LEPF-0010A

Final Report

Water quality simulation modeling using SWAT in the Sandusky Watershed

PI: Kevin P. Czajkowski
Co-I: Sabine Grunwald, University of Florida

Correspondent's Address:
Kevin Czajkowski
Department of Geography and Planning
The University of Toledo
Toledo, OH 43606
419-530-4274 (o)
fax: 419-530-7919
kczajko@utnet.utoledo.edu

University of Toledo Administrative Contact:
Dr. Frank Calzonneti Vice Provost For Research
Office of Research
The University of Toledo
2801 West Bancroft Street
Toledo, Ohio 43606
Phone: (419) 530-4749
Fax: (419) 530-2841
FCalzon@UTNet.UToledo.Edu

This project, "Water quality simulation modeling using SWAT in the Sandusky Watershed", was completed by Kevin Czajkowski of the University of Toledo and Sabine Grunwald from the University of Florida. All data set and modeling objectives of the project were met under the current grant. In the following report, we lay out the tasks that were involved in the project, discussion development of the GIS data layers and then present results from the SWAT model runs.

Task 1:
We assembled GIS-data layers for the land cover, digital elevation model (DEM), soils, and weather to use within SWAT (Soil and Water Assessment Tool).

Task 2:
We simulated transport of suspended sediment and agrichemicals to assess non-point source and point source pollution in the Sandusky Watershed utilizing SWAT for the years 1999-2000 focusing on suspended sediment, Nitrate-nitrogen, Nitrite-nitrogen,

Ammonia-nitrogen, Total-phosphorus, Soluble phosphorus, Soluble pesticides, and Pesticides (atrazine, metolachlor).

Task 3:
We compared water quality data provided by the Water Quality Laboratory (WQL), Heidelberg College with simulated data.

Task 4:
We produced maps covering the Sandusky Watershed showing the spatial distribution of water quality impact for different stressors.

Task 5:
We are currently sharing this information with other researchers, government agencies and local action groups working on the Sandusky watershed, namely Anthony Friona from the Army Corps. of Engineers, Thomas Croley from NOAA's Great Lakes Environmental Research Lab (GLERL), Christopher Riddle from the Sandusky River Watershed Coalition and Chengshen He of Western Michigan University.

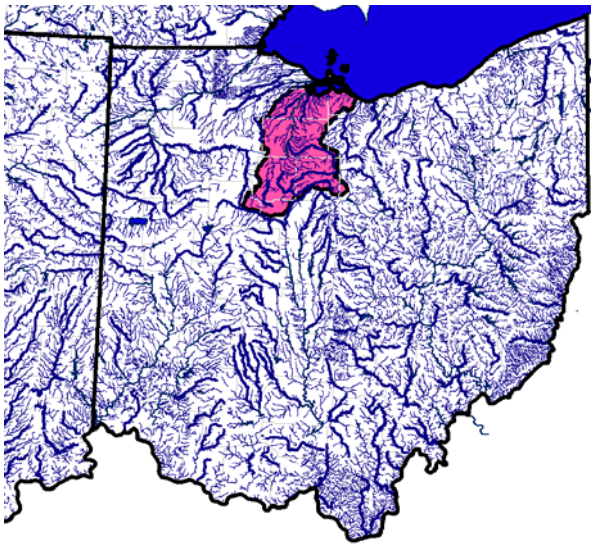
Land Use/Land Cover Using Multi-spectral Image Classification

The Sandusky River Drainage Basin has 1.17 million acres. The basin drains directly into Lake Erie. The basin has an approximate population of 225,000 people. The drainage basin covers 5 counties in north central Ohio. The region is best characterized as rural with a few smaller cites and towns, the largest being Fremont, OH. There is extensive agricultural land use in this region. Sugar beets, Tomatoes, cabbage, corn, soy, and wheat make up the bulk of the crops grown. There are some grazing for cattle, sheep, horses, and alpacas, but these land uses are limited in scope in this region. Row crops make up a bulk of the farming. Large tracts of wetlands and riparian vegetation and trees are also found in the drainage basin.

