WATER RESOURCES RESEARCH, VOL. 42, XXXXXX, doi:10.1029/2005WR004033, 2006

Soil moisture estimation in a semiarid watershed using RADARSAT-1 satellite imagery and genetic programming

Ammarin Makkeasorn,[1] Ni-Bin Chang,[2] Mark Beaman,[3] Chris Wyatt,[4] and Charles Slater[4]

Received 14 February 2005; revised 11 April 2006; accepted 9 May 2006; published XX Month 2006.

[1] 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 for they may exhibit all-weather observation capability. 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, including slope, aspect, vegetation cover, and soil permeability, with the well-calibrated SAR data. Research findings indicate that the novel application of GP was 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.

Citation: Makkeasorn, A., N.-B. Chang, M. Beaman, C. Wyatt, and C. Slater (2006), Soil moisture estimation in a semiarid watershed using RADARSAT-1 satellite imagery and genetic programming, *Water Resour. Res.*, *42*, XXXXXX, doi:10.1029/2005WR004033.

1. Introduction

[2] 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. Multitemporal 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 [see also *Yeh et al.*, 1998]. The surface soil moisture measurement, however, is very difficult to obtain over a large area because of 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.

[3] 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

[1]Environmental Engineering Department, Texas A&M University, Kingsville, Texas, USA.
[2]Civil and Environmental Engineering Department, University of Central Florida, Orlando, Florida, USA.
[3]The Conrad Blucher Institute for Surveying and Science, Texas A&M University, Corpus Christi, Texas, USA.
[4]Alaska Satellite Facility, Geophysical Institute, University of Alaska, Fairbanks, Alaska, USA.

0043-1397/06/2005WR004033$09.00

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 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 [*Alaska Satellite Facility* (*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, 1997], on a regional scale because of its large footprint. It is well suited to large-scaled, hydrological applications.

[4] Because of 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.

[5] 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 on the basis of traditional statistical regression theory. Studies using neural network models and inversion approaches to retrieve soil moisture on the basis of passive microwave remotely sensed data can be found elsewhere [*Narayanan and Hirsave*, 2001; *Del Frate et al.*, 2003; *Wigneron et al.*, 2003]. Up to this point, there is an exceptional difficulty to derive 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 from 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 semiarid 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 1 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.

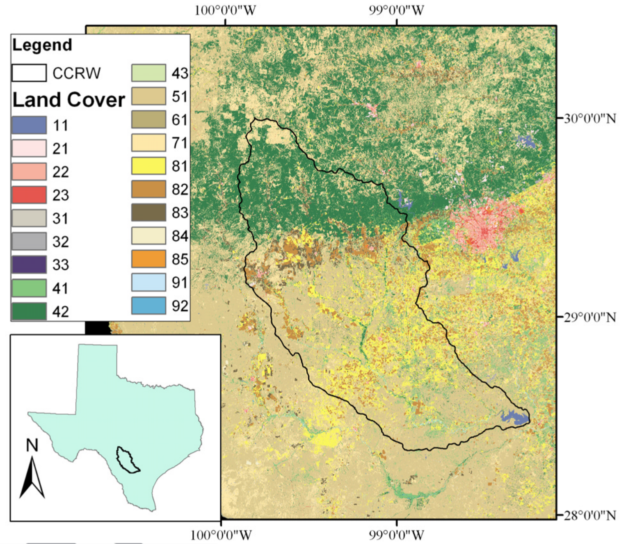

Figure 1. A description of the land covers in CCRW (see also Table 1).

2. Study Area

[6] The Choke Canyon Reservoir Watershed is composed of several land use/land cover types. Farming and livestock husbandry are major land use patterns in the past few

Table 1. Classification System Used for National Land Cover Data (NLCD)

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

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 1 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 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.

[7] Landscape in south Texas, however, is intimately tied with the geological structure. Figure 2 shows the geographic environments and geological features of the CCRW. The Choke Canyon Reservoir Watershed (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 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). Right above the Choke Canyon Reservoir 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 3). 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

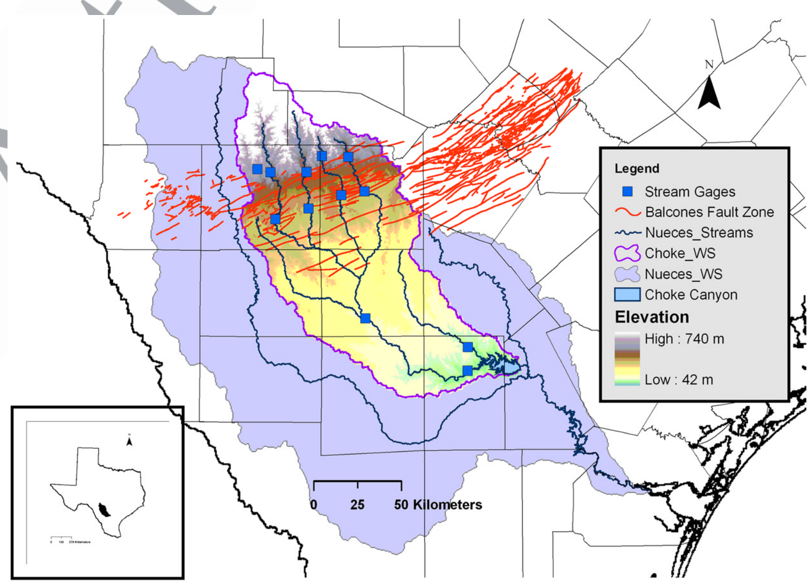

Figure 2. Fault lines, where water recharges to the underground water aquifer, shown in red. The Texas Hill Country comprises hills and valleys located above the Balcones zone. The differences of slope above and below the fault zone are obvious. Streams flow southward to the east and are merged together 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 downstream immediately before the streams flow into the CCRW.

