University of Alaska Fairbanks, Geophysical Institute. ASF-SD-003.

Pardo-Iguzquiza, E., 1998. Optimal selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and simulated annealing. Journal of Hydrology, 210 (1 – 4), 206 – 220.

Pickett, E.E. and Whiting, R.G., 1981. The design of cost-effective air quality monitoring networks. Environmental Monitoring and Assessment 1, 59 – 74.

PRISM Group, 2006. PRISM products, Oregon State Service, Oregon State University, URL:http://www.ocs.orst.edu/prism/products.

Romshoo, S.A., Nakaekawa, T., Koike, M., and Musaike, K., 1999. Soil Moisture Determination Under Different Field Conditions Using a Scatterometer and Space Borne SAR Systems", Retrieved from http://gisdevelopment.net/aars/acrs/1999/ps3/ps300b.shtml, accessed in September 2003.

Roth, C.H., Malicki, M.A., and Plagge, R., 1992. Empirical evaluation of the relationship between soil dielectric constant and volumetric water content as the basis for calibrating soil moisture measurements by TDR. Journal of Soil Science 43, 1 – 13.

Salgado, H., Genova, L., Brisco, B., and Bernier, M., 2001. Surface soil moisture estimation in Argentina using RADASAT-1 imagery. Canadian Journal of Remote Sensing 27 (6), 685 - 690.

Sokol, J., Pultz, T.J., Deschamps, A., and Jobin, D., not dated. Polarimetric C-Band Observations of Soil Moisture for Pasture Fields. Centre for Remote Sensing, Natural Resource Canada, Ottawa, Ontario, Canada.

Song, J., Wesely, M.L., Lemone, M.A., and Grossman, R.L., 2000. Estimating watershed evapotranspiration with PASS. Part II: moisture budgets during drydown periods. Journal of Hydrometeorology 1 (5), 462 – 473.

Spectrum Technologies, Inc., not dated. Field Scout™ TDR 300 Soil Moisture Meter, User's Manual: catalog # 6430FS.

Texas Water Development Board (TWDB), 2003. Texas Precipitation, GIS Data, TWDB Mapping Website. Retrieved from http://www.twdb.state.tx.us/mapping/gisdata.asp accessed in October 2003.

Timmerman, J.G., Adriaanse, M., Breukel, R.M.A., van Oirschot, M.C.M., and Ottens, J.J., 1997. Guidelines for water quality monitoring and assessment of transboundary rivers. European Water Pollution Control 7 (5), 21 – 30.

Topp, G.C., Davis, J.L., and Annan, A.P., 1980. Electromagnetic determination of soil water content: Measurements in coaxial transmission lines. Water Resources Research 16 (3), 574 – 582.

Tseng, C.C. and Chang, N. B., 2001. Assessing relocation strategies of urban air quality monitoring stations by GA-based compromise programming. Environment International 26, 523 – 541.

U.S. Geological Survey (USGS) 2005. Real-time data for Texas, URL: http://waterdata.usgs.gov/tx/nwis/rt, accessed in April 2005.

U.S. Geological Survey (USGS), 2001. National Land Cover Data (NLCD), http://landcover.usgs.gov/classes.html, last update July 12, 2001, accessed in October 2003.

Ulaby, F., 1974. Radar measurement of soil moisture content. IEEE Trans. Antennas Propag. 22 (2), 257– 265.

Venkatram, A., 1988. On the use of Kriging in the spatial analysis of acid precipitation data. Atmospheric Environment 22 (9), 1963 – 1975.

Walker, J.P., Willgoose, G.R., and Kalma, J.D., 2001. One-dimensional soil moisture profile retrieval by assimilation of near-surface measurements: A simplified soil moisture model and field application. Journal of Hydrometeorology 2 (4), 356 – 373.

Wan, Z., 1999. MODIS land-surface temperature algorithm theoretical basis document, version 3.3, NAS5-31370, Institute for Computational Earth System Science, University of California, Santa Barbara, CA, 77.

Western, A.W. and Grayson, R.B., 1998. The Tarrawarra data set: soil moisture patterns, soil characteristics and hydrological flux measurements. Water Resources Research 34 (10), 2765 – 2768.

Western, A.W., Duncan, M.J., Olszak, C., Thompson, J., Anderson, T., Grayson, R.B., Wilson, D., and Young, R., 2001. Calibration of CS615 and TDR instruments for MARVEX, Tarrawarra and Point Nepean soils. In: Dowding, C.H., (Ed.), TDR 2001: The Second International Symposium and Workshop on Time Domain Reflectometry for Innovative Geotechnical Applications, Infrastructure Technology Institute at Northwestern University, Evanston, Illinois, 95 – 108.

Wilson, D.J., Western, A.W., Grayson, R.B., Berg, A.A., Lear, M.S., Rodell, M., Famiglietti, J.S., Woods, R.A., and McMahon, T.A., 2003. Spatial distribution of soil moisture over 6 and 30 cm depth, Mahurangi river catchment, New Zealand. Journal of Hydrology 276, 254 – 274.

Wu, C.C. and Chang, N.B., 2003. Grey input-output analysis and its application for environmental cost allocation. European Journal of Operation Research 145, 175 – 201.

Zadeh, L.I., 1965. Fuzzy Sets. Information and Control 8, 338 – 353.

Zalewski, M., 2000. Ecohydrology the scientific background to use ecosystem properties as management tool toward sustainability of freshwater resources. Ecological Engineering 16, 1 – 8.

Zalewski, M., Janauer, G.A., and Jolankai, G., 1997. Ecohydrology. A New Paradigm for the Sustainable Use of Aquatic Resources. UNESCO IHP Technical Document in Hydrology No. 7. IHP–V Projects 2.3/2.4, UNESCO Paris, 58.

CHAPTER 5
FORECASTING STREAMFLOW WITH THE AID OF NEXRAD, SEA SURFACE TEMPERATURES, AND METEOROLOGICAL CHARACTERISTICS USING GENETIC PROGRAMMING

Introduction

Fluctuations of global climate change challenge scientists and engineers to estimate and forecast the magnitude and timing of stream discharges with higher accuracy. The availability of adequate fresh water is a fundamental requirement for the sustainability of human and terrestrial landscapes. Thus, the importance of understanding and improving predictive capacity regarding all aspects of the global and regional water cycle is certain to continue to increase. One fundamental component of the water cycle is streamflow. Streamflow is related to fresh water availability for human, animal, and plant populations, and to the incidences of natural hazards, such as flood and drought, that occur abruptly and may result in loss of human and animal life and damages to human properties. Flood alert systems hold the highest possibility of reducing the damages from the floods. On the other hand, drought analysis also counts on appropriate forecasts of stream flow. Stream flow prediction therefore provides crucial information for adaptive water resources management. Recent leaps in remote sensing and artificial intelligence technologies empower such an effort.

Many hydrological and statistical techniques of streamflow prediction were widely used in water resources management around the world [Sherman, 1932; Clark, 1945; Bender and Simonovics, 1994; Khavich and Benzvi, 1995; Singh, 1995; Beven, 2000; Mohammadi and Eslami, 2006]. There are stream discharge models addressing climatic variations, which were designed as physically-based models such as "SHE" [Abbott, 1966; Abbott et al., 1986a and

1986b; Erlich, 1998], conceptual models such as IHACRES [Jakeman et al., 1990] and CMD-IHACRES [Evans, 2003]. The physically-based models intensively involve continuum mechanics of water transport. They are computationally demanding. Schreider et al. [1995] mentioned that these conceptual models could work very well with good accuracy despite the highly computational needs of the physically-based models. The conceptual rainfall-runoff model IHACRES has been used in many streamflow studies because it establishes relationships between rainfall, evaporation and streamflow data [Dye and Croke, 2003; Evans, 2003]. Yet the availability of complete meteorological data, such as rainfall, evaporation, and soil moisture, are lacking mainly because they are generally point measurements [Evans, 2003; Anctil et al., 2004a]. In most practices, precipitation, evaporation, and soil moisture data used in any model are spatially interpolated from a network of point measurements. The inaccuracy composed of interpolation and randomness of such samples is shown to have effects on the hydrologic prediction uncertainty [Vieux and Bedient, 2004]. The relationships between the meteorological data and the streamflow are non-linear [Young and Beven, 1991; Schreider et al. 1995]. Besides, tropical sea-surface temperature (SST) indicates connections to tropical Pacific and Indian Ocean SSTs, at a short lead time of a month or so which could be relevant for flood forecasting [Chowdhury and Ward, 2007].

Since a few years ago, the neural networks (NNs) models had become one of the most promising tools for stream flow forecasting [Dibike and Solomatine, 2001; Chang et al., 2002; Anctil et al., 2004a; Anctil and Tape, 2004; Chang et al., 2004; Wu et al., 2005; Sahoo and Ray, 2006]. However, most of the works presented in the literature is focused in watersheds, which generally utilizes stream flow as inputs records rather than other hydro-meteorological variables [Pulido-Calvo and Portela, 2007]. It was shown that the NNs models were the more accurate

ones than the others. Similar works were done by Lachtermacher and Fuller (in Hipel et al., 1994a). Zhu et al. (in Hipel et al., 1994b) compared a backpropagation neural network time-series stream forecasting model against ARMA, Markov, Nonparametric, FGN, and FDIFF models. The backpropagation NN model yielded a slightly better performance over other models.

Recent advances in small-scaled, high performance computers have opened the way for high intensity computational algorithms, such as genetic programming, neural network, and support vector machine [Holland, 1975; Koza, 1992; Patterson, 1996; Chang and Chen, 2000; Asefa et al., 2006]. Cramer (1985) presented the applications and capabilities of the evolutionary algorithm of simple sequential programs, which stated the promising approach of data mining and solution seeking. Specifically, use of the Sea Surface Temperature Anomalies (SSTA) improved flow volume predictions may capture some global climate change impacts [Neal et al., 2002; Asefa et al., 2006]. Genetic Programming (GP) algorithm is proposed in this study to generate complex models of streamflow prediction based on geologically and topologically site-selected meteorological data, radar-based rainfall intensity (i.e., NEXRAD), Sea Surface Temperature (SST), and USGS stream data. The meteorological data used in this study include precipitation, soil moisture, soil temperature, air temperature and relative humidity. These are related to El Nino-southern oscillation (ENSO). Effective lead-time stream flow forecast is one of the key aspects of successful water resources management based on an enlarged hydrometeorological datasets. Efforts to improving existing methods and developing new methods of stream flow prediction in this chapter in the nexus of artificial intelligence and high performance computing may support the adaptive water resources management at all scales spatially and temporally.

Study Area

The Choke Canyon Reservoir Watershed (CCRW) is a portion of the Nueces River Basin, South Texas. It is composed of several land use and land cover types covering an area of approximately 15,000 km^2. The major uses of the land are agricultures and livestock. Intensive uses of groundwater for irrigation are highly concentrated in the middle and lower areas of the basin. Geography of the area strongly influences the special hydrology of the watershed. In the upper portion of the watershed the steep slopes and arid terrain of the Balcones Escarpment rise into the Edwards Plateau. This area is the location of the very unique formation of karst aquifer called Edwards Aquifer [Eckhardt, 2005]. A karst formation refers to carbonate rock of limestone and dolomite where groundwater solutes the matrix of rock creating fractures of holes [White, 1988]. The unique hydrogeology of the Edwards Aquifer began millions of years ago when the limestone was exposed into the surface. It was then extensively eroded creating fractures and sinkholes. The fractures and the sinkholes are openings that form a subsurface drainage system. Overtime the carbonic acid from the atmosphere causes the honeycombed porous formation of the limestone rock to be enlarged. Some sinkholes can be as large as 60 feet in diameter [Eckhardt, 2005]. Groundwater can flow in karst aquifer turbulently with velocity approaching those surface streams [Center for Cave and Karst Studies, 2005]. A well-developed karst aquifer could transport water as fast as several miles per day.

The Edwards Aquifer is between 300-700 feet thick composed of porous limestone. The San Antonio segment is approximately 160 miles long and between five and 40 miles wide at the surface. The aquifer is characterized into three zones: the contributing zone, the recharge zone, and the artesian zone. The contributing zone of CCRW is the area above the fault zone (see Figure 1). It is the drainage area of the aquifer where the precipitation is gathered and the water

run offs into streams that flow over fairly impermeable limestone until they reach the recharge zone. The red lines in Figure 5-1 represent the recharge zone of the Edwards Aquifer where faulted and fractured limestone outcrop at the land surface. As the streams cross the Edwards Aquifer Recharge zone, large amounts of water flow into the underground aquifer and sinkholes. The streams lose a significant portion of their flow through the faults and solution cavities in the limestone formations. Some streams may dry eventually as the water recharges into the aquifer. Placement of USGS Stream Gages above and below the fault zone helps to quantify the water that recharges into the aquifer [USGS, 2005a]. Two USGS stream gages are also place at the exits of the watershed to quantify the water released into the CCRW, see also Figure 5-1.