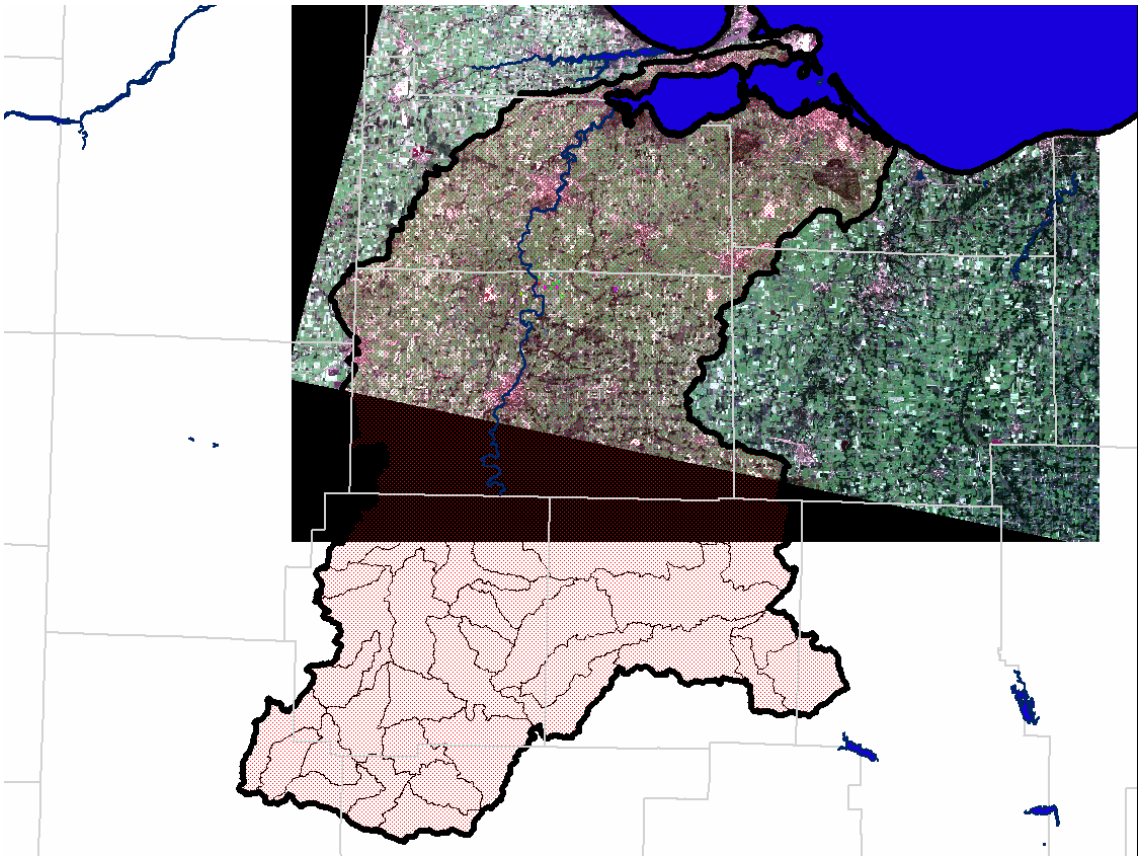
This project used data imaged from the Landsat 7 Enhanced Thematic Mapper Plus (ETM+). All images were either obtained from the UT archive or downloaded via the Ohio Library and Information Network (OhioLINK's) online data warehouse. OhioLINK, currently purchases all available Landsat 7 imagery taken over Ohio for the OhioView Remote Sensing Consortium if the sky is less than 30% cloud covered. OhioView is a consortium of eleven universities in Ohio whose goal is to spread the use of remote sensing data. This imagery extends back to the beginning of the Landsat 7 collection period, June 1999. This multi-spectral imagery is made freely available through the OhioLINK and OhioView websites, http://dmc.ohiolink.edu/GEO/LS7 http;//www.ohioview.org. Landsat 7 records 8 spectral bands. A spectral band is the discrete range of the electromagnetic spectrum that a sensor records. For instance, band 1 on the ETM+ ranges from 450 nm to 515nm. The spectral resolution of the ETM+ ranges from 450 nm to 1250 nm. The spatial resolutions for this scanner are 30m for the multi-spectral, 60m for the thermal and 15m for the pan-chromatic band. The Landsat 7 satellite orbits the earth every 99 minutes. The satellite covers the same point on the earth every 16 days, at approximately the same time of day. The Landsat program was started in 1972, and is designed for a "long term acquisition" of the earth's resources. Each image represents a swath width of 185 km. The spatial extent of the image combined with the

synoptic and repetitive coverage make this the ideal sensor with which to classify agriculture and non-agricultural land uses.