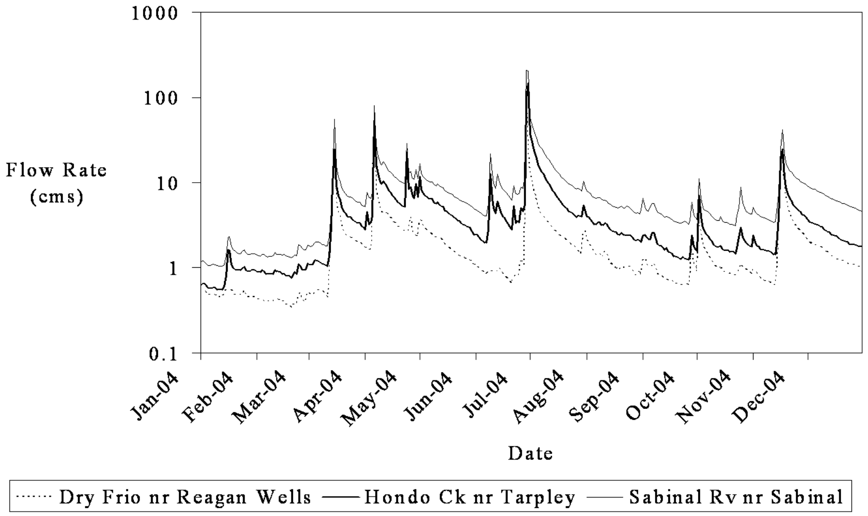

Figure 3. A log-plot of mean streamflows measured at the USGS stations that are located at Dry Frio River near Reagan Wells, Hondo Creek at Tarpley, and Sabinal River at Sabinal. The high flow rates occurred in April 2004 (wet month) throughout the time frame of the SAR data acquisition on 19 April 2004. In September 2004 the flow rates were very low nearly at the base flow, which was considered as a dry month.

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.

3. Methodology

3.1. Field Data Collection

[8] Modeling the soil moisture in this study requires emphasizing the efforts of data synthesis of 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 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 1 and Figure 1). It looks like there are only 7 fields sampled in the ground truth data acquisitions in Figures 4 and 5. 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

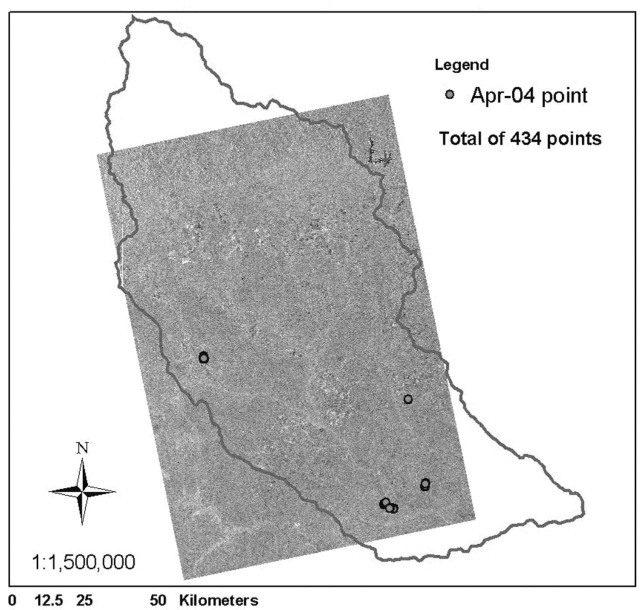

Figure 4. Four hundred and thirty-four ground truth measurement points collected within 24 hours before and after the SAR data acquisition on 19 April 2004. The measurements were done on flat bare soil, high-density mesquite trees, deciduous forest, and grassland.

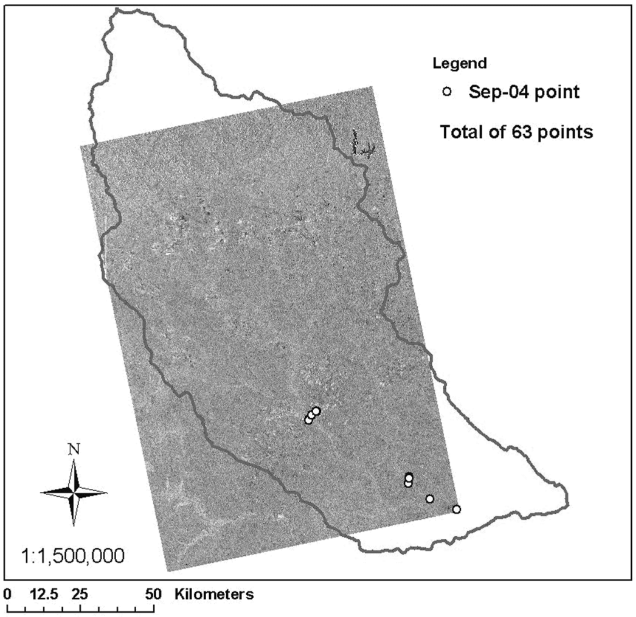

Figure 5. Sixty-three ground truth measurement points collected within 8 hours before the SAR data acquired on 12 September 2004. The ground truths were done on evergreen forest, raw crop, brush land, high-density mesquite trees, and grassland.