Although there are many conduits and large caverns in the Edwards limestone, the aquifer is not a good storage because of the high transmissivity. However, most of the water is traveling in small pore spaces within the rock. Water enters the aquifer easily in the recharge zone, but the underground drainage is generally inadequate to hold all the water in large rain events [Eckhardt, 2005]. Recharge conduits and sinkholes quickly become filled up with water. This is one reason why the region floods so easily. Downstream of the Balcones fault zone the landscape tends to flatten out as the streams continue southward and eastward into the South Texas Brush Country where slopes range from 0 to 10 percent. These areas are mostly used for agricultural practices and livestock. Large amount of groundwater is withdrawn for irrigation. Since there are not many available surface water sources, major source of water uses is the ground water.
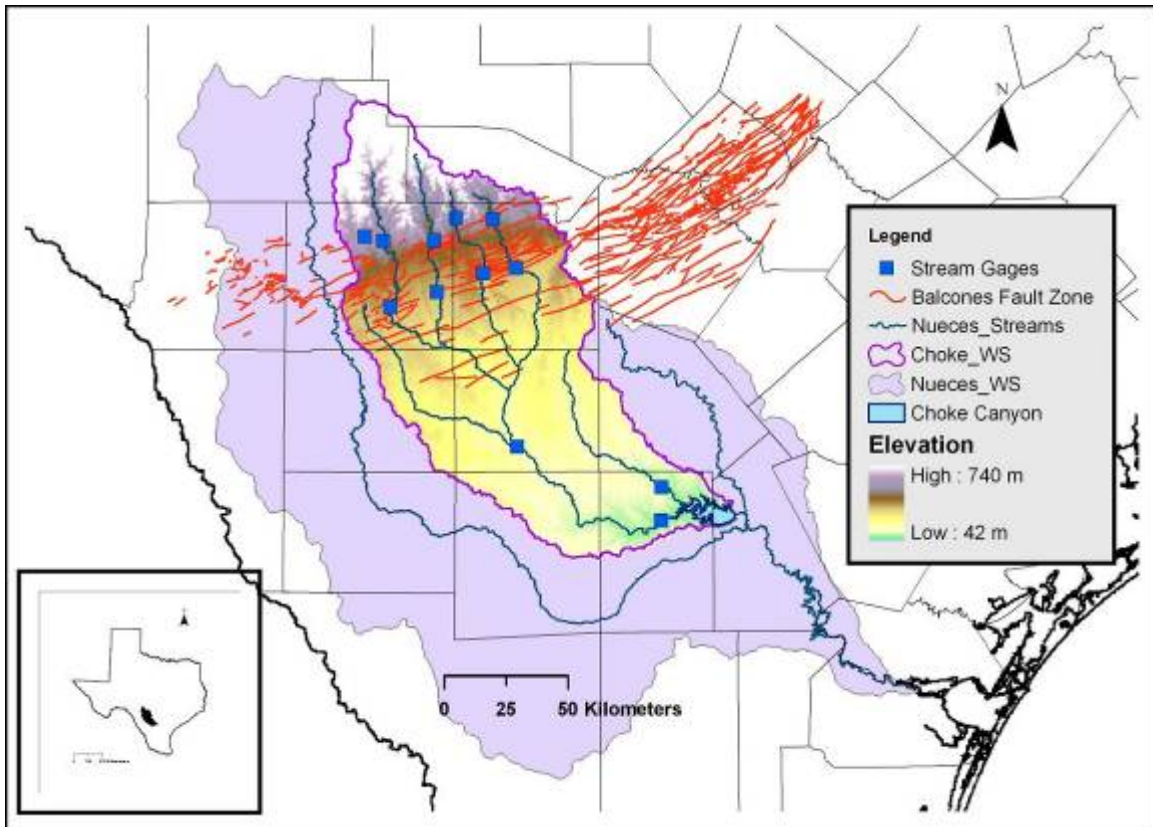
Figure 5-1. The Choke Canyon Reservoir Watershed (CCRW) is located in the South Texas. The elevation changes between 740 meters and 42 meters from top to bottom. Faults exist in the upper area where the water recharges into the groundwater aquifer. USGS stream gages are located above and below the fault lines to capture the amount of water that loses in the fault zone. Two USGS stream gages are located at the exit of the watershed to measure the amount of water released into the CCRW.

Data

Two groups of data used in this study include the existing national data from USGS Water Data for Nation, National Weather Service (NWS), National Data Buoy Center (NDBC), and our data collected from three weather stations deployed in the study area. The National Water Information System (NWIS) (http://waterdata.usgs.gov) provides surface stream discharges. The NWS provides precipitation data obtained from the NEXt generation RADar (NEXRAD). The NEXRAD offers spatio-temporal precipitation data which provide advantages over the point measurements of rain gauges. The NDBC provides Sea Surface Temperature (SST) of the seas and oceans around the United States. Our research team has three meteorological measuring stations located precisely at strategic locations derived from Grey Integer Programming (GIP). All of the datasets were collected or obtained on a daily basis between December 16, 2004 and May 5, 2005 daily. Figure 5-2 presents the locations of USGS stream gages and the three weather stations deployed in the watershed.
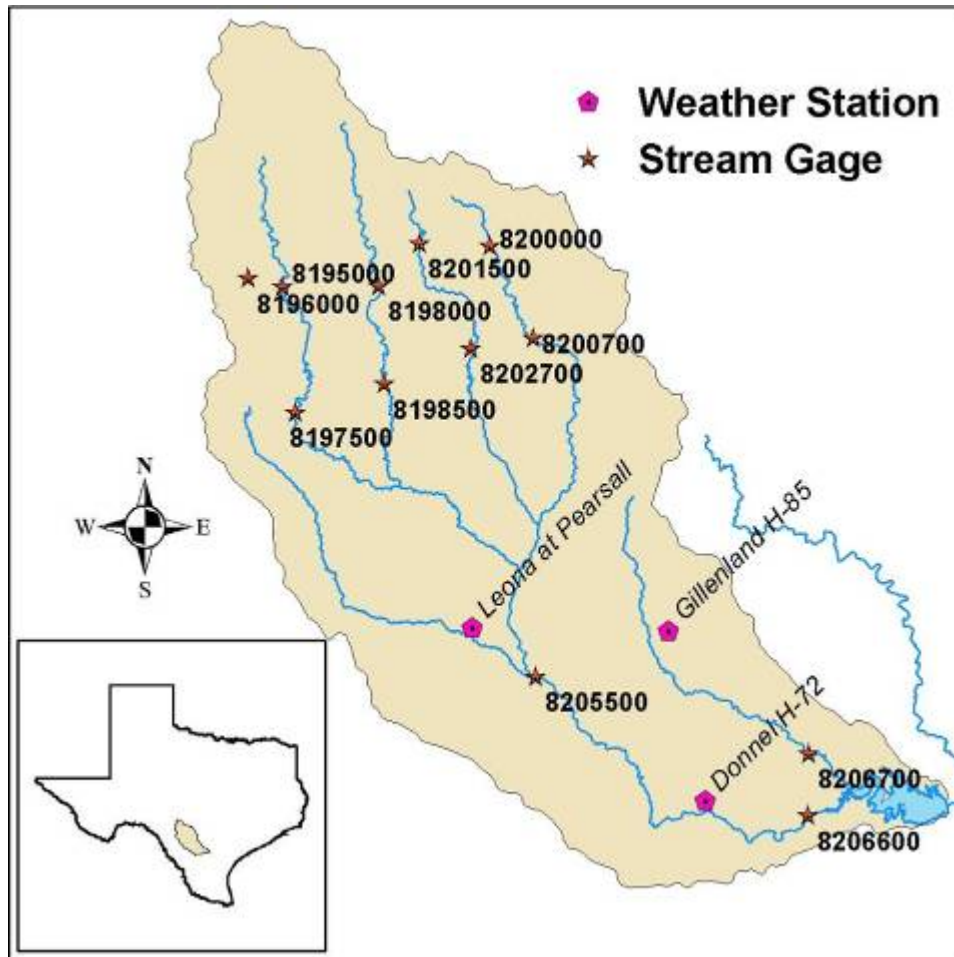
Figure 5-2. Pre-existing weather stations

Three weather stations are located in the area of missing stream gages to capture the characteristics of the watershed to enhance streamflow prediction.

USGS surface water data

Nationally, USGS surface-water data include stream levels, streamflow (discharge), reservoir and lake levels, surface-water quality, and rainfall. The data are collected by automatic recorders across the U.S. Once a complete day of readings are received from a site, daily summary data are generated and stored in the database [USGS, 2005b].

A total of 12 stream gages are located in the CCRW. Five stream gages are placed above the faults and seven stream gages are located below the fault zone. The four gages below the fault zone are the USGS 08197500, USGS 08198500, USGS 08202700, and USGS 08200700.

An additional stream gage station (USGS 0825500) is placed right in the middle of the watershed. At the exits of the watershed two USGS stream gages are used to measure the total discharge right above the Choke Canyon Reservoir, which are the USGS 08206600 and the USGS 08206700. Data of the stream flow are published by the USGS and are available at http://waterdata.usgs.gov/tx/nwis/sw. Daily summary discharges are used in this study.

NEXRAD precipitation data

Rainfall data are the acquired from NEXRAD. NEXRAD are a network of high-resolution Doppler radars that can detect precipitation and its movements over time. The returned radar signals are processed into rainfall intensity imagery as grids. The rainfall intensities over areas are converted into grids. Each grid covers 16 km^2 (4km by 4km). The daily data are published in inch per day per 16 km^2 (Figure 5-3). In order to match the precipitation rate to the discharge rate, each grid of NEXRAD data was converted from inch per day to cubic foot per second (cfs) as following:

$$1\frac{in.}{day} \times 16km^2 \times 10,763,867.31\frac{ft^2}{km^2} \times 0.083333\frac{ft}{in} \times \frac{1}{86,400}\frac{day}{\sec ond} = 166.11\frac{ft^3}{\sec} \qquad (5-1)$$

The sum of rainfall of all grids inside the watershed represents the average daily rainfall in cubic foot per second (cfs) (Figure 5-4). Every rainfall event that occurred inside the watershed between the studying time periods was concluded and exported into GIS. The NEXRAD rainfall data are published and available to public at http://www.srh.noaa.gov/rfcshare/p_download /p_download.php.

166
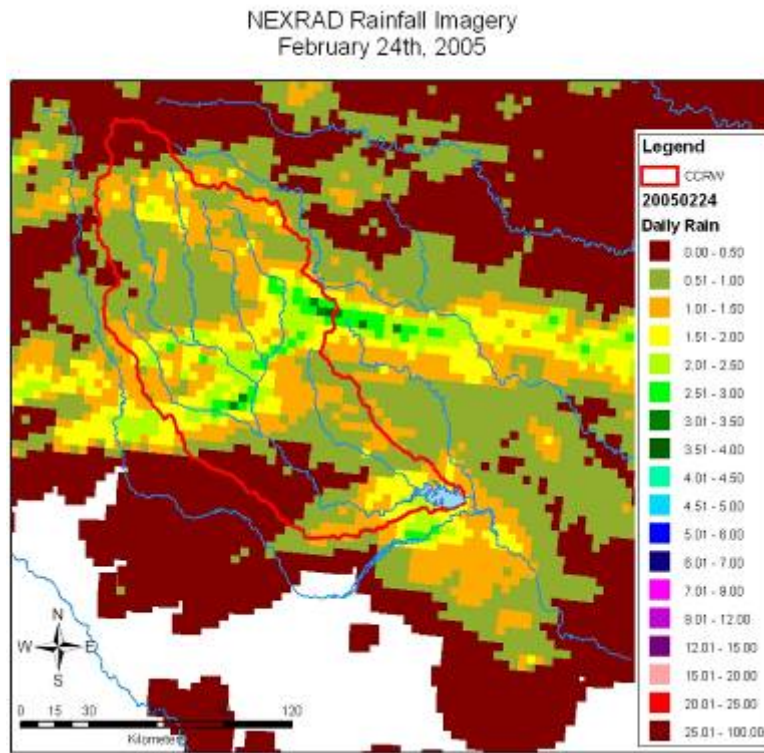
Figure 5-3. NEXRAD rainfall imagery acquired on February 24th, 2005 in the south Texas shows spatial distribution of daily rainfall intensity. Each pixel represents a 4-km grid (16 square kilometer).
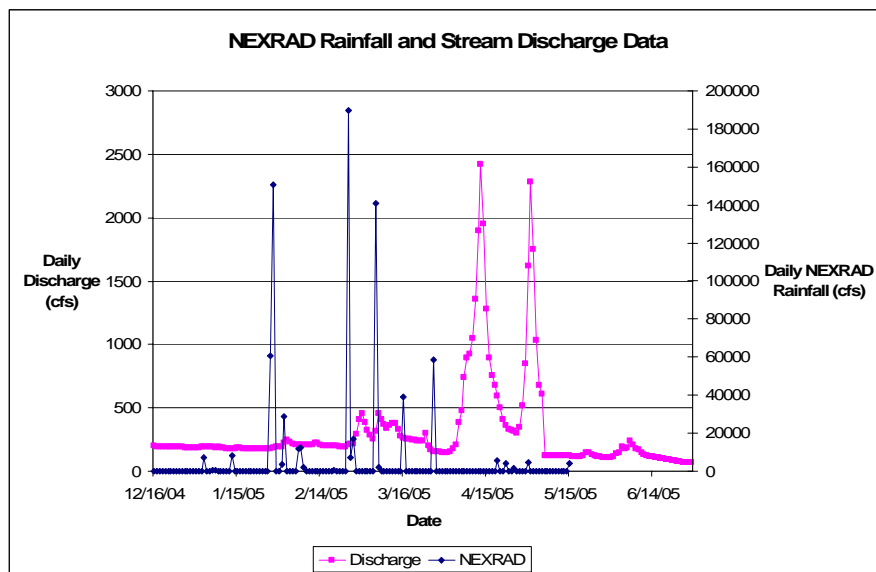
Figure 5-4. The relationship between NEXRAD data and stream discharge in CCRW is showing that the highest peak discharges occurred several days after the highest peak rainfall event.

