

Drought Detection of the Last Century: An MRF-based Approach

Qiang Fu * Arindam Banerjee * Stefan Liess† Peter K. Snyder†

Abstract

Droughts are one of the most damaging climate-related hazards. The late 1960s Sahel drought in Africa and the North American Dust Bowl of the 1930s are two examples of severe droughts that have an impact on society and the environment. Due to the importance of understanding droughts, we consider the problem of their detection based on gridded datasets of precipitation. We formulate the problem as the one of finding the most likely configuration of a Markov Random Field and propose an efficient inference algorithm. We apply this algorithm to the Climate Research Unit precipitation dataset spanning 106 years. The empirical results show that the algorithm successfully identifies the major droughts of the twentieth century in different regions of the world.

1 Introduction

Droughts are one of the most damaging climate-related hazards and the consequences are often abrupt, severe and potentially catastrophic to both society and the environment. Droughts may lead to reductions in water supply, diminished power generation, disturbed riparian habitats as well as a host of other associated economic, political and social activities [18]. A frequently cited example is the decades long Sahel drought [5, 8] starting in the late 1960s, which led to widespread famine, ecosystem degradation and dispersion of its inhabitants. Other examples include the Dust Bowl event in the central US [15] in the 1930s, which is marked by sudden reductions in precipitation.

In the climate science community, the cause of droughts has been extensively studied. For example, [13] identify the physical basis for long-term droughts and [4] analyze how the increase in deserts influences climate, e.g., a reduction in precipitation in a general circulation model (GCM). In addition to the general theory developed for droughts worldwide, some work has been focused on specific drought events. For instance, [15] perform model simulations to identify the mechanisms contributing to the Dust Bowl. [5] show how the interaction between climate and vegetation

prolongs the Sahel drought. [3] analyze the Sahel drought simulation with GCMs and projection what may happen in the future.

Even though the importance of understanding droughts cannot be overstated, there are few rigorous and systematic tools to detect them. The current standard of drought detection is the Palmer Severity Drought Index (PSDI) [7]. The PSDI is based on a supply-and-demand model of soil moisture. It uses a 0 as normal, and drought is shown in terms of negative numbers. However, the utility of the Palmer index is weakened by the arbitrary nature of Palmer’s algorithms, including the technique used for standardization. In a recent paper, [11] analyze global historical rainfall observations and detect regions that have undergone large, sudden decreases in rainfall. Their algorithm identifies the potential regions of abrupt rainfall changes using a wavelet-based method. However, the algorithm is not fully automatic as it requires a manual inspection step to remove the potential droughts with low magnitude of rainfall change and short span of persistence.

In contrast, machine learning and data mining methods have been successfully applied to application domains, such as computer vision, natural language processing and others. With a significant increase in the number of climate datasets available, we believe that relevant machine learning and data mining algorithms can also be applicable to climate science. In this paper, we propose a drought detection algorithm based on the well studied Markov Random Field (MRF) model [17].

We formulate the detection problem as the one of finding the most likely configuration of a binary MRF, where each node can only take two values: 1 means the node is in a drought state and 0 means a normal state. Since the gridded precipitation dataset we use is spatio-temporal, i.e., the dataset contains precipitation observations of the globe over a period of time, we construct a 3-dimensional grid graph as the underlying dependency graph for the MRF. More specifically, for a particular time, we model the dependency using a 4-nearest neighbor grid, where each node represents a location. The 3-dimensional grid can be viewed as a replication of 2-dimensional grids and the nodes representing the same location are connected together.

*Department of Computer Science and Engineering, University of Minnesota, Twin Cities.

†Department of Soil, Water and Climate, University of Minnesota, Twin Cities.

We design the potential functions carefully from the climate datasets to ensure spatio-temporal consistency, i.e., the neighboring nodes in the 3-dimensional grid are encouraged to take the same value.

Our goal is to estimate the binary value each node takes based on the MRF and we consider that the nodes with value 1 are in drought states. However, in general the integer programming problem of finding the most likely configuration is computationally intractable and people often resort to relaxation and obtain approximate solutions [17]. Throughout the paper, we use the Linear Programming (LP) relaxation [16], which has been extensively studied in the MRF literature. Instead of solving the LP directly, we adopt the idea of proximal minimization [1] from the optimization literature and propose an efficient inference algorithm. After the algorithm terminates, we round the relaxed fractional solution and obtain the integer solution. We then identify major droughts which are spatially widespread over long duration based on the integer solution.

We apply our drought detection algorithm on the Climate Research Unit (CRU) precipitation dataset [10] over 106 years (1901-2006). The drought detection problem on this dataset is of large scale, since the underlying dependency graph has over 7 million nodes and the number of configurations is more than $2^{7,000,000}$. Our algorithm is fully automatic and solves the problem efficiently, i.e., the algorithm converges within one and a half hours in a Linux workstation. The empirical results show that the algorithm successfully detects the major droughts of the twentieth century, including the Dust Bowl in the 1930s and the drought in the Sahel starting in the late 1960s. We compare our algorithm with the drought detection algorithm in [11] and find that both algorithms detect similar droughts.

The rest of the paper is organized as follows: In Section 2, we give a brief review of MRFs and propose a MAP inference algorithm. We outline the MRF-based drought detection algorithm in Section 3. We show the experimental results on the CRU dataset in Section 4 and conclude in Section 5.

2 Markov Random Fields and MAP Inference

We start by introducing some basic background on Markov Random Fields. An MRF is defined on an undirected graph $G = (V, E)$, where V is the vertex set and E is the edge set. Each node $u \in V$ has a random variable X_u associated with it, which can take value x_u in some discrete space $\mathcal{X} = \{1, \dots, k\}$. Concatenating all the random variables $X_u, \forall u \in V$, we obtain an n dimensional random vector $X = \{X_u | u \in V\} \in \mathcal{X}^n$. We assume that the distribution P of X is a Markov Random Field [17], meaning that it factors according

to the structure of the undirected graph G as follows: With $f_u : \mathcal{X} \mapsto \mathbb{R}, \forall u \in V$ and $f_{uv} : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}, \forall (u, v) \in E$ denoting nodewise and edgewise potential functions respectively, the distribution takes the form:

$$P(x) \propto \exp \left\{ \sum_{u \in V} f_u(x_u) + \sum_{(u,v) \in E} f_{uv}(x_u, x_v) \right\}.$$

An important problem in the context of MRF is that of *maximum a posteriori* (MAP) inference, which is to compute the configuration x^* with the largest probability:

$$x^* \in \operatorname{argmax}_{x \in \mathcal{X}^n} \exp \left\{ \sum_{u \in V} f_u(x_u) + \sum_{(u,v) \in E} f_{uv}(x_u, x_v) \right\}.$$

The above optimization problem is equivalent to the following integer programming problem:

$$(2.1) \quad x^* \in \operatorname{argmax}_{x \in \mathcal{X}^n} \left\{ \sum_{u \in V} f_u(x_u) + \sum_{(u,v) \in E} f_{uv}(x_u, x_v) \right\}.$$

