# Spatially Penalized Regression for Extremes Dependence Analysis and Prediction: Case of Precipitation Extremes

Debasish Das[1,2], Auroop R. Ganguly[1], Snigdhansu Chatterjee[3], Vipin Kumar[4], Zoran Obradovic[2]
[1]Department of Civil and Environmental Engineering, Northeastern University, Boston, MA, USA
[2]Center for Data Analytics and Biomedical Informatics, Temple University, Philadelphia, PA, USA
[3]School of Statistics, University of Minnesota, Minneapolis, MN, USA
[4]Computer Science and Engineering, University of Minnesota, Minneapolis, MN, USA
d.das@neu.edu, a.ganguly@neu.edu, chatterjee@stat.umn.edu, kumar@cs.umn.edu, zoran.obradovic@temple.edu.

## ABSTRACT
The inability to predict precipitation extremes under non-stationary climate remains a crucial science gap. Precipitation is not a state-variable within climate models, exhibits space-time heterogeneities, and is subject to thresholds and intermittences. Atmospheric variables in the spatiotemporal neighborhood, like temperature, humidity and updraft velocity, are often better predicted than precipitation from these models, and may have information relevant for precipitation extremes. Model-simulated atmospheric variables have been used to enhance model-predicted precipitation extremes in two ways: statistical downscaling routinely uses regression methods including neural networks and recently physics-based formulations have been developed. The former may not generalize under non-stationary climate while the latter is more interpretable but may not be able to discover or leverage the full information content in atmospheric covariates. We propose robust data-mining strategies to complement these approaches. The challenges are to discover spatiotemporal neighborhoods of influence, extract dependence structures, and determine predictive power, under non-stationary climate. We have developed a data-dependent method to discover sparse spatiotemporal dependence structure using spatially-penalized elastic net regression focused on extremes of target variables. The approach addresses neighborhood discovery, dependence discovery and predictive modeling of precipitation extremes. The methods show promise, specifically to improve our understanding of precipitation extremes and hence inform stakeholders and policy-makers in the water sector. In addition, further developments may generalize to problems in multi-physics simulations and to other complex, nonlinear and spatiotemporal dynamical systems where extremes are of interest.

## Keywords
Precipitation, Climate Change, Extremes Regression, Elastic Net, Sparse Modeling..



ACM SustKDD'12, August 12, 2012, Beijing, China Copyright 2012 ACM 978-1-4503-1558-6/12/08 ....$10.00

## 1. CLIMATE SCIENCE GAP AND A NEW DATA MINING SOLUTION

One of the largest gaps in climate science relevant for informing stakeholders and policy makers is the inability to develop credible projections for extreme events and regional change at spatiotemporal scales that matter to decision-makers. Precipitation and their extremes are particularly relevant in this context because of the deep science challenges as well as the widespread impacts on flood hazards and water resources decisions. The data mining literature in climate applications have tended to focus on teleconnections (long-range spatial dependence), especially on oceanic influence over regional land climatology [18, 19, 20]. However, while teleconnections are important, local and regional atmospheric conditions typically tend to dominate in the context of climate-related extremes. This is an area where novel data mining approaches are motivated to address science gaps that are relevant to global societal priorities.

The importance of adapting to a non-stationary climate has been emphasized in the context of water resources [21]. However, precipitation and regional climate prediction have been highlighted as two of four real holes in climate science [5], which makes the prediction of precipitation or their extremes at local to regional scales a major challenge.

The climate system is nonlinear dynamical (even chaotic) and non-stationary, while projections are sought for long lead-times. Thus, physics-based climate models, which in turn have become very computationally expensive, are essential. Purely data-guided methods are not best suited for multi-step ahead prediction under these circumstances, thus the problem is not cast here as a standard data-mining prediction problem. On the other hand, climate models by themselves may not be adequate, especially for critical challenges like precipitation extremes. Large-scale climate models solve a system of partial differential equations (PDEs) based on first principles, but also contain parameterizations for processes that are not so well understood. Unfortunately, processes pertaining to precipitation are among the least well understood and precipitation is not a state variable in the PDEs. In addition, precipitation is known to be extremely variable in space and time and the underlying processes are subject to thresholds and intermittences. However, as pointed out in the literature [1-3, 22-24], precipitation extremes tend to have a dependence on atmospheric variables ranging from temperature, humidity and precipitable water, to updraft velocity and horizontal wind components. These atmospheric variables, which can be thought of as potential covariates for precipitation extremes, are often better predicted that precipitation itself. Thus, there have been (somewhat counter-intuitive) suggestions that precipitation extremes may be more predictable than precipitation mean processes, simply because the extremes may relate more directly to covariates that are better predicted from models. The prediction problem in the context of precipitation extremes therefore translates to extraction of information content from these covariates and translating them to predictive insights. The prediction in conditional on the availability of model-simulations all the way to the prediction horizon of the models, which implies that the standard multi-step-ahead formulations used in data-mining are not well-suited. Rather, the issue is one of functional mapping from the covariates which are better predicted by models to the extremes of precipitation which are less well-predicted but more desired. The fundamental challenge is to ensure that the functional mappings generalize under non-stationary conditions over long lead times.

