Using Historical Data for Retrospective Prediction of Rainfall

UMBC REU Site: Interdisciplinary Program in High Performance Computing

Ephraim Alfa[1], Huiyi Chen[2], Kristen Hansen[3], Mathew Prindle[4],

RA: Sai Popuri[1] and Nadeesri Wijekoon[1], Faculty mentor: Kofi Adragni[1], Client: Amita Mehta[5]

[1]UMBC, [2]UC Merced, [3]University of Minnesota, Morris, [4]California State University Sacremento, [5]JCET

Motivation

The Missouri River Basin (MRB) is an important food-producing region in the United States and Canada. Climate variability and water availability affect crop production in this region because most of the area in not irrigated. Past climate data have been recorded at various locations in the basin. A reduction of the dimensionality of the data was undertaken. The resulting reduced data are used to retrospectively predict rainfall.

Missouri River Basin

Methodology

MADE: Minimum Average Deviance Estimation

▶ Given $Y \in \mathbb{R}$ and $X \in \mathbb{R}^p$, MADE aims at finding a sufficient dimension reduction $B^T X$ so that

$$Y \perp X \mid B^T X.$$

▶ Assume $Y|X$ is of exponential family distribution

$$f(Y \mid X) = \exp\{[Y\vartheta(X) - b(\vartheta(X))]/a(\phi)\} f_0(X)$$

▶ Locally model $\vartheta(X)$ at X_i as

$$\vartheta(X_i) = \alpha + \gamma^T B^T (X_i - X)$$

▶ Evaluate a local likelihood as

$$L_X = \sum_{i=1}^{n} w_{0i}(X) \log f(Y_i|\alpha, \gamma, X_i)].$$

▶ Obtain the deviance function $Q(B, \alpha, \beta)$

▶ B is estimated by optimizing the deviance function over the Stiefel manifold.

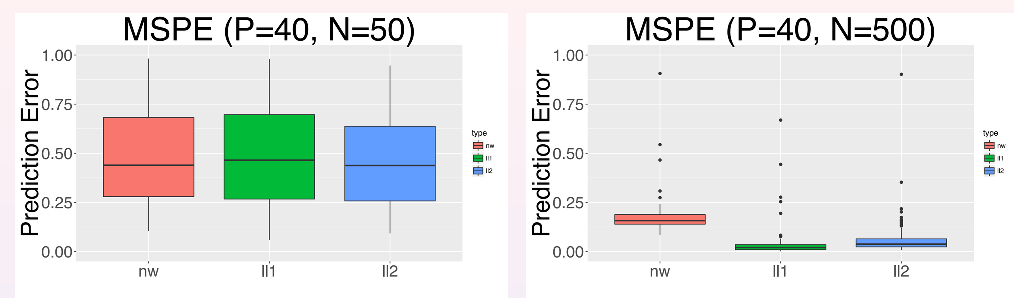
▶ Three prediction methods are compared:

- Nadaraya-Watson $\hat{E}(Y|X^*) = \sum_{i=1}^{n} w_{i*} Y_i$
- MADE prediction I: $\hat{E}(Y|X^*) = g^{-1}(\hat{\alpha})$
- MADE prediction II:

$$\hat{E}(Y|X^*) = g^{-1}\left(\sum_{i=1}^{n} (\hat{\alpha} + \hat{\gamma}^T \hat{B}^T (X_i - X^*)) w_{i*}\right)$$

▶ Mean Squared Prediction error: $\frac{1}{n_e} \sum_{i=1}^{n_e} (Y_i - \hat{Y}_i)^2$

Simulation Study

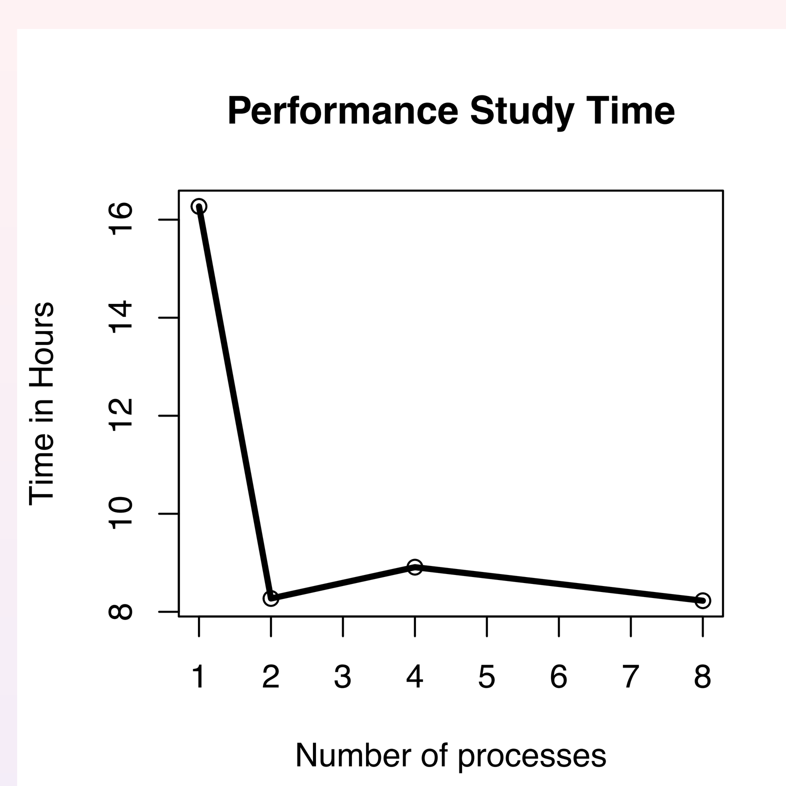


- The prediction error for MADE decreased as the number of observations and covariates increased.
- The Nadaraya Watson estimator performed worse than both MADE specific estimators when the number of observations was highest

Simulation Table

Median Prediction Error for LL1			
p	**n=50**	**n=250**	**n=500**
5	0.050	0.036	0.040
10	0.078	0.036	0.011
20	0.097	0.139	0.059
40	0.597	0.065	0.020

Performance Study



- Code was parallelized using snow and snowslurm packages on R (statistical computing software)
- Data sampled has 500 observations per sample, 40 covariates
- Time to run one hundred samples serially took over 16 hours
- Almost perfect efficiency from 1 to 2 processes, followed by a plateau
- Implemented on the maya cluster at UMBC

Application

- The Missouri River Basin (MRB) data is time-series data collected over a period of 56 years from JCET which consist of precipitation measurements and other climate variables in the MRB.
- The data set contains over 18,000 observations that represents different locations from 1949-2005.
- The data consist of latitude, longitude, date recorded, precipitation measurement, etc.
- Tested a single location over ten years (1950–1960) of precipitation data with three climate variables and two monthly precipitation lag terms
- Ran a leave one out cross validation with the LL1 estimator
- Had an ending result of 0.1136 for the mean squared prediction error
- This suggests MADE was successful in reducing the dimension of X from 5 to 1 while still retaining regression information

Conclusions

- Minimum Average Deviance Estimation (MADE) specific estimators are more accurate than the Nadaraya-Watson prediction estimator
- Minimum Average Deviance Estimation is most effective on data with many covariates and observations
- Further study is warranted into the use of MADE on time series data

References

- KP Adragni, AM Raim, E. Al–Najjar, (2016) Minimum Average Deviance Estimation for Sufficient Dimension Reduction. (Unpublished)
- Full technical report: HPCF–2016–11 hpcf.umbc.edu > Publications

Acknowledgments

- REU Site: hpcreu.umbc.edu
- NSF, NSA, DOD, UMBC, HPCF, JCET