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.

[9] 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 Hégarat-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.

3.2. Genetic Programming

[10] The well-known approach invented by *Koza* [1992, 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.

[11] 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].

[12] 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.

3.3. Integrated Framework

[13] Figure 6 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 (i.e., SAR pixel size), 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 CCRW is composed of many frames, mosaicking was performed to

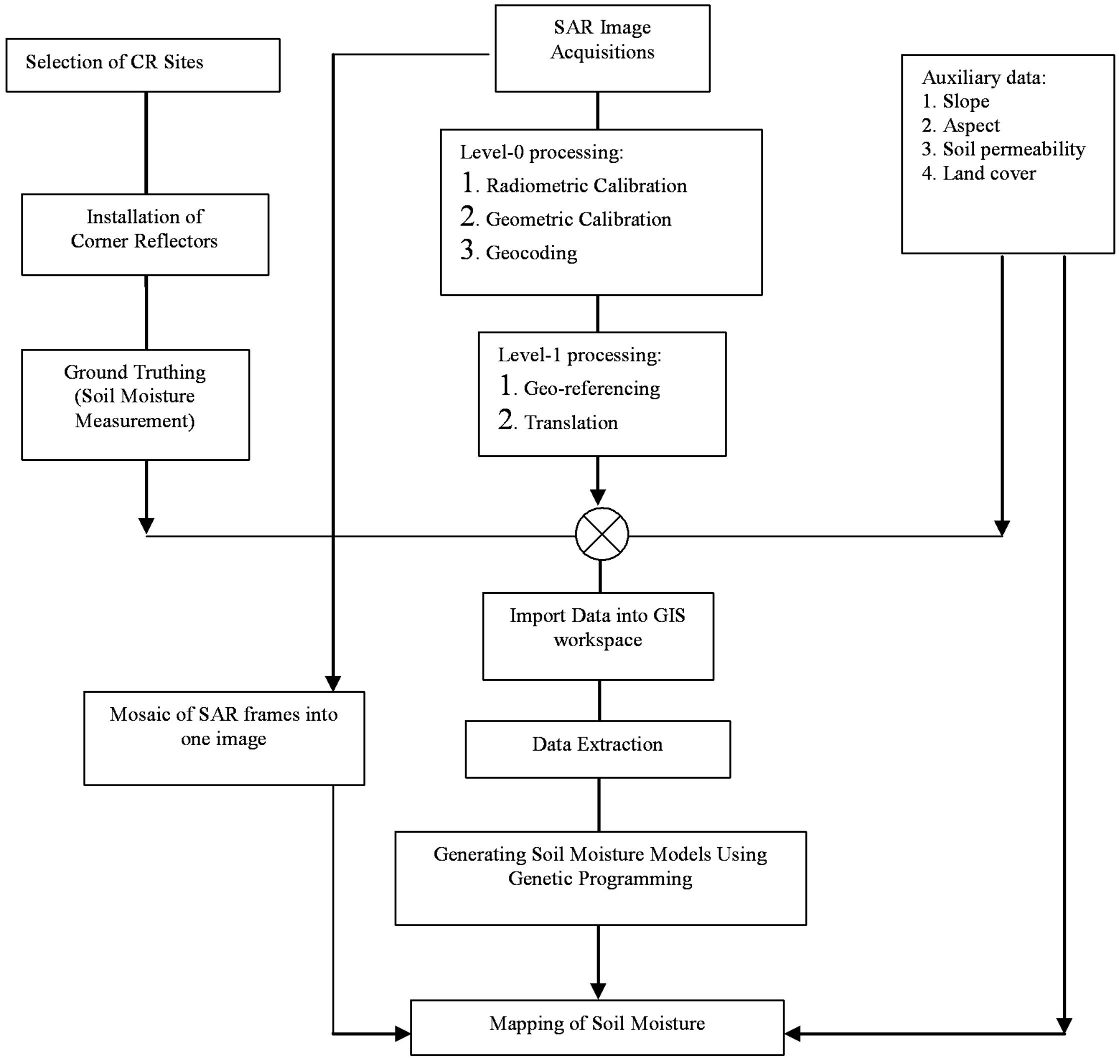

Figure 6. Work flowchart of this study.

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.

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

[15] 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 the generated GP model.

4. Data Synthesis and Processing

4.1. SAR Data

[16] 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.

[17] 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 [*TSS*, 1996; *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

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

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:

$$DN = (\sigma_0 * 10) + 255 \tag{2}$$

[18] 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.

4.2. Slope and Aspect Data

[19] 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 Hégarat-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.

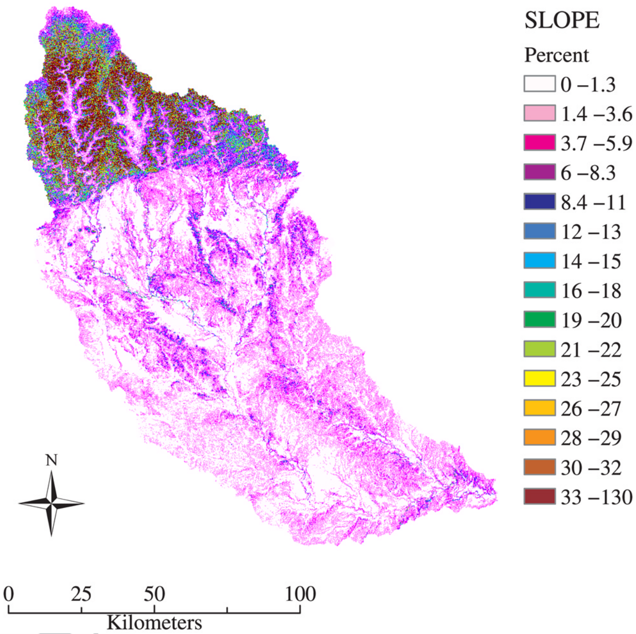

Figure 7. Slope map showing the variation of slopes in the CCRW. Most area below the Balcones zone is very flat and is used mostly for agriculture and livestock. The upper area is the Edwards Plateau that comprises hills.