The complexity of (2.1) depends critically on the structure of the underlying graph. When G is a tree structured graph, the MAP inference problem can be solved efficiently via the max-product algorithm [9]. However, for an arbitrary graph G , the MAP inference algorithm is usually computationally intractable. The intractability motivates the development of algorithms to solve the MAP inference problem approximately. In this paper, we focus on the Linear Programming (LP) relaxation method [16]. The LP relaxation of MAP inference problem is defined on a set of pseudomarginals μ_u and μ_{uv} , which are non-negative, normalized and locally consistent:

$$(2.2) \quad \mu_u \geq 0, \quad \forall u \in V,$$

$$(2.3) \quad \mu_{uv} \geq 0, \quad \forall (u, v) \in E,$$

$$(2.4) \quad \sum_{x_u \in \mathcal{X}_u} \mu_s(x_u) = 1, \quad \forall u \in V,$$

$$(2.5) \quad \sum_{x_u \in \mathcal{X}_u} \mu_{uv}(x_u, x_v) = \mu_v(x_v), \quad \forall (u, v) \in E, x_v \in \mathcal{X}_v.$$

We denote the polytope defined by (2.2)-(2.5) as $L(G)$ and the LP relaxation of MAP inference problem (2.1) becomes maximizing the following LP:

$$(2.6) \quad \langle \mu, f \rangle = \sum_{u \in V} \sum_{x_u} \mu_u(x_u) f_u(x_u) + \sum_{(uv) \in E} \sum_{x_u, x_v} \mu_{uv}(x_u, x_v) f_{uv}(x_u, x_v),$$

subject to the constraint that $\mu \in L(G)$.

2.1 Proximal Methods Although standard LP solvers can be used to solve the optimization problem (2.6), they are usually inefficient compared to the algorithms which exploit the graph structure [19]. In this section, we briefly introduce the proximal maximization algorithm [1] which can take advantage of the graph structure and is guaranteed to converge to the global maximizer of (2.6).

Instead of solving the constrained LP (2.6) directly, the proximal maximization method solves a sequence of maximization problems:

$$(2.7) \quad \mu^{t+1} = \operatorname{argmax}_{\mu \in L(G)} \left\{ \langle \mu, f \rangle - \frac{1}{w^t} D_h(\mu \| \mu^t) \right\},$$

where the subscript $t = 1, 2, \dots$ denotes the iteration number, w^t is a positive constant and $D_h(\mu \| \nu)$ is a Bregman divergence [2] between μ and ν induced by the strictly convex function h :

$$D_h(\mu \| \nu) = h(\mu) - h(\nu) - \langle \nabla h(\nu), \mu - \nu \rangle.$$

Now we study the choice of h that will be used in the sequel. To take the advantage of the graph structure, in principle, we can decompose the graph into N (overlapping) parts and assign each part a strictly convex function $h_i, i = 1, \dots, N$. Let μ_i denote the components of μ that belong to part i and $h = \sum_{i=1}^N h_i$. The Bregman divergence then becomes:

$$D_h(\mu \| \nu) = \sum_{i=1}^N h_i(\mu_i) - h(\nu_i) - \langle \nabla h_i(\nu_i), \mu_i - \nu_i \rangle.$$

For the sake of simplicity, we focus on a straightforward decomposition: we decompose the graph into $|V|$ nodes and $|E|$ edges and the strictly convex function h_i is the negative entropy of the pseudomarginals:

$$h_u = \sum_{x_u} \mu_u(x_u) \log \mu_u(x_u), \forall u \in V,$$

$$h_{uv} = \sum_{x_u, x_v} \mu_{uv}(x_u, x_v) \log \mu_{uv}(x_u, x_v), \forall (u, v) \in E.$$

Then the Bregman divergence is the Kullback-Leibler (KL) divergence across all the nodes and edges:

$$D_h(\mu \| \nu) = \sum_{u \in V} D_{h_u}(\mu_u \| \nu_u) + \sum_{(uv) \in E} D_{h_{uv}}(\mu_{uv} \| \nu_{uv}),$$

where $D_{h_u}(\mu_u \| \nu_u) = \sum_{x_u} \mu_u(x_u) \log \frac{\mu_u(x_u)}{\nu_u(x_u)} + \mu_u(x_u) - \nu_u(x_u)$. We note that this is also the setting of one of the proximal algorithms in [12].

2.2 Bregman Projection In this section, we show how to solve the optimization problem (2.7) by performing the Bregman projection [2]. We observe that (2.7) can be solved by first obtaining the solution $\mu^{t+1,0}$ to the unconstrained problem of (2.7) and then projecting $\mu^{t+1,0}$ to $L(G)$. To be more specific, we have:

$$(2.8) \quad \mu^{t+1,0} = \operatorname{argmax}_{\mu} \left\{ \langle \mu, f \rangle - \frac{1}{w^t} D_h(\mu \| \mu^t) \right\},$$

$$(2.9) \quad \mu^{t+1} = \operatorname{argmin}_{\mu \in L(G)} D_h(\mu \| \mu^{t+1,0}).$$

When h is the negative entropies on node and edge pseudomarginals, (2.8) has a closed form solution. Taking derivatives and setting them to zeros yields:

$$\mu_u^{t+1,0}(x_u) = \mu_u^t \exp(w^t f_u(x_u)),$$

$$\mu_{uv}^{t+1,0}(x_u, x_v) = \mu_{uv}^t \exp(w^t f_{uv}(x_u, x_v)).$$

Unfortunately, (2.9) does not have a closed form solution and the projection is usually computed iteratively. In particular, we show that the projection can be obtained by performing a sequence of cyclic Bregman projections.

We note that the polytope $L(G)$ can be viewed as an intersection of the equality constraints, i.e., (2.4) and (2.5) (The inequality constraints (2.2) and (2.3) are taken care of automatically when h is the negative entropy function.). It is easy to see that the number of constraints is $NC = k|V| + 2k|E|$. Denote each equality constraint as $C_i, i = 1, \dots, NC$ and define

$$(2.10) \quad \mu = \Pi_{C_i}(\nu)$$

as the operation of projecting ν onto the constraint C_i . The Bregman projection algorithm projects $\mu^{t+1,0}$ sequentially onto each constraint C_i of $L(G)$ in a cyclic manner, i.e., starting from $\nu = \mu^{t+1,0}$, we perform the following operation repeatedly until convergence:

$$\mu = \Pi_{C_i}(\nu),$$

$$\nu = \mu,$$

where i is the constraint index and $i = 1, \dots, NC, 1, \dots, NC, 1, \dots$. It can be shown [2] that the above cyclic Bregman projection converges to the projection defined in (2.9). It is important to point out that when the constraint set includes inequality constraints, the Bregman projection has to be followed by a correction step [2]. We avoid this by choosing the KL divergence as the Bregman divergence.

