2008 IEEE International Conference on Data Mining Workshops

# Kernels for the Investigation of Localized Spatiotemporal Transitions of Drought with Support Vector Machines

*Matthew W. Collier[1] and Amy McGovern[2]*

## Abstract

*We present and discuss several spatiotemporal kernels designed to mine real-life and simulated data in support of drought prediction. We implement and empirically validate these kernels for Support Vector machines. Issues related to the nature of geographic data such as autocorrelation and directionality are investigated.*

## 1. Introduction

It is the purpose of this work to increase the toolset available for use in spatiotemporal data mining scenarios through the development of geographically sensitive kernels for Support Vector Machine (SVM) data mining. We are motivated by drought, an application that is meaningful to society and varies in both temporal and geographic space. Drought is a common natural hazard that, on average, results in multibillion dollar disasters nearly every year in the United States alone [7]. The recent Intergovernmental Panel on Climate Change (IPCC) report states that the severity of the impacts of drought will may be increasing through the effects of a changing climate [5]. Second, it is clear that drought varies in both space and time. A casual look at the archived maps produced by the US Drought Monitor (as in Figure 1) will show drought as a phenomenon that exhibits dynamical space-time behavior through merges, splits, growth, & decline, and provides a rich set of behaviors that researches may query [18].

Our general development process is two-fold. First, new kernels will be created that exhibit spatiotemporal biases. Our current work is development of kernels that are sensitive to autocorrelation, and directionality of spatial phenomenon (e.g. orientation). For instance, kernels that take advantage of autocorrelation by emphasizing local structure are implemented. Furthermore, spatial kernels that emphasize directionality such as might exist in an east-to-west flow pattern are created and tested as outlined in the methods section below. Second, once the constructions are achieved, are tested against the expected effects in geotemporal data space regarding autocorrelation, and directionality.

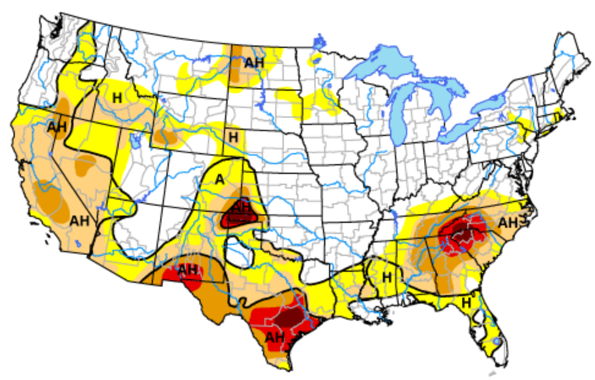

**Figure 1 - The US Drought Monitor map, valid July 8th, 2008 [16]. The darker red regions indicate more intense conditions of drought while the yellow regions are merely abnormally dry. Go to http://www.drought.unl.edu/dm/ for current maps and a full discussion of the Drought Monitor.**

The field of geospatial technologies is one of the top ten upcoming industries of this age. Valued at over $10 billion annually and growing at over 10% per year [6], the application space in terms of both volume and diversity has great potential. With private GPS technology, there is an explosion of geo-enabled data becoming available. [19, 14]. Although much of this geographical data is time stamped and thus could potentially be embedded within spatiotemporal processes and patterns, the potential to mine this data for patterns and processes remains largely untapped due to the nascency of spatiotemporal data mining techniques.

[1] Department of Geography, University of Oklahoma. Contact: mwc@ou.edu
[2] Department of Computer Science, University of Oklahoma. Contact: amcgovern@ou.edu

978-0-7695-3503-6/08 $25.00 © 2008 IEEE
DOI 10.1109/ICDMW.2008.71

## 2. Review

### 2.1. Geographic Data

The exploration of phenomenon such as drought distributed in geographic space presents certain unique challenges and opportunities. These data are subject to the effects of sampling, scaling, periodicity, fractal dimension, direction, and autocorrelation [6].

Recording geographic data necessarily implies that some form of sampling has been used as simplifying choices have been made. In terms of the drought data that we are using, the samples are of interpolated drought intensity values across a regular geographic grid of latitude and longitude similar to that illustrated in Figure 2. Thus, our samples consist of one half degree on a side quadrilaterals oriented in a North-South, and East-West fashion. For latitudes similar to those of the United States, this implies roughly 50 km resolution on the ground in both latitude and longitude. With this sample size, we cannot detect spatial structure less than about 100 km in diameter.

