Approximating covariance structures for large-N applications: a simulations study

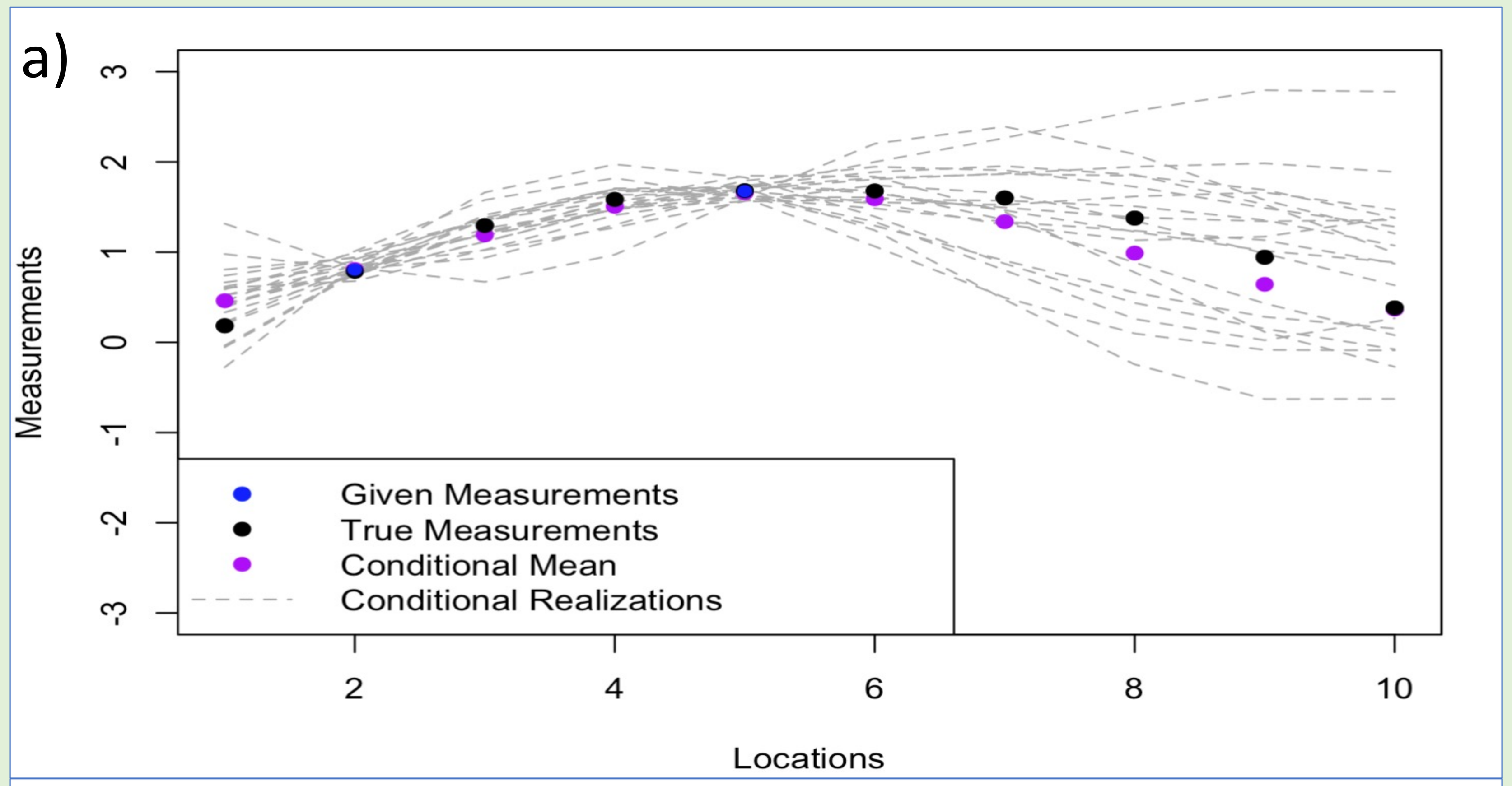
Nathaniel Fogg
BREE Research Program, Research Mentor: Maike Holthuijzen