We now show that the projection (2.10) onto each constraint has closed form solutions. We use the notation $\mu^{t+1,p}$ to denote the value of μ^{t+1} after $\mu^{t+1,0}$ has been projected to $L(G)$ p times according to the

cyclic projection scheme. Consider the constraint on node u : $\sum_{x_u} \mu_u(x_u) = 1$ and let λ_u be the Lagrangian multiplier associated with the constraint. The KKT condition is:

$$\nabla_h(\mu_u^{t+1,p+1}(x_u)) = \nabla_h(\mu_u^{t+1,p}(x_u)) + \lambda_u^{t+1,p+1} .$$

Expanding the derivatives and performing some simple algebra yields:

$$\begin{aligned} \mu_u^{t+1,p+1}(x_u) &= \mu_u^{t+1,p}(x_u) \exp(\lambda_u^{t+1,p+1}) , \\ \exp(\lambda_u^{t+1,p+1}) &= \frac{1}{\sum_{x_u} \mu_u^{t+1,p}(x_u)} . \end{aligned}$$

Then it follows the normalization update for each node:

$$\mu_u^{t+1,p+1}(x_u) = \frac{\mu_u^{t+1,p}(x_u)}{\sum_{x_u} \mu_u^{t+1,p}(x_u)} .$$

Similarly, we can derive the update for each edge

$$\begin{aligned} \mu_{uv}^{t+1,p+1}(x_u, x_v) &= \mu_{uv}^{t+1,p}(x_u, x_v) \sqrt{\frac{\mu_u^{t+1,p}(x_u)}{\sum_{x_v} \mu_{uv}^{t+1,p}(x_u, x_v)}} , \\ \mu_u^{t+1,p+1}(x_u) &= \sqrt{\mu_u^{t+1,p}(x_u) \sum_{x_v} \mu_{uv}^{t+1,p}(x_u, x_v)} . \end{aligned}$$

In summary, the MAP inference algorithm (Algorithm 1) is a double loop algorithm: In the outer loop, the algorithm performs proximal maximization and in the inner loop, the algorithm performs cyclic Bregman projection. It is shown in [1, 12] that for appropriate choice of weight sequence $\{w^t\}$, Algorithm 1 has super-linear convergence rate.

3 MRF-based Drought Detection Algorithm

In this section, we show how droughts are detected using the MAP inference algorithm presented in Section 2.

3.1 Designing the Potential Functions Climate datasets are usually spatio-temporal datasets in that they have climate variable observations over the globe for a period of time. Suppose a precipitation dataset has yearly precipitation averages over a $m \times n$ global grid over T years, i.e., the resolution is $\frac{180}{m}$ degree latitude $\times \frac{360}{n}$ degree longitude. To use the MRF model, the underlying graph structure has to be determined. To model the spatio nature of the dataset, we use a $m \times n$ 4-nearest neighbor grid for each year. To model the temporal nature, we construct a 3-dimensional grid by connecting the nodes representing the same location in the T 2-dimensional grids. Figure 1 shows the graph structure for $m = 3, n = 4, T = 2$.

Algorithm 1 MAP inference algorithm with proximal maximization and Bregman projection

Input: potential functions $\{f_u, f_{uv}\}$, weight sequence $\{w^t\}$ and the number of integer values k

Output: pseudomarginals μ

Initialization: $\mu_u^1(x_u) = \frac{1}{k}$ and $\mu_{uv}^1(x_u, x_v) = \frac{1}{k^2}$

Outer Loop: For $t = 1, 2, \dots$ until convergence

 Compute μ^{t+1} by the inner loop

 1. Initialization:

$$\mu_u^{t+1,0}(x_u) = \mu_u^t(x_u) \exp(w^t f_u(x_u)) ,$$

$$\mu_{uv}^{t+1,0}(x_u, x_v) = \mu_{(uv)}^t(x_u, x_v) \exp(w^t f_{uv}(x_u, x_v)) .$$

 2. Inner Loop : For $p = 0, 1, \dots$ until convergence

 For each node

$$\mu_u^{t+1,p+1}(x_u) = \frac{\mu_u^{t+1,p}(x_u)}{\sum_{x_u} \mu_u^{t+1,p}(x_u)} .$$

 For each edge

$$\mu_{uv}^{t+1,p+1}(x_u, x_v) = \mu_{uv}^{t+1,p}(x_u, x_v) \sqrt{\frac{\mu_u^{t+1,p}(x_u)}{\sum_{x_v} \mu_{uv}^{t+1,p}(x_u, x_v)}} ,$$

$$\mu_u^{t+1,p+1}(x_u) = \sqrt{\mu_u^{t+1,p}(x_u) \sum_{x_v} \mu_{uv}^{t+1,p}(x_u, x_v)} .$$

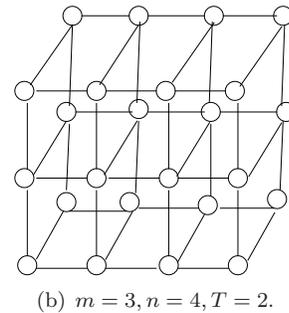
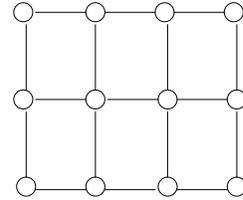


Figure 1: The graph structure for climate datasets used in this paper.

To facilitate the discussion in the sequel, we introduce some notations first: We denote the precipitation observation at location u at time t as y_u^t . The node-wise potential of location u at time t is $f_u^t(x_u^t)$. For two neighboring locations at time t , we denote the pairwise potential as $f_{uv}^t(x_u^t, x_v^t)$. For the same location u at time $t-1$ and t , we denote the pairwise potential as $f_u^{t-1,t}(x_u^{t-1}, x_u^t)$.

We are now ready to define the potential functions based on the 3-dimensional grid. To detect drought regions, we set $k=2$, where $x_u^t=1$ means that location u at time t is in a drought state, i.e., abnormal state, and $x_u^t=0$ means a normal state. To define the nodewise potential function for each location u , we partition the observations $y_u^t, t=1, \dots, T$ into two parts and we consider the observations below the $p\%$ percentile as abnormal and the rest as normal. We also compute $\mu_u^{abnormal}$ and μ_u^{normal} , the mean of the abnormal and normal observations respectively. Then the nodewise potential function comes from the log-likelihood of a Gaussian distribution

$$\begin{aligned} f_u^t(x_u^t=1) &= \log \mathcal{N}(y_u^t | \mu_u^{abnormal}, \sigma_u^2), \\ f_u^t(x_u^t=0) &= \log \mathcal{N}(y_u^t | \mu_u^{normal}, \sigma_u^2), \end{aligned}$$

where σ_u is the standard deviation of the observations at location u .

We define the pairwise potential functions to encourage label consistency, i.e., the potential value is higher if neighboring nodes take same values. Specifically, we set the pairwise potential as follows:

$$f_{uv}^t(x_u^t, x_v^t) = \begin{cases} C_1 > 0, & \text{if } x_u^t = x_v^t; \\ 0, & \text{Otherwise.} \end{cases}$$

and

$$f_u^{t-1,t}(x_u^{t-1}, x_u^t) = \begin{cases} C_2 > 0, & \text{if } x_u^{t-1} = x_u^t; \\ 0, & \text{Otherwise.} \end{cases}$$

Intuitively, the higher C_1 is, the more likely the neighboring nodes in the 2-dimensional grids are to take same values. Similarly, the higher C_2 is, the nodes representing the same location at consecutive time intervals are to take same values.