Scale, of potentially several meanings in the geographic world [6], in the present study refers to the level of spatial detail that can be discerned in data. Scale is directly impacted by the sampling as the finest level of discernable detail will be no smaller than twice the nearest distance between sample points. However, it is also impacted by the size of any window used for a neighborhood view of the data. Referring to our drought data for an example, if we let a moving window of nine grid cells determine the output value of a new map our results will be different than if we looked at a different scale of say, twenty five grid cells. Also, scale changes when re-sampling through aggregation. This is an important operation that must often be performed on geographic data to ensure that dissimilar data can be appropriated overlaid.

Geographic data may also have repetitive features such as streams that occur every few miles that drain parallel valleys. This repetitive nature may even extend into the realm of fractal geometry. The classic geographic example of a fractal object is of a coastline. Depending on the length of the stick you use to measure the coast, you'll come up with different answers. The shorter the stick, the more variations in the coast that are accounted for and the longer the measurement becomes.

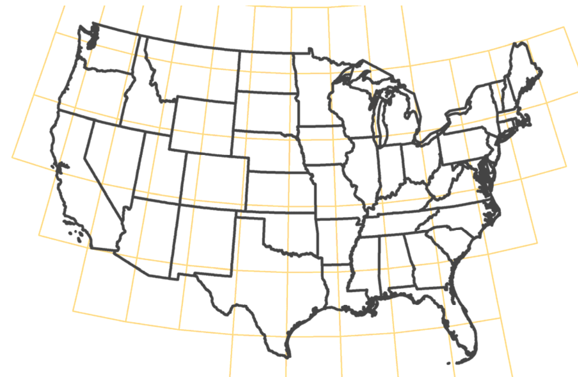

**Figure 2 - A five degree geographical grid overlaid on the contiguous United States. The actual grids used in our study were one half of a degree on a side for the real world SPI data. For the simulated drought data, both one half & two and a half degrees grid spacings were used. The grid above is not aligned with our actual grids and is displayed for illustrative purposes only.**

While these critical characteristics of geographical data must always be considered as potential influences on the outcome of any geographical study, they are not the focus of this study. The last two characteristics, direction and autocorrelation are treated in more detail in the present work.

Direction matters in geography because the real world is almost never isotropic. Distributions and flows can align themselves to wind patterns, geologic depositions, human constructed highways, and other isotropy altering structures.

Autocorrelation is the tendency for values sampled "close" to one another to be similar. In fact, in the geographic world this is codified by Tobler's 1$^{st}$ Law of Geography which states that all things connect in some way, but close objects tend to share more similarities [6]. Some authors have noted that some might say that without autocorrelation, geography wouldn't even matter at all [9]. Thus, while in the traditional world of statistics, autocorrelation is to be eschewed, in the geographic world it reveals underlying spatial structure and may thus be useful. Also, in traditional statistics we are often dealing with only samples of the space in question. In geography the samples are often the entire population of interest and thus no loss of information occurs through the presence of autocorrelation [9]. In the example of our drought data, the grid cells are exhaustive and mutually exclusive in tiling geographic space and thus consist of the population of interest. Unfortunately, autocorrelation depends very strongly on sampling and scaling of geographic data. This can render the detection and discussion of spatial autocorrelation an art at best. We proceed with the idea that the autocorrelation is present and is part of the predictive power of the data, ignoring its presence and making no attempt at removing it. Preliminary experiments with the drought data indicated that removing the autocorrelation through sampling did not affect the results.

In summary, while not unique to geography, these aspects of geographic data and more make geographic data what it is, complex, and thus a rich source of information to the data mining community. We will specifically focus on the autocorrelation and direction aspects of geographic data in this study, hoping that others (or ourselves in the future) will attempt further studies on all of the discussed characteristics of geographic data and the effects they have on data mining and machine learning.

### 2.2. Geographic Kernels

Geographic kernels are, through their design, sensitive to data in geographic space. That is, data that is a function of latitude and longitude, or more generally Y, and X. By sensitive we mean that the spatial structure in the data takes part in the predictive power of the learning machine and structure may be potentially revealed by the appropriately chosen kernel.

Gaussian kernels should be, by their very nature sensitive to autocorrelation effects. This is because as the distance falls off, less weight is given to each input. The standard Gaussian kernel is well known in the machine learning world and uses the distance between point locations calculated as a difference between two radial vectors.

Recently, Pozdnoukhov and Kanevski [11] used kernel dictionaries wherein several Gaussian radial basis functions are made available to an SVM, each of which has a different characteristic radius. In effect, their SVM can probe across various geographic scales in order to find the kernel best suited for the data at hand.