Sea surface temperature (SST) data

Sea Surface Temperature (SST) data are collected by many groups. The data are published at the National Data Buoy Center (NDBC) and are available to public at http://www.ndbc.noaa.gov. Three buoy datasets were used in this studying including buoy 42038 in the Gulf of Mexico, 46047 in the Pacific of Southern California, and 41002 in the Atlantic of South Carolina. Locations of the three Buoys are presented in Figure 5-5. The three locations of SST are converted to indices as following:

$$I_{sst-1} = \frac{SST_{Gulf}}{SST_{Atlantic}} \tag{5-2}$$

$$I_{sst-2} = \frac{SST_{Gulf}}{SST_{Pacific}} \tag{5-3}$$

The indices are created under the assumption that the hot air would raise up to the high altitude, and the cold air would move in to replace the hot air. Thus, if the SST in the Gulf of Mexico is higher than the SST in the Atlantic, the cold air over the Atlantic will transport moisture to Texas. Similarly, if the Pacific cold air moves toward Texas, moisture from the Pacific would transport to Texas. These measures could reveal long-term effects of global climate and the large-scaled transport across longitudes.

Figure 5-5. Sea surface temperatures from the three buoys
are included in the inputs to aid in stream forecast. The buoys collect sea surface temperature in the
Atlantic Ocean, the Gulf of Mexico, and the Pacific Ocean. NEXRAD gridded rainfall data have a
coverage over the region. Each grid covers 16 square kilometer area.

Meteorological monitoring stations

In addition to the stream gages, three meteorological measuring stations were monitoring

changes at three locations in the CCRW from December 2004 to May 2005. The meteorological

measurements collected at those stations include the air temperature, relative humidity, soil

moisture, soil temperature, and the precipitation. The locations of the stations are delicately

chosen by performing Grey Integer Programming (GIP) analysis [Chang et al., 1996a and 1996b;

Huang and Huang, 1996; Chen and Chang, 2000; Wu and Chang, 2003]. The main purpose of

the stations is to fill in the information of the gap in between the inflow gages and the discharge

169

gages, and the inland temperature and the SST. Figure 5-2 shows the locations of the stream

gages and the meteorological measuring stations in the study area. Figure 5-6 shows the plots of

precipitation measured from the three weather stations and the total discharge. It obviously

shows that the peak of discharge occurred after each precipitation. Figure 5-7 shows the plots of

soil moisture measured from the three weather stations. There is also a pattern of increasing and

decreasing of discharge that follows the pattern of the soil moisture. **Error! Reference source**

not found. shows the plots of relative humidity, and Figure 5-9 shows the plots of air

temperature. The humidity and the air temperature change daily and do not seem to have any

correlation with the discharge. However, there might be some kind of correlation among them

that the genetic programming or the neural networks may be able to capture, which might

improve the accuracy of the prediction.
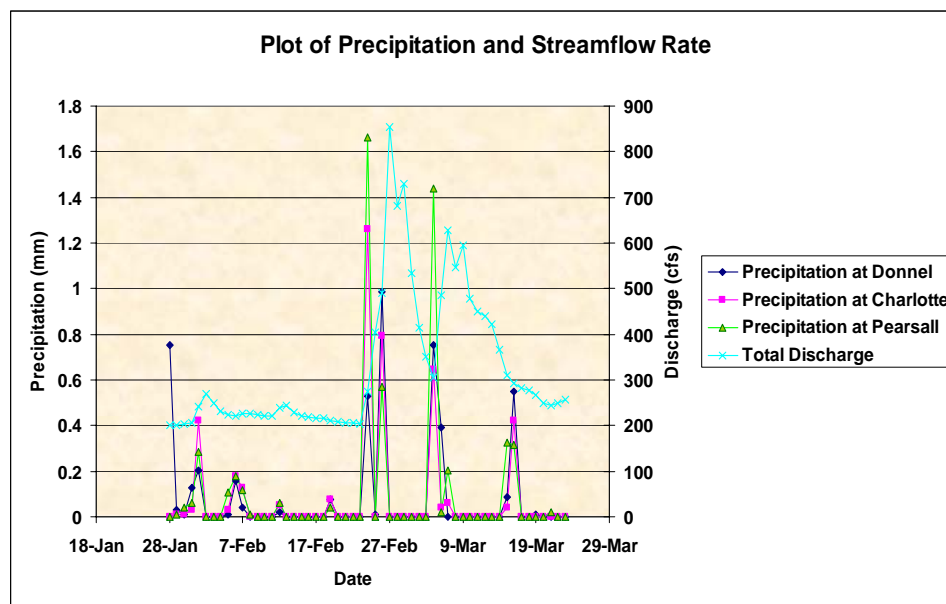
Figure 5-6. Plot of precipitation of the three weather stations
and the total discharge measured at the exits of the watershed. It obviously shows that the peak
discharges occurred after each precipitation. However, the last precipitation, which is the smallest one,
was not followed by a high flow. A small peak of flow may delay longer because of the small amount of
rain.
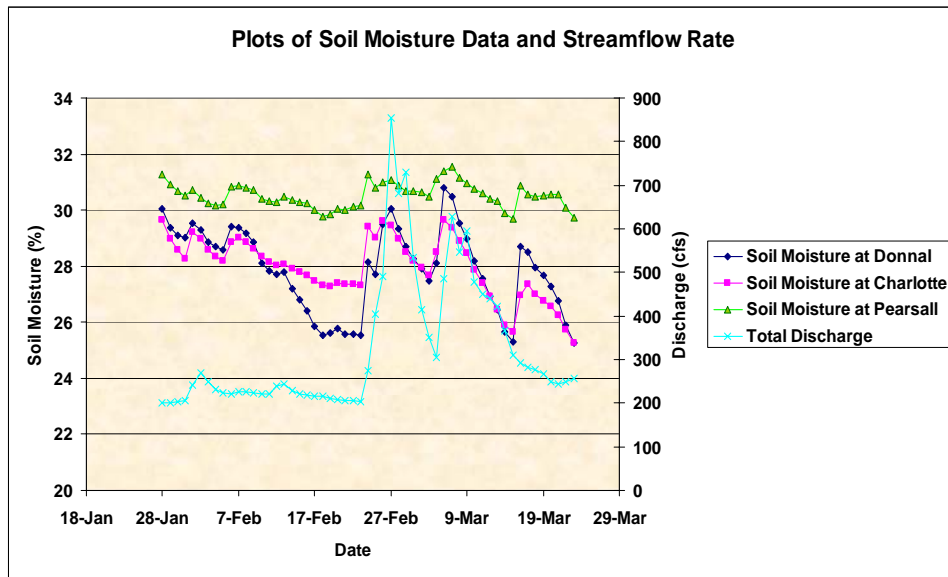
Figure 5-7. Plot of soil moisture data acquired from the weather stations

The soil moisture data of the three weather stations seem to have the same pattern. They increased and decreased together most of the time, however, the values of soil moisture were different.
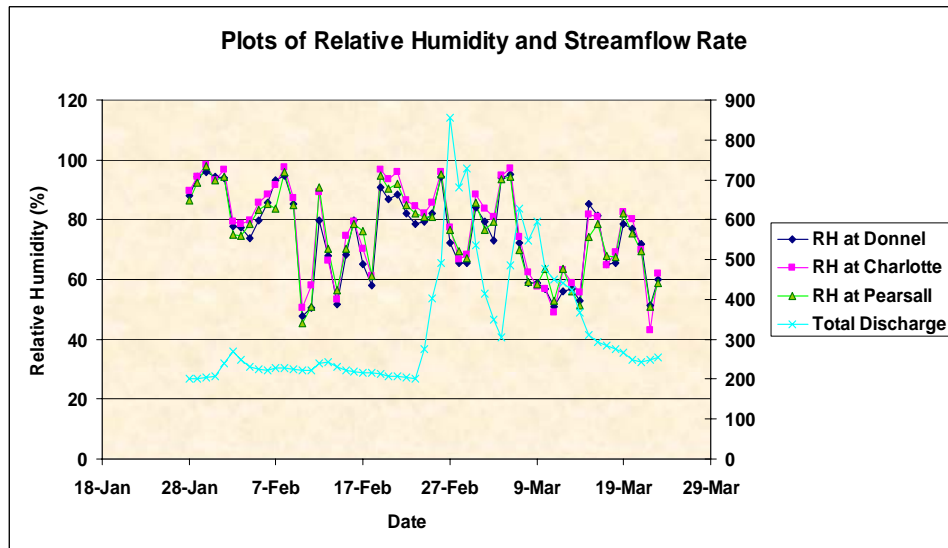
Figure 5-8. Plot of relative humidity and streamflow

The changes of discharge do not seem to follow the changes of the relative humidity on either of them. The relative humidity at the three weather stations are changing together with slight differences.
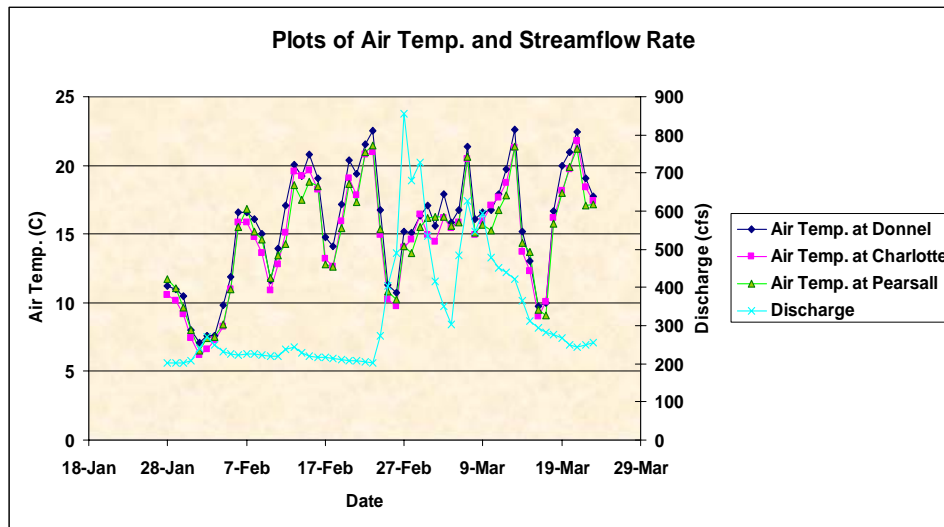
Figure 5-9. Plot of air temperature and streamflow
The air temperatures at three weather stations are slightly different, but share the same pattern of changes
overall. The discharge does not seem to follow the changes of the air temperature.

Methodology

The impact of global climate change on stream flow forecasting has been on the rise with

a great potential to improve the adaptive water resources management. Models in this study were

created under the assumption that SST of the oceans and the Gulf of Mexico are two driving

forces of the precipitation at the global scale, while the character of the stream discharge is also

influenced by the rainfall and the soil characteristics regionally at the same time. The USGS

stream gages therefore imply the capacity of streams to hold and transport the surface water.

NEXRAD rainfall represents the water input. Soil moisture content implies the water storage

capacity in the soil at critical location hydrologically. The SSTs and air temperatures imply

driving forces of rainfall indirectly. Creating stream flow forecasting models by incorporating

driving forces of precipitation and implicit watershed and climate characteristics including

NEXRAD, SST indices, soil moisture, soil temperature, air temperature, and relative humidity

should demonstrate a more transparent process of the stream flow prediction. Such a set of

172

hypothesis has also sparked a marked interest of finding out the nonlinear structure of stream flow variations among those valid driving forces on a short-term basis due to global climate change.

For the purpose of comparison, a baseline scenario is prepared based on the neural network time-series model using the discharge data only. Three groups of data sets, including S1, S2, and S3, are then created as summarized in Table 5-1. The S1 dataset includes three USGS discharge datasets and the NEXRAD rainfall. The S2 dataset include S1 dataset plus the SST indices (2) and (3). The S3 dataset include S1, S2, and fifteen meteorological parameters collected from three weather stations. The S1 dataset is used to test how models response to the USGS discharges data and the NEXRAD rainfall data. The S2 dataset is purposely designed to test the influences of SST indices on the performance of the forecasting models. The S3 dataset is used to test the effects of meteorological data on the performance of the forecasting models. Neural network time-series model is used as the based model for comparisons. The NN time-series is derived using the STATISTICA Neural Network software package from StatSoft (2005).

Table 5-1. Summary of the datasets used in different scenarios

	S1 Dataset	S2 Dataset	S3 Dataset	NN Time-Series
Three USGS Stream Data	✓	✓	✓	
NEXRAD	✓	✓	✓	
Two SST Indices		✓	✓	
Fifteen Hydro-Meteorological Parameters			✓	
Total Discharge				✓

In addition to including three different datasets addressing the scenarios in forecasting practices, three forecasting time periods were also designed to test the robustness of models.