Recent studies [25,26] have investigated precipitation extremes and their attributes under changing climate based on global climate models (GCM) simulations [25,26]. The studies indicate that the larger uncertainties remain at regional and local scales. Dynamical downscaling based on regional climate models (RCM), while higher-resolution and physics-based, suffers from complex parameterizations and difficult boundary conditions. Statistical downscaling, which have used models ranging from simple linear regression to artificial neural networks, suffers from lack of interpretability. The downscaling approaches may not generalize well to non-stationary conditions owing to the complexity of parameterizations and/or the lack of physical consistency. One promising recent approach has been the development of physics-based approaches which attempt to relate the atmospheric covariates with precipitation extremes through what could be viewed as hypothesis-guided approaches. The physics operates at different scales or accounts for different processes than are handled within the large-scale computational models of climate, hence their added value [1,3,22]. While these approaches have demonstrated significant promise, they may not be able to leverage the full information content in atmospheric covariates and translate these to predictive insights, primarily because they have to rely on known physics-based hypotheses. The best approach would need to leverage the information content in the covariates through both the physics-based hypotheses and the data, while keeping the functional mappings between the covariates and the precipitation extremes interpretable and without losing the ability to generalize to non-stationary conditions. We propose sparse and spatially-penalized extremes regression as a way to fill this gap. Model parsimony is embedded into our formulations through sparse regularization and spatial penalties to reduce spurious or overly specific relations that may not generalize. Our data-guided approaches can be viewed as complementary to the physics-based approaches for relating covariates to extremes, perhaps eventually leading to hybrid approaches to handle this difficult challenge. This paper is focused on demonstrating the value of our approaches to extract the functional forms and developing predictive insights from model simulations. We focus on reanalysis data [17], which are climate reconstructions developed by assimilating multiple remote and in-situ sensor data into meteorological models, and do not validate using climate model simulations in the future. In addition, we do not yet consider vertical profiles of the atmospheric column, but attempt to show proof-of-concept based on spatiotemporal neighborhoods.

Based on the above, the data-mining goals can be summarized:

(1) Developing a data-driven model that focus exclusively on describing the extreme values of a target variable as a linear combination of the atmospheric covariates.
(2) Enforcing the sparsity constraint in the linear model using sparse regularization techniques in order to facilitate the emergence of a spatiotemporal neighborhood of influence directly from the data.
(3) Introducing an adaptive spatial penalty within the sparse regularizers in order to enforce a spatiotemporal dependence structure that is not overly complex and remains interpretable to domain scientists in climate, meteorology and hydrology.

The rest of this paper is organized in the following order. In second section, we explain the data-mining problem in detail and in third section we provide related background. In the fourth section, we explain the notations and assumptions used to describe the problem and in fifth section we present our experimental results with interpretations. In section six, we conclude and discuss future work.

## 2. PROBLEM DEFINITION AND RELATED BACKGROUND

### 2.1 Dependence Discovery

We posit that not all the climatic variables that are well-predicted will contain useful information about the precipitation. Therefore we need to find out a set of variables $\{X_1, X_2 \ldots X_p\}$ that contain information about target variable $Y$ (precipitation here) out of a larger pool of variables $\{X_1, X_2 \ldots X_P\}$ where $P > p$. As a first step, we consider linear dependence structures and leave nonlinear dependence analyses to future research. Dependence in climate data may often be reasonably well captured through linear or quasi-linear structures.

### 2.2 Neighborhood Discovery

Precipitation generation processes are inherently multi-scale in nature, all the way from localized severe thunderstorms to the propagation of mesoscale fronts at regional scales and all the way to the influence of larger-scale climate oscillators. For the purposes of data-mining, we may differentiate between the influences of local or regional spatiotemporal neighborhoods versus long-range dependence or teleconnections, specifically owing to ocean-driven natural climate oscillators. The former is typically more dominant for climate extremes including precipitation, which is our focus here. The neighborhood may depend on the selected location and prevailing climate and wind conditions. Here we select the neighborhood based on data rather than enforcing a specific shape or size *apriori*.

However, in our proposed method we have combined both these problems into a single sparse regression learning problem with a spatial penalty.

### 2.3 Predicting Extremes

The data-mining literature has focused more on the prediction of frequent patterns with a recent emphasis on anomaly analyses. However, a thorough treatment of extremes, or the tails of distributions, has been lacking. One consideration is the definition of extremes and their attributes. In the context of precipitation extremes, percentile-based [1-2] and extreme value theoretic definitions [25, 26] have been used. The specific definitions of extremes are expected to impact the predictive modeling and corresponding insights. A fundamental issue is that extremes cannot be expected to follow the distribution of the original precipitation time-series since they represent distributional tails.

Thus, transformations need to be constructed based on the statistical properties of the extreme values to make these values amenable to predictive modeling. Finally, for precipitation extremes, there is a need to be cognizant of the domain knowledge available [1-2], while keeping the problem definition open to novel data-guided insights given the nature of the science gaps.

A data-driven solution for discovering the variables those influence precipitation extremes can be described as follows. Let us denote precipitation extremes at a grid indexed by $(i,j)$ on a certain geographical region of interest by $Y^{ij}$. Also, let us denote all (say $N$) candidate variables at $(i,j)$-th grid by $V^{ij} = \{V_1^{ij}, V_2^{ij} \ldots V_N^{ij}\}$ and variables at all grids by $V = \{V^{mn}\ \forall (m,n)\ \in S\}$ where $S$ is the set of all grid-points within the region under consideration. We can combine variable and neighborhood selection into a single problem described as: For each variable $Y^{ij}$ we are required to find a set of variables/node $NE^{ij} = \{v_k : v_k \in V\}$ so that $Y^{ij}$ is linearly dependent only on $NE^{ij}$ and nothing else. It has been shown before [7,8,9] that $L_1$-regularized linear regression algorithms can produce sparse solutions by learning a regression model and at the same time discards the irrelevant features by forcing their corresponding coefficients to zero. In the next two sub-sections we will briefly discuss these methods.

## 2.4 The Elastic Net

One way of dropping uninformative regressors is to use L1-regularization on regressors' coefficients which results in a LASSO (least absolute shrinkage selection operator) estimator and has the following form.

$$\min_{\beta} \mathrm{RSS}(\beta) + \lambda ||\beta||_1^1 \qquad (1)$$

An important feature of the L1 penalty is that some coefficient estimates can be exactly zero. The parameter λ controls how many coefficients will be zero (as λ goes higher, more coefficients become zero). However, LASSO is not without its drawback. Conceptually there are two problems as highlighted by Zou and Hastie (2005). First, if there are $L$ variables and $D$ examples and $L > D$, LASSO can select at most $D$ variables. Second, if there is a group of variables with high pairwise correlations, LASSO tends to select only one variable from the group and does not care which one. It turns out that a convex combination of L2 and L1 penalties solve these problems. The result is the `Elastic Net' (EN) estimator [8].