## 3. Methods

### 3.1. Data: Simulated & Real

Our real world drought data is available online from the IRI/LDEO Climate Data Library[3]. The original dataset covers the world in geographical extent and from January 1901 to December 1998 in temporal extant and was developed from meteorological data gridded by New et. al. [8]. The gridded data is at one half degree spacing in both longitude and latitude and covers all major land masses except Antarctica. Of particular note in geographic space is the convergence of the East-West width of each grid cell. While each cell is of fixed width in longitude, the linear distance on the ground diminishes to zero at the poles. Data spacing in the temporal extent is one month and covers 1176 months in duration. Since the gridded values represent interpolation from real world meteorological stations that may not have been in existence over the entire temporal extend of the dataset, not every grid cell contains data for all times.

For our study, a subset of the global data was created in spatial extent that only covers the contiguous United States. This dataset, of less than 70MB in size, is more manageable and is chosen as representative of the approach.

This real data consists of values of the Standardized Precipitation Index (SPI). While literally hundreds of drought indices have been defined due to the complexity of the phenomenon, this particular index stands out as particularly useful for studies in geographic space. This is because it has been shown to be spatially invariant [3]. By way of example, an SPI value of negative one in Eureka, Utah has the same meaning to as a value of negative one for the SPI in Thibodeaux, Louisiana. This is despite the fact that the two locations have dramatically different topography and climate. The reason for the invariance is that the SPI is actually a probability distribution for rainfall that is based on local rainfall statistics. This characteristic of the SPI is of great importance in our extended future work because we wish to compare and contrast drought across geographic regions.

It is desirable to mine the real world data to see what types of structures and predictions can be made regarding drought. However, real world data may be incomplete and contain various amounts of noise. For this reason, simulated drought data was also constructed so that we could compare algorithmic performance against data whose properties are presumably well-known, or at the least complete and mathematically predictable.

We constructed artificial drought data to cover spatiotemporal extents similar to the real world ata. The data consists of a sinusoidal wave traveling across the region from an epicenter near the eastern edge of the geographical extent. The functional from of the simulated drought's intensity can be seen in this equation,

$$I(r,t) = \sin(kr - \omega t), \qquad \text{eq. (1)}$$

[3] http://iridl.ldeo.columbia.edu/SOURCES/.IRI/.Analyses/.SPI/

where r is the distance from the epicenter, k is the spatial wave number (or, $2\pi/\lambda$ where $\lambda$ is the spatial wavelength), and $\omega$ is the angular frequency (or, $2\pi\nu$, where $\nu$ is the temporal frequency). The temporal frequency (0.02 months) and spatial wavelength (10 degrees) were chosen to move a simulated drought event across a region of about 1000 kilometers in diameter in the time span of a few years (a few tens of months) in order to mimic typical movements of drought as delineated by the US Drought Monitor. Our simulated data varies between +/- 1.0 in value. This is of the same order of magnitude as the real SPI data which has values roughly between +/- 2.3 and is unitless [4].

Finally, several simulated datasets were created by varying the levels of noise from a Gaussian noise with mean zero and a standard deviation of one. While extreme values may be present in this distribution, 'typically' they are not. This noisy data is intended to provide an idea of how robust our algorithmic approach may be in the presence of data imperfections. The amount of noise was varied by dividing the output of the generator by the values of infinity (effectively no noise), 30, 10, 3, 1, 1/3, and 1/10.

## 3.2. Drought

A careful definition of what is meant by drought is required in any study. Conceptual drought characteristics vary by geographic region, by cultural response, and by the affected sector. The major sectors are meteorological, agricultural, hydrological, and socioeconomic [17]. Furthermore, an operational definition must be established for computational purposes. The meteorological sector has a long history of excellent and extensive data capture. Furthermore, it is simpler to work with meteorological drought since the complicated impacts to human and natural systems are avoided. For these reasons, droughts as defined from the meteorological sector are chosen for this study.

The specific question we ask of our SVM is whether or not the next month will be drier or wetter than the current month at the center cell of the geographic window representing the input vectors. In other words, is $SPI_{t+1}$ greater than or less then $SPI_t$ in value at the center of the geographic window? Thus, the question asks whether or not a grid cell is moving towards or away from a condition of drought.

## 3.3. Kernel variations

As noted earlier, the purpose of this study is to vary kernels in order to search for spatial structure with SVMs. Two specific types of spatial structure in drought data will be probed in this work: autocorrelation, directionality. Each of these is best explained in terms of a concrete example. As mentioned before, spatial autocorrelation is an integral part of geographic space. A Gaussian weighted linear kernel is one way to mathematically code Tobler's 1st law and such a weighted kernel may be used to probe the response of the drought predictions based on local data providing a stronger influence on the outcome of the prediction..