The Sandusky drainage basin is represented in 2 Landsat scenes. In the Worldwide Reference System (WRS) images path 19 row 32 and path 19 row 31 cover the Sandusky drainage basin.



The Sandusky Drainage Basin, in pink, in North Central Ohio



The Basin over the Landsat 7 image Path 19 Row 32.

The second step involved an inspection of the available images to determine which images met a pre-specified set of requirements. The primary consideration for inclusion in the stack was that the images included in the temporal stack needed to be relatively free of cloud cover. All of the Landsat 7 images taken of path 19 row 32 from July 1999 were evaluated for inclusion in the temporal stack image. Images obscured with clouds over a significant portion of the area of interest--within the basin-- were excluded. All of the images made between March and November evaluated. This range of dates was designed to include the entire growing season. Multiple years were used in this initial stack. Again, the point of this classification was to isolate regions of agriculture and non-agricultural land uses. It is assumed that an area that is used for agriculture in 1999 is still likely to be an agricultural field in 2001.

After each image was checked to insure sufficiently low cloud cover the sample sites were located within the boundary of the northern most image, path19 row 32. The southern half of the drainage basin, in the Landsat 7 image P 19 R 31, consequently did not have any corresponding ground truth data. To compensate for this absence of data the lower Sandusky image was classified using training sets extracted from the northern image.

Alignment of the spatial features between different dates is of paramount importance to insure a good classification. Any spatial variation between images will cause the classification algorithm to not properly classify the image stack. Landsat images made available through OhioView come geographically corrected and have a spatial error of less than one pixel. These images are projected in to the Universal Transverse Mercator (UTM) zone 17.

Crop Type from Multi-temporal Multi-spectral image classifications

The type of crop planted by a farmer can significantly impact the amount of runoff of a given location. To accurately map the land use in terms of individual crop type it became necessary to explore a more complex classification and information extraction series of techniques. Based on the information created in the first phase of this project it was obvious that a single Landsat multi-spectral scene would not yield the desired level of detail for the SWAT modeling. We developed crop rotation including soybeans, corn and wheat. Beans, corn, wheat cannot be differentiated with a single multi-spectral image scene. For this portion of the study it was necessary to involve a many image or multi-temporal approach in identifying crop types. Multiple scenes or multitemporal satellite images can be used to detect the seasonal variation characteristics unique to each of the target species and thus be used to differentiate crop type.

Atmospheric correction

In order to use a series of images together in a classification, atmospheric effects of scattering and absorption must be taken into effect. Dark object subtraction uses the darkest object in the scene to adjust the brightness values of all of the pixels in the image. The dark object is considered the base or zero value in terms of Electro-magnetic energy reflectance. Any value over zero of that feature in the image is considered to be due to atmospheric scattering. The difference is considered the atmospheric scattering component and is then applied to all of the pixels in the image in a subtraction of the

scattering value. ENVI by RSI Kodak was used to implement the dark object subtraction of all of the bands of all of the images.

The Dark object subtraction assumes uniform scattering throughout the scene. Though this is a rather broad assumption it did not detract from our choice to use this method. The region being processed is comparatively small so that this assumption can be considered valid. It is a simple straightforward method that proves easy to implement and reduces complex and time-consuming computations associated with more involved atmospheric modeling methods. Other more involved techniques may prove to increase the accuracy of the final product, however the cost in terms of implementing more complex atmospheric modeling should be weighed carefully against expected returns in increased accuracy.

Each band was atmospherically corrected using the Dark object subtraction. It should be noted that dark objects may not remain the darkest object through the entire spectrum. Careful selection for each band and each date must be made to insure that a good dark object is chosen. The resulting appearance of a Dark subtracted image is similar to a contrast stretched image. The former changes the digital value of each pixel where as a contrast changes only the appearance of the image.

Minimum Noise Fraction.

Processing data that is made up of many discrete bands is difficult. Many bands pose problems related to the storage and processing of the data. Hyperspectral data, images made up of many—generally over 30—discrete bands pose unique challenges in terms of storage and processing. Given that the data held in the many bands of a single image scene is very often redundant and thus contributes very little to the final data extraction process a method that can reduce the scale and noise of a complex, many-layered data set is an indispensable step in the image processing scheme of hyperspectral imagery.