[20] On the basis of such rationale, Figure 7 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 8 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.

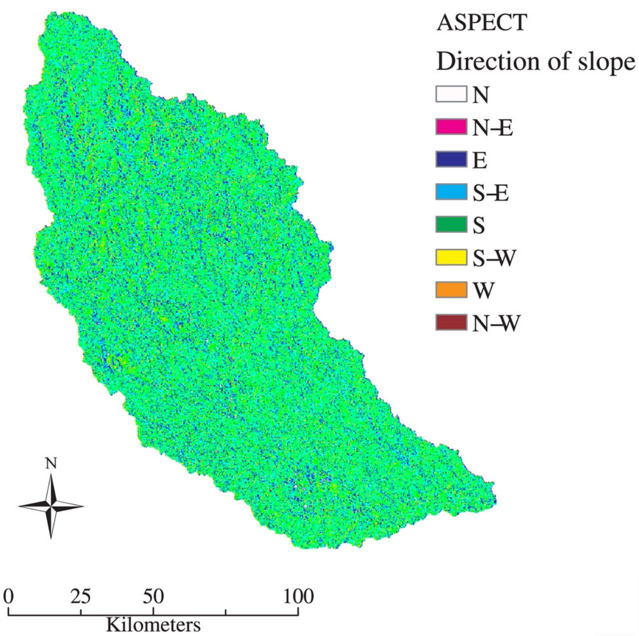

Figure 8. Aspect map showing the major directions of slopes are facing toward the south and the southeast. It is the actual direction that the streams flow toward the Gulf of Mexico.

4.3. Soil Permeability Data

[21] 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 9. The development of soil moisture model would benefit from incorporating soil permeability, measured in inch/hr, along with some other geoenvironmental variables and SAR.

4.4. NDVI Data

[22] 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.

[23] 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

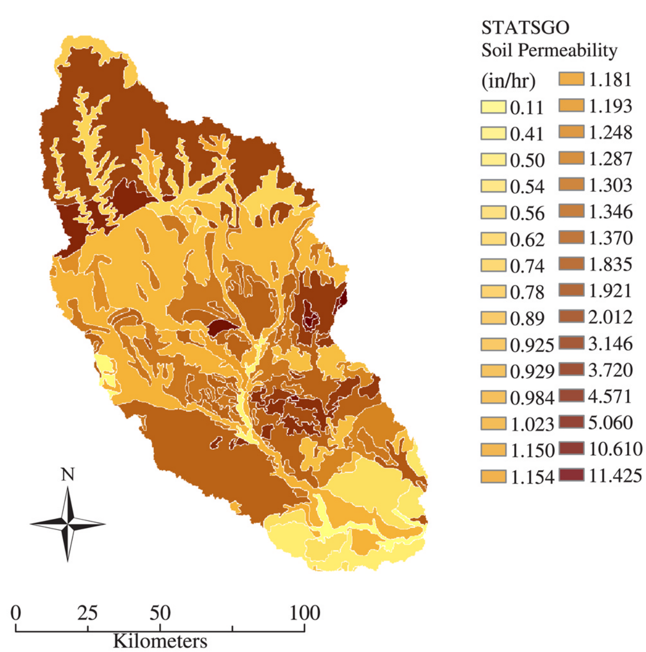

Figure 9. Thirty-one classes of STATSCO soil classifications found in the watershed. The average permeability (inch/hr) of the soils was included in the derivation of the soil moisture model.