To introduce the Gaussian into our geographic window, it is necessary to go beyond the Gaussian radial basis vectors which work between points and find a way for Gaussians to work in an areal fashion. We start with the well-known x-squared (polynomial of degree 2) kernel and modify it so that it can sense spatial structure by using a Gaussian Hat (GHat). For example, in our approach the grid cell of interest is taken along with its eight surrounding cells to form a 9x1 re-dimensioned input vector to the SVM with the vector's label coming from the future state of the central cell. We weight the input from these cells with the Gaussian function according to their distance from the center cell. A typical GHat for a 3x3 grid cell geographic window might look like Figure 3 below,

| | | |
|---|---|---|
| **0.513** | **0.717** | **0.513** |
| **0.717** | **1.000** | **0.717** |
| **0.513** | **0.717** | **0.513** |

**Figure 3 – Example of a Gaussian hat construction for a three by three geographic window. Equation 2 was used for this illustration with A set to 1 and $s^2$ set to 3. More generally, equation 2 may be used to fine tune the Gaussian hat for a particular purpose.**

where the weighting is given as in the following equation,

$$G_{ij} = A\exp[\, -d^2/s^2 \,], \qquad \text{eq. (2)}$$

where d is the distance from the center cell measured between cell centers, A is an amplitude scaling constant, and the ij subscript on G refers to the cell's location with respect to the central cell. While s is typically thought of as related to the standard deviation of the data, any useful value may be used to vary the shape of the Gaussian surface.

This type of Gaussian Hat is easily extended to larger windows around the cell of interest, e.g. 5x5 or larger or even non-square windows. The GHat is introduced to the x-square kernel through the well-known property that as long as the new kernel is still an inner product, it is still a valid kernel as described in Bishop [1]. We specifically use the property found in equation 3 below and given in Bishop,

$$k(\mathbf{x},\mathbf{x'}) = \mathbf{x}^T\mathbf{G}\mathbf{x}, \qquad \text{eq. (3)}$$

where the boldface **x**'s are appropriately re-dimensioned input vectors taken from an m by n window in geographic space and the boldface **G** is the appropriately re-dimensioned matrix derived from a spatial context similar to Figure 3 above.

For this study, the parameters of the Gaussian are chosen to maintain the shape of the surface relative to the size of the geographic window. In other words, if the window is three by three grid cells in size, and if a divisor of $s^2 = 3$ (as in equation (2) above) has been chosen, then if we change the window to a five by five grid, then $s^2$ becomes 5 as well.

An example of directionally biased geographic data would be isotherms (or, more generally isolines) on a weather map. Similar temperature values occur near the isotherm and change more rapidly as a function of orthogonal distance from the isotherm. Locally, the similarity along one direction is the spatial property we seek. It is expected that spatial bias can be revealed through the use of non-square windows. For instance, we define a three by seven window as containing three cells in the Y (or longitude) direction, and seven cells in the X (or latitude) direction.

To introduce this functionality, we develop a bi-variate GHat kernel by properly tuning its parameterization to respond strongly to biases lying in a given direction, thus revealing underlying spatial structure. To do this requires two parameters: one controlling the elongation in the East-West direction (EW), and one controlling the elongation in the North-South direction (NS). The mathematical form is illustrated here,

$$W = A\exp[\, -(y^2/b^2 + x^2/a^2)], \qquad \text{eq. (4)}$$

where A is defined the same as in equation (2), b is the parameter controlling the elongation in the Y direction, and a controls the elongation in the X direction. The completely general form of the bi-variate Gaussian is more complicated than what is shown here. However by choosing to look only along the cardinal directions, the cross terms disappear and we are left with the equation above. In Figure 4 below, we see an isoline of a bi-variate Gaussian that is oriented along an EW direction (assuming North is up). An SVM using this kernel may be expected to perform better prediction than one using a NS biased kernel in the presence of data with EW structure.

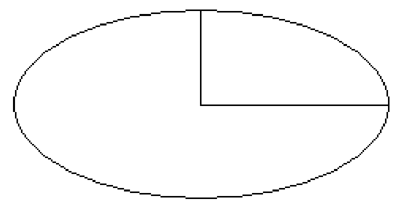

**Figure 4 - Isoline of a bi-variate Gaussian is illustrated alone with the semi-major and semi-minor axes. The kernel has varying spatial extents along orthogonal directions.**

### 3.4. The Algorithmic Approach

We chose to implement the least squares support vector machine algorithm as outlined in Suykens, et. al. [15], but without the sparse approximation. Python is the language of choice for implementation because of its rapidly growing influence in the Geographic Information Systems world, and its ability to import the SPI data which is in netCDF [12] format through the use of freely available, third-party modules. The last one fifth of each dataset was used for testing purposes. In other words, if we had 100 years of data, the first 80 would be used to train the SVM and the last 20 would be used to test the fit. With 1176 months of real SPI data and 1200 months of simulated drought data, each test set had over 200 examples in it.