3.2 Obtaining the Integer Solution from the Pseudomarginals After the potential function is defined, we can run the MAP inference algorithm presented in Section 2 and compute the pseudomarginals μ . To find the drought regions, we need to decode the fractional solution μ and obtain the integer configuration \mathbf{x} . The value each node takes tells us whether the location is in a drought state or not.

We employ a simple node-based rounding scheme to interpret the pseudomarginals μ :

$$(3.11) \quad x_u = \operatorname{argmax}_{x' \in \mathcal{X}_u} \mu_u(x').$$

We apply (3.11) to each node and obtain the corresponding integer solution.

Now each node has an integer value associated with it: If $x_u^t=1$, we consider that the location u at time t is in a drought state, otherwise it is in a normal state.

3.3 Drought Detection from the Integer Solution Once we have the integer solution, we can detect droughts based on it. Since a drought can be defined both spatially and temporally, we define it as a set of neighbouring nodes in the three-dimensional dependency graph (Figure 1(b)) whose states are drought. Thus, the drought detection problem becomes one of finding sets of neighbouring nodes with drought states. To accomplish this goal, we first construct a three-dimensional adjacency graph and calculate the connected components of this graph. We treat each connected component as a drought. In the adjacency graph, two nodes are connected if both nodes are in drought states, i.e., we simply remove the edges from the dependency graph if at least one node is not in the drought state. A connected component is defined as a subgraph of the adjacency graph in which any two vertices are connected to each other by paths and which is connected to no additional vertices. Since a connected component may only contain a few locations and may not be long in duration, we only pick the sizable components and consider them as major droughts. The details as how to select connected components are in Section 4.

3.4 Practical Issues Algorithm 1 is a general MAP inference algorithm and can be applied to estimate the mostly likely configuration for any pairwise MRF. In practice, we find several ways to speed up the algorithm for our application.

The first option is that, instead of a uniform initialization on μ , we can initialize the pseudomarginals based on the precipitation value. To be more specific, we set $\mu_u^t=1$, if the climate variable on node u at time t is below the $p\%$ percentile and $\mu_u^t=0$, otherwise. We attribute the speed up to the conjecture that the value-based initialization is more close to the optimal solution than the uniform initialization.

The other option is to divide the globe into several disjoint parts, run the MAP inference algorithm on each part and combine the results. Since we can estimate the configuration for each part simultaneously, we can gain significant speed up due to parallel computing. Some climate datasets have disjoint parts in nature.

For example, since the CRU only has precipitation over land, North America, South America and Australia are isolated from the rest of the world.

Finally, we choose the weight scalar $\{w^1, \dots, w^T\}$ such that $w^T \rightarrow 0$. In this case, we observe a super-linear convergence rate if we terminate the inner loop when the change of μ between two consecutive iterations is less than 10^{-3} .

4 Experimental Results

In this section we show the drought regions detected by the algorithm using the CRU dataset, which has monthly precipitation from the years 1901-2006. The dataset (Figure 2) is of high gridded spatial resolution (0.5 degree latitude \times 0.5 degree longitude) and only includes the precipitation over land (67420 locations with precipitation records). To eliminate the monthly variance, we convert the monthly dataset to a yearly dataset by calculating the average precipitation over 12 months for each year.

We first apply the drought detection algorithm over the United States and the Sahel region and show that our algorithm successfully discovers the dust bowl in the 1930s and the prolonged drought in the Sahel starting in the late 1960s. We then apply the algorithm to the global dataset and report the droughts that we identify.

For all the experiments, we set $C_1 = 0.5$, $C_2 = 3$, $w^1 = 1$ and $w^{t+1} = 0.8w^t$. We terminate both the inner and outer loop when the change of μ between two consecutive iterations is less than 10^{-3} .

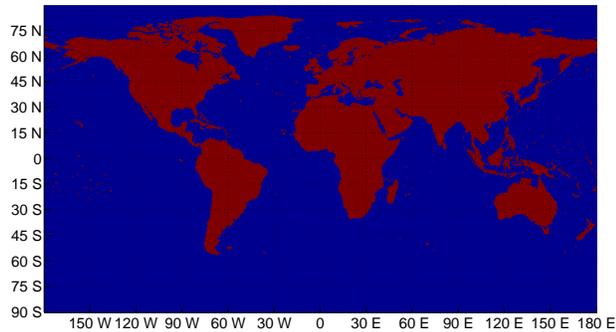


Figure 2: The CRU dataset is a highly gridded dataset containing precipitation for land locations only (red region).

4.1 The Dust Bowl The Dust Bowl, in the 1930s, was one of the most devastating droughts of the past century in the Great Plains region of the United States. The severe drought affected almost two-thirds of the country and was infamous for the numerous dust storms

that occurred [15].

In this subsection, we show how our approach can be used to detect the Dust Bowl drought from the CRU precipitation dataset. We first extract the precipitation of the United States and run the drought detection algorithm on this sub-dataset with $p = 15$. After the algorithm outputs the pseudomarginals, we compute the integer solution. Figure 3 shows the number of locations with drought state detected by our algorithm. We compare our drought detection algorithm with a simple threshold algorithm: If the precipitation in a location is lower than the $p\%$ percentile, we consider it in a drought state. Otherwise, it is in a normal state. It is very obvious from Figure 3 that our algorithm detects a drought in the US in the 1930s, while the threshold algorithm does not provide any meaningful results.

We then compute the connected components from the integer solution and find that the component starting from the year 1928 corresponds to the Dust Bowl region. This connected component spans over 13 years (1928-1940). The drought region we detect is shown in Figure 4. For each location in the drought region, we count the number of years it is included in the connected component and divide this number by the number of years that the component spans. Figure 4 shows the resulting ratio associated with each location in a color. We find that the Dust Bowl map is similar to the one in [15].

We also carefully examine the solution provided by the threshold algorithm. Since the threshold algorithm fails to capture the spatial and temporal consistency, we find that the nodes with drought states are isolated from each other and do not form sizable connected components. In contrast, the MRF model encourages neighboring nodes to be in same states and, as a result, the solution by the drought detection algorithm has less number of ‘drought’ locations, as shown in Figure 3.

To show the drought region detected by the algorithm is valid, we also draw a time series plot (Figure 5) of the average precipitation of the Dust Bowl region we identify. It is clear to see that a sudden reduction in precipitation occurred around 1930 and the drought lasts for about 10 years.