Like LASSO, the EN simultaneously shrinks the coefficient estimates and performs model selection. The LASSO penalty is convex, but not strictly convex. Strict convexity enforces the grouping effect so that predictors with similar properties will have similar coefficients. The EN objective function is

$$\min_{\boldsymbol{\beta}} \mathrm{RSS}(\boldsymbol{\beta}) + \lambda_1 ||\boldsymbol{\beta}||_1^1 + \lambda_2 ||\boldsymbol{\beta}||_2^2 \qquad (2)$$

The EN penalty is thus a convex combination of the LASSO and the ridge penalty and is strictly convex when $\lambda_2/(\lambda_1 + \lambda_2) > 0$. The relative importance of L1 and L2 parts of the regularizer depends on the values of $\lambda_1$ and $\lambda_2$ respectively. A computationally appealing property of the EN is that it can be reformulated as a LASSO problem and hence solved using LASSO algorithms [8].

Efron et al. [15] showed that LASSO is in fact special case of what is known as Least Angle Regressions (LARS) algorithm which is extremely efficient with complexity comparable to OLS estimate. Recently Friedman, Hastie and Tibshirani (2009) [16] developed an algorithm called 'glmnet' for solving generalized linear models with convex penalties which include both LASSO and EN which outperforms LARS in terms of speed. This algorithm uses cyclic coordinate descent, computed along a regularization path.

# 3. NOTATIONS AND ASSUMPTIONS

Let us assume that there are $N$ potential climate variables that are being considered for having possible influence on precipitation extremes. We will henceforth call these climate variables as *covariates*. But the values of these covariates at different grid-points are needed to be considered as actual features for neighborhood discovery as discussed earlier. So, henceforth we will regard the time-series of covariate values at different grid locations as separate variables and call them *features*.

Let us assume that for a particular grid-point $(i,j) \in S$ ($S$ is the set of all grid-points within the spatial region being considered), the daily precipitation time-series is given by $P^{ij} = \{p_1^{ij}, p_2^{ij}, \ldots p_T^{ij}\}$ where $T$ is the number of observations made during the period for which data is available (or for the time-frame under consideration). Let us also denote the time-series for potential features as $V_k^{ij} = \{v_{k1}^{ij}, v_{k2}^{ij}, \ldots v_{kT}^{ij}\}$ where $k$ denotes individual covariates and therefore ranges from 1 to $N$. Let us build the extremes time-series from $P^{ij}$ by picking the extremes (using some pre-defined definition of extreme) from the time-series.

Let us denote the new series as $E^{ij} = \{e_1^{ij}, e_2^{ij}, \ldots e_D^{ij}\}$ where $D(i,j)$ is the number of precipitation extremes occurred at grid-point $(i,j)$. Now, let us assume that $e_d^{ij} = p_t^{ij}$ (i.e. the $d$-th extreme precipitation occurred on the $t$-th day of the daily precipitation time-series). Now for each $d$ in $1 \ldots D(i,j)$, let us build the set given by $\{e_t^{ij}, \mathbf{X}_t^{ij}\}$ where $\mathbf{X}_t^{ij} = \{v_{k(d-1)}^{mn}, v_{k(d-2)}^{mn} \ldots v_{k(d-\Delta)}^{mn}\ ;\ k \in \{1..N\}, \forall \{m,n\} \in S\}$. Here $\Delta$ is the number of days prior to an occurrence of precipitation extreme from which covariates will be considered for possible influence on a precipitation extreme. So, once these spatial and temporal distributions of covariate values are considered, the total number of features become $L = N$(number of covariates) x $|S|$ (number of grids) x $\Delta$.

The main assumption here is that the spatio-temporal dependence structure between precipitation extremes and the covariates remains unaltered over time. Although this might not hold when the period under consideration is in the order of thousand years but this is a reasonable assumption for a shorter time-period. We assume the dependence structure might vary over space and our model should be capable of accommodating that change.

# 4. METHODOLOGY

The overall problem described above can be regarded as a feature selection problem where a few features will be selected out of a set $L = N\mathrm{x}|S|\mathrm{x}\Delta$ possible candidate features. Since we have included all the grid-points within the set of candidate features as a neighbor for each grid-point, we can safely assume that most of these features will be irrelevant for each of the grid-points (our method will still work even for the unlikely case of all features being relevant), although the set of candidate features is same for all the grid-points, the set of irrelevant features can be different (but might be overlapping) for different grid-points. We are further interested in exploiting the information content in the covariates within the discovered neighborhood for which we need to train a predictive model. For our problem, there might be multiple correlated features (due to spatial and temporal correlation) and more relevant features than the number of available data-points. So, we used elastic net [8] to achieve sparse linear models.

An alternative approach can be applying a feature selection algorithm to select relevant features and thereby train a linear predictive model based on the selected features. This alternative was not considered here since most feature selection methods do not work when number of features is larger than the number of data-points. Furthermore elastic net is much faster than this alternative since it encapsulates both feature selection and model estimation in a single optimization problem that can be very efficiently solved with state-of-the-art available techniques.