## 4. Results & Discussion

### 4.1. The Question Matters

In the process of running scenarios we encountered a most interesting observation. In running the Least Squares Support Vector Machine (LSSVM) as described by Suykens (2000) against random data generated across a normal distribution, we were witnessing much higher skills than we anticipated

| Threshold | True Positive | False Positive | False Negative | True Negative | HK Score |
|---|---|---|---|---|---|
| 0.0 | 46 | 60 | 66 | 68 | -0.06 |
| -0.1 | 86 | 88 | 35 | 31 | -0.03 |
| -0.3 | 138 | 100 | 1 | 1 | 0.00 |
| -0.7 | 176 | 64 | 0 | 0 | 0.00 |
| -1.0 | 199 | 41 | 0 | 0 | 0.00 |
| -1.3 | 222 | 18 | 0 | 0 | 0.00 |
| -1.7 | 231 | 9 | 0 | 0 | 0.00 |
| -2.0 | 234 | 6 | 0 | 0 | 0.00 |
| -2.3 | 238 | 2 | 0 | 0 | 0.00 |
| -2.7 | 240 | 0 | 0 | 0 | NaN |

**Table 1 - HK Skill as a function of the question asked on random data. If the future drought state is lower than the threshold, the current label is negative (otherwise positive).**

| Threshold | True Positive | False Positive | False Negative | True Negative | HK Score |
|---|---|---|---|---|---|
| 0.0 | 136 | 75 | 16 | 8 | -0.02 |
| -0.1 | 151 | 75 | 5 | 4 | 0.11 |
| -0.3 | 168 | 67 | 0 | 0 | NaN |
| -0.7 | 190 | 45 | 0 | 0 | NaN |
| -1.0 | 200 | 35 | 0 | 0 | NaN |
| -1.3 | 210 | 25 | 0 | 0 | NaN |
| -1.7 | 226 | 9 | 0 | 0 | NaN |
| -2.0 | 229 | 6 | 0 | 0 | NaN |
| -2.3 | 233 | 2 | 0 | 0 | NaN |
| -2.7 | 235 | 0 | 0 | 0 | NaN |

**Table 2 - HK Skill as a function of the question asked on real SPI drought data. If the future drought state is lower than the threshold, the current label is negative (otherwise positive).**

Using the Hanssen and Kuipers (HK) discriminant[4], which determines how well we are splitting the observed positives and negatives, we found skills of 0.48 where 0.00 should demonstrate no skill. We tested the code against a well-known matlab implementation available on Suykens' website and found that our code was giving the same answers as an acceptable standard implementation of the algorithm.

We thus concluded that it is not our implementation of the LSSVM that is causing the issue. In fact, it appears as though the issue is the question itself that we are asking. The question was not appropriate for the data. In our study, we were attempting to predict a future state of drought in the central cell of a gridded geographic window by looking at the spatial pattern of drought around the central cell. The specific question being asked was, "Given the pattern of drought conditions at a particular time, will drought conditions increase or decrease in the next time step?" Upon reflection, it would appear as though this question can be asked with skill against a random data set for one simple reason. Since the data is a random number (varying between about +/2.3) the LSSVM can guess that the next value will be closer to zero and be right a significant number of times.

We then asked a more neutral question, "Will the drought value be above zero on the next time step?" and found that the algorithm only had a skill of 0.06 in answering that question. This is understandable as the LSSVM could never determine with any probability that the next value would be greater than or less than zero. As we varied the question to seek for drought values above some nonzero threshold, say 1.7, the algorithm continued to provide little to no skill. However, as can be seen in Table 1, the number of predicted no's slowly decreased, and eventually the number of false positives also decreased until the algorithm was always predicting better conditions. This is to be expected because the more extreme the threshold, the more often the machine will predict that conditions will be better. Due to the way the HK skill test discriminates between the two columns of the contingency table, the HK value simply hovers near zero. These experiments suggest that the LSSVM, which is very good at recognizing patterns, is simply finding the structure in the data with relation to the question being asked.