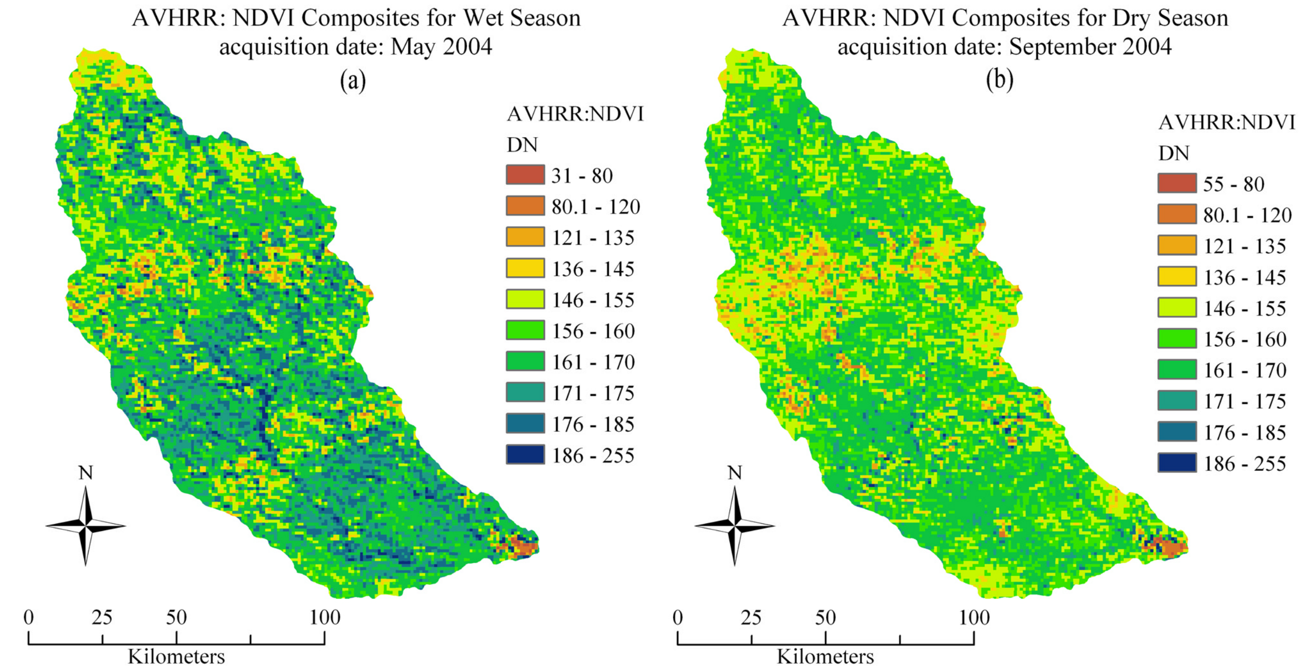

Figure 10. Maps of AVHRR NDVI images for (a) May 2004 and (b) September 2004 showing high values of greenness along the river corridor (shown in dark blue). The plowed land and farms are likely presented in orange and yellow. Texas brush land, ranches and grassland are shown in dark and light greens. Red area in the maps indicates the water body of the Choke Canyon Reservoir.

[*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 Figures 10a 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.

4.5. Model Development for Soil Moisture Estimation

[24] Overall, the backscatter might be more strongly influenced by roughness than soil moisture. This is because grain size is an order of magnitude or more 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.

[25] 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.

[26] 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 submeter 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 GRID-SPOT 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:

$$VWC = fn(V0, V1, V2, V3, V4) \tag{3}$$

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).

5. Results and Discussion

[27] 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:

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

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

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

$$A3 = A4 - |\mathrm{Sin}(A5)|$$

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

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

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

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

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

$$A9 = 2 \cdot (V4) + \mathrm{Cos}(0) + 1.03$$

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).

[28] 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.

Table 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

[29] 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 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–data set pair-wise. Figure 11a 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 11b 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.

[30] Research findings also indicate that the ground truth measurements with larger value of soil moisture are likely to generate disturbance in model development (see Figures 11a and 11b) 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

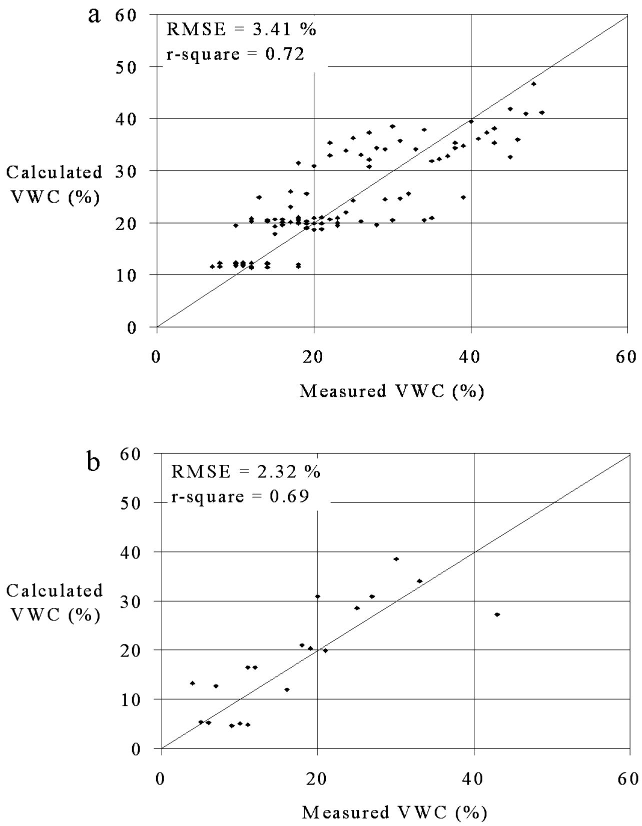

Figure 11. (a) Plot of measured versus calculated soil moisture of the unseen samples collected in April 2004. The unseen samples used in the calculations are independent from the samples used to calibrate the model. (b) Plot of measured versus calculated soil moisture of the unseen samples collected in September 2004. The unseen samples used in the calculations are independent from the samples used to calibrate the model.

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.

[31] With the aid of soil moisture model derived by the GP technique, Figures 12a and 12b present the soil moisture estimations watershed wide. The grid cell resolution of Figures 12a and 12b is 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 12a shows the map of soil moisture in April 2004 that has a mean soil moisture value of 16.3% by volume. Figure 12b 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.