## 4.1 Dependence Estimation Using Elastic Net

We used top-M approach for selecting extremes from the daily precipitation time-series. Specifically, we select the top M highest independent daily precipitation events from each year. By independent events, we mean that there should be at least one dry day in between any two of the selected precipitation events. As discussed earlier, precipitation extremes selected this way does not follow Gaussian distribution, but they follow Generalized Extreme Value distribution, PDF of which is given by

$$f(x;\mu,\sigma,\xi)=\frac{1}{\sigma}\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{\left(-\frac{1}{\xi}\right)-1}\times \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{\left(-\frac{1}{\xi}\right)}\right\} \quad (3)$$

Now, let us fit a GEV distribution on $E_{ij}$. Let us assume the resulting distribution is given by $GEV(e_{ij};\ \xi_{ij},\sigma_{ij},\mu_{ij})$ . In order to make these values to conform to Gaussian distribution, we used the following transformation on $E_{ij}$.

$$y^{ij} = \Phi^{-1}(GEV(e^{ij};\ \xi^{ij},\sigma^{ij},\mu^{ij})) \quad (4)$$

where $\Phi^{-1}$ is the inverse normal distribution with zero mean and unit variance. We then solve the following elastic net optimization problem for each location $(i,j) \in S$

$$\widehat{\boldsymbol{\beta}}^{\mathbf{ij}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{d=1}^{D(i,j)} \left(y_d^{ij} - \mathbf{X_d^{ij}}^{\mathbf{T}}\boldsymbol{\beta}^{\mathbf{ij}}\right)^2 + \lambda_1||\boldsymbol{\beta}^{ij}||_1^1 + \lambda_2||\boldsymbol{\beta}^{ij}||_2^2 \quad (5)$$

## 4.2 Spatial Penalty

In equation (5) we have a formulation of the problem of finding a spatio-temporal dependence structure between precipitation extremes and regional covariates in terms of coefficients $\boldsymbol{\beta}$. But in its current form it is missing important domain knowledge well-known in Geography which says "Everything is related to everything else, but near things are more related than distant things". Currently, we are giving equal importance to covariates belonging to all neighbors as a potential feature irrespective of its distance from the grid-point for which the dependence structure is being estimated. We address this problem by letting the multipliers $\lambda_1$ and $\lambda_2$ be functions of $\beta_p^{ij}$ and depend directly on the normalized geodesic distance of the associated grid-point from which the corresponding feature belong. So the new formulation of the problem will be

$$\widehat{\boldsymbol{\beta}}^{\mathbf{ij}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{d=1}^{D(i,j)} \left(y_d^{ij} - \mathbf{X_d^{ij}}^{\mathbf{T}}\boldsymbol{\beta}^{\mathbf{ij}}\right)^2 + \sum_{l=1}^{L}\lambda_{1l}|\beta_l^{ij}| + \sum_{l=1}^{L}\lambda_{2l}|\beta_l^{ij}|^2 \quad (6)$$

where $\lambda_{1l} = \lambda_{10}.(g_l/g_l^{max})$

and $\lambda_{2l} = \lambda_{20}.(g_l/g_l^{max})$

and $g_l$ is the geodesic distance of the grid-point from which the $l$-th feature belong. Note that, there are total of $|S|$ grid-points and each grid-point generates $N\Delta$ potential features making the total number of potential features $N\Delta|S|$. Therefore for each elastic net model corresponding to each of $|S|$ different grid-points we have $|S|$ different values of the multiplier $\lambda$ and each of them will be repeated $N\Delta$ times.

# 5. EXPERIMENTAL RESULTS

## 5.1 Dataset

The precipitation data that we used for this study is originated from NCEP-NCAR reanalysis project [17] which is publicly available for download. This dataset is constructed by fusing and assimilating measurements from heterogeneous remote and in-situ sensors within the physics-based climate models. Measurements are provided for points (grid cells) at a resolution of $2.5^o$ x $2.5^o$ on a latitude-longitude spherical grid. We used daily forecasts starting from 1948 until 2010 for the following variables as potential covariates: i. Temperature (surface level); ii. Sea-level Pressure (surface level); iii. Relative Humidity (surface level); iv. Pressure (surface level); v. Precipitable Water (Entire Atmosphere); vi. Horizontal Wind Speed (North-south). (surface level); vii. Horizontal Wind Speed (East-west). (surface level); viii. Updraft Velocity (Omega) (surface level).

Precipitation rate is available at a finer resolution (192x94 grid-points over whole globe instead of 144x73, which is the resolution for other variables). So, we had to interpolate it down to the resolution of the other variables.

## 5.2 Experimental Set-up

In order to reduce the computational cost, we did not use the global dataset. Rather we focused our analysis more at regional level and therefore applied our algorithm on different regions in North America instead of the whole globe.

We used the top-M approach for selecting extremes to create the dataset for precipitation extremes. We selected highest 15 independent precipitation events from each year at each location and considered each extreme as one instance of the target variable. We defined precipitation events as independent if they are separated by at least one dry day in between them. The number 15 has been chosen after consultation with the climate scientists as a typical value used by them. Now for each instance of the extreme, the potential features are selected as values of each covariate at each of the grid-points within the target region (which is different for different experiments) on the same the extreme occurred and on previous two days. The potential features selection process is described in Figure 1. Three days of covariate values were chosen since the empirical study showed very little correlation among precipitation extremes and other covariates beyond two previous days.

Since we have 63 years of data, we have a total of 63x15 = 945 data-points. Among them we used first 700 points for training and rest of them for testing. We have applied our model on four different regions, namely North-west, South-west, North-east and South-east US. We present the numbers of grid-points and numbers of potential features in each region in Table 1.