They are based on 3 days, 7 days, and 30 days time period respectively. A total of 12 scenarios designed for the ultimate modeling analysis thus include a selective coverage of scenarios such as GP-S1-3day, GP-S2-3day, GP-S3-3day, GP-S2-7day, GP-S3-7day, GP-S1-30d, GP-S2-30d, GP-S3-30d, NN-S1-30d, NN-S2-30d, NN-S3-30d, and NN time-series models.

Data scaling is a very important step before the models can be formulated and developed. All the input variables were standardized by subtracting mean and divided by the standard deviation (Iglewicz, 1983). This would generate a set of standard normal random variables with mean '0' and standard deviation '1'. Preparation of inputs for calibration and verification of models requires splitting the dataset into two sub-datasets in model development. The first sub-dataset contains about two-third of the original data and used for the purpose of calibration of stream flow forecasting models. The second sub-dataset was used solely for the purpose of verification, which is unseen by the model during the development. The independent variables in the sub-dataset for verification were separated from the dependent variable (the discharge). Once the models in the calibration stage were created, the independent variables prepared for verification were then fed into the models. The ultimate outputs in verification stage were compared against the observed discharge for final evaluations.

During the preparation of the inputs, the dependent variable (the discharge) is offset from the independent variables (i.e., 3 USGS variables, NEXRAD rainfall, 3 Buoys variables, and 15 meteorological variables) by the period of time of the prediction. For instance, to prepare an input dataset for a 3-day forecasting model, a set of independent variables are coupled with the dependent variable three days forward. Two-third of the offset dataset was assigned to training dataset for calibration, and the rest of the data was assigned to the unseen dataset for verification.

Many models were developed to forecast the streamflow. In our study, neural network multilayer perceptron (MLP) model was trained to generate a time-series stream forecasting model as a base model. A three-step lookback and one-step lookahead was used to train the MLP network. The resulting MLP network was then used to project the next 30 days of discharge as a base model. Three MLP-NN models were also trained based on the datasets S1, S2, and S3 for 3-day stream forecasts. These three MLP networks were created as cross references for GP-derived stream forecasting models leading toward the final screening of the best one for future applications.

In any circumstances, it is necessary to have ways to evaluate the developed models based on their performance. Many methods of the evaluation and selection were published in WMO (1986). Several methods of evaluating the stream forecasting models were summarized by Tao and Lennox in Hipel et al. (eds.) [1994c]. A good model offers the least differences of the calculated forecasting discharges to the observed discharges. Five numeric evaluation criteria used in this case study include:

1. Ratio of standard deviation of predicted to observed discharges: The ratio of standard deviation of predicted and observed discharges would indicate a better model as it approaches to 1.

$$CO = \sqrt{\frac{\sum\left(y_f - \bar{y}_f\right)^2}{\sum\left(y_o - \bar{y}_o\right)^2}} \qquad (5\text{-}4)$$

2. Root-mean-square error (RMSE): The RMSE would indicate a better model as the value approaches zero.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}S.E.}{N}} \qquad (5\text{-}5)$$

3. Ratio of the mean error to the mean observed discharge: The ratio of the mean error to the mean observed discharge would indicate a better model when it approaches zero.

$$R = \frac{\sum (y_f - y_o)}{N \cdot \overline{y}_o}$$ (5-6)

4. Square of the Pearson product moment correlation coefficient: The square of the Pearson product moment correlation coefficient would indicate a better model as it approaches 1.

$$r^2 = \left[\frac{\sum (y_o - \overline{y}_o)(y_f - \overline{y}_f)}{\sqrt{\sum (y_o - \overline{y}_o)^2 \sum (y_f - \overline{y}_f)^2}} \right]^2$$ (5-7)

5. Mean of Percent Error (PE): The PE indicates a better model when its value approaches zero.

$$P.E. = \frac{\sum_N \left[\frac{(y_f - y_o)}{y_o} * 100 \right]}{N}$$ (5-8)

where, y_o and y_f are the observed and forecasted discharges, respectively. N is the total number of data points involved. The bar above each parameter indicates the arithmetic mean.

Start

Results and Discussion

Genetic programming model

Millions of GP-based models were created during the evolutionary process and only the top twenty models can be selected for further evaluation. The degrees of complexity of each scenario vary from very high to mild level. Many GP-derived models were rejected due to their overfitting or poor fitness. Less complex structure models may have higher survivorship for they avoid having the overfitting issues. A suite of GP-derived forecasting models with respect to flood forecasting lead time of 30-day, 7-day, and 3-day time period are shown in (5-9), (10), and (11) as follows.

30-day Forecasting Model:

$$\text{30-day Forecasting Discharge} = 2 \cdot \left(2^{A1} - 1\right) \cdot \left(V4\right)^2$$

(5-9)

$$A1 = \left|Sin\left(A2\right)\right| + V2 - A3$$

$$A2 = Sin\left(2 \cdot V10 + 3.075786\right) + V2$$

$$A3 = \left[- Sin\left(\left(-V1\right) \cdot \left(V8\right) \cdot \left(V16\right) - V15\right) \cdot \left(2^{A4} - 1 + V2\right)\right]$$

$$A4 = \sqrt{\left|Sin\left(A5\right) + V9\right|}$$

$$A5 = \left(\frac{Sin\left(\left(-V1\right)\left(V8\right)\left(V16\right) - V15\right)}{1.29105} - 1.0309 + V2 - V4\right)^2 + V2$$

where,

V1	stream data at gage id# 8197500	
V2	stream data at gage id# 8198500	
V4	$\dfrac{SST_{Gulf}}{SST_{Atlantic}}$	
V8	soil temperature at Donnel site	

End

V9   precipitation at Donnel site
V10 volumetric water content at Donnel site
V15 volumetric water content at Gillenland site
V16 air temperature at Leona site

7-day Forecasting Model:

$$\text{7-day Discharge (cfs)} = \left\lfloor \left| 2^{B2} - 1 \right| - 2 \cdot V4 \right\rfloor \cdot \left(- V15 \right)$$

(5-10)

$$B2 = -0.6205 \cdot \left\{ 2^{B3} - 1 - \left[\operatorname{Sin} \left(-B4 \right) - V12 \right]^2 \right\}$$

$$B3 = \left[\operatorname{Sin} \left(-B4 \right) - V12 \right]^2 \cdot 2^{-V15}$$

$$B4 = \left(2^{0.69816 \cdot (B5)} - 1 \right) \cdot \left(2^{-V15} \right)$$

$$B5 = -1.9659 \cdot (B6) + V12 - 0.9 - 2 \cdot V15 \cdot \operatorname{Sin} \left(V15 \right)$$

$$B6 = \left[\operatorname{Sin} \left(V15 \right) \cdot \left(-V15 \right) \right] - V4$$

where,

$$V4 \quad \frac{SST_{Gulf}}{SST_{Atlantic}}$$

V12 relative humidity at Gillenland
V15 volumetric water content at Gillenland site

3-day Forecasting Model:

$$\text{3-day forecasting discharge} = 2^{C1} - 1.3729$$

(5-11)

$$C1 = 2^{C2} - 1$$

$$C2 = 2^{C3} - 1$$

$$C3 = 2^{C4} - 1.3729$$

$$C4 = 2^{C5} - 1$$

$$C5 = \left\| |C6| - C7 + V4 - V19 \right| - C7 \right|$$

$$C6 = (-1.2029) \cdot \left[-\left(C8^4\right) - V15 \right]^2 \cdot (V6)$$

$$C7 = \left[(C12 - C11) \cdot (C11) \right] \cdot \left[-\left(C8^4\right) - V15 \right]^2$$

$$C8 = \operatorname{Cos}^2(C9) - \operatorname{Sin}^2\left(|C10 - V15|\right)$$

$$C9 = \left[\frac{C11 + 0.3363}{V0} \right] \cdot \left[(C12 - C11) \cdot (C11) \right]$$

$$C10 = (0.1932) \cdot \left[(2 \cdot C13) + V18 \right]$$

$$C11 = \operatorname{Sin}\left[(C14 - V15) \cdot (V20) \right]$$

$$C12 = (C13 - C15) \cdot (C15)$$

$$C13 = \left[C19 - \operatorname{Sin}(C16) \right] \cdot \operatorname{Sin}(C16)$$

$$C14 = (-0.0291) \cdot \operatorname{Sin}(2 \cdot C17)$$

$$C15 = \operatorname{Sin}\left(|C10 - V15|\right)$$

$$C16 = C18 - V15 - 0.0290$$

$$C17 = \operatorname{Sin}^2\left(|C10 - V15|\right)$$

$$C18 = \left\{ V5 \cdot \operatorname{Cos}\left[\operatorname{Sin}(C20) \right] \right\} - (0.7150 \cdot V5) + V4$$

$$C19 = \frac{-\operatorname{Sin}(C20) \cdot \operatorname{Sin}(C20)}{\operatorname{Cos}\left[\operatorname{Sin}(C20) \right]}$$

$$C20 = 2^{(-V12)(V20 + 1.4686)} - 1$$

where,

$$V4 \quad \frac{SST_{Gulf}}{SST_{Atlantic}}$$

$$V5 \quad \frac{SST_{Gulf}}{SST_{Pacific}}$$

V6   air temperature at Donnel site
V12   relative humidity at Gillenland site
V15   volumetric water content at Gillenland site
V18   soil temperature at Leona site
V19   rain at Leona site
V20   VWC at Leona

Table 5-2. Frequency analysis of input parameters used in the 30-day, 7-day, and 3-day stream forecast during the evolutionary process.

Input	description	Frequency (30-day model)	Frequency (7-day model)	Frequency (3-day model)
V0	NEXRAD	53%	3%	33%
V1	USGS 8197500	47%	17%	17%
V2	USGS 8198500	100%	47%	17%
V3	USGS8205500	43%	30%	3%
V4	SSTI Gulf/Atlantic	100%	100%	37%
V5	SSTI Gulf/Pacific	50%	53%	20%
V6	Air Temp. Donnel	43%	13%	40%
V7	RH Donnel	57%	10%	27%
V8	Soil Temp Donnel	30%	80%	83%
V9	Rain Donnel	47%	30%	17%
V10	VWC Donnel	67%	17%	40%
V11	Air Temp. Gillenland	37%	50%	37%
V12	RH Gillenland	43%	63%	43%
V13	Soil Temp Gillenland	3%	40%	50%
V14	Rain Gillenland	30%	40%	13%
V15	VWC Gillenland	60%	60%	80%
V16	Air Temp. Leona	43%	37%	10%
V17	RH Leona	37%	73%	13%
V18	Soil Temp Leona	47%	37%	40%
V19	Rain Leona	20%	43%	13%
V20	VWC Leona	43%	30%	83%

All of independent variables are used during the GP-based evolutionary process. However, not all of the independent variables are selected for each GP-based model developed. Some of the input parameters were discarded during the evolutionary process. Table 5-2 presents the frequency-of-use analysis to show that how often of each parameter was used in the top 20

models selected from at least 20 million models generated by the GP-based evolutionary process. The frequency-of-use analysis in Table 2 also shows that the USGS stream gage #8198500 and the SSTI Gulf-Atlantic are used in all of the top 20 models (i.e., 100%). In the end, the GP-based model chose a few independent variables to predict the stream discharge as shown in equation (9). Those variables include precipitation, soil temperature, air temperature, and the streamflow. The USGS stream data, SSTI and meteorological parameters demonstrate more influences in the short-term prediction in this case. NEXRAD rainfall data that were used 50% during the time of evolutionary process, however, did not survive in the selected modeling structure. This indicates that the NEXRAD data are also important, but has less influence in this short-term forecasting with respect to the data sets collected in the past 5 months because such a small set of input data points may not reveal the total variations of rainfall data to the level that the GP evolutionary process can capture its inherent influence. SST index of the Gulf and the Pacific Ocean that was used by 50% of the populations was also omitted. Again, the influence of the SST Pacific index might not be strong enough for the short-term forecasting in this case. Correlations between the SSTs of the preceding stream flow forecasting are weaker but notable. It does imply that SST might be influential in long-term forecasting analyses.

Sea Surface Temperature Index (V4) is a multiplier of the equation (8). V4 is the ratio of SST in the Gulf to the SST in the Atlantic Ocean. The value of this index will be greater than 1 when the SST in the Gulf is greater than the SST in the Atlantic Ocean. If the value of V4 is greater than 1, it implies that the wind would move from the Atlantic Ocean to the Gulf, and transporting water vapor (clouds) in the process. Thus, the V4 would imply the availability of water in the air before it precipitates down. Therefore, as the value of V4 increases, the discharge should increase.

The streamflow at the gage #8198500 (V2) is treated as additives in many places in the functional structure of the models. The stream gage #8198500 is located upstream immediately below the fault zones, which far away from the discharge point nearby the reservoir. This stream gage is normally used incorporable with other gages above the fault zone to measure the surface water lost in the recharge. Therefore, the measured streamflow at this gage represents the surface water that would flow downstream to the discharge point. As the value of V2 increases, the discharge should increase accordingly.