[32] 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

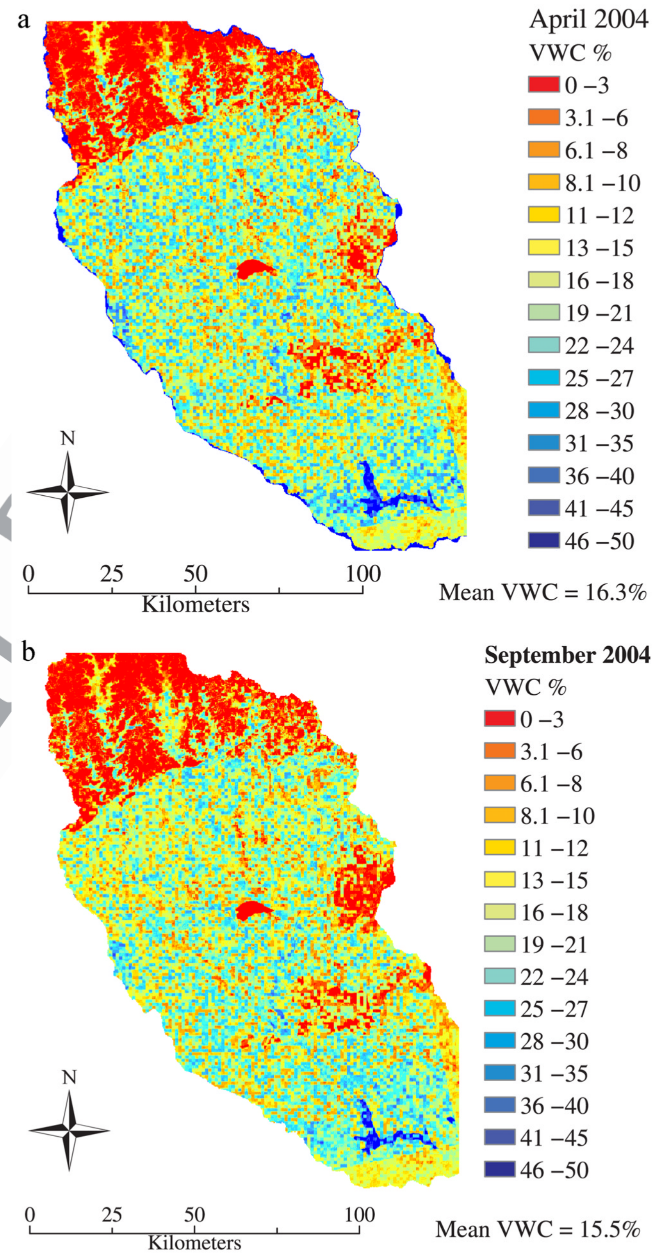

Figure 12. (a) Soil moisture map derived from RADARSAT-1 SAR data acquired in April 2004 and the supplemental data. The average volumetric soil moisture is 16.3%. The low estimates of soil moisture at the top area are due to the high percent slopes. The patch of low soil moisture at the center in the 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. (b) Soil moisture map derived from RADARSAT-1 SAR data acquired in September 2004 and the supplemental data. The average volumetric soil moisture is 15.5%. The extreme values of percent slope and soil permeability affect the estimation of soil moisture similarly to Figure 12a.

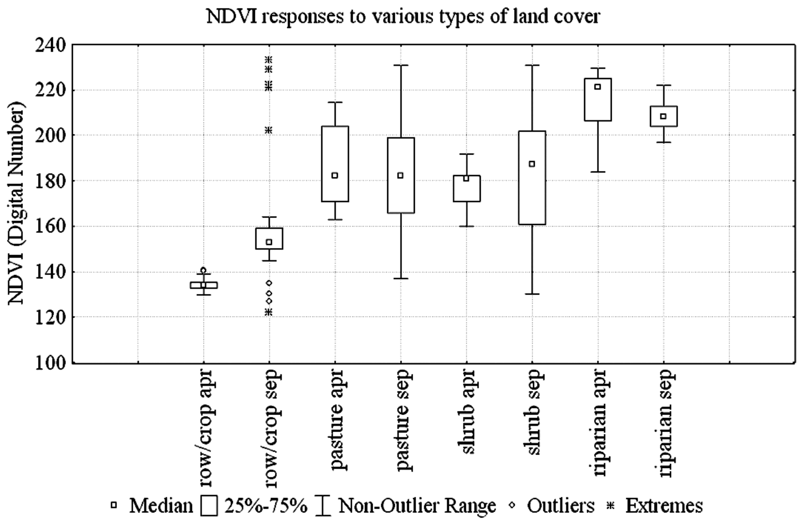

Figure 13. Box plots presenting NDVI values corresponding to various types of land cover over a time period from April to September 2004. NDVI of row/crop increased. NDVI of pasture changed insignificantly over the time period. The shrub area responded to a wider range of NDVI values in September than in April. The NDVI of the riparian area dropped over the time period.

soil permeability increases. This is a reason for Figures 12a and 12b 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 9. Similarly, the patches of high soil moisture areas at the bottom of Figures 12a and 12b are due to the very low soil permeability as evidenced in Figure 9 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 Figures 12a and 12b because of the high values of slope as evidenced in Figure 7.

[33] 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.

[34] 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 12 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.

[35] 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 13. 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.

[36] 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

Table 3. Statistical Evaluation of Regression Models (Unseen Data Are Used)

Model	R-Square APR Data	R-Square SEP Data	RMS Error, % vwc APR Data	RMS Error, % vwc 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 regression model 1	0.27	0.36	11.5	9.4
Nonlinear regression model 2	0.30	0.33	11.8	8.8

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:

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

Nonlinear regression 1:

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

Nonlinear regression 2:

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

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).