Principal components Analysis provides a means of reducing the size of the data while focusing the variation of the many bands into a smaller dimensioned image data set. A further refinement on the PCA technique combines 2 PCA in a cascading manner that significantly reduces noise and concentrates the bulk of the inter-band variability into far fewer bands. This process is called Minimum Noise Fraction (MNF) and is a statistical noise reduction technique and key component of hyperspectral image processing. The MNF was used to process the Multi-temporal multi-spectral imagery created for this project.

Multi-temporal data too has many bands of data much of it fairly redundant in terms of inter-band variation. To reduce both the size and noise of the data a MNF was run on each of the multi-temporal stacks. A subjective analysis was done on the resulting Eigen—a measure of the contribution to the variation of the resulting PCA or MNF band. These values are used to determine the threshold of the resulting MNF bands on which to perform the supervised classification.

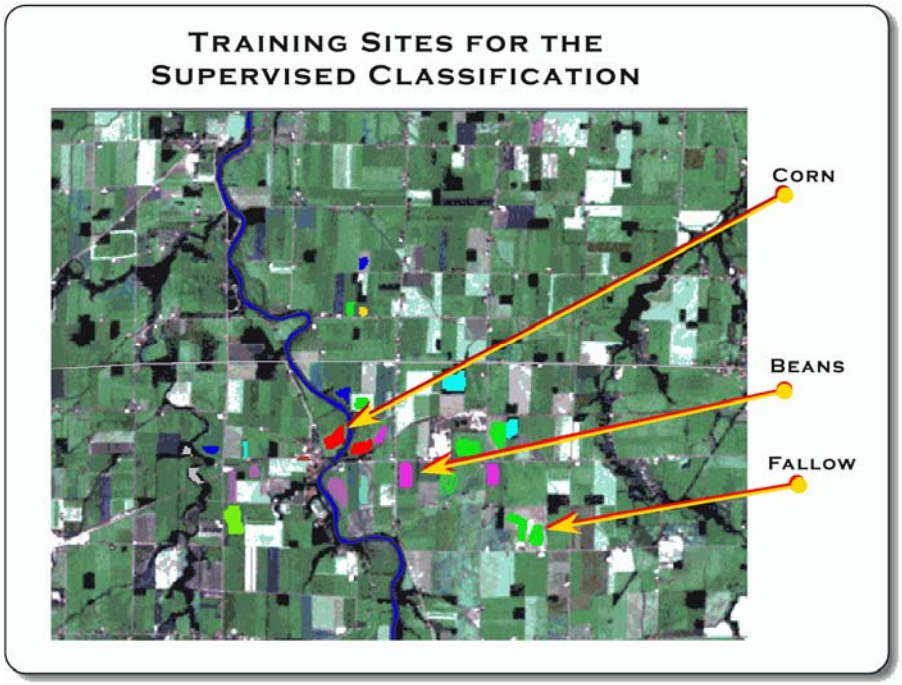
Each of the stacks resulted in a different number of bands in the final MNF file. This was due in part because the Eigen values differ from year to year on the multi-temporal stack. A natural break in the Eigen values was sought so as to determine the threshold or cutoff for usable MNF bands. More images in a given year increased the number of bands in the MNF. The MNF output for a given year with the higher number of bands was the initial output of the MNF transformation and will have a total band number divisible by 5 (because the input images all contained 5 bands each) in the same

year the MNF file with the lesser number of bands was the spectral subset of only the useful bands in that image. The MNF puts all the noise from the image in the higher resulting bands which is considered the noise component and discarded.

Supervised classification

Extracting the information from the imagery involved a supervised classification in much the same manner as was done in the first phase of the project. Training for the supervised classification were pulled from a variety of sources. Twenty-six ground validation points representing the target agricultural classes were recorded with a digital camera taken *in situ* and located with a GPS. Further classification for non-agricultural sites were added using the high spatial resolution component of the Landsat image (band 8) to identify, roads, quarries, and water features. The *in situ* samples represented a wide range of agricultural crops and land use types found in the Sandusky Drainage Basin. This information was then imported into ArcView GIS so that sample point could be overlain the image for a systematic training of the supervised classification in ENVI.

The number of pixels contained in a training set was representative of the size of the urban area being classified. All sets within a single urban classification were +/- 20 pixels of each other. Most sets resided between 60 and 100 pixels.



These and many other sites were empirically derived and used to associate those pixel values with a particular feature type.