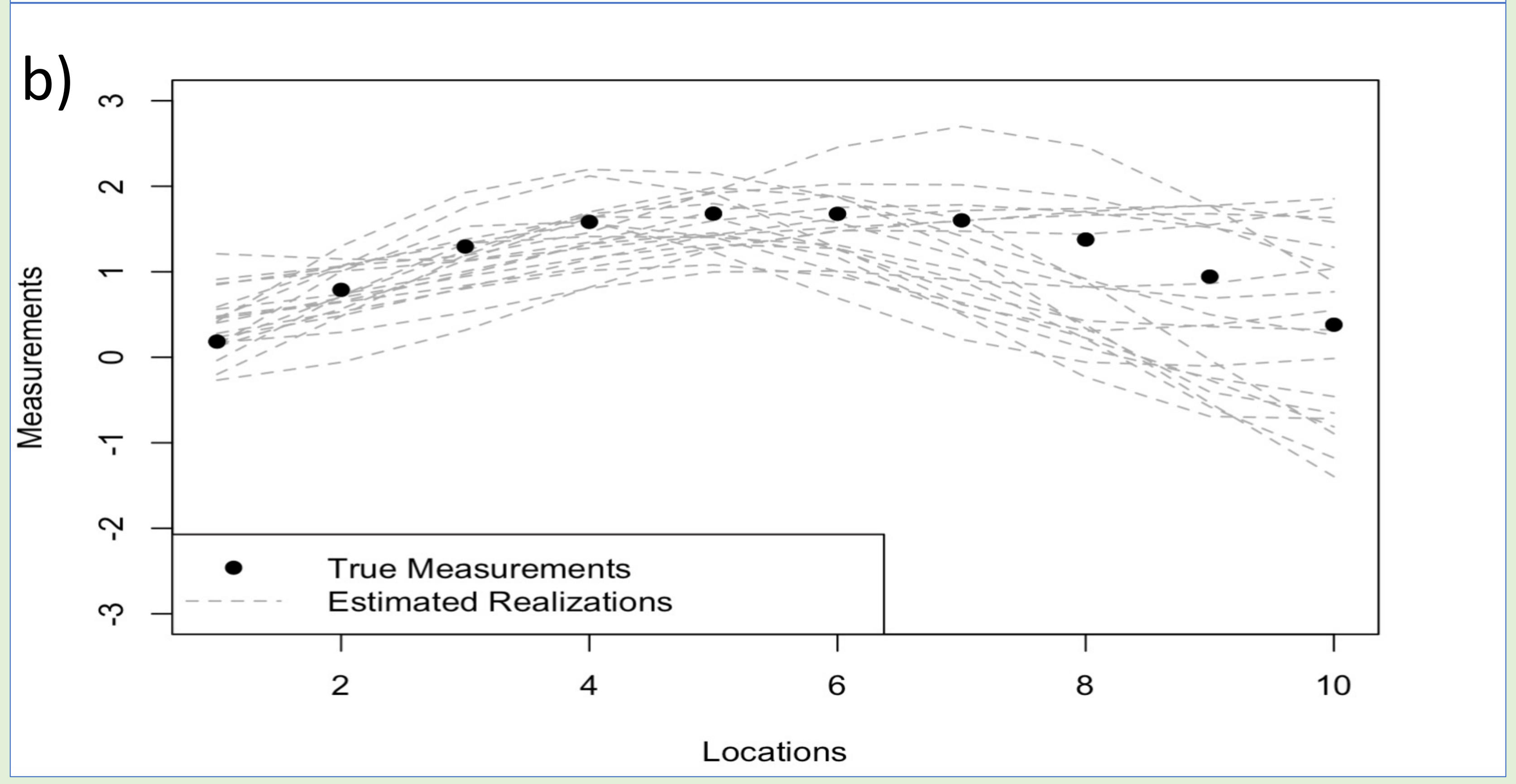
Funding provided by NSF Grant OIA-1556770