We used the 'glmnet' package designed by Friedman, Hastie and Tibshirani [16] to implement elastic net. In this package, two hyper-parameters $\lambda_1$ and $\lambda_2$ are replaced by just one parameter $\lambda$ and a mixing coefficient $\alpha$, so that

$$\lambda_1 = \lambda.\alpha \text{ and } \lambda_1 = \lambda.(1-\alpha)$$

The package provides option for choosing differential penalty factor $\lambda$ for different components of $\boldsymbol{\beta}$. So in one experiment we used fixed value of $\lambda$ with no spatial penalty, whereas in a separate experiment we introduced spatial penalty by using

$$\lambda = \lambda_0 . (g_l / g_l^{max}) \qquad \text{(Refer to equation (6))}$$

**Table 1: Number of grid-points and potential features in the target regions considered**

| | NW US | SW US | NE US | SE US |
|---|---|---|---|---|
| # Grid-points | 72 | 21 | 30 | 42 |
| # Features | 1728 | 588 | 720 | 1008 |

We chose α = 0.5 (it was observed that the value of α does not influence the end results when it is within a range 0.5±0.25.) and estimated $\lambda_0$ using cross-validation.

In order to determine whether the selected covariates carry any useful information about the precipitation extremes, we designed a baseline experiment where the target values (precipitation extremes) were shuffled randomly before training the elastic net model for each grid. We call this "null experiment" and perform this experiment several times for each run of our proposed experiment. We claim that if we can achieve a better accuracy than this null experiment that is enough proof that the covariates carry some information. A set of similar experiments, each starting from one of the following different subsets of all potential features, were performed for comparison.

i) All covariates except precipitable water for all 3 days and all grid-points.
ii) Only precipitable water for the day when the extreme occurred in all grid-points.
iii) All covariates for the day when the extreme occurred in all grid-points.
iv) All covariates for the previous 2 days when the extreme occurred in all grid-points.
v) Only precipitable water for the day when the extreme occurred and in the grid extreme occurred.
vi) Precipitable Water + Updraft Velocity + Relative Humidity + Longitudinal Wind (V-wind) for all 3 days and all grid-points (these covariates were chosen since they dominate in terms of number of non-zero beta values).

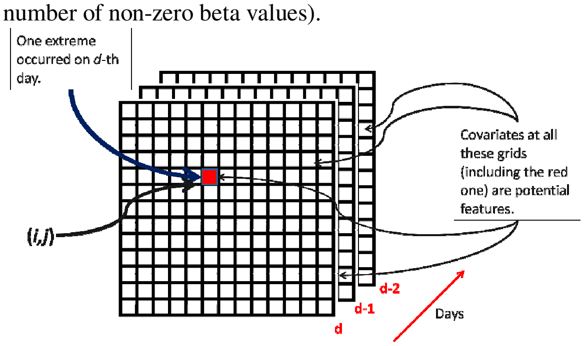



**Figure 1. Showing the candidate features for an instance of precipitation extreme**

## 5.3 Results and Discussions

We have estimated the $R^2$-accuracies of the linear models trained using the elastic net for each grid-point and compared them with the results of the "null experiment". The spatial distributions of the accuracies are shown in Figure 2 with and without spatial penalty for NW and SW US. The spatial penalty does not appear to alter the accuracies of the model. Both for NW and SW US, the maximum accuracy reaches as high as 0.7, which is significant given the complexity of the problem. We can use these accuracies as an indication of our confidence on the dependence structures obtained for the corresponding grid-point. A low accuracy may imply spurious dependence structures.

In Figure 3, we show the cumulative distribution of accuracies in the grid-points within the target regions when we start from different subsets of potential features as described in previous section. The figure shows that irrespective of the starting subset of covariates, our method almost always performs better than the "null experiment", implying that the covariates do have information-content. However, the information content varies depending on the combination of covariates and target region. Figure 3 suggests that (a) covariates from neighboring grid-points improve the accuracy over covariates from just the grid where the extremes have occurred, (b) covariates from the day the extreme has occurred contains significantly more information compared to the previous days, although previous days do contain information (more than the null distribution), and (c) the east coast is more difficult to predict than west coast. The insights both confirm current climate knowledge and offer new insights to climate science.

In this particular application, the distribution of **β**-values are of equal, if not more, importance as the accuracy of the prediction models. We can represent the non-zero **β**-values as edges connecting two nodes where one node represents the precipitation extremes in the grid-point on which the elastic net model is currently being trained and the other node is one of the potential features belonging from one of all the available grid-points (this includes the grid-point on which model is being trained). So, if there are |*S*| total grid-points in the target region, we will have a total of 3x8x|*S*| possible **β**-values (however, most of them will be zero for a sparse model) for each grid. Again, we have one such model for each of the |*S*| grid-points. So, altogether there can be total of possible $24|S|^2$ **β**-values or edges. Figure 4 shows these edge distributions, as a function of the distance between the grid-points they connect, before and after using spatial penalties. We only present this for the NW US due to lack of space, but this kind of analysis can be done for any target region. The distance will be zero for a non-zero **β** that connects with a variable in the same grid-point where the model is being trained. The plots are separated according to the covariates they correspond to. We can see that adding the spatial penalty results in more parsimonious models which are more easily interpreted by the domain scientists, while accuracies of the models remain intact. Some of the interesting information available from these plots about NW US are as follows: (a) winds, both vertical and horizontal, influence the precipitation extremes from a large number of neighboring grid-points, (b) pressure from neighboring grid-points has very small influence on precipitation extremes, and (c) both temperature and precipitable water have more localized influence on precipitation extremes. The insights, which are exemplary rather than exhaustive, range from known (c) or intuitive (a: horizontal) to relatively novel (b) or counter-intuitive (a: vertical). We present an example of the spatial dependence obtained from our analyses. Figure 5 presents the actual distribution of the edges for different covariates and for one of the grid-points in NW US that attained maximum accuracy both before and after using spatial penalty. Here, all the edges originate from the grid-point on which the model is being trained. If a number appears on the originating grid-point, that means there is edge connecting the