[4] The HK Skill (also called TSS) is described at: http://www.bom.gov.au/bmrc/wefor/staff/eee/verif/verif_web_page.html

We followed up using the question of whether or not the drought value was below zero or not on the next time step on the real drought data. Results in Table 2 show that the questions are interacting with the data in similar fashion to the random data described above. Furthermore, we found that the skill remained near zero (absolute value less than 0.10) for the square geographic window sizes we tested (3, 5, 7, 9, and 11 grid cells in width).

## 4.2. Autocorrelation

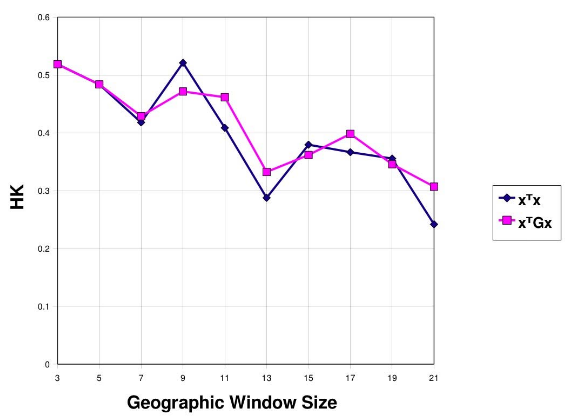


**Figure 5 - Skill at predicting real-world drought as a function of geographic window size. Both regular linear kernels and linear kernels with Gaussian weighting are shown.**

We provide, in Figure X, our results on real drought data while using the standard linear kernel as compared to using our Gaussian Hat kernel. The fact that the skill begins at about 0.5 for small geographic window size and then slowly decreases implies that the results of the random data cases described above are independent from this case with real data and using the original question. In other words, the findings above of skill for the random data case were only possible because of structure in the data with respect to the question being asked. In Figure 5, we feel that the skill being demonstrated is a direct result of the LSSVM being able to discern patterns in the data with respect to the question being asked as well. Specifically, the question being asked in Figure 5 is the original question of "Given the pattern of drought conditions at a particular time, will drought conditions increase or decrease in the next time step?" It is appears that the Gaussian Hat kernel, in general, provides greater skill over larger geographical window sizes than the simple linear kernel. Furthermore, it seems as though Tobler's First Law of Geography also holds true as the smaller geographic window size does a better job at predicting the future state of drought at this particular location for both styles of kernel.

## 4.3. Directionally Biased Kernels

We applied directional biases to our kernels by creating sample vectors out of differing rectangular windows around the central grid point. For instance, windows of 3x5 or 9x5 might be chosen around and including the central grid point to construct the sample vector. When directionally biased kernels were applied to our pure sine world data, the effects of the biasing went unseen because of the efficiency of the SVM's in finding the patterns. In other words, we always found perfect skill in prediction on the sine world data whether or not biases were used.

We overcame this impediment by introducing noise to the pure sine world so that the SVM must try harder to find the patterns. We generated Gaussian noise with mean of zero and standard deviation of one, divided it by three, and then added it to the pure sine data at one half degree grid spacing. We then ran the algorithm on 120 geographic locations along an East-West transect while using both EW and NS biased kernels. We used windows of 3x7 and 7x3 in size. The HK skill dropped to values between about 0.3 and 0.5 so that we had non-perfect skills to compare. Since the use of these kernels is aimed at finding structure in geotemporal data, it makes little sense to use statistical tests that rely upon assumptions about the data. Consequently, we chose the Wilcoxon Signed Rank Test [10] to test our hypotheses where, $H_o$: The EW biasing of the kernels makes no difference to HK skill of the SVM, and $H_a$: The EW biased kernels will provide greater skill for the SVM.

That these hypotheses make sense can be appreciated by envisioning riding on a wave of drought. If you look along the crest (NS biased kernel) of a drought moving in the EW direction, you will see little variation. In essence, the world looks flat and there is less information available to predict the future. However, if you look along the direction of travel of the wave, you start to see rising and falling structure that can help to predict the future state of your location.

For 120 samples, we compute the appropriate test statistic and find that z = -1.7672, leading to a p-value of less than 0.0392. Thus, we clearly reject the null hypothesis at a level of 95% confidence. The EW biasing does indeed reveal more information about the underlying structure in the data.