4.2 The Sahel Region In this subsection, we show the results of detection of the 30-year drought in the Sahel starting in the late 1960s [5]. We extract the precipitation of the Sahel region and run the drought detection algorithm on this sub-dataset. We also use the 15% percentile. Figure 6 shows the number of locations with drought states detected by both algorithms. It is not surprising that the base line algorithm detects many more drought locations than our algorithm. We

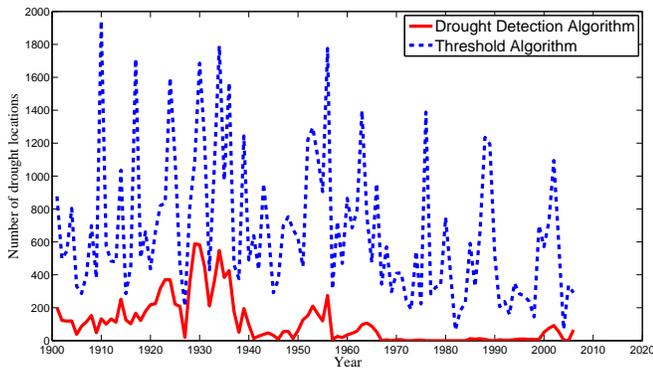


Figure 3: The number of locations with drought states in the United States detected by the drought detection and threshold algorithms.

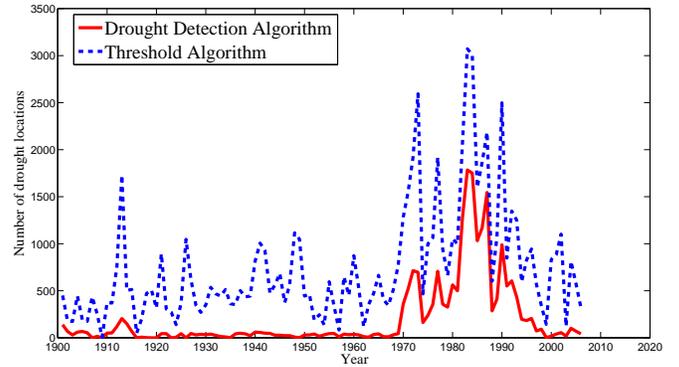


Figure 6: The number of locations with drought states in the Sahel region detected by the drought detection and threshold algorithms.

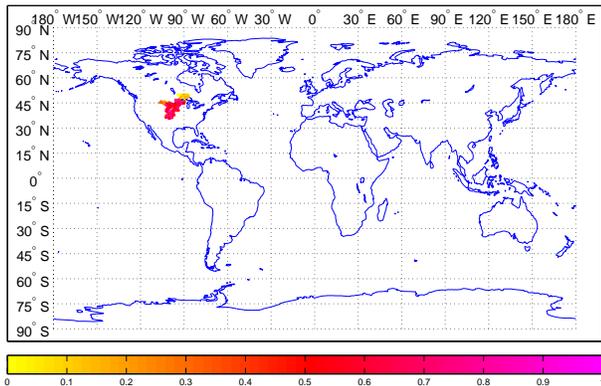


Figure 4: The dust bowl drought region, which corresponds to the connected component starting in 1928.

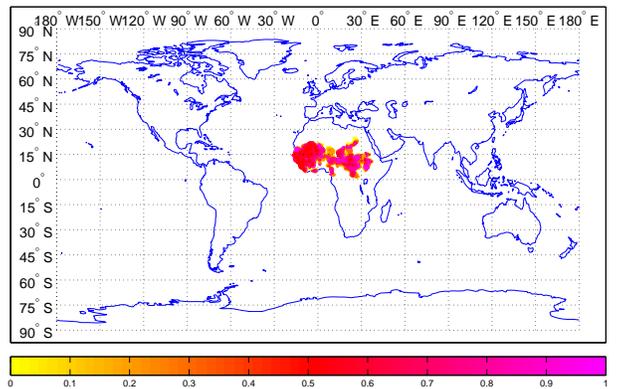


Figure 7: The prolonged drought in the Sahel region, which corresponds to the largest connected component starting in 1968.

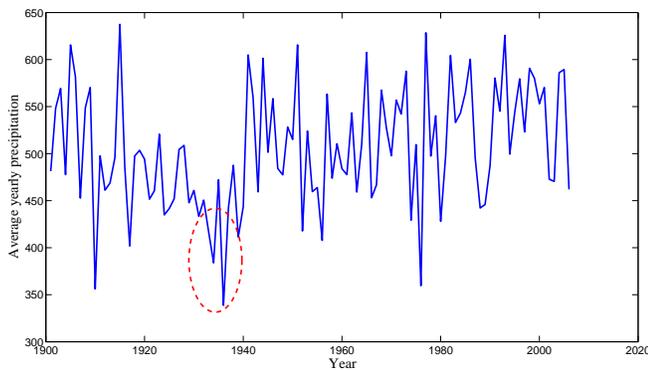


Figure 5: The time series plot of the average precipitation of the Dust Bowl region. The red ellipse shows the sudden reduction in precipitation in the 1930s.

find that the largest component corresponds to the prolonged drought in the Sahel region starting in the late 1960s. The connected component spans 31 years (1968 to 1998). The drought region we detect is shown in Figure 7 and the color code is the same as the one used in Figure 4. We find that the drought map is similar to the one in [8]. The time series plot (Figure 8) shows the 30-year precipitation reduction in the area.

4.3 Global Data Finally, we apply our drought detection algorithm to the entire CRU dataset. We want to emphasize that since the CRU dataset has a high resolution and the underlying 3-dimensional grid has 7,146,520 nodes (67,420 nodes per year \times 106 years), the drought detection problem is of large scale.

Since the global precipitation exhibits large variance, a uniform percentile may not be sufficient. To take this into account, for each location, we compute

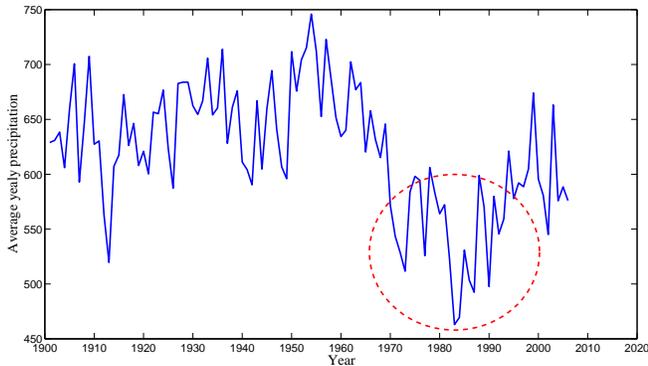


Figure 8: The time series plot of the average precipitation of the Sahel drought region. The red ellipse shows the decades long reduction in precipitation.

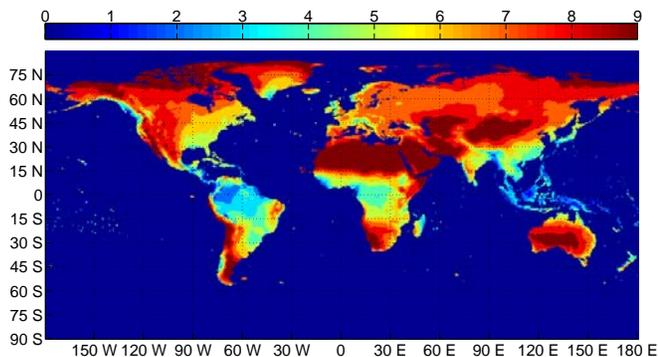


Figure 9: The k-means clustering on the medians with $k = 9$. Each color represents a different cluster. The cluster in dark red (cluster index 9) indicates the lowest precipitation while the blue cluster (cluster index 1) indicates the highest precipitation.

the median precipitation over the 106 years and run the k -means clustering algorithm on the medians with $k = 9$. Figure 9 shows the 9 clusters over the globe. Intuitively, the locations in the clusters with low precipitation are more likely to experience droughts than those in the clusters with high precipitation. Thus, we sort the clusters according to the mean precipitation in descending order and set $p = 15$ for the first three clusters, $p = 10$ for the next three clusters and $p = 15$ for the rest.