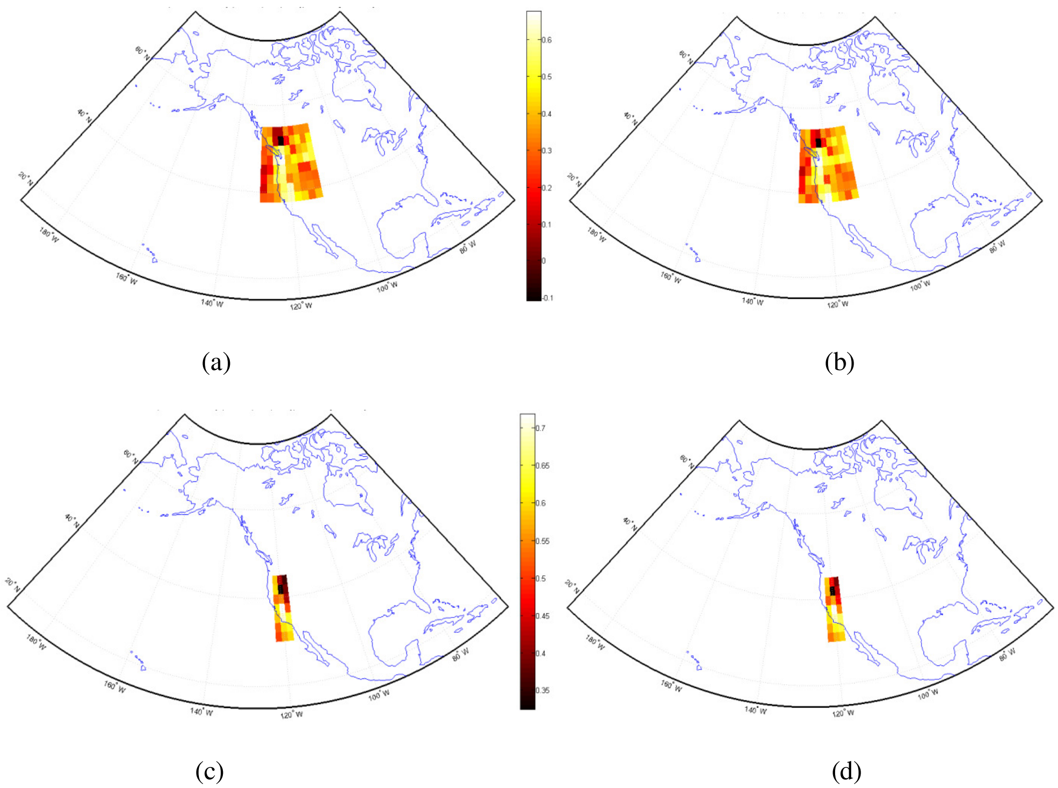



**Figure 2: Spatial distribution of accuracies in (a) North-west US before using spatial penalty (average accuracy = 0.366); (b) North-west US after using spatial penalty (average accuracy = 0.39); (c) South-west US before using spatial penalty (average accuracy = 0.541); (d) South-west US after using spatial penalty (average accuracy = 0.53);**

corresponding covariate in the same grid-point. We can see, even after adding spatial penalty, there are still few edges that connect grid-points that are far apart although most edges are now short. This is a result that might be interesting and may lead to new insights.

# 6. RELATED WORK

Data mining techniques has only been recently applied to the climate applications. Especially, $L_1$-regularized sparse algorithms have been successfully applied for climate modeling before [6, 10]. In [10] Lozano et. al. used group elastic net for causal modeling of climate change attribution. They were only interested in finding the variables that influence some statistical property (namely, return level) of the temperature extremes. But they do not consider the spatial and temporal pattern in the dependence structure and therefore their approach does not involve finding a neighborhood of influence. Secondly, they assumed a uniform dependence structure over the entire region they considered. We have relaxed this constraint by letting the dependence structure vary over space. Furthermore, they are only interested in attribution, not in prediction of extremes, whereas we do both.

In a second paper [11] Liu et. al. considered the same problem, but now they used multiple time-series of observations of the same set of variables available from different sources and learned a relational graph between them using a hidden MRF and sparse regularization. This approach assumed the similar set of constraints assumed in [10] mentioned before. They have not considered extremes either.

In [6] Chen et. al. used graphical lasso to learn sparse graphical models between different atmospheric variables for a fixed time and space and they let these graphs vary over space and time using kernel weighted covariance matrix. But, they neither considered variable values in the space-time neighborhood nor they considered extremes. In a more recent work [27], sparse group lasso is used to select climate variables where values of a single climate variable with a certain temporal lag at all grid-points within a certain spatial neighborhood are considered a group. So a feature (a climate variable at any of the neighborhood grid-point) cannot be selected unless the group (the climate variable itself) from which it belongs is selected. Sparsity is enforced both at the group level and the individual feature level. Our method does not enforce any group structure and are permitted to select a variable value from a grid-point even if it does not select the same variables from other grid-points.

# 7. IMPACT AND FUTURE WORK

We have introduced a method for finding spatiotemporal dependence of precipitation extremes on regional atmospheric covariates using the elastic net and exploited the dependence to develop a predictive model for the extremes. The novelty lies in being able to directly train a linear model exclusively on extremes rather than on average values and achieving a prediction accuracy that is significant for application domain. This method can be extended for other domains, including but not limited to multi-physics simulations (e.g., astrophysics or biology) and/or complex and nonlinear spatiotemporal systems (e.g., turbulence in computational fluid dynamics), where extremes are considered more important or interesting than average behavior (e.g., biology, finance, healthcare). Ultimately, we have been successful in achieving our stated goals of: (1) discovering spatiotemporal dependence structures of precipitation extremes on regional