With this confidence, we next turn our attention to real drought data and apply the directionally biased kernels to the SPI data along three contiguous transects

**Figure 6 - Three latitudinal transects laying across the southern United States between North Carolina in the East (at the right) to Arizona in the West (at the left). The upper transect is located at 35.75 degrees North, the middle at 35.25 degrees North and the lower at 34.75 degrees North latitude. The gray circles represent locations where a North-South biased kernel gave better skill, while the black circles represent locations where an East-West biased kernel was better.**

at latitudes of 34.75, 35.25, and 35.75 degrees north latitude. These transects are also spaced at one half degrees in longitude and run from North Carolina in the eastern United States to Arizona in the western United States. The results may be seen in Figure 6 which may be imagined to overlay a map of the United States with North being up and East to the right. In this figure, the size of the dot represents the absolute value of the difference between the HK skills for EW and NS biased kernels. If the skills are the same, no dot appears. If the skill score favors the EW kernel, then the dot appears black, and if the skill score favors the NS kernel, the dot is gray. Several contiguous regions appear to exist wherein EW or NS structure to dominate the drought landscape for a degree or two in spatial extent.

## 5. Conclusions & Future Work

Perhaps the most startling realization concerning our results was the fact that the questions matter. Interaction between the questions and data itself creates structure that may lead to perfectly correct, yet useless conclusions. This occurred when we asked a perfectly reasonable question about real drought data against a random data set. This implies that it is always wise to vary questions, parameters, and algorithms and verify that answers are consistent with theoretical results. This is not always an easy task, but carefully constructed artificial datasets can aide greatly in the task.

Our results from varying the geographic window size along with the Gaussian weighting both point to the usefulness of autocorrelation in the data to improving the predictability of drought. This is for two reasons. First, when the geographic window was increased in size with either linear or linear combined with Gaussian hat kernels, predictability of future drought waned. This implies Tobler's 1st Law is in effect and that we should focus our efforts at using the nearby information to predict drought. Second, when Gaussian hats were used, the ability of the kernel to perform better at larger distances from the central grid cell improved. Once again, this arises from the suppression of distant noise by the Gaussian hat. This implies that we may use a larger window size for some particular purpose, yet still emphasize the nearest and most meaningful data through the Gaussian hat. The tradeoffs afforded by these two opposing effects offers a richer selection of parameters with which to tune geographically sensitive kernels.

Results from the application of directionally biased kernels show how to further enrich the geographical sensitivity of kernels as EW and NW biased kernels showed a statistically significant difference in the abilities to predict future states of simulated drought in data with known underlying structure. We carried that idea to the next step and used two kernels biased in orthogonal directions to reveal potential underlying structure in real drought across the United States. Future work will include mapping of entire regions and potentially continents in this fashion. Additionally, with greater areal coverage provided by these maps, more powerful spatial statistics may be applied to the results to confirm the existence of coherently regions of drought prediction skill.

In making the above maps, it is desirable to have full control over the directionality responses. In other words, we not only want to know there is an EW structure, we wish to assign a direction to this structure. Is drought propagating from the East? Or, from the West? Additionally, the inter-cardinal direction will be used. We will continue further research into greater directionality control of the kernels.

Several other ideas present themselves immediately for research beyond that outlined above: one of these ideas involves further exploration of the temporal dependence of the kernel spaces as variation of the temporal extent of the kernel may yield better predictability in the drought data. In other words, we would like to explore the ability of the presence of multiple previous time snapshots [2] in the kernels for prediction of the next step's state. Appropriately re-dimensioned vectors will be constructed from more than one previous time step. Indeed, following Şen [13], it is very likely that drought is better modeled by including its persistent nature in the prediction process.

Two more ideas for future work involve comparison of our results against other known algorithms and also against other datasets. First, we wish to run a series of comparison experiments by using another algorithm such as Markov Chains to compare its predictive power against the SVM approach we have used here. This series of comparisons could reveal whether or not a spatial approach (using the geographic neighborhood) of a grid cell location is more skillful in predicting future states of drought than by simply using one spatial dimension. Second, we wish to run these experiments against a different drought index, the Palmer Drought Severity Index (PDSI). While the SPI has shown to be spatially invariant, because of its dependence upon local time series data, it may not be stationary in view of global climate change. The PDSI value, while not spatially invariant does use temperature as one of its independent variables. Thus, it may provide better predictive power in the long-run under global climate change.

Another future goal is improved computational performance in the SVM. Currently, about 13 seconds may be required to run through the algorithm for a single geographic location. While this is sufficient for initial exploration of kernels and parameters, it is inadequate for our ultimate tasks of creating maps, testing many kernel variations, and using multiple datasets. For instance, to make a single map of the United State at one half degree resolution requires roughly 7200 locations to be tested. Implementation of sparse approximation as presented by Suykens et. al [15] may aid us in this task. Also, exploration of the gamma parameter space, which fine tunes the learning rate may help here as well.