After obtaining the integer solution from the pseudomarginals, we compute the connected components. Since a significant drought should be spatially widespread over a long duration, we first select the largest 200 connected components and then further pick among them the ones which last for more than 5 years. We consider the resulting connected components as ma-

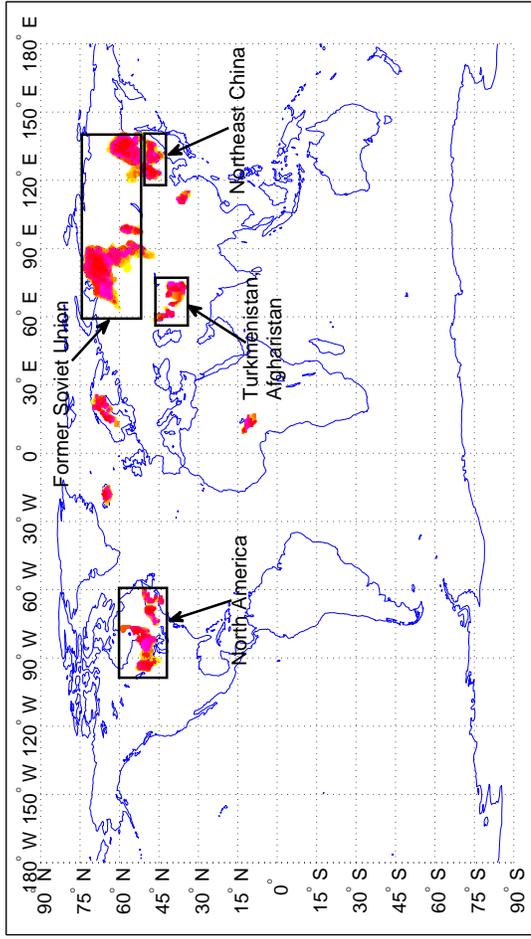
ior droughts.

Our algorithm runs efficiently on the CRU dataset and converges within one and a half hours in a Linux workstation. The significant droughts are shown in Figure 10-11 and each sub-figure shows the droughts beginning in a particular decade. Besides the three-decade drought in the Sahel region starting in the late 1960s and the Dust Bowl in the 1930s, the algorithm also detects the drought in the southwest US and northern Mexico in the 1950s, the region’s most severe drought of the 20th century [6]. Other detected strong droughts include: the drought in northeastern China in the 1920s, the drought in Kazakhstan in the 1930s, the drought in west Europe in the 1940s, the drought in Iran in the 1950s, the drought in eastern India and Bangladesh in the 1960s and the drought in southern Africa in the 1980s. We find that most of the droughts have a duration of at least 10 years. We also find that the drought regions are mostly located in the arid and semi-arid regions and this observation is consistent with many climate modeling studies [5, 14].

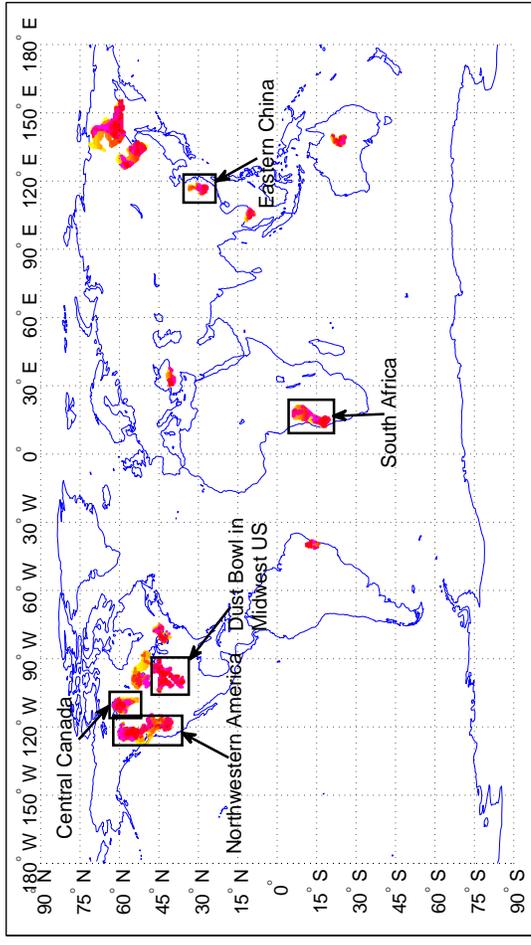
In [11], the authors list the top 30 regions of the world with abrupt decreases in rainfall during the 20th century. Our algorithm discovers all the droughts in the list, except for a drought in Ukraine and two droughts in Australia. The droughts found by both algorithms are shown by black rectangles in Figure 10 and 11.

5 Conclusions

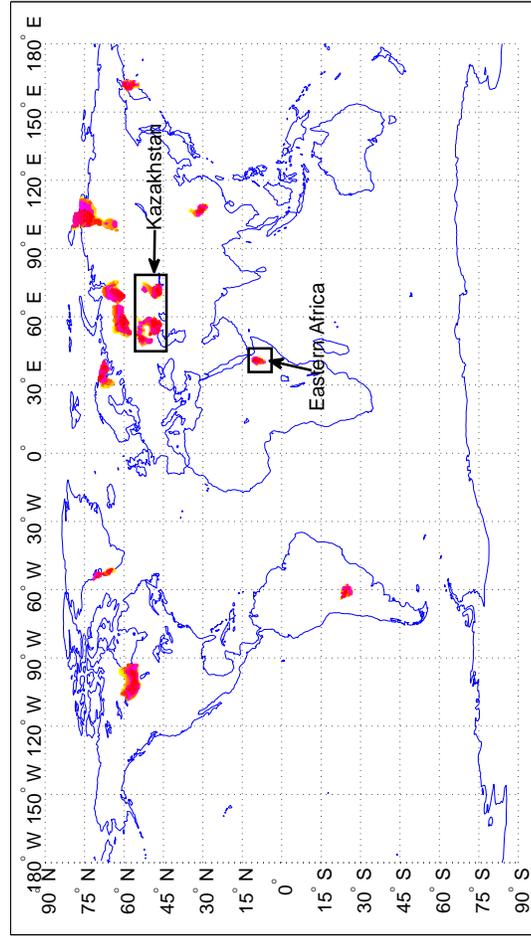
Due to the importance of understanding droughts, we consider the problem of their detection in this paper and develop a fully automatic drought detection algorithm. We formulate the problem as the one of estimating the most likely configuration of a binary MRF, where each node can be in either a drought or normal state. We adopt the proximal maximization and Bregman cyclic projection scheme for the MAP inference task. To maintain spatio-temporal consistency, we design the potential functions to encourage the neighbouring nodes to take same values. We run the algorithm on the high resolution CRU precipitation dataset and it efficiently solves this large scale problem with over 7 million variables. The empirical results shows that we successfully identify some well-documented drought regions of the last century in different parts of the world. We want to emphasize that even though we mainly run our algorithm on a precipitation dataset, the methodology is applicable to other climate variables as well. We plan to extend the current algorithm to handle multi-variate climate datasets in the future, e.g., incorporating the soil moisture variable to the model. We are also interested in applying the algorithm to the model output datasets.