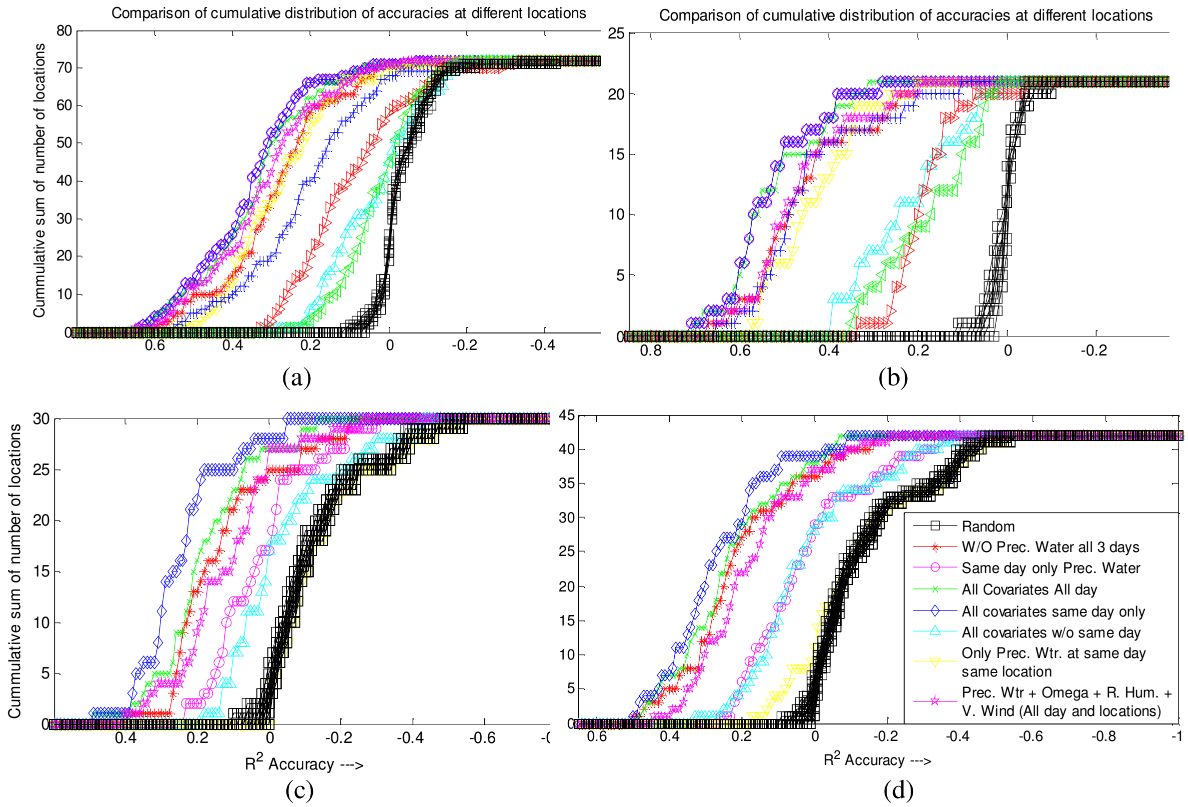



**Figure 3: Cumulative distribution of accuracies (best viewed in colors) in (a) North-west US; (b) South-west US; (c) North-east US; (d) South-east US; ($R^2$-accuracy along x-axis and cumulative number of grid-points that exceeds a corresponding accuracy are plotted along y-axis**

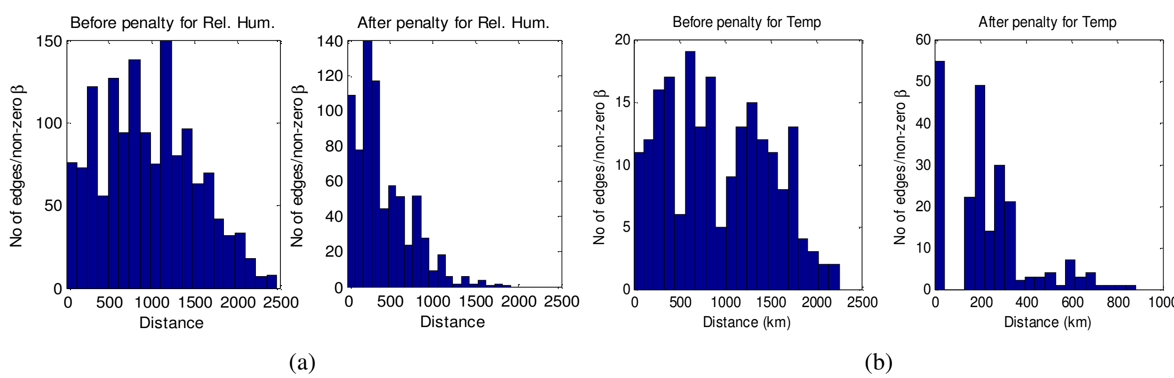
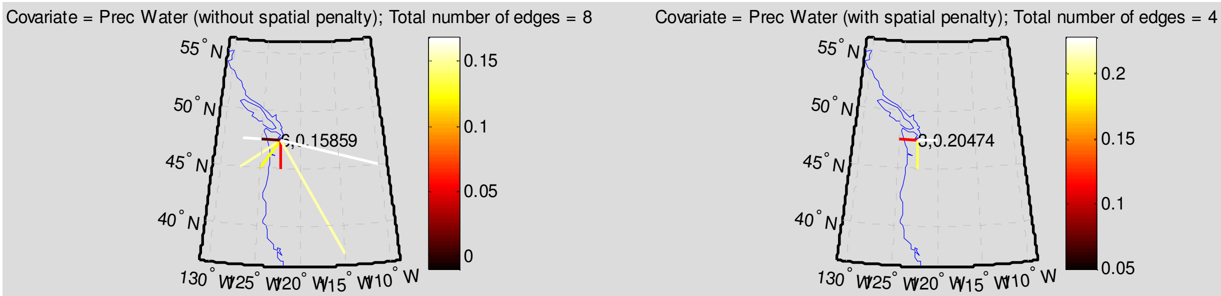



**Figure 4: The distribution of edges before and after spatial penalty for (a) Relative humidity; (b) Temperature for NW US**

atmospheric covariates, and (2) demonstrating the value of data-driven approaches to extract the information about precipitation extremes from model-simulated extremes and translating the information to enhanced predictive models. The methods we have proposed, specifically for sparse extremes regression with a spatial penalty, are applicable to this problem and may generalize to other domains. Future research needs to consider non-linear dependencies inherent in the climate system, include atmospheric covariates in the vertical layer and incorporate the physical relations that have been developed in climate science, perhaps as pre-processors to the data algorithms. Combining the grid-based regression models and letting them share information is another direction. Statistical properties (including uncertainty quantification) of the sparse regression models that focus exclusively on extremes need to be examined. Combining the spatiotemporal neighborhood-based predictions with teleconnections, specifically the influence of ocean-based oscillators, could be a way forward for precipitation extremes analysis.