[37] 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 3 concludes the evaluation of these models on the basis of the same criteria. Figure 14 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

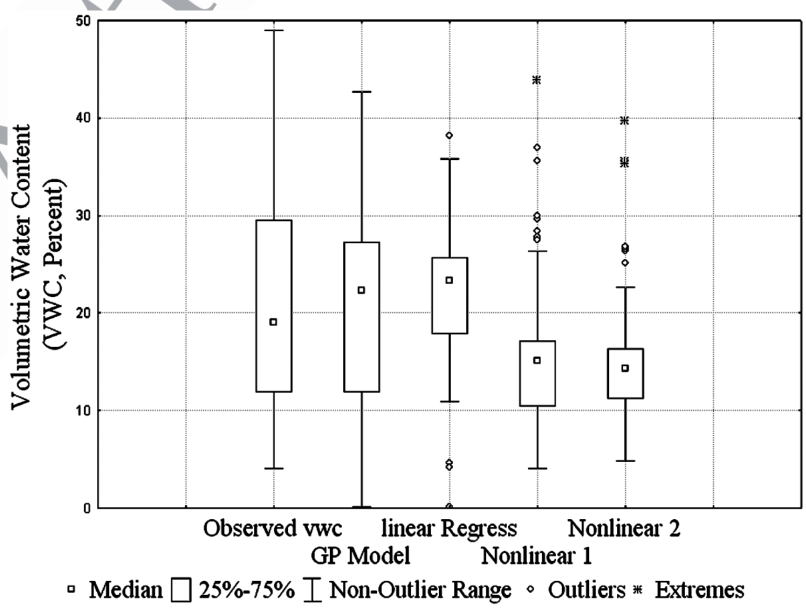

Figure 14. Box plots of observed and calculated soil moisture generated from different models. The result of GP model represents very similarly to the observed data. Its interquartile range is almost as wide as the interquartile range of the observed data, while the other models only result in a half of the interquartile range. Obviously the linear regression model results in the overestimation, and the two nonlinear models result in the underestimations.

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.

6. Conclusion

[38] 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.

[39] 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: (1) the application of new technologies for remote sensing hydrologic quantities for terrestrial hydrologic interpretation; (2) completion of studies on appropriate spatial and temporal sampling scales of new synergy of optical and microwave sensors for satisfying specific scientific objectives; and (3) 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.

[40] 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.

[41] **Acknowledgments.** The authors extend their deep gratitude to the NASA Grant support (NAG13-03008) sponsored by Glenn Research Center, NASA. The authors also are sincerely grateful for these invaluable comments raised by anonymous referees that contribute to the quality improvement of this paper.

References

Alaska Satellite Facility (1999), RADARSAT-1 standard beam SAR images, report, Geophys. Inst., Univ. of Alaska, Fairbanks.

Alaska Satellite Facility (2002), ASF interferometric SAR processor (AISP) calibration report, version 4.0, Geophys. Inst., Univ. of Alaska, Fairbanks.

Back, T., U. Hammel, and H.-P. Schwefel (1997), Evolutionary computation: Comments on the history and current state, *IEEE Trans. Evol. Comput.*, *1*(1), 3–17.

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

Chang, N. B., and W. C. Chen (2000), Prediction of PCDDs/PCDFs emissions from municipal incinerators by genetic programming and neural network modeling, *Waste Manage. Res.*, *18*, 341–351.

Cramer, N. L. (1985), A representation for the adaptive generation of simple sequential programs, paper presented at International Conference on Genetic Algorithms and Their Application, Am. Assoc. for Artif. Intel., Cambridge, Mass.

Das, B. M. (1999), Soil classification, in *Fundamentals of Geotechnical Engineering*, pp. 35–44, Brooks/Cole, Pacific Grove, Calif.

Del Frate, F., P. Ferrazzoli, and G. Schiavon (2003), Retrieving soil moisture and agricultural variables by microwave radiometry using neural networks, *Remote Sens. Environ.*, *84*, 174–183.

Distributed Active Archive Center (2005), AVHRR Normalized Difference Vegetation Index (NDVI) composites, U.S. Geological Survey EROS Data Center, Sioux Fall, S. D. (Available at http://edc.usgs.gov//products/landcover/ndvi.html)

D'Souza, G., A. S. Belward, and J. P. Malingreau (1993), *Advances in the Use of NOAA AVHRR Data for Land Applications*, *Remote Sensing*, vol. 5, Springer, New York.

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

Earth System Science Center (2004), STATSGO, soil information for environmental modeling and ecosystem management, Pa. State Univ., University Park. (Available at http://www.essc.psu.edu/soil_info/index.cgi?index.html)

ESRI (2004a), Spatial analyst: Deriving slope, in *ArcGIS 8.3 Help File*, Redlands, Calif.

ESRI (2004b), Spatial analyst: Aspect, in *ArcGIS 8.3 Help File*, Redlands, Calif.

Fogel, L. J., A. J. Owens, and M. J. Walsh (1966), *Artificial Intelligence Through Simulated Evolution*, John Wiley, Hoboken, N. J.

Francone, F. D. (1998), Discipulus™ software owner's manual, version 3.0 draft, Mach. Learning Technol., Inc., Littleton, Colo.

Freeman, A. (1992), SAR calibration: An overview, *IEEE Trans. Geosci. Remote Sens.*, *30*(6), 1107–1121.

Gagne, C., and M. Parizeau (2004), Genericity in evolutionary computation software tools: Principles and case-study, *Tech. Rep. RT-LVSN-2004-01*, Lab. de Vision et Syst. Numer., Univ. Laval, Quebec, Que., Canada.