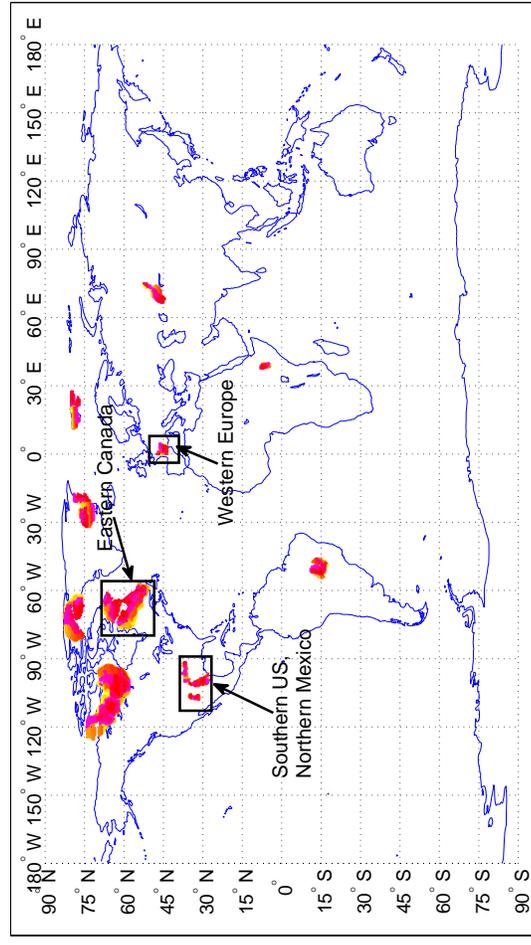
(a) Major droughts starting within the period 1905-1920.



(b) Major droughts starting within the period 1921-1930.



(c) Major droughts starting within the period 1931-1940.



(d) Major droughts starting within the period 1941-1950.

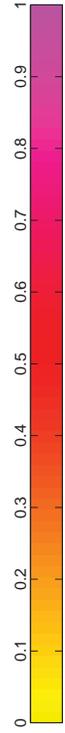
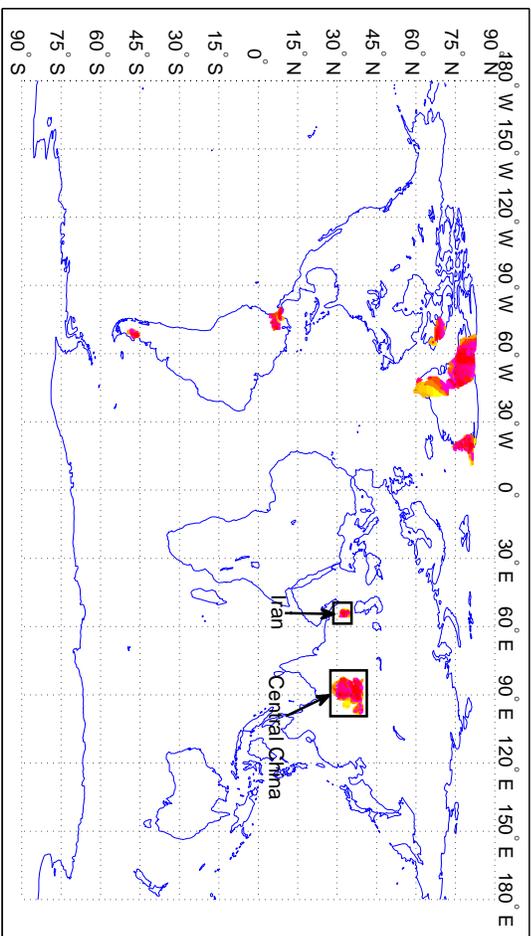
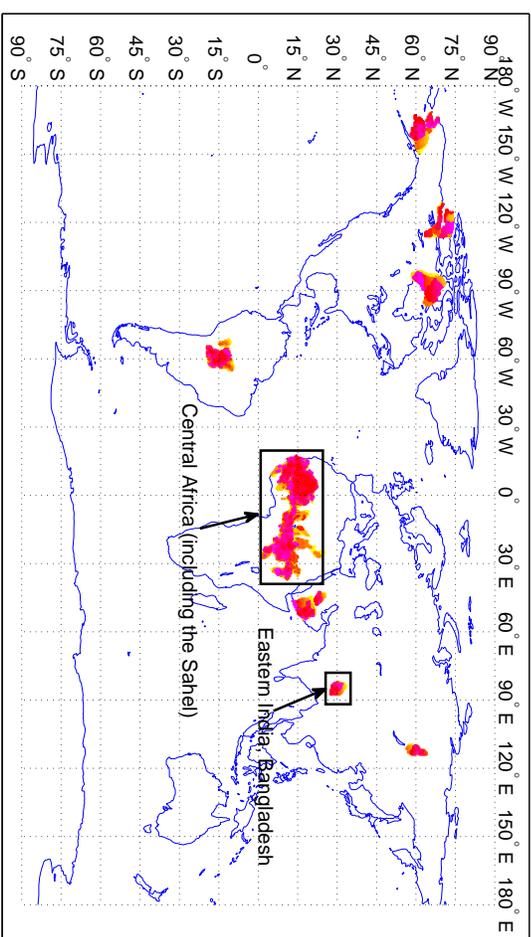


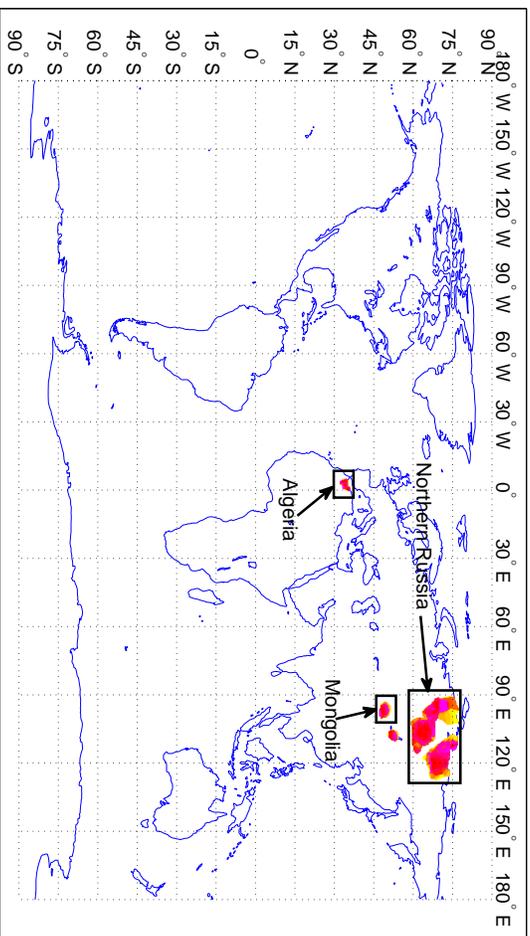
Figure 10: The drought regions detected by our algorithm. Each sub-figure shows the droughts starting from a particular decade. The regions in black rectangles indicate the common droughts found by [11].