## Acknowledgements

This research is supported by NSF grants IIS-1029711 and SES-0851705, a grant to ARG from the Nuclear Regulatory Commission and ARG's start-up funds from Northeastern Univ. Please forward any questions related to this publication to *a.ganguly@neu.edu.*




(a) (b)

**Figure 5: Dependency structure of precipitation extremes on precipitable water at a grid-point that attained maximum accuracy in North-west US (see figure 2(a) and 2(b)) (a) without and (b) with spatial penalty. Colors of the edges indicate their strength (*β*-value). If a number appears on the originating grid-point, that means there is edge connecting the corresponding covariate in the same grid-point.**

## 8. REFERENCES


[1] M. Sugiyama, H. Shiogama, and S. Emori. Precipitation extreme changes exceeding moisture content increases in MIROC and IPCC climate models. *Proc. Nat. Acad. Sci.* USA, 107(2):571–575, 2010.

[2] P. A. O'Gorman and T. Schneider. The physical basis for increases in precipitation extremes in simulations of 21st-century climate change. *Proc. Nat. Acad. Sci.* USA, 106(35):14773–14777, 2009.

[3] P. A. O'Gorman and T. Schneider. Scaling of precipitation extremes over a wide range of climates simulated with an idealized GCM. *Journal of Climate* 22, 5676-5685

[4] S. Coles. An Introduction to Statistical Modeling of Extreme Values. *Springer-Verlag*. 2001.

[5] Q. Schiermeier. The real holes in climate science. *Nature* 463, 284-287 (2010).

[6] X. Chen, Y. Liu, H. Liu, J. G. Carbonell. Learning Spatial-Temporal Varying Graphs with Applications to Climate Data Analysis. *AAAI*, 2010.

[7] N. Meinshausen and P. Buhlmann. High-dimensional graphs with the Lasso. *Ann. Statist.* 34, 1436–62, 2006.

[8] H. Zhou and T. Hastie, Regularization and variable selection via the elastic net. *J. Roy. Statist. Soc. Ser.* B 67 301–320, 2005.

[9] R Tibshirani, Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser.* B 58 267–288. 1996.

[10] A. C. Lozano, H. Li, A. Niculescu-Mizil, Y. Liu, C. Perlich, J. Hosking and N. Abe, Spatial-temporal causal modeling for climate change attribution. In *ACM SIGKDD*, 2009.

[11] Y. Liu, A. Niculescu-Mizil, A.C. Lozano and Y. Lu Learning Temporal Causal Graphs for Relational Time-Series Analysis, in *ICML*, 2010.

[12] J. Fan, R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* 96, 1348-1360, 2001

[13] W. Fu, Penalized regressions: The bridge vs. the Lasso. *Journal of Computational and Graphical Statistics* 7 (3), 397-416, 1998.

[14] H. Liu, X. Chen, J. Lafferty and L. Wasserman, Graph-valued regression. *NIPS*, 2010

[15] B. Efron, T. Hastie, I. Johnstone, and R.Tibshirani, (2004), 'Least angle regression', *The Annals of Statistics* 32, 407–499

[16] J. Friedman, T. Hastie and R. Tibshirani, Regularized Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33(1), 2010.

[17] Kalnay et al., The NCEP/NCAR 40-year reanalysis project, *Bull. Amer. Meteor. Soc.*, 77, 437-470, 1996.

[18] M. Steinbach, P.-N. Tan, V. Kumar, S. Klooster, C. Potter, Discovery of climate indices using clustering, in *KDD* 2003, pp. 446-455, 2003.

[19] K. Steinhaeuser, N. V. Chawla and A. R. Ganguly, Complex Networks as a Unified Framework for Descriptive Analysis and Predictive Modeling in Climate Science. *Statistical Analysis and Data Mining*, 2010.

[20] J. Kawale et. al. Data Guided Discovery of Dynamic Climate Dipoles *NASA Conference on Intelligent Data Understanding (CIDU)*, 2011.

[21] P. C. D. Milly. Stationarity Is Dead: Whither Water Management?. *Science*, 2008.

[22] C. J. Muller and P. A. O'Gorman. An energetic perspective on the regional response of precipitation to climate change. *Nature Climate Change*, 2011.

[23] S. C. Liu et al. Temperature dependence of global precipitation extremes. *Geophysical Research Letters*, vol. 36, L17702, 2009.

[24] R. P. Allan and B. J. Soden. Atmospheric Warming and the Amplification of Precipitation Extremes. *Science*, 2008.

[25] V. V. Kharin et. al. Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. *Journal of Climate*, 20, 1419—1444, 2007.

[26] S. C. Kao, and A. R. Ganguly, Intensity, duration, and frequency of precipitation extremes under 21st-century warming scenarios, *Journal of Geophysical Research - Atmospheres*, 2011.

[27] S. Chatterjee, K. Steinhaeuser, A Banerjee, S. Chatterjee, A. Ganguly. Sparse Group Lasso: Consistency and Climate Applications. *SDM* 2012.