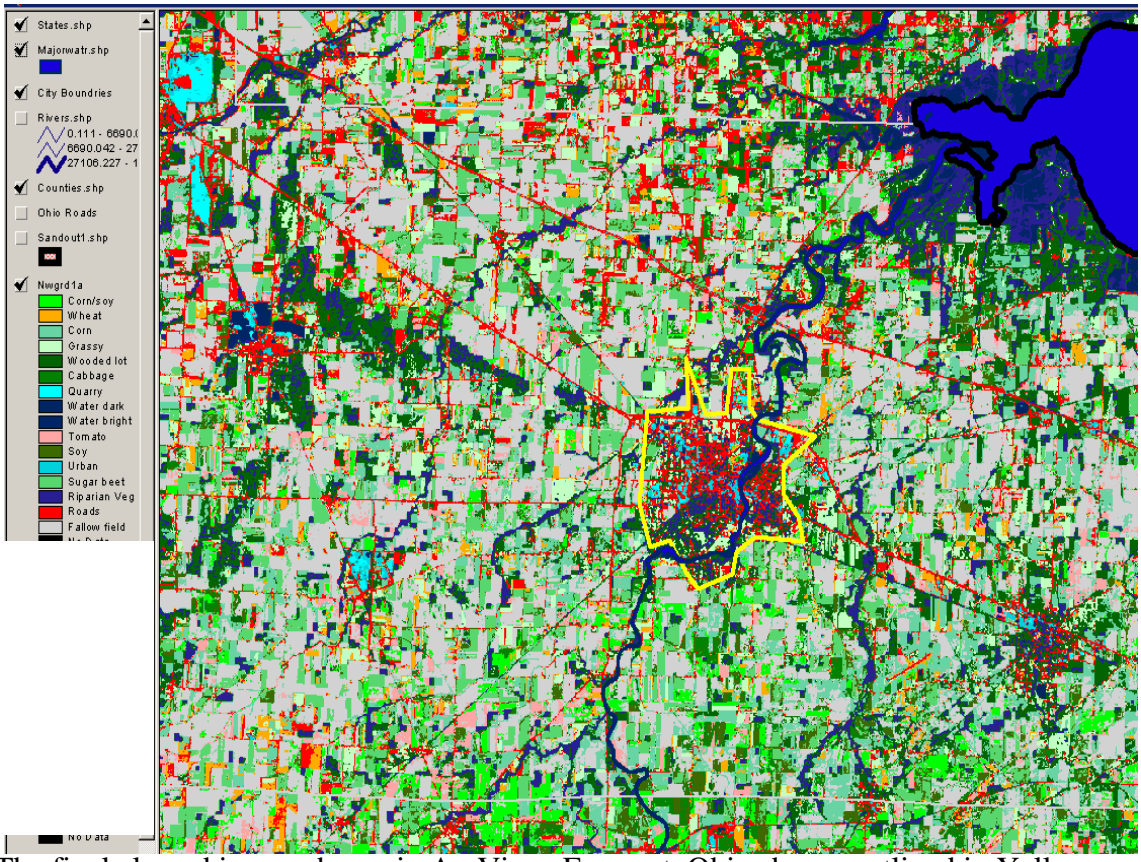
Each image was spatially subset to exclude areas that fell outside of the drainage basin. The image was spectrally subset as well. Bands 6 and 8 the thermal and pan-chromatic bands, were removed from each of the component images. A boundary was defined on each of these component images that represented the extent of the Basin within the image. The component images were stacked into a single image using ENVI. The resulting image had 35 bands. Each date included in the stack was represented by 5 of the bands. Seven images were used in this final image. A supervised classification was

performed on the image. The 26 agricultural sites were extracted from the GIS and incorporated in the classification routine. A minimum distance classification was used with 16 classes.