Two parameters (V10 and V15) reflect the point measurements of volumetric water content (VWC) measured at the Donnel and Gillenland sites, respectively. Both were subject to the Sine function. It implies that GP-derived models may capture the periodical characteristic via the use of the parameters V10 and V15. Plots of forecasting discharges derived from NN time-series and GP-derived models are presented with the observed discharges in Figure 5-10 collectively. The forecasting period starts from April 21, 2005. NN Time-series model presents relatively worse prediction as the time goes far forward in which the r-square value of the NN time series model can only reach 0.0001 in the case of 30-day prediction. The GP-derived models show very promising performance though. **Error! Reference source not found.** also shows 30-day prediction of discharges based on the NNs models. The NNs model that uses S3 dataset suffers from the curse of dimensionality or over-parameterization [Gaume and Gosset, 2003]. The dimensions of the input parameters are so large (21 parameters) with only 140 data points that this NN model even results in negative discharges predicted (see Figure 5-11).
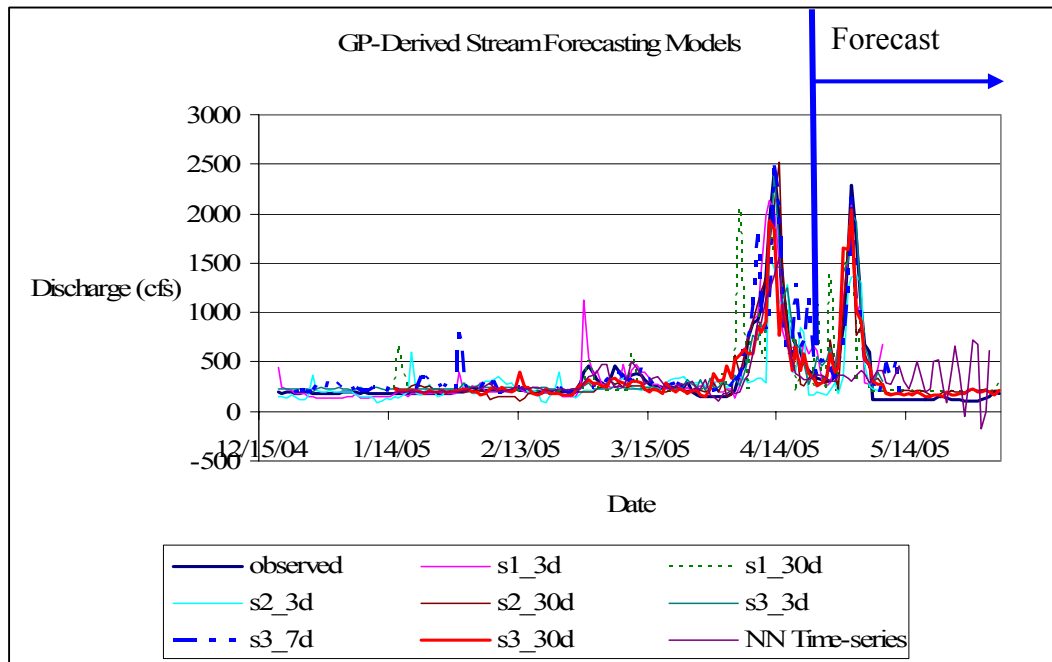
Figure 5-10. GP-derived stream forecasting model
can capture the peak flows where the time-series model underestimated the discharges. This advantage of
the GP-derived model is very critical for water management and early-warning system. The NN time-
series shows unstable forecast as the time of forecast extents further into the future.
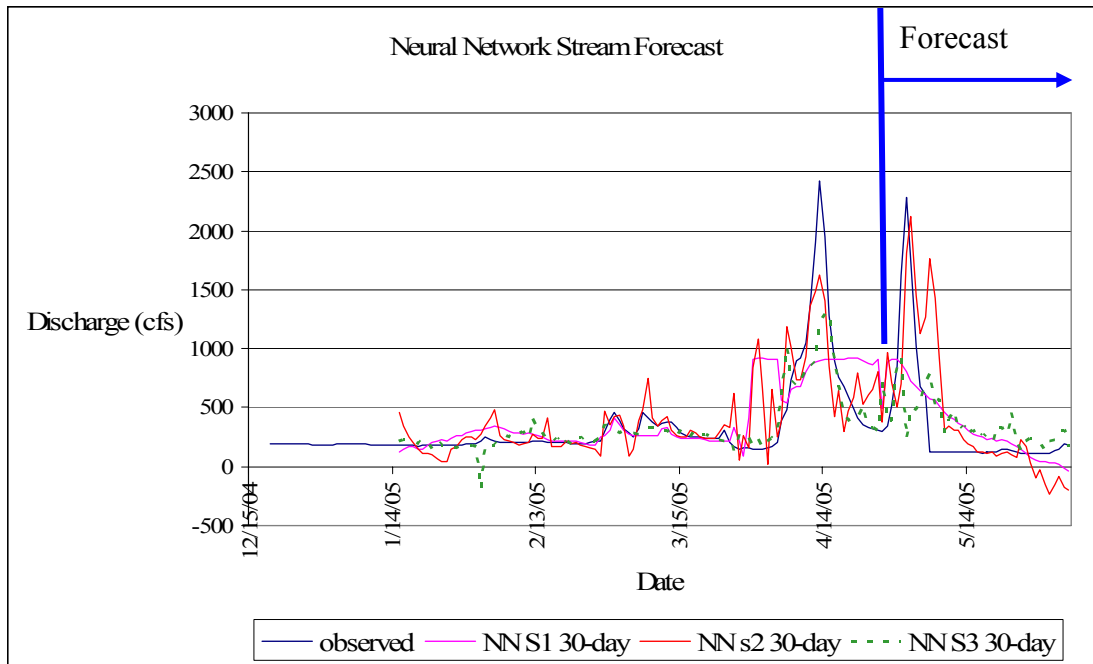
Figure 5-11. Neural networks' stream forecasts
Neural networks are used to predict discharges using datasets S1, S2, and S3. The prediction starts from
April 21, 2005. The neural network that uses the S3 dataset suffers from the lack of data points. The
forecasting discharge starts to the negative values after awhile.

Summary of results

For the final screening of all models, table 3 provides an exhaustive matrix that incorporates all models developed with respect to all indices calculated. The models derived from GP algorithms perform very well in predicting streamflow. Evaluation measures include the square of Pearson product moment of correlation coefficient (r-square) (Figure 5-12), RMSE and PE (Figure 5-13), the ratio of the standard deviation of predicted to observed values (CO) and the ratio of mean error to the mean observation (R) (see Figure 5-14). Plots of the square of Pearson product moment of correlation coefficient are compared in Figure 5-12. The r-square value increases as the input dataset moves from S1 to S3 in all GP-derived models which indicate that the S3 dataset improves the forecasting models. In contrast, the r-square of the NN model based on dataset S3 was very low, indicating the poor performance of the model. The NN

184

models suffers from the increases in dimension of variables evidenced by the decrease of r-square value as the dimension of the inputs gets larger from S1 to S3. R-square and RMSE show an agreement that training the models with S3 dataset improves the forecasting accuracy (see Figure 5-12 and Figure 5-13). The ratio R (Figure 5-14) show a trend of improved performance of many models, particularly 7-day GP, 30-day GP, and 30-day NN models. The ratio CO (Figure 5-14) presents a vast deviation amongst models.

Table 5-3. Summary of the statistical evaluation

Model	r-square	RMSE	R	CO	PE
GP s1 3-day	0.75	216	0.63	1.04	30
GP s2 3-day	0.83	230	0.52	0.98	73
GP s3 3-day	0.9	169	0.69	1.16	40
GP s2 7-day	0.76	232	0.47	1.07	26
GP s3 7-day	0.79	216	0.36	0.96	31
GP s1 30-day	0.54	334	0.49	0.91	38
GP s2 30-day	0.84	201	0.47	0.78	36
GP s3 30-day	0.84	191	0.31	0.9	26
NN Time-series 3-day	0.7143	34.8	0.09	0.61	10
NN Time-series 7-day	0.1397	64.1	0.13	0.39	12
NN Time-series 30-day	0.0001	502	0.93	0.36	36
NN s1 30-day	0.31	404	0.79	0.65	100
NN s2 30-day	0.45	453	0.85	1.2	70
NN s3 30-day	0.15	439	0.72	0.4	38

The RMSE, r-square, PE, and CO values indicate that GP-derived models are generally more accurate than the NNs models in term of the prediction accuracy. Especially, the NNs models show very poor correlation when having larger dimension of inputs (S3 dataset) (see Figure 5-13). The poor performances of NNs models were due to suffering from the curse of dimensionality when the dimension of the inputs is big and the total number of data points is relatively small [Anctil et al., 2004b]. Plots of both RMSE and PE show general agreements that the increase of dimension of the inputs may significantly decrease the forecasting errors. Even the NNs models show smaller error when using S3 dataset compared to using the S1 dataset

(Figure 5-13). GP-derived models seem to get more benefits when increasing the dimension of inputs. It results in significantly smaller forecasting errors when using S3 dataset to replace S1 dataset.

The near-term forecasting models may obtain relatively higher accuracy and smaller errors than the far-term forecasting models. Nonetheless, the 30-day GP-derived forecasting model may still provide very high accuracy (r^2 = 0.84) at considerably small error (PE = 7 percent). This fact is relevant, and it reinforces the purpose of early warning system. Within such a highly nonlinear system, however, the statistical time-series model shows very poor performance when forecasting 30-day ahead. Overall, the findings reinforce that meteorological parameters collected within the watershed of concern can obviously improve the credibility of streamflow forecasting models. Figure 5-15 presents the scatter plot between the predicted and the observed discharges of the unseen dataset. It shows a very good correlation. The plot is tacking along the 45° perfect line very well up to the very high values. This plot shows that the GP-derived models do not suffer much from a small dataset.
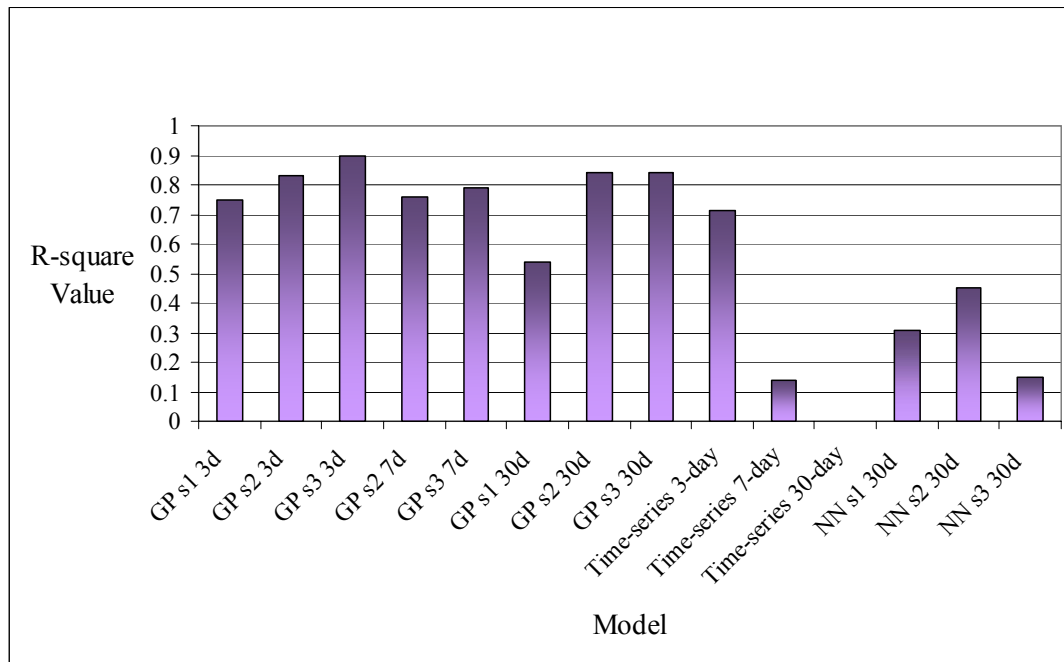
Figure 5-12. Plots of the square of Pearson product moment of correlation coefficient are compared. The r-square value increases as the input dataset moves from S1 to S3 in all GP-derived models. NNs models suffer from the increases in dimension of variables indicated by the decrease of r-square value as the dimension of the inputs gets larger.
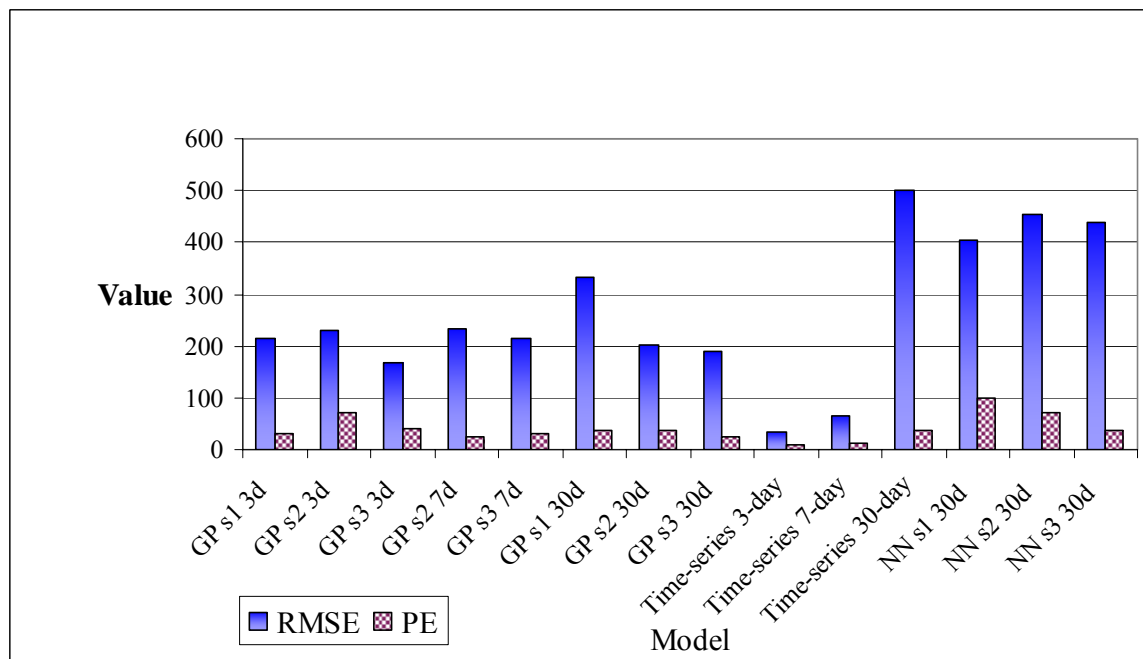
Figure 5-13. Plots of Root-Mean-Square Error (RMSE) and Percent Error (PE) show agreements that as the dimension of the inputs increases, the errors reduce. Even the NNs show smaller error when using S3 dataset compared to using the S1 dataset.
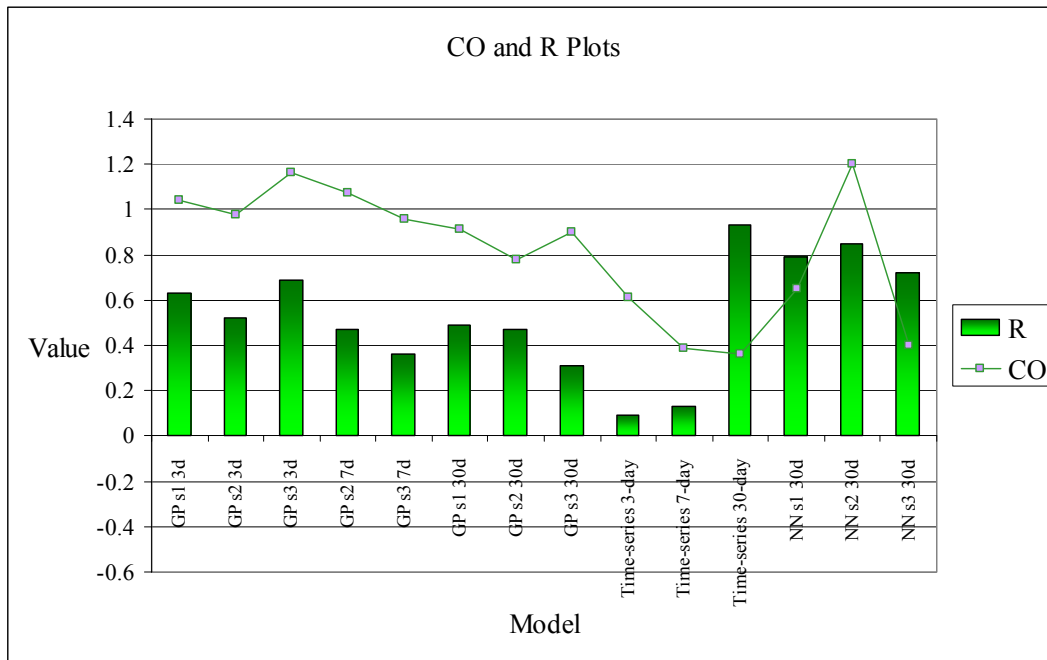
Figure 5-14. The ratio of standard deviation of predicted to observed discharges (CO) and the ratio of mean error to mean observed discharges (R)