As can be seen from the potential future work discussed above, the application of geotemporally sensitive kernels in SVM machines to real-world data is tremendous. There is much work left to be done and the authors invite all interested parties to continue to develop these kernel ideas and applying them to other domains of knowledge. All of our data and code is available at http://idea.cs.ou.edu/ .

## 6. Acknowledgements

The authors would like to thank the following individuals for valuable conversations regarding this work: Michael Hayes, Nic Grounds, Nathan Hiers, Rodger Brown, David Gagne, and Andrew Spencer. However, any errors found therein are the sole responsibility of the authors.

This material is based upon work supported by the National Science Foundation under Grant No. NSF/IIS/0746816, NSF/CISE/REU 0453545 a seed grant from the University of Oklahoma's College of Engineering, the Oklahoma Space Grant, and NASA Grant number NNG05GN42H.

## 7. References

[1] Bishop, C.A., (2007). Pattern Recognition and Machine Learning, Springer.

[2] Grenon, P., & Smith, B., (2004). SNAP and SPAN: Towards Dynamic Spatial Ontology, Spatial Cognition and Computation, **4**(1):69-104.

[3] Guttman, N.B., (1997). Comparing the Palmer Drought Index and the Standardized precipitation Index, *Journal of the American Water Resources Association,* **34**:113-121.

[4] Guttman, N.B., (1999). Accepting the Standardized Precipitation Index: A Calculation Algorithm, *Journal of the American Water Resources Association*, **35**(2):311-322.

[5] IPCC, (2007). *Intergovernmental Panel on Climate Change Fourth Assessment Report, Climate Change 2007: Synthesis Report*. Available on the web, last accessed August 10th, 2008 from http://www.ipcc.ch/pdf/assessmentreport/ar4/syr/ar4_syr_spm.pdf.

[6] Longley, P.A., Goodchild, M.F., Maguire, D.J., & Rhind, D.W. (2005). *Geographic Information Systems and Science*. Wiley UK.

[7] Lott, N., and Ross, T., (2006). Tracking and Evaluating U.S. Billion Dollar Weather Disasters, *86th AMS Annual Meeting*, 28 Jan - 3 Feb, 2006, Atlanta, GA. Available on the web, last accessed on August, 10th, 2008 at http://lwf.ncdc.noaa.gov/oa/reports/billionz.html.

[8] New, M. G., M. Hulme and P. D. Jones, 2000: Representing twentieth-century space-time climate variability. Part II: Development of 1901-1996 monthly grids of terrestrial surface climate. *Journal of Climate*, **13**: 2217-2238.

[9] O'Sullivan, D.O., & Unwin, D.J., (2003). *Geographic Information Analysis*. John Wiley & Sons, Hoboken, New Jersey.

[10] Ott, Lyman R., (1993). *An Introduction to Statistical Methods and Data Analysis, 4th ed*., Duxbury, Belmont, California.

[11] Pozdnoukhov, A., & Kanevski, M., (2008). Multi-scale support vector algorithms for hot spot detection and modeling. *Stochastic Environmental Research and Risk Assessment*, **22**:647-660.

[12] Rew, R. K. & Davis, G.P., (1990). NetCDF: An Interface for Scientific Data Access, *IEEE Computer Graphics and Applications*, **10**(4):76-82.

[13]  en, Z., (1990). Critical Drought Analysis by Second-Order Markov Chain, *Journal of Hydrology*, **120**:183-202.

[14] Sui, D., (2008). Is Neogeography Hype or Hope? The 'G' in GIS, *Geoworld*. March 2008, pp 16-17.

[15] Suykens, J.A.K., Lukas, L., & Vandewalle, J., (2000). Sparse Approximation Using Least Squares Support Vector Machines. IEEE International Symposium on Circuits and Systems, May 28-31, 2000, Geneva, Switzerland.

[16] Tinker, Rich., *http://www.drought.unl.edu/dm/monitor.html*, Climate Prediction Center/NOAA, Lincoln, NE, Last Accessed July 11th, 2008.

[17] Wilhite, D.A., & Glantz, M.H., (1985). Understanding the Drought Phenomenon: The Role of Definitions, *Water International*, **10**:111-12.

[18] Yuan, M., & McIntosh, J., (2002). A Typology of Spatiotemporal Information Queries. *Mining Spatiotemporal Information Systems*. K. Shaw, R. Ladner, & M. Abdelguerfi eds. Kuwer Academic Publishers. Pp 63-82.

[19] Yuan, M., Buttenfield, B., Gahegan, M.N., & Miller, H. (In Press). Chapter 14 in *Research Challenges in Geographic Information Science*, R. McMaster and L. Usery eds. John Wiley & Sons.