A list of the classes included in the classification

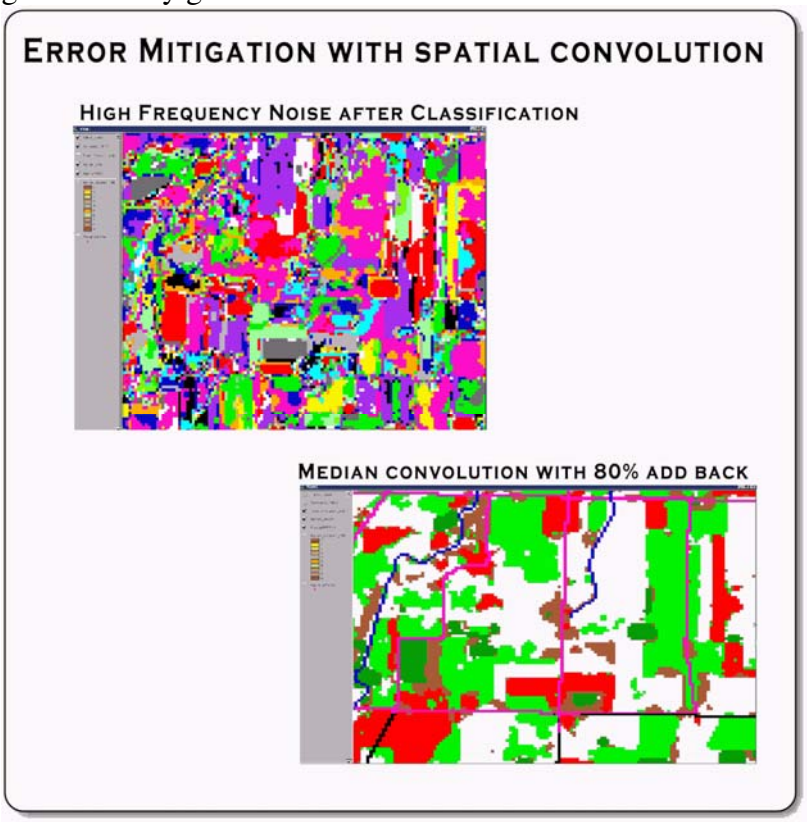
1. Corn/Soy
2. Wheat
3. Corn
4. Grass
5. Wooded lot
6. Cabbage
7. Quarry
8. Dark water
9. Bright water
10. Tomato
11. Soy
12. Urban
13. Sugar beet
14. Riparian Vegetation
15. Roads
16. Fallow fields

The resulting image was the converted into an image file format native to ArcView GIS. The image in ArcView GIS was then converted into a Spatial Analyst Grid file. The GRID format is the native GIS raster format for ArcView.



The final classed image shown in ArcView. Fremont, Ohio shown outlined in Yellow.

The ground control information was then applied to the supervised classification routine in RSI ENVI. A maximum likelihood classification was used to cluster pixels into specified classes. Each class represents a specific type of crop. This is possible because unlike the single date image pixel values, multi-temporal pixels vary with the seasonal variation specific to individual crop types. The resulting classification of the image stack required some post processing to reduce errors associated with the classification process. Atmospheric anomalies, shadows, location and feature type all can contribute to a misclassification of the resulting thematic pixel. A convolution routine was used to help mitigate some of these errors. High frequency noise may appear as a single urban classed pixel placed in the middle of a cornfield. It is very unlikely that this pixel is truly representative of the actual ground feature. By comparing each pixel with its neighbors a more accurate classification may be achieved. Though some subjective consideration must be applied for each classed image, it was found that a median convolution with an 80% add back of original pixel values was the best at reducing high frequency classification errors without obliterating patterns of features known to empirically exist, resulting in an overly generalized feature set.



An example of the effect of a median convolution.

Delineation of Subwatersheds

The delineation of watersheds and subwatersheds. The delineanation is created by using the digital elevation. SWAT then creates the number of subwatersheds that are specified. Attached is a powerpoint file showing the watershed delineation, the existing delineations courtesy of USGS and the delineations from SWAT. The SWAT version has more detail, probably since the 30 meter DEM was used to delineate each

subwatershed. We can of course test both versions when we begin modeling the water quality.

Weather Data

Input of weather station data for the simulation period was gathered from of the NOAA weather data from 1/1/1989 - 12/31/2001 to cover the 3 different time periods where we have land use data. Over time the stations have moved so dual entries have been used for some of the weather stations. SWAT was run on a daily time step so the weather data was inputted daily. From the NOAA stations we collected observed min temp, max temp and precipitation for every day.

In addition SWAT needs less commonly measured variables such as wind speed and humidity, these are simulated. In SWAT there is a national dataset of weather stations with the monthly min, max and average for all the weather variables that SWAT needs. Using this information we simulated the weather. For both observed and simulated weather data the nearest station to the center of each subwatershed is used to characterize the weather for the subwatershed.