The ratio of standard deviation of predicted to observed discharge (CO) shows that GP-derived models performed better when using a larger dimension of inputs. The ratios of mean error to mean observation (R) indicate the decreasing error value when using S3 dataset compared to using S1 dataset in the NN.
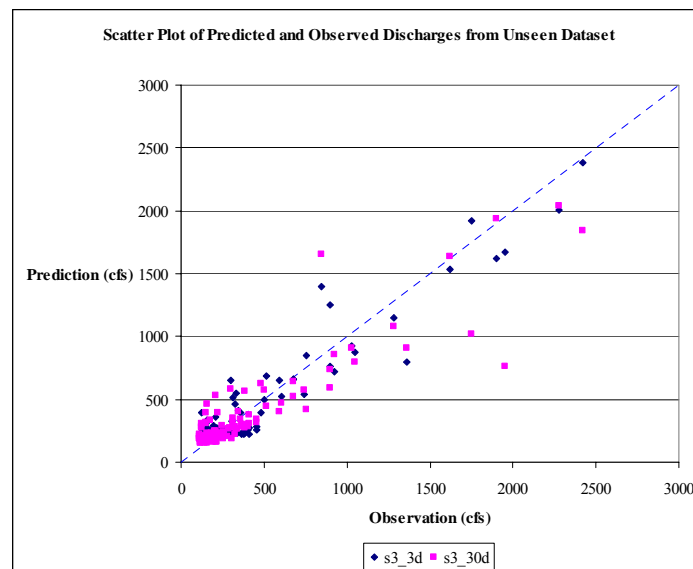
Figure 5-15. Scatter plot between the predicted and observed discharges produced from the unseen dataset.

This plot verifies that the correlation between the observation and the prediction of discharges is very good. The values are tacking along the perfect line of correlation (45° line).

Conclusions

A new approach for stream flow forecasting was introduced by incorporating the conventional stream data with NEXRAD rainfall data, SST buoy data, and meteorological parameters measured in the watershed to aid in predicting streamflow using genetic programming. The GP-derived models have shown that the inclusion of SST in the Gulf of Mexico and the Atlantic Ocean significantly improve the forecast of streamflow in the semi-arid watershed of coastal bend, Texas. Meteorological parameters enhanced streamflow prediction as well. With the expenses of processing a larger amount of data required to train the models, meteorological data and SST were found helpful in the streamflow prediction. With the aid of computational and monitoring power, deriving highly complex nonlinear functions for forecasting the streamflow with high confidence can be achieved. Improvement of streamflow forecasting was accomplished via expanding the network of meteorological sensors to capture more missing characteristics basin wide. Genetic programming revealed highly complex nonlinear structure of the underlying functions. The robustness of the GP did not suffer the curst of dimensionality as much as the NNs models did. The evolutionary feature of the GP allows the nonlinear structures embedded in the model to be shown via the natural selection algorithm that provided significant impacts on the formulation and development of the model. It is believed that more accurate near-term and far-term stream flow prediction can therefore be produced with the proposed GP modeling technique and a slightly enlarged dataset. Future work will focus on long-term stream flow forecasting using integrated GP and multifractals approach, a new statistical physics techniques, which permit - over wide ranges of scale - to easily handle extreme variability, intermittency and long range correlations in the hydrometeorological fields [Whitaker et al., 2001; Schertzer et al., 2006].

References

Abbott, M.B., 1966. An introduction to the method of characteristic. American Elsevier, New York.

Abbott, M.B., Bathurst, J.C., Cunge, J.A., O'Connell, P.E., and Rasmussen, J., 1986a. An Introduction to the European Hydrological System – Système Hydrologique Européen, "SHE", 1: History and Philosophy of a physically-based distributed modeling system. J. Hydrology 87, 45–59.

Abbott, M.B., Bathurst, J.C., Cunge, J.A., O'Connell, P.E., and Rasmussen, J., 1986b. An Introduction to the European Hydrological System – Système Hydrologique Européen, "SHE", 2: Structure of a physically-based distributed modeling system. J. Hydrology 87, 61–77.

Anctil, F. and Tape, D.G., 2004. An exploration of artificial neural network rainfall-runoff forecasting combined with wavelet decomposition. Journal of Environmental Engineering and Science 3, S121-S128 Suppl. 1.

Anctil, F., Michel, C., Perrin, C., and Andreassian, V., 2004a. A Soil Moisture Index as an Auxiliary ANN Input for Stream Flow Forecasting. Journal of Hydrology 286 (1-4), 155-167.

Anctil, F., Perrin, C., and Andreassian, V., 2004b. Impact of the Length of Observed Records on the Performance of ANN and of Conceptual Parsimonious Rainfall-runoff Forecasting Models. Environmental Modeling and Software 19 (4), 357-368 2004b.

Asefa, T., Kemblowski, M., McKee, M., and Khalil, A., 2006. Multi-time Scale Stream Flow Predictions: The Support Vector Machines Approach. Journal of Hydrology 318 (1-4), 7-16.

Bender, M. and Simonovic, S., 1994. Time Series Modeling for Long-term Stream Flow Forecasting. Journal of Water Resources Planning and Management, ASCE 120 (6), 857-870.

Beven, K.J., 2000. Rainfall-Runoff Modelling: The Primer. John Wiley & Sons, West Sussex, England.

Center for Cave and Karst Studies, 2005. Karst. Western Kentucky University, USA. Retrieved from URL:http://www.dyetracing.com/karst/karst.html.

Chang, F.J., Chang, L.C., and Huang, H.L., 2002. Real-time Recurrent Learning Neural Network for Stream-flow Forecasting. Hydrological Processes 16 (13), 2577-2588.

Chang, L.C., Chang, F.J., and Chiang, Y.M., 2004. A two-step-ahead recurrent neural network for stream-flow forecasting. Hydrological Processes 18 (1), 81-92.

Chang, N.B. and Makkeasorn, A., 2007. Optimal Site Selection of Hydrological Monitoring Stations Using Satellite Remote Sensing and Grey Integer Programming. Journal of Hydrology, in review.

Chang, N.B., and Chen, W.C., 2000. Prediction of PCDDs/PCDFs emissions from municipal incinerators by genetic programming and neural network modeling. Waste Management & Research 18, 341–351.

Chang, N.B., Wen, C.G., Chen, Y.L., and Yong, Y.C., 1996a. A grey fuzzy multiobjective programming approach for the optimal planning of a reservoir watershed. Part A: theoretical development. Wat. Res. 30(10), 2329 – 2334.

Chang, N.B., Wen, C.G., Chen, Y.L., and Yong, Y.C., 1996b. A grey fuzzy multiobjective programming approach for the optimal planning of a reservoir watershed. Part B: application. Wat. Res. 30 (10), 2335 – 2340.

Chen, H.W. and Chang, N.B., 2000. Prediction analysis of solid waste generation based on grey fuzzy dynamic modeling. Resources, Conservation and Recycling 29, 1-18.

Chowdhury, M.R. and Ward, M.N., 2007. Seasonal flooding in Bangladesh - Variability and Predictability, Hydrological Processes 21 (3), 335-347.

Clark, C.O., 1945. Storage and the Unit Hydrograph. Transactions of the American Society of Civil Engineers 110, 1419-1446.

Cramer, N.L., 1985. A representation for the adaptive generation of simple sequestial programs. Proceeding of the International Conference on Genetic Algorithms and Their Application, 183-187.

Dibike, Y.B. and Solomatine, D.P., 2001. River Flow Forecasting Using Artificial Neural Networks. Physics and Chemistry of the Earth Part B-Hydrology Oceans and Atmosphere 26 (1),1-7 2001.

Dye, P.J. and Croke, B.F.W., 2003. Evaluation of streamflow predictions by the IHACRES rainfall-runoff model in two South African catchments. Environmental Modelling & Software, 18, 705 – 712.

Eckhardt, G.A., 2005. The Edwards Aquifer Homepage. Retrieved from http://www.edwardsaquifer.net/aquifer.html in July 2005.

Erlich, M., 1988. On the modelling of the infiltration process in arid zones for irrigation project purposes with the aid of the "Système Hydrologique Européen" (SHE). Agricultural Water Management 13, 195 – 210.

Evans, J.P., 2003. Improving the characteristics of streamflow modeled by Regional Climate Models. J. Hydrology 284, 211 – 227.

Francone, F.D., 1998. Discipulus™ Software Owner's Manual, version 3.0 DRAFT. Machine Learning Technologies, Inc., Littleton, CO, USA.

Gaume, E. and Gosset, R., 2003. Over-parameterisation, A Major Obstacle to the Use of Artificial Neural Networks in Hydrology ? Hydrology and Earth System Science 7 (5), 693-706.

Heywood, M.I. and Zincir-Heywood, A.N., 2002. Dynamic page-based linear genetic programming. IEEE Trans. On Systems, Man and Cybernetics – Part B, Cybernetics 32 (3), 380 – 388.

Hipel, K.W., McLeod, A.I., Panu, U.S., and Singh, V.P., (eds.), 1994a. Stochastic and statistical methods in hydrology and environmental engineering. volume 3, 229-242.

Hipel, K.W., McLeod, A.I., Panu, U.S., and Singh, V.P., (eds.), 1994b. Stochastic and statistical methods in hydrology and environmental engineering. volume 3, 205-216.

Hipel, K.W., McLeod, A.I., Panu, U.S., and Singh, V.P., (eds.), 1994c. Stochastic and statistical methods in hydrology and environmental engineering. volume 3, 77-85.