Abstract

Covariance structures are necessary for inference and prediction in spatiotemporal modeling applications. However, for large N problems, the direct calculation of covariance matrices is computationally infeasible. Here, we implement and provide theoretical justification for a method previously described by Douglas Nychka that approximates covariance structures in Gaussian process models. In two simulations, we generated realizations from a conditional Gaussian distribution using 1) the approximation technique and 2) direct calculation. We compared structures of the estimated empirical covariance and directly computed covariance matrices to determine the accuracy of the approximation method. We found the covariance structures from both methods were similar. Importantly, we showed that increasing the number of realizations from the approximation method by a factor of 4 resulted in a twofold increase in accuracy. The estimation method thus allows for accurate prediction and inference in problems where direct calculation of covariance structures would be computationally costly or not possible.

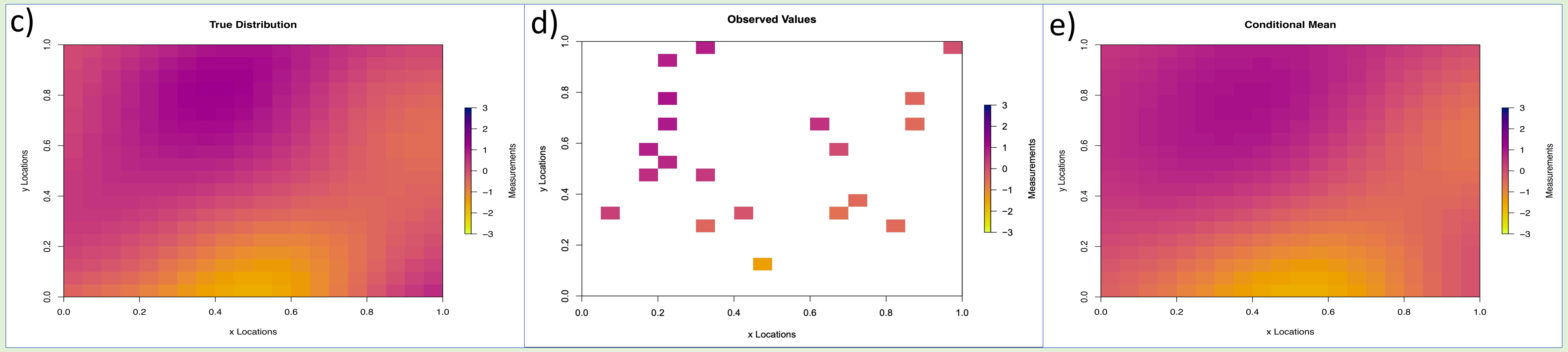
One-dimensional Simulations



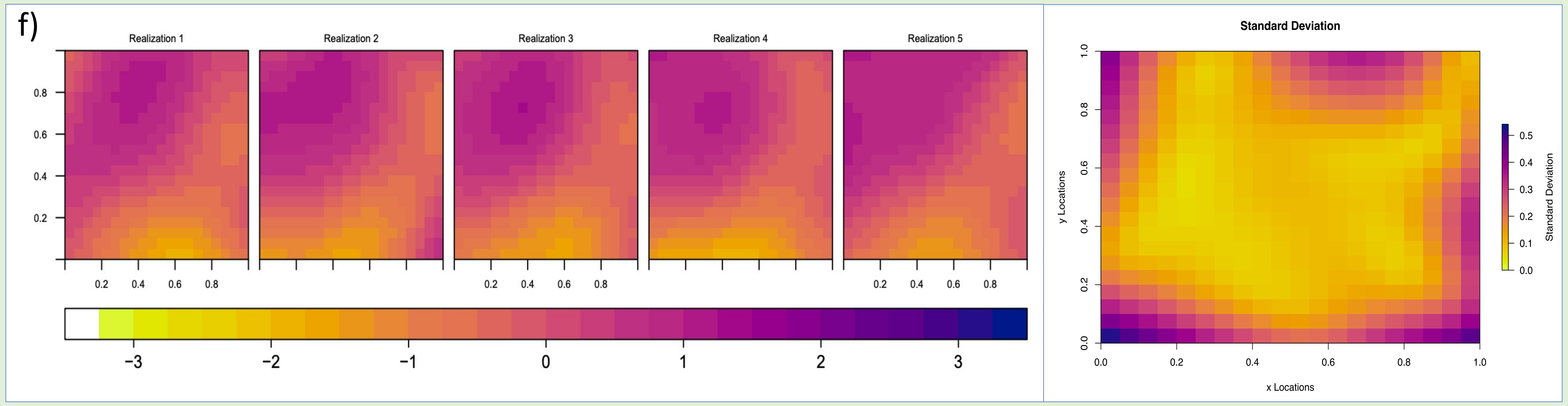


