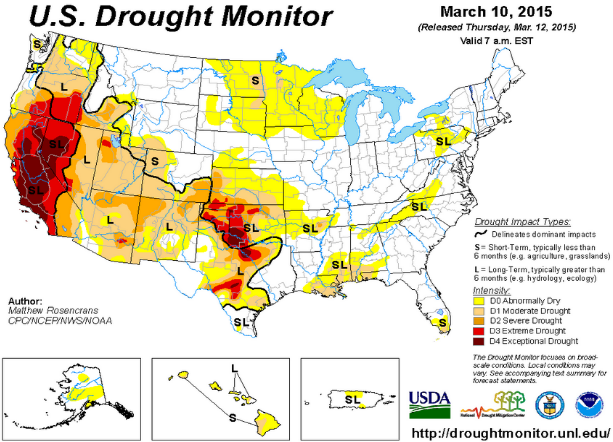
Model Reduction Techniques for Spatiotemporal Data Analysis in Drought Modeling

MARILYN VAZQUEZ, MUHAMMAD BAQUI
ADVISOR: MARIA EMELIANENKO

Department of Mathematics
George Mason University
mvazque3@masonlive.gmu.edu

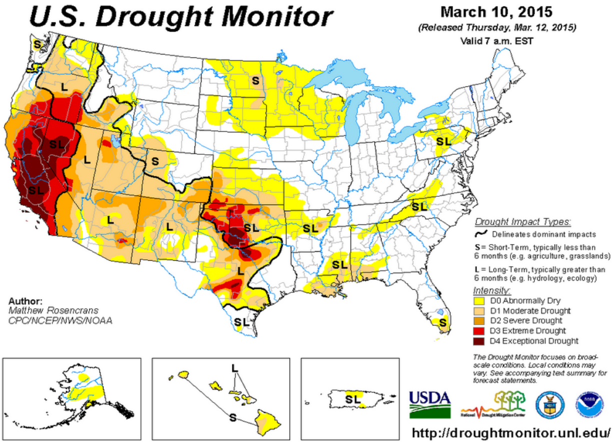
STudent REsearch TalkS
April 17, 2015

Interesting Questions



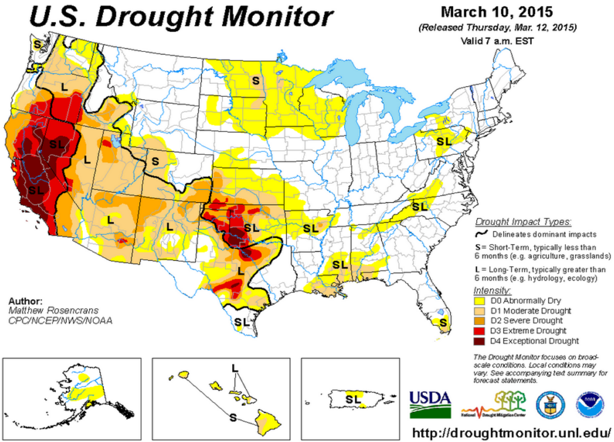
Are the droughts increasing in coverage and duration or are they pretty much stationary?

Interesting Questions



Will analysis on compressed data yield similar results compared to actual data?

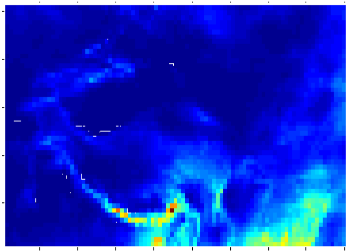
Interesting Questions



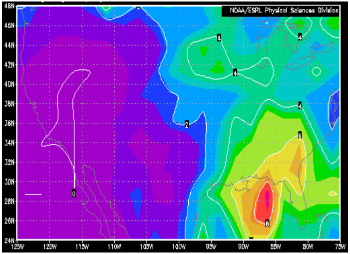
What is the best distribution of gauges to efficiently measure precipitation?

Challenges: Different Data from Different Sources

Precipitation for June 1994



Data from MERRA database

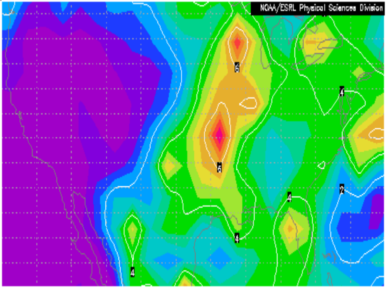


Data from NOAA

Image provided by Physical Sciences Division, Earth System Research Laboratory, NOAA, Boulder, Colorado, from their Web site at http://www.esrl.noaa.gov/psd/.

Challenges: Large Data

Precipitation for June 2000

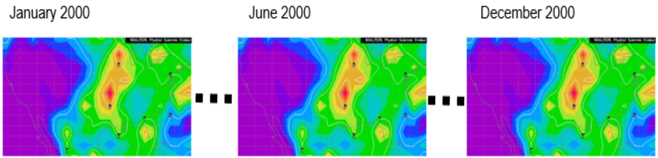


90 by 51 pixels (4590 pixels)

Image provided by Physical Sciences Division, Earth System Research Laboratory, NOAA, Boulder, Colorado, from their Web site at http://www.esrl.noaa.gov/psd/.

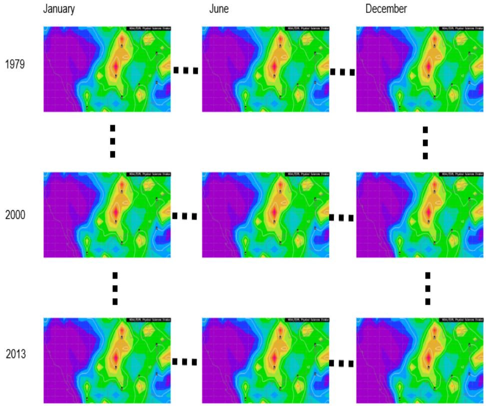
Challenges: Large Data

12 months



90 by 51 by 12 pixels (55080 pixels)

Challenges: Large Data



90 by 51 by 12 by 35 pixels (1927800 pixels)

Data

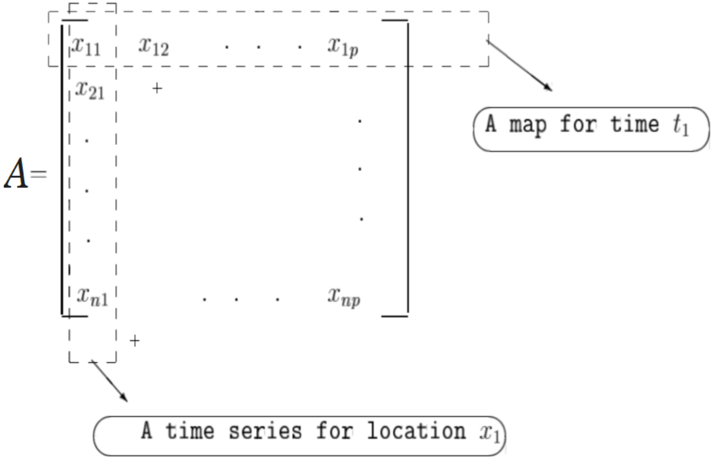


Figure : Each row is one map and each column is the time series observation for a given location. [1]

Interesting Questions

- Are the droughts increasing in coverage and duration or are they pretty much stationary?
- If we compress the data and analyze it, will it give us similar results to analyzing the whole data?
- What is the best distribution of gauges to efficiently measure precipitation?

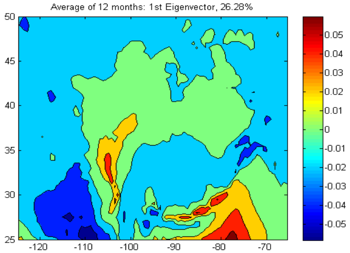
Eigenvector maps and Variation

$A = U\Sigma V^T$

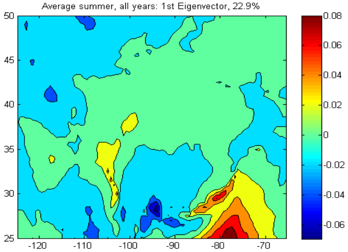
Classical calculation of SVD [1]

1. Form the matrix $A^T A$
2. Find eigenvalues $\lambda_j = \sigma_j^2$ for singular values and normed eigenvectors i.e. $||v_j|| = 1$ to form right unitary matrix
3. Project eigenvectors $u_j = \frac{1}{\sigma_j} A v_j$ to form left unitary matrix

Preliminary Results



(a)



(b)

Figure : The contour eigenvector map corresponding to the first eigenvalue for an average of (a) all months and (b) summer months only.

Preliminary Results

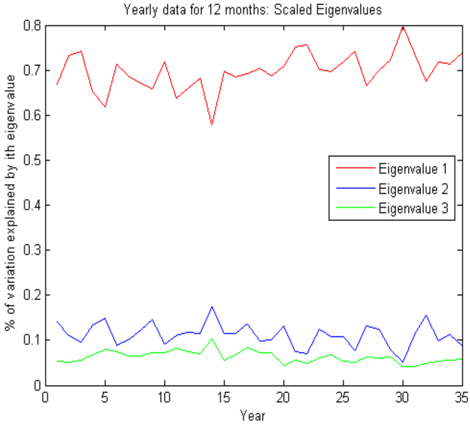


Figure : The percentage of variation that the first 3 eigenvalues represent for the 35 years. This is calculated as follows: $\frac{\lambda_k}{\sum_j \lambda_j}$

Interesting Questions

- Are the droughts increasing in coverage and duration or are they pretty much stationary?
- If we compress the data and analyze it, will it give us similar results to analyzing the whole data?
- What is the best distribution of gauges to efficiently measure precipitation?

SVD and Compression

$$A = \sigma_1 u_1 v_1^T + \cdots \sigma_n u_n v_n^T$$

$$\epsilon = \sum_{j=1}^{n} |a^{[j]} - P_{\psi,d} a^{[j]}|^2$$

where, $P_{\psi,d} a^{[j]} = \sum_{i=1}^{d} c_{ji}$

$$c_{ji} = v_i^T a^{[j]}$$

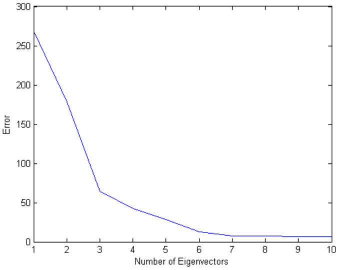
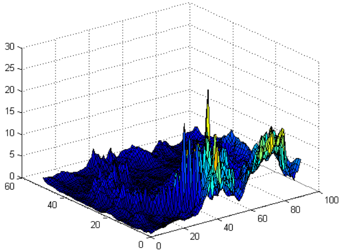
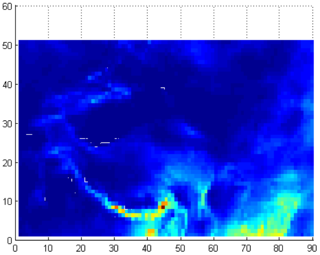


Figure : The error of including d many eigenvectors in the reconstruction of A

Compressing the Intensity Domain



(a)

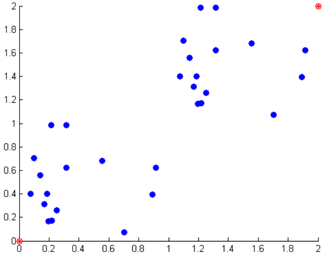


(b)

Figure : June 2000 precipitation (a) side view and (b) top view.

K-means algorithm

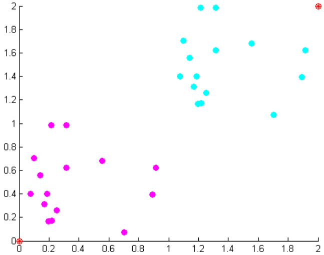
Step 1: Minimize distance [2]



$$C_i = \{x_k \in S : |x_k - z_i| \leq |x_k - z_j| \text{ for } j = 1, \cdots, d \text{ and } j \neq i\}$$

K-means algorithm

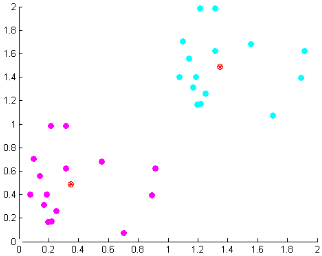
Step 1: Minimize distance [2]



$$C_i = \{x_k \in S : |x_k - z_i| \leq |x_k - z_j| \text{ for } j = 1, \cdots, d \text{ and } j \neq i\}$$

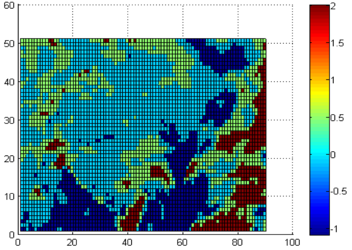
K-means algorithm

Step 2: Center [2]

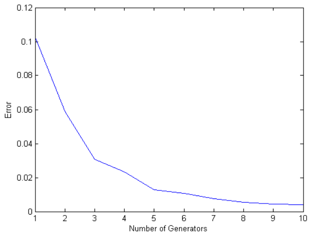


$$\hat{z}_i = \frac{1}{m_i} \sum_{x_k \in C_i} x_k$$

Results



(a)



(b)

Figure : (a)MATLAB's built-in k-means function applied on detrended precipitation data for a particular June. Number of generators = 4.(b) Error of representing the data with 1 through 10 generators.

New Focus

- Are the droughts increasing in coverage and duration or are they pretty much stationary?
- Which parameters contribute most to modeling drought?
- If we compress the data and analyze it, will it give us similar results to analyzing the whole data?
- What is the best distribution of gauges to efficiently measure precipitation?

Where to place the gauges?

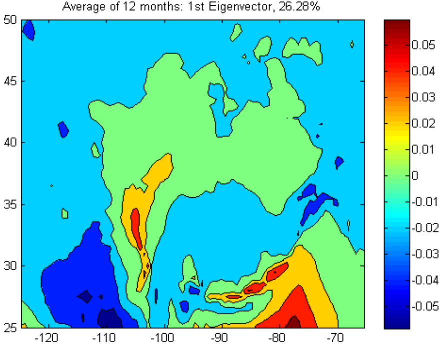


Figure : First eigenvector map for average over all months.

Centroidal Voronoi Tessellation

Tessellation instead of cluster [2]

$$V_i = \{x \in W : |x - z_i| \leq |x - z_j| \text{ for } j = 1, \cdots, d \text{ and } j \neq i\}$$

Center of mass instead of mean

$$\hat{z}_i = \frac{\sum_{x \in V_i} \rho(x) x}{\sum_{x \in V_i} \rho(x)}$$

Minimize cost function instead of distance

$$\epsilon = \sum_{i=1}^{d} \sum_{x \in V_i} \rho(x) |x - z_i|^2$$

Centroidal Voronoi Tessellation

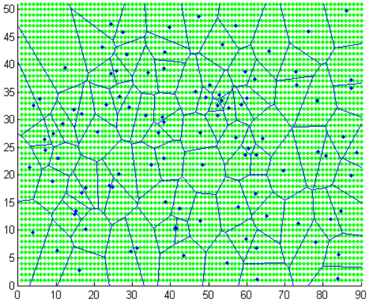


Figure : Random points used as generators of the tessellation.

Centroidal Voronoi Tessellation

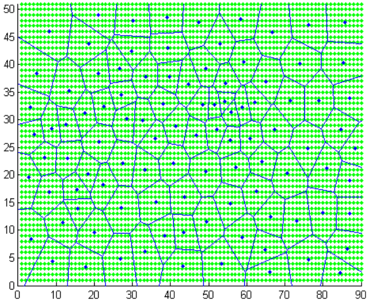


Figure : Second iteration.

Centroidal Voronoi Tessellation

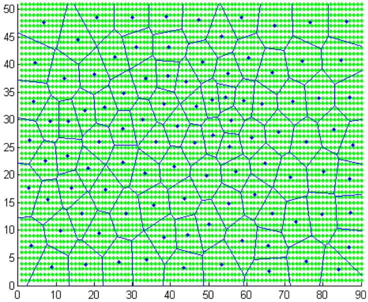


Figure : Fifth iteration.

Centroidal Voronoi Tessellation

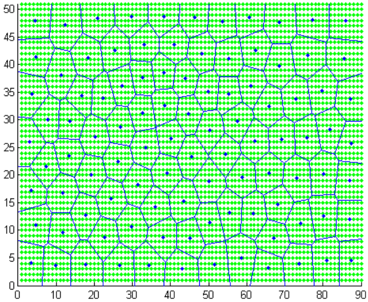
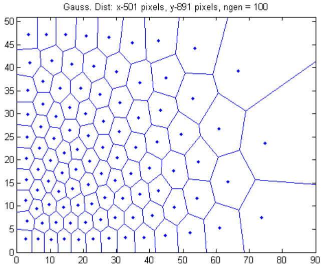
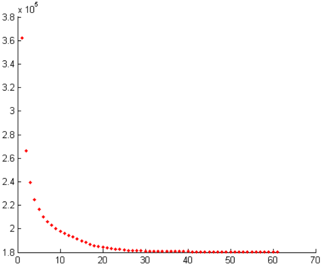


Figure : 29th iteration.

Centroidal Voronoi Tessellation



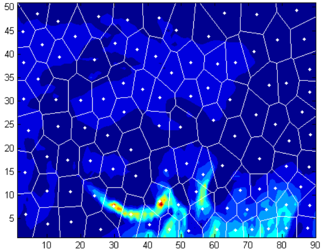
(a)



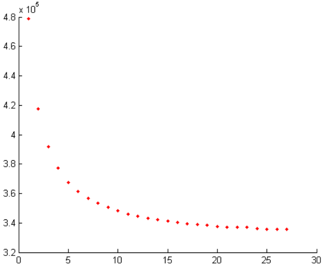
(b)

Figure : (a) The CVT using an exponential density function.(b) A regular cost function.

CVT and Gauge Locations



(a)



(b)

Figure : (a) The CVT on top of the first eigenvector map for average over 12 months (26.28%).(b) The cost function (not normalized).

References

- H Bjornsson.
  A manual for eof and svd analyses of climatic data.
  CCGCR Report, 1997.

- John Burkardt, Max Gunzburger, and Hyung-Chun Lee.
  Centroidal voronoi tessellation-based reduced-order modeling of complex systems.
  SIAM Journal on Scientific Computing, 28(2):459–484, 2006.

- Max Gunzburger and Janet Peterson.
  Reduced-order modeling of complex systems with multiple system parameters.
  In Ivan Lirkov, Svetozar Margenov, and Jerzy WaÅŻniewski, editors, *Large-Scale Scientific Computing*, volume 3743 of *Lecture Notes in Computer Science*, pages 15–27. Springer Berlin Heidelberg, 2006.