Holland, J.M., 1975. Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor, MI, USA.

Huang, Y.P. and Huang, C.C., 1996. The integration and application of fuzzy and grey modeling methods. Fuzzy Sets and Systems 78, 107 – 119.

Iglewicz, B., 1983. Robust scale estimators and confidence intervals for location, in Hoaglin, D.C., Mosteller, M. and Tukey, J.W., eds., Understanding Robust and Exploratory Data Analysis. Wiley, NY, USA.

Jakeman, A.J., Littlewood, I.G., and Whitehead, P.G., 1990. Computational of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments. Journal of Hydrology 117, 275-300.

Khavich, V. and Benzvi, A., 1995. Flash Flood Forecasting-Model for the Ayalon Stream, Israel. Hydrological Sciences Journal 40 (5): 599-613.

Koza, J.R., 1992. Genetic Programming: On the Programming of Computers by Means of Natural Selection. MIT Press, Cambridge, MA, USA.

Koza, J.R., 2004. URL: http:genetic-programming.org, last update 16 September 2004.

Mohammadi, K., Eslami, H. R., and Kahawita, R., 2006. Parameter Estimation of an ARMA Model for River Flow Forecasting Using Goal Programming. Journal of Hydrology 331 (1-2), 293-299.

Neal, E.G., Walter, M.T., and Coffeen, C., 2002. Linking the Pacific Decadal Oscillation to Seasonal Stream Discharge Patterns in Southeast Alaska. Journal of Hydrology 263 (1-4): 188-197.

Patterson, D., 1996. Artificial Neural Networks. Prentice Hall, Singapore.

Pulido-Calvo, I. and Portela, M.M., 2007. Application of Neural Approaches to One-step Daily Flow Forecasting in Portuguese Watersheds. Journal of Hydrology 332 (1-2), 1-15.

Sahoo, G.B. and Ray, C., 2006. Flow Forecasting for a Hawaii Stream Using Rating Curves and Neural Networks. Journal of Hydrology 317 (1-2), 63-80.

Schertzer, D., Bernardara, P., Biaou, A., Tchiguirinskaia, I., Lang, M., Sauquet, E., Bendjoudi, H., Hubert, P., Lovejoy, S., and Veysseire, J.M., 2006. Extremes and Multifractals in Hydrology: Results, Validations and Prospects. Houille Blanche-Revue Internationale de L Eau (5), 112-119.

Schreider, S.Y., Jakeman, A.J., Falkland, A., and Knee, R., 1995. Streamflow prediction for the Queanbeyan River at Tinderry, Australia. Environment International 21 (5), 545 – 550.

Sherman, L.K., 1932. Streamflow from rainfall by the Unit Hydrograph Method. Engineering News Record 108, 501-505.

Singh, V.P., 1995. Computer Models of Watershed Hydrology. Water Resources Publications, Highland Ranch, CO.

Song, D., Heywood, M.I., and Zincir-Heywood, A.N., 2003. A linear genetic programming approach to intrusion detection, E. Cantú-Paz et al. (Eds.): GECCO 2003, LNCS 2724, pp. 2325-2336, Springer-Verlag Berlin Heidelberg.

StatSoft, Inc., 2005. STATISTICA (data analysis software system), version 7.1. www.statsoft.com.

U.S. Geological Survey, 2005a. USGS Activities in Texas: Edwards Aquifer Online Resources. URL:http://tx.usgs.gov/aquifer/edwards.html.

U.S. Geological Survey, 2005b. Water Resources: Surface-Water Data for Texas. Retrieved from URL:http://waterdata.usgs.gov/tx/nwis/sw.

Vieux, B.E. and Bedient, P.B., 2004. Assessing urban hydrologic prediction accuracy through event reconstruction. Journal of Hydrology 299, 217-236.

Whitaker, D.W., Wasimi, S.A., and Islam, S., 2001. The El Nino-Southern Oscillation and Long-range Forecasting of Flows in the Ganges. International Journal of Climatology 21 (1), 77-87.

White, W.B., 1988. Geomorphology and Hydrology of Karst Terrains. Oxford, Oxford University Press, 464.

World Meteorological Organization, 1986. Intercomparison of models of snowmelt runoff. Geneva, Switzerland, 646.

Wu, C.C. and Chang N.B., 2003. Grey Input-Output Analysis and its Application for Environmental Cost Allocation. European Journal of Operational Research 145, 175 – 201.

Wu, J.S., Han, J., Annambhotla, S., and Bryant, S., 2005. Artificial Neural Networks for Forecasting Watershed Runoff and Stream Flows. Journal of Hydrological Engineering 10 (3), 216-222.

Young, P. and Beven K., 1991. Computation of the instantaneous Unit Hydrograph and identifiable component flows with application to two small upland catchments – comment. J. Hydrology 129, 389 – 396.

CHAPTER 6
GENERAL CONCLUSIONS AND RECOMMENDATION

<u>Conclusions</u>

Results demonstrate that the use of genetic programming and multi-dimensional datasets is reasonable feasible in estimating soil moisture map, detecting riparian vegetation, and forecast stream discharge with very good accuracy. National institutes supply continuous data such as USGS National Water Information System, NOAA Satellite and Information Service, the National Buoy Data Center, and the PSU Center for Environmental Informatics Database, for example. The availability of data offers an opportunity to observe the global environment toward the sustainable management of resources.

The first study presents a systematic data synthesis and analyses that establish the foundation for the multitemporal soil moisture estimation in the study area. It uniquely demonstrates the use of remote sensing of hydrologic fluxes, states, and parameters, including combined active microwave and optical observations, to improve the understanding of the soil moisture variability in the terrestrial hydrosphere. The genetic-programming-derived soil moisture model is proved useful to identify the correlations between soil moisture measurements, SAR backscatter coefficient, geographical and topographical features at a watershed scale The GP model exhibits a credible record supported statistically by R-square value of 0.72 and RMSE of 3.4 based on the April 2004 data set, and R-square value of 0.69 and RMSE of 2.3 based on the September 2004 data set. Such a case study in Texas promotes the scientific justification of new measurements, involving satellites and artificial intelligence algorithms that potentially support several key scientific regimes:

i. the application of new technologies for remote sensing hydrologic quantities for terrestrial hydrologic interpretation;

ii. completion of studies on appropriate spatial and temporal sampling scales of new synergy of optical and microwave sensors for satisfying specific scientific objectives; and

iii. enhancement the information on flood and drought prediction systems indirectly, which is deemed ecologically important for the basin management authority, especially in the semiarid coastal region, south Texas.

The quantification of soil moisture will be used to estimate water availability of the watershed in general seasons in order to assist hydrologists, engineers, and stakeholders in managing water resources in this semiarid watershed. Such development also serves the scientific basis in the future for observing and modeling large scale terrestrial water-storage dynamics with emphases on how these processes are affected by the heterogeneity of soil, vegetation, precipitation, and topography and even their interaction with various biogeochemical cycles.

The second study developed a unique RICAL algorithm that was proved useful in detecting the riparian vegetation in the semi-arid watershed in south Texas. The RICAL outputs verified that soil moisture information has only minimal influence for improving the classification of riparian buffer zones. It is believed that the redundancy of vegetation index used in the development of the soil moisture model and the vegetation index used in the riparian classification causes the rejection of the hypothesis during the evolutionary process. The GP was proved robust and powerful for the classification of riparian buffers. In addition, the white box

characteristic of the GP offers the opportunity to analyze highly nonlinear structure subject to evolutionary numerical rules and constraints.

The finding was in agreement with many other studies that SAVI performs better with sparse vegetation cover and semi-arid areas. Based on the data mining feature embedded in the GP it was concluded that the Landsat 5 TM band 1 (corresponding to visible blue spectral) greatly improved the classification accuracy. It was believed that the shadows and background effects of riparian forests were the key of the usefulness of the Landsat band 1 for detecting riparian buffers. The change detection of riparian buffers can temporally and spatially be made and used in many applications for water resources management at the watershed scale.

The third study demonstrated how the deployment of the eco-hydrological monitoring stations can be enhanced by implementing the Grey Integer Programming (GIP) Site-Selection Algorithm. The analysis integrated space-borne satellite sensor data, meteorological data, and topography map to evaluate optimal locations for environmental-related monitoring stations in the Choke Canyon Reservoir Watershed, South Texas. The GIP technique was used to sort out locations in the watershed that are significant in providing high-quality of hydrologic parameters. The GIP model selected reasonable locations that are well distributed over the watershed along the rivers.

The forth study introduced a new approach of stream forecast by incorporating the conventional stream data with meteorological parameters measured in areas of the watershed, NEXRAD rainfall data, and the sea surface temperature around the United States. The results showed that the sea surface temperature in the Gulf of Mexico and the Atlantic Ocean improved the forecast of streamflow in Texas. It was also shown that only the streamflow data could not

197

provide enough information for forecasting streamflow in the area of so much diversity of land cover, activities, and soil types. Meteorological data was found helpful and engage the important role in the streamflow prediction. It was also shown that GP derived models did not suffer from the curse of dimensional inputs as much as the neural networks did when a relatively small number of data points was much less than the number of independent variables. R-square of 0.84 with a percent error of 1 was obtained from a GP-derived stream forecasting model using data from USGS stream gages, NEXRAD rainfall, and the sea surface temperature.

<u>**Recommendation**</u>

Even though streamflow can be predicted very well for a 30-day forecast in this study, precipitation remains the true driving force of the stream and availability of inland water resources. The effort of being able to predict rainfall events would be wonderfully useful in many hydrological and related applications. While precipitation is considered as random event, the true driving forces of precipitation are known and can be measured. It has been understood that energies from the Sun is the main driving force of the Earth's climate. Also the gravitational system of the sun, the earth, and the moon influences the earth's climate. In modeling the streamflow, we could only capture the effects of the watershed's characteristics after precipitation. However, if we could find out the relationships between the Earth's climate, the sun's released energies, and the gravitational forces between the sun, the earth, and the moon, it is feasible to understand the change of Earth's climate more than we do now.

The evaporation induced by the solar energy could be predicted by measuring the energies released from the Sun. The energies released from the Sun could be predicted from

monitoring and measuring the sunspots, the coronal loops, the solar flares and the magnetic reconnection on the Sun.

Once the water is evaporated to becoming clouds, winds have to transport the water vapors to the lands. Observing the gravitational forces between the Earth, the Sun, the Moon, might help us understand more about the wind. It is normally understand that the wind is driven by the differences of air temperature and pressure in different locations. However, the gravitational forces could also be another key to unlock a door to understand more about the wind. Since the gravitational forces can alter the tidal and sea levels, the clouds are affected as well.

Another influence of the gravitational system affects the Earth's climate is the influence on Earth's magnetic field. The Earth's core and flowing magma underneath the shell are also subjected to the gravitation forces. The relative location of Earth in the Solar system would be a measurement that leads us to understand more about the changes of our magnetic field. The relationship between the magnetic field and the climate will provide a deeper understanding of the Earth's climate as well.

Measuring the driving forces of the Earth's climate is a great challenge. Scientists around the world have been monitoring the Earth globally. Gathering all the data and find out all the relationships between them would indeed be the almost impossible challenge. However, we have an advantage of being created by the Mother Nature, and eventually we may be able to understand parts of nature just because we are parts of her.

APPENDIX A
GREY INTEGER PROGRAMMING ALGORITHM
(HUANG AND MOORE, 1993)

Grey Linear Programming has the following standard format:

$$\text{Max } f^{\pm} = C^{T\pm} X^{\pm} \tag{1}$$

Subject to:

$$A^{\pm} X^{\pm} \leq B^{\pm} \tag{2}$$

$$x_j^{\pm} \geq 0, \ x_j^{\pm} \ \varepsilon \ X^{\pm}, \ \forall j = 1,...,n \tag{3}$$

Where:

$$C^{T\pm} = \left[c_1^{\pm}, c_2^{\pm},..., c_n^{\pm} \right],$$

$$X^{T\pm} = \left[x_1^{\pm}, x_2^{\pm},..., x_n^{\pm} \right],$$

$$B^{T\pm} = \left[b_1^{\pm}, b_2^{\pm},..., b_m^{\pm} \right],$$

$$A^{\pm} = \left\{ a_{ij}^{\pm} \right\}, \qquad \forall i = 1,...,m, \ j = 1,...,n.$$

For the grey numbers $c_j^{\pm}, a_{ij}^{\pm}$, and $b_i^{\pm}$, we have:

$$c_j^{\pm} = \left[c_j^{-}, c_j^{+} \right], \qquad \forall j \tag{4}$$

$$a_{ij}^{\pm} = \left[a_{ij}^{-}, a_{ij}^{+} \right], \qquad \forall ij \tag{5}$$

$$b_i^{\pm} = \left[b_i^{-}, b_i^{+} \right], \qquad \forall j \tag{6}$$

Since some grey parameters exist in the objective function and constraints, the optimal solution of model equations (1) to (3) will be:

$$f^{*\pm} = [f^{*-}, f^{*+}] \tag{7}$$

$$X^{*\pm} = \left[x_1^{*\pm}, x_2^{*\pm},..., x_n^{*\pm} \right] \tag{8}$$

$$x_j^{*\pm} = \left[x_j^{*-}, x_j^{*+} \right], \qquad \forall j \tag{9}$$

Method of Solution
Model equation (1) to (3) can be converted from a grey problem (uncertain) to a white problem (certain) in the following way:

$$\text{Max } f_m^{\pm} = C_m^{T\pm} X_m^{\pm} \tag{10}$$

Subject to: $\quad A_m^{\pm} X_m^{\pm} \leq B_m^{\pm} \tag{11}$

$$x_{j\,m}^{\pm} \geq 0, x_{j\,m}^{\pm} \in x_m^{\pm}, j = 1,...,n \tag{12}$$

$$C_m^{T\pm} = \left[c_{1\,m}^{\pm}, c_{2\,m}^{\pm},..., c_{n\,m}^{\pm} \right],$$

$$X_m^{T\pm} = \left[x_{1\,m}^{\pm}, x_{2\,m}^{\pm},..., x_{n\,m}^{\pm} \right],$$

$$B_m^{T\pm} = \left[b_{1\,m}^{\pm}, b_{2\,m}^{\pm},..., b_{n\,m}^{\pm} \right],$$

$$A_m^{\pm} = \left\{ a_{ij\,m}^{\pm} \right\}, \qquad \forall i = 1,...,m, \ j = 1,...,n.$$

$c_{j\,m}^{\pm}, a_{ij\,m}^{\pm}$, and $b_{i\,m}^{\pm}$ are the whitening values of $c_j^{\pm}, a_{ij}^{\pm}$, and $b_i^{\pm}$, respectively. Therefore a set of

whitening solution $f_m^{*\pm}$ and $x_m^{*\pm}$, which are included in the optimal grey solutions $f^{*\pm}$ and $x^{*\pm}$, can be derived by solving the model defined in equations (10) to (12).

