Proceedings of the 8th International Conference on Civil Structural and Transportation Engineering (ICCSTE'23)
Ottawa, Canada – June 04-06, 2023
Paper No. 110
DOI: 10.11159/iccste23.110

Bayesian Belief Network and Optimal Learning Analysis of Historical Flood Level Data for the Mississippi Watershed Under Data Paucity Conditions

Nicholas V. Scott[1] and Dietrick Lawrence[2]
[1]Riverside Research, Open Innovation Center, Dayton Research Center
2640 Hibiscus Way, Beavercreek, Ohio, 45431, USA
nscott@riversideresearch.org; dlawrence@riversideresearch.org
[2]Riverside Research, Modelling and Development Laboratory
1902 Fox Drive, Champaign, IL, 61820, USA

Extended Abstract

Floods along the Mississippi watershed cause large amounts of property damage and inflict great costs to human life [1]. With drastic changes in climate, characterization of flood events has taken on great importance to hydrologists, civil engineers, and water resource scientists tasked with understanding its dynamics, characterizing its spatial statistical structure, and indirectly safeguarding human life. Bayesian belief networks and optimal learning analysis are applied to sparse flood height data captured at specific local sites providing spatio-temporal statistical insight into flooding event structure occurring in the Mississippi watershed over large time scales. The objectives of this preliminary analysis are twofold addressing how to gain statistical process insight given data paucity. The first objective is to gain understanding of the spatial correlation of riverine site flooding data for the purpose of establishing Bayesian priors for future Bayesian hierarchical modelling of flood event structure. The second objective is to provide guidance to leadership as where to sample in space and time over the array in the future given a preconceived notion of mean and maximum spatial flood event structure.

Time series of flood heights from twenty sites distributed over the Mississippi watershed stretching from Minnesota to the Louisiana delta region at the Gulf of Mexico, obtained from the United States Geological Survey (USGS) database, consisted of 1967 temporal points for each riverine site spanning the years of 1785 to 2021. Approximately 10 % of the points in each time series were nonzero where sampling of flood height levels was irregularly sampled in time but cotemporally across space. This extreme case of data paucity suggested treatment of flood height as an extreme value variable. The Chow-Liu algorithm-based Bayesian belief network (BBN) for the flood height array spanning the complete data time span demonstrates a complex tree structure comprised of riverine site links with overall weak mutual information values [2,3]. Five riverine links possessed mutual information strengths greater than 0.01 representing predominantly local spatial connections in the middle portion of the array. State levels associated with marginal probability distributions for the BBN nodes experience extreme state modulations which are strongly locally bounded when evidence is provided at the five nodes. These characteristics are all due to data paucity. The BBN provides limited understanding of global propagation of evidence but does provide spatial Bayesian prior state understanding for future Bayesian hierarchical modelling as more data is accrued in the future.

With the strong need for more data to perform robust Bayesian modelling of the flood height level monitoring system, optimal learning simulations were performed on the twenty-site array. Optimal learning is used to gain insight into where data samples should be captured in space and time given a preconceived understanding of appropriate mean flood height levels [4]. Four simulation studies were performed where four different multivariate objective functions and prior belief combinations were used. The aim was to estimate the optimal site sample choices at a series of time steps for goal state function attainment. The knowledge gradient, the principle which guides the Bayesian myopic choice at each time step where correlated beliefs between the twenty-site multivariate array are modulated, is used in the optimization process [5].

Initial results demonstrate that convergence to the different multivariate objective functions from different prior states, measured by the root mean square value, occurs on the order of two hundred steps for all simulations. On the other hand, different levels of convergence exist with the largest root mean square values or errors existing for the simulations modeling the approach towards an extreme value objective function from an extreme value-based prior. All results are consistent with data paucity conditions suggesting that robust estimates of reasonable objective functions require abundant measurements at specific measurement sites designated by the knowledge gradient.

***Keywords*:** flood height, data paucity, Mississippi watershed, Bayesian belief network, Chow-Liu algorithm, marginal probability distribution, optimal learning, knowledge gradient

References

[1] J. Barry, *Rising Tide: The Great Flood of 1927 and How It Changed America*. New York, New York: Simon and Schuster, 1998.
[2] L. E. Sucar, *Probabilistic Graphical Models: Principles and Applications*. London, UK: Springer-Verlag, 2015.
[3] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* San Mateo, CA, UK: Morgan Kaufmann, 1988.
[4] W. B. Powell and I. O. Ryzhov, *Optimal Learning.* San Mateo, CA, UK: Morgan Kaufmann, 1988.
[5] P. Frazier, W. Powell, and S. Dayanik, "The Knowledge-Gradient Policy for Correlated Normal Beliefs", *INFORMS Journal on Computing*, 2009, vol. 21(4), pp. 599-613.