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

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

Gupta, R. K., T. S. Prasad, and D. Vijayan (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 A. N. Zincir-Heywood (2002), Dynamic page based linear genetic programming, *IEEE Trans. Syst. Man Cybernet., Part B*, *32*(3), 380–388.

Holland, J. M. (1975), *Adaptation in Natural and Artificial Systems*, Univ. of Mich. Press, Ann Arbor.

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

Jeremy, N. (2002), SAR geometry and backscatter, paper presented at Technical Seminars, Alaska Satell. Facil., Univ. of Alaska, Fairbanks. (Available at http://www.asf.alaska.edu/educational/seminarpdf/2002/sar_geometry.pdf)

Koza, J. R., M. A. Keane, M. J. Streeter, W. Mydlowec, J. Yu, and G. Lanza (2003), *Genetic Programming IV: Routine Human-Competitive Machine Intelligence*, Springer, New York. (Available at http://www.genetic-programming.org/gp4chapter1.pdf)

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

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

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

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

Lu, Z. (2005), Satellite radar remote sensing: Applications to the study of Earth sciences and natural resources, paper presented at Technical Seminars, U.S. Geol. Surv./Earth Resour. Obs. and Sci., Sioux Falls, S. D. (Available at http://www.asf.alaska.edu/educational/seminarpdf/2005/satellite_radar_remote_sensing.pdf)

Moeremans, B., and S. Dautrebande (2000), Soil moisture evaluation by means of multi-temporal ERS SAR PRI images and interferometric coherence, *J. Hydrol.*, *234*, 162–169.

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

Narayanan, R. M., and P. P. Hirsave (2001), Soil moisture estimation models using SIR-C SAR data: A case study in New Hampshire, USA, *Remote Sens. Environ.*, *75*, 385–396.

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

Nolan, M. (2003), DInSAR measurement of soil moisture, *IEEE Trans. Geosci. Remote Sens.*, *41*(12), 2802–2813.

Olmsted, C. (1993), Alaska Satellite Facility scientific SAR user's guide: ASF-SD-003, 53 pp., Geophys. Inst., Univ. of Alaska, Fairbanks.

Owe, M., A. T. C. Chang, and R. E. Golus (1988), Estimating surface soil moisture from satellite microwave measurements and a satellite derived vegetation index, *Remote Sens. Environ.*, *24*, 331–345.

Rathert, D. (2003), GRIDSPOT, ESRI Support Cent., ESRI, Redlands, Calif. (Available at http://arcscripts.esri.com/details.asp?dbid=12773)

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

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

Sarabandi, K. (1994), Calibration of a polarimetric synthetic aperture radar using a known distributed target, *IEEE Trans. Geosci. Remote Sens.*, *32*(3), 575–582.

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

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

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

Small, D., F. Holecz, E. Meier, D. Nuesch, and A. Barmettler (1997), Geometric and radiometric calibration of RADARSAT images, paper presented at Geomatics in the Era of RADARSAT, Can. Inst. of Geomatics, Ottawa, Ont., Canada.

Song, D., M. I. Heywood, and A. N. Zincir-Heywood (2003), A linear genetic programming approach to intrusion detection, in *GECCO 2003, LNCS 2724*, edited by E. Cantú-Paz et al., pp. 2325–2336, Springer, New York.

Spectrum Technologies, Inc. (2004) User's manual, catalog #6430FS: Field Scout™ TDR 300 soil moisture meter, Plainfield, Ill. (Available at http://www.specmeters.com/pdf/6430FS.pdf)

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

Trimble Navigation Limited (2004), GEO XT, datasheet and specifications, Sunnyvale, Calif. (Available at http://www.trimble.com/geoxt.shtml)

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

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

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

Weier, J., and D. Herring (1999), Measuring vegetation (NDVI and EVI), *Earth Obs.* (Available at http://eobglossary.gsfc.nasa.gov//Library/MeasuringVegetation/index.html)

Wigneron, J. P., J. C. Calvet, T. Pellarinb, A. A. Van de Griend, M. Berger, and P. Ferrazzoli (2003), Retrieving near-surface soil moisture from microwave radiometric observations: Current status and future plans, *Remote Sens. Environ.*, *85*, 489–506.

Williams, J. (2004), Introduction to uncalibrated data, calibration/validation, report, Alaska Satell. Facil., Geophys. Inst., Univ. of Alaska, Fairbanks.

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

Yeh, G. T., H. P. Cheng, J. R. Cheng, H. C. Lin, and W. D. Martin (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), *Tech. Rep. CHL-98-19*, U.S. Army Corps of Eng., Waterw. Exp. Stn., Vicksburg, Miss.

Zribi, M., and M. Dechambre (2002), A new empirical model to retrieve soil moisture and roughness from C-band radar data, *Remote Sens. Environ.*, *84*, 42–52.

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

M. Beaman, The Conrad Blucher Institute for Surveying and Science, Texas A&M University, Corpus Christi, TX 78412, USA.

N.-B. Chang, Civil and Environmental Engineering Department, University of Central Florida, 4000 University Boulevard, Orlando, FL 32816, USA. (nchang@mail.ucf.edu)

A. Makkeasorn, Environmental Engineering Department, Texas A&M University, Kingsville, TX 78363, USA.

C. Slater and C. Wyatt, Alaska Satellite Facility, Geophysical Institute, University of Alaska, Fairbanks, AK 99775, USA.