For n grey coefficients $c_j^{\pm}\,(j=1,2,...,n)$ in the objective function, if k_1 of them are positive, and k_2 coefficients $c_j^{\pm} \leq 0\,(j=1,2,...,k_2)$, where $k_1+k_2=n$ (the model does not include the situation where the two bounds of $c_j^{\pm}$ have different signs). Thus, we can develop the following expressions for the upper and lower bounds of $f^{\pm}$:

$$f^+ = c_1^+ x_1^+ + c_2^+ x_2^+ + ... + c_{k_1}^+ x_{k_1}^+ + c_{k_1+1}^+ x_{k_1+1}^- + ... + c_n^+ x_n^- \tag{13}$$

$$f^- = c_1^- x_1^- + c_2^- x_2^- + ... + c_{k_1}^- x_{k_1}^- + c_{k_1+1}^- x_{k_1+1}^+ + ... + c_n^- x_n^+ \tag{14}$$

Based on equation (13), relevant constraints can be given as:

$$a_{i1}^- x_1^+ + a_{i2}^- x_2^+ + ... + a_{ik_1}^- x_{k_1}^+ + a_{ik_1+1}^+ x_{k_1+1}^- + ... + a_{in}^+ x_n^- \leq b_i^+ \tag{15}$$

Similarly, based on equation (14), relevant constraints are:

$$a_{i1}^+ x_1^- + a_{i2}^+ x_2^- + ... + a_{ik_1}^+ x_{k_1}^- + a_{ik_1+1}^- x_{k_1+1}^+ + ... + a_{in}^- x_n^+ \leq b_i^- \tag{16}$$

For whitening solutions $x_{jm}^{*\pm}$, we have $x_{jm}^{*\pm}\,\varepsilon\,x^{*\pm}$. Therefore:

$$x_j^+ \geq x_{jm}^{*\pm}, \qquad\qquad j=1,2,...,k_1 \tag{17}$$
$$x_j^- \leq x_{jm}^{*\pm}, \qquad\qquad j=k_1+1, k_1+2,...,n \tag{18}$$
$$x_j^- \leq x_{jm}^{*\pm}, \qquad\qquad j=1,2,...,k_1 \tag{19}$$
$$x_j^+ \geq x_{jm}^{*\pm}, \qquad\qquad j=k_1+1, k_1+2,...,n \tag{20}$$

Thus, the model defined by equation (1) to (3) can be divided into two sub models:

$$Max\, f^+ \tag{21}$$
$$\text{Subject to: (3), (15), (17), and (18)} \tag{22}$$
$$Max\, f^- \tag{23}$$
$$\text{Subject to: (3), (16), (19), and (20)} \tag{24}$$

The model defined by equations (21)-(22)and (23)-(24) are linear programming models with a single objection function, Therefore, f^{*+}, x_j^{*+} ($j=1,2,...,k_1$) and x_j^{*-} ($j=k_1+1,k_1+2,...,n$) can be solved by model equations (21)-(22), and f^{*-}, x_j^{*+} ($j=k_1+1,k_1+2,...,n$) and x_j^{*-} ($j=1,2,...,k_1$) can be solved by model equations (23)-(24). Thus, the solutions of the GLP model equations (1) to (3) are:

$$f^{*\pm} = \left[f^{*-}, f^{*+}\right] \tag{25}$$
$$x_j^{*\pm} = \left[x_j^{*-}, x_j^{*+}\right] \quad \forall j \tag{26}$$

Where $f^{*\pm}$ and $x_j^{*\pm}$ are all grey numbers.

Solutions of the GLP model include decision variables ($x_j^{*\pm}$, $\forall j$) and the relevant objective value ($f^{*\pm}$). The decision variable solutions are expressed as $x_j^{*\pm} = \left[x_j^{*-}, x_j^{*+}\right]$, $\forall j$, which means that the maximum possible value of $x_j^{*\pm}$ is x_j^{*+} (upper limit), and the minimum is x_j^{*-} (lower limit). The solutions can be directly applied to decision making, with the values being adjusted within the grey intervals in the final decision scheme.

The solution of the objective function is important for assessing decision efficiencies. It is expressed as $f^{*\pm} = \left[f^{*-}, f^{*+}\right]$, which means that the maximum objective value is f^{*+} (upper limit), and the minimum is f^{*-} (lower limit). The upper and lower limits of the objective function value correspond to different distributions of decision variables. The adjustment of decision variables within their grey intervals will lead to the variation of objective function value within its corresponding grey interval.

APPENDIX B
SAMPLE OF GENETIC PROGRAMMING OUTPUT

Two sets of data were introduced to the genetic programming. The two sets of data were calculated from the following functions. They are the inverted form of each other:

$$y = x^2 - 1$$

and
$$x = \sqrt{y+1}$$

First, the genetic programming produced a large number of programs. The best program that correnponded to the $y = f(x)$ is shown in the following chain of functions:

Define: $f[0] = f[1] = f[2] = 0$

Instruction 0: $f[0] = 0 + x$

Instruction 1: $f[0] = x * x$

Instruction 2: $f[1] = f[0] = x^2$; $f[0] = tmp = 0$

Instruction 3: $f[0] = Cos\ (f[0]) = Cos\ (0) = 1$

Instruction 4: $f[2] = 0 + f[0] = 0 + 1 = 1$

Instruction 5: $tmp = f[1] = x^2$; $f[0] = tmp = x^2$

Instruction 6: $f[0] = f[0] - f[2] = x^2 - 1$ (perfect solution!)

The genetic programming was tested if it could develop the invert form of the above function, which is the $x = f(y)$. After thousand of functions were produced through the evolutionary process, the best set of functions was selected as following:

Difined initial value: $f[0] = 0$

Instruction 0: $f[0] = \cos(f[0]) = 1$

Instruction 1: $f[0] = f[0] + y = 1 + y$

Instruction 2: $f[0] = \text{sqrt}(f[0]) = \sqrt{1 + y}$ (perfect solution, again!)

This experiment proves that the search through evolutionary process based on the genetic algorithm can obtain the invert form of a function. This experiment reveals the robustness and the whitebox characteristic of the genetic programming.

APPENDIX C
PERMISSION TO REPRODUCE PUBLISHING MATERIALS

The written permission to reproduce publishing materials that are contained in Chapter 2 was granted in the form of electronics mail.

permission
From: **Michael Connolly** (MConnolly@agu.org)
Sent: Tue 5/29/07 8:15 AM
To: ammarin_r@hotmail.com

We are pleased to grant permission for the use of the material requested
for inclusion in your thesis. The following non-exclusive rights are
granted to AGU authors:

- All proprietary rights other than copyright (such as patent rights).
- The right to present the material orally.
- The right to reproduce figures, tables, and extracts, appropriately
cited.
- The right to make hard paper copies of all or part of the paper for
classroom use.
- The right to deny subsequent commercial use of the paper.

Further reproduction or distribution is not permitted beyond that
stipulated. The copyright credit line should appear on the first page
of the article or book chapter. The following must also be included,
"Reproduced by permission of American Geophysical Union." To ensure
that credit is given to the original source(s) and that authors receive
full credit through appropriate citation to their papers,
we recommend that the full bibliographic reference be cited in the
reference list. The standard credit line for journal articles is:
"Author(s), title of work, publication title, volume number, issue
number, page number(s), year. Copyright [year] American Geophysical
Union."

If an article was placed in the public domain, in which case the words
"Not subject to U.S. copyright" appear on the bottom of the first page
or screen of the article, please substitute "published" for the word
"copyright" in the credit line mentioned above.

Copyright information is provided on the inside cover of our journals.
For permission for any other use, please contact the AGU Publications
Office at AGU, 2000 Florida Ave., N.W., Washington, DC 20009.

Michael Connolly
Journals Publications Specialist

>
> -------- Original Message --------
> Subject: would like to reprint the article WRR004033 in my dissertation.
> Date: Fri, 25 May 2007 16:48:22 -0400
> From: Ammarin Makkeasorn <ammarin_r@hotmail.com>
> To: <service@agu.org>
> CC: Ni-Bin Chang <nchang@mail.ucf.edu>
>
>
>
> Hi,

1 of 1   7/12/2007 10:59 AM

Also, many contents from the other chapters have been submitted for reviews as followings:

- The content in the chapter 3 has been submitted for publication as: Makkeasorn, A. and Chang, N. B., "Data Mining Analysis for the Change Detection of Riparian Buffer Zone with Multiple Remote Sensing Imageries in a Semi-arid Watershed," Journal of Environmental Management, in review, Feb., 2007.
- The content in the chapter 4 has been submitted for publication as: Chang, N. B. and Makkeasorn, A. "Optimal Site Selection of Hydrological Monitoring Stations Using Satellite Remote Sensing and Grey Integer Programming," Journal of Hydrology, in review, Sep., 2006.
- The content in the chapter 5 has been submitted as: Makkeasorn, A. and Chang, N. B., "Stream Flow Forecasting by Remote Sensing-based Artificial Intelligence Models with Global Climate Change Pattern," Journal of Hydrology, in review, May., 2007.

APPENDIX D
REVISED LANDSAT-5 TM RADIOMETRIC CALIBRATION DATA

The following data are obtained from the work of Gyanesh Chander and Brian Markham (2003), published as:

Chander, G. and Markham, B., 2003. Revised Landsat-5 TM radiometric calibration procedures and postcalibration dynamic ranges. IEEE Trans. On Geoscience and Remote Sensing 41 (11), 2674 – 2677.

Table D-1. L-5 TM postcalibration dynamic ranges for US processed NLAPS data

Band	These parameters' value is only applied for the Landsat-5 TM data after May 5th, 2003 Spectral radiances, LMIN and LMAX in W/(m²·sr·µm)			
	$LMIN_\lambda$	$LMAX_\lambda$	$G_{rescale}$	$B_{rescale}$
1	-1.52	193.0	0.762824	-1.52
2	-2.84	365.0	1.442510	-2.84
3	-1.17	264.0	1.039880	-1.17
4	-1.51	221.0	0.872588	-1.51
5	-0.37	30.2	0.119882	-0.37
6	1.2378	15.303	0.055158	1.2378
7	-0.15	16.5	0.065294	-0.15

where

$$G_{rescale} = \frac{LMAX_\lambda - LMIN_\lambda}{Q_{cal\,max}}$$

$$B_{rescale} = LMIN_\lambda$$

$Q_{cal\,max}$ = maximum quantized calibrated pixel value (DN = 255) corresponding to $LMAX_\lambda$.

Table D-2. TM solar exoatmospheric spectral irradiances

Units: ESUN = W/(m²·µm)		
Band	Landsat-4	Landsat-5
1	1957	1957
2	1825	1826
3	1557	1554
4	1033	1036
5	214.9	215.0
7	80.72	80.67

Table D-3.
Earth-Sun distance in Astronomical Units

Day-of-Year (Julian Date)	Distance (Astronomical Units)
1	0.9832
15	0.9836
32	0.9853
46	0.9878
60	0.9909
74	0.9945
91	0.9993
106	1.0033
121	1.0076
135	1.0109
152	1.014
166	1.0158

Day-of-Year (Julian Date)	Distance (Astronomical Units)
182	1.0167
196	1.0165
213	1.0149
227	1.0128
242	1.0092
258	1.0057
274	1.0011
288	0.9972
305	0.9925
319	0.9892
335	0.986
349	0.9843
365	0.9833

Table D-4. TM thermal band calibration constants

Units	W/(m^2·sr·μm)	Kelvin
Constant	K1	K2
Landsat-4	671.62	1284.30
Landsat-5	607.76	1260.56