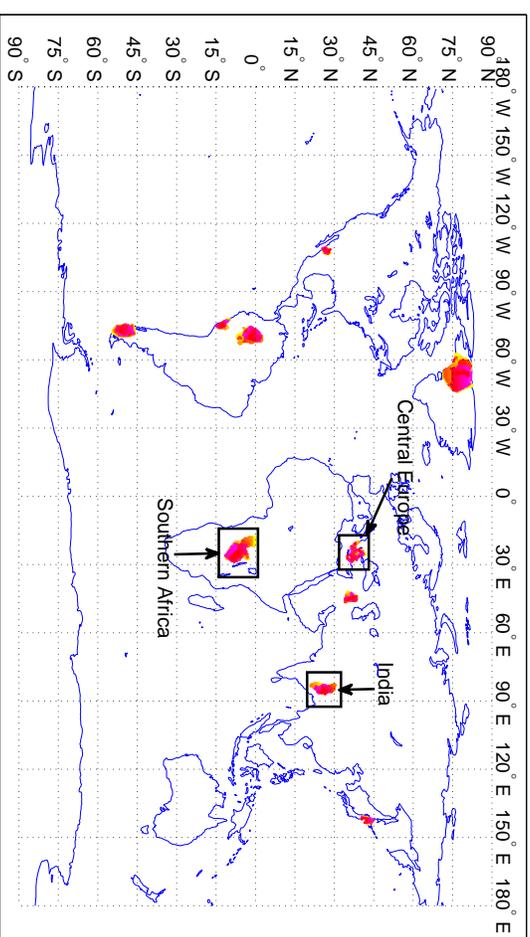
(a) Major drought starting within the period 1951-1960.



(b) Major droughts starting within the period 1961-1970.



(c) Major droughts starting within the period 1971-1980.



(d) Major droughts starting within the period 1981-1995.

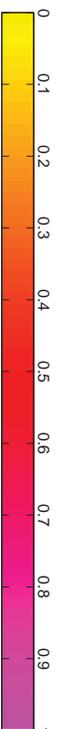


Figure 11: The drought regions detected by our algorithm. Each sub-figure shows the droughts starting from a particular decade. The regions in black rectangles indicate the common droughts found by [11].

Acknowledgements: This research was supported in part by NSF Grants IIS-1029711, IIS-0916750, IIS-0812183, and NSF CAREER Grant IIS-0953274. The authors are grateful for the technical support from the University of Minnesota Supercomputing Institute (MSI).

References

- [1] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and distributed computation: numerical methods*, Athena Scientific, Boston, MA, 1997.
- [2] Y. Censor and S. Zenios, *Parallel optimization: Theory, algorithms, and applications*, Oxford University Press, 1998.
- [3] K. H. Cook, *Climate science: The mysteries of sahel drought*, Nature Geoscience **1** (2008), no. 10, 647–648.
- [4] P. A. Dirmeyer and J. Shukla, *The effect on regional and global climate of expansion of the world’s deserts*, Quarterly Journal of the Royal Meteorological Society **122** (1996), no. 530, 451–482.
- [5] J. A. Foley, M. T. Coe, M. Scheffer, and G. Wang, *Regime Shifts in the Sahara and Sahel: Interactions between Ecological and Climatic Systems in Northern Africa*, Ecosystems **6** (2003), 524–532.
- [6] F. K. Fye, D. W. Stahle, and E. R. Cook, *Paleoclimatic analogs to twentieth-century moisture regimes across the united states*, Bulletin of the American Meteorological Society **84** (2003), no. 7, 901–909.
- [7] N. B. Guttman, *Comparing the palmer drought index and the standardized precipitation index1*, JAWRA Journal of the American Water Resources Association **34** (1998), no. 1, 113–121.
- [8] M. Hoerling, J. Hurrell, J. Eischeid, and A. Phillips, *Detection and Attribution of Twentieth-Century Northern and Southern African Rainfall Change*, Journal of Climate **19** (2006), no. 16, 3989–4008.
- [9] F. R. Kschischang, B. J. Frey, and H. A. Loeliger, *Factor graphs and the sum-product algorithm*, IEEE Transactions on Information Theory **47** (2001), no. 2, 498–519.
- [10] T. D. Mitchell, T. R. Carter, P. D. Jones, M. Hulme, and M. New, *A comprehensive set of high-resolution grids of monthly climate for Europe and the globe: the observed record (1901-2000) and 16 scenarios (2001-2100)*, vol. Working Paper 55, Tyndall Centre for Climate Change Research, 2004.
- [11] G. T. Narisma, J. A. Foley, R. Licker, and N. Ramankutty, *Abrupt changes in rainfall during the twentieth century*, Geophysical Research Letters **34** (2007), L06710+.
- [12] P. Ravikumar, A. Agarwal, and M. J. Wainwright, *Message-passing for graph-structured linear programs: Proximal methods and rounding schemes*, Journal of Machine Learning Research **11** (2010), 1043–1080.
- [13] M. Scheffer, S. Carpenter, J. A. Foley, C. Folke, and B. Walker, *Catastrophic shifts in ecosystems*, Nature **413** (2001), no. 6856, 591–596.
- [14] M. Scheffer, M. Holmgren, V. Brovkin, and M. Claussen, *Synergy between small- and large-scale feedbacks of vegetation on the water cycle*, Global Change Biology **11** (2005), no. 7, 1003–1012.
- [15] S. D. Schubert, M. J. Suarez, P. J. Pegion, R. D. Koster, and J. T. Bacmeister, *On the cause of the 1930s dust bowl*, Science **303** (2004), 1855–1859.
- [16] M. J. Wainwright, T. S. Jaakkola, and A. S. Willsky, *MAP estimation via agreement on (hyper)trees: Message-passing and linear-programming approaches*, IEEE Transactions of Information Theory **51** (2005), no. 11.
- [17] M. J. Wainwright and M. I. Jordan, *Graphical models, exponential families, and variational inference*, Foundations and Trends in Machine Learning **1** (2008), no. 1-2, 1–305.
- [18] C. A. Woodhouse and J. T. Overpeck, *2000 years of drought variability in the central United States*, Bulletin of the American Meteorological Society **79** (1998), 2693–2714.
- [19] C. Yanover, T. Meltzer, and Y. Weiss, *Linear programming relaxations and belief propagation: an empirical study*, Journal of Machine Learning Research **7** (2006), 2006.