Observed, noisy, values at locations 2 and 5 (blue) and 'true' values (black) simulated from a Gaussian process. The conditional mean at locations 1-10 is constructed by conditioning on values at locations 2 and 5. Twenty conditional realizations are shown using (a) the conditional Gaussian formula and (b) the approximation method.

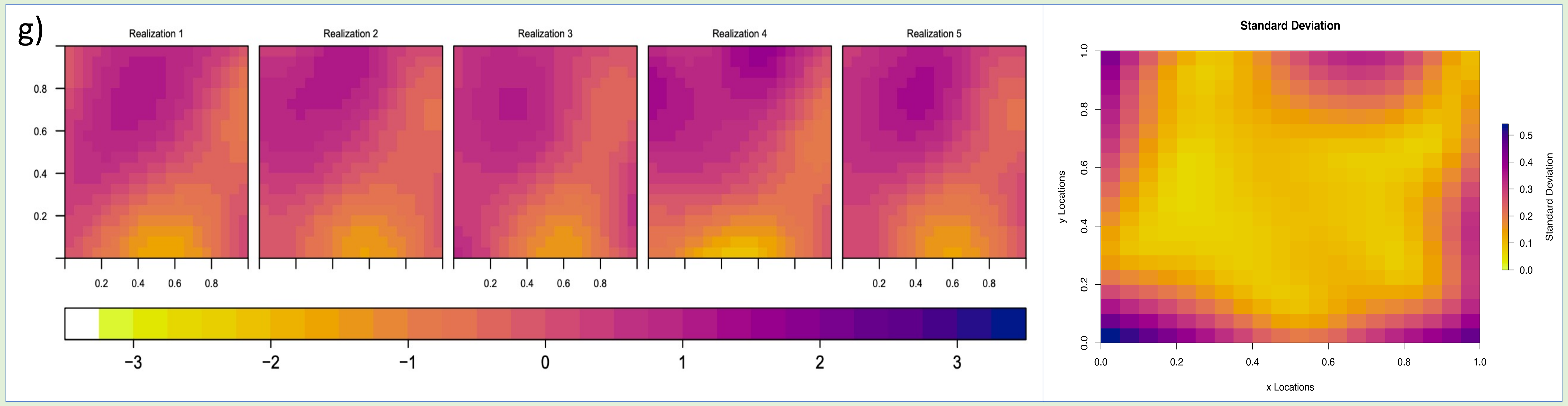
Two-dimensional Simulations



Realizations Using The Gaussian Conditional Formula:



Realizations Using The Estimation Method:



A Gaussian process is generated over a grid of 400 points (c), 20 of which are conditioned on. Values for the 20 conditional points are shown in (d). The conditional mean is calculated and plotted on the same grid (e). 5 realizations of the process calculated with the conditional Gaussian formula are shown in (f), and 5 realization of the process calculated with the approximation method are shown in (g). Standard deviation plots in (f) and (g) are calculated using 250 realizations.

The University of Vermont · NSF · Vermont EPSCoR · BREE Basin Resilience to Extreme Events in the Lake Champlain Basin · CWDD Center for Workforce Development & Diversity

Paper describing the estimation method: