Invited Discussion*

Bo Li† and Lyndsay Shand‡,§

1 Overview

We admire the authors for developing this computationally efficient Bayesian method to estimate the nonstationary correlation structure in large spatial data, without relying on a restrictive parametric model. The method is presented as a nonparametric extension of the Vecchia approach (Vecchia, 1988) and is based on the ordered conditional independence assumption that holds or approximately holds for many data sets arising from a Gaussian random process. The conditional independence leads to a sparse precision matrix and consequently a sparse Cholesky factorization. It has been shown that an n-variate Gaussian model can be expressed as a series of linear regression models with the nonzero elements in the Cholesky factor matrix as the regression coefficients (Huang et al., 2006). This enables us to estimate the Cholesky factorization and thus the precision matrix through Bayesian regression. The authors carefully studied the properties of unknown parameters and selected independent conjugate normal-inverse-gamma (NIG) priors that lead to closed-form posteriors and thus further improve computational efficiency.

There have been various approaches to nonstationary or nonparametric covariance modeling for spatial or spatiotemporal data. Kidd and Katzfuss (2022) (referred to KK22 thereafter) has provided an excellent review of previous literature. In addition to all methods reviewed for nonstationary covariance modeling in the Introduction, another semiparametric approach for nonstationary covariance modeling is through dimension expansion (Bornn et al., 2012; Shand and Li, 2017). Bornn et al. (2012) also requires replicates to estimate the spatial model.

KK22 thoroughly discussed many aspects of their method, including the theory, computation complexity, solutions for the presence of noise or trend, data ordering, conditioning-set size, and how to adapt priors to allow posterior converging to a covariance structure other than Matérn. We find several aspects very interesting and worth additional discussions. In the following, we adopt all notations from KK22.

*Li's research is partially supported by NSF-DMS-1830312 and -2124576. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government. SAND2022-0656 J.

†Department of Statistics, University of Illinois Urbana-Champaign, libo@illinois.edu
‡Department of Statistics, University of Illinois Urbana-Champaign
§Sandia National Laboratories, Albuquerque, New Mexico, lshand@sandia.gov

2 Further considerations

2.1 Requirement for replicates

The KK22 method was developed for spatial data with independent replicates and requires an ordering of spatial locations. KK22 used a maximin ordering that makes locations in $g_m(i)$ all have a roughly similar distance to the location $\mathbf{s}_i$ and this distance decreases systematically as i increases. Observing how the mean and variance of the d_i in $\mathbf{D}$ decrease exponentially with i given an underlying isotropic Matérn covariance structure (Sháfer et al., 2021a,b), the authors developed an inverse gamma prior for d_i with an appropriate form for α_i and β_i. Via the same scheme, they chose the normal prior for $\mathbf{u}_i$ with a tailored form for the correlation matrix. These carefully chosen conjugate priors ensure fast posterior sampling of unknown parameters, making the proposed method computationally efficient. The posteriors hold the nice property that the estimated covariance matrix contracts around the true covariance matrix as the number of replicates N increases. The numerical results showed that the Kullback-Leibler (KL) divergence of the KK22 method is lower than other methods in comparison, even for a very small N relative to the number of spatial locations n. In particular, the method works more efficiently than the maximum likelihood (ML) method for small N due to the inclusion of prior information, and then performs similarly to ML without a surprise when N increases. This conclusion holds for large n as well.

Given that this method demonstrates reasonable performance when N is as low as 3, do we really need replicates, i.e., $N > 1$, for the proposed method to be valid? This seems a rather constraining requirement. It is very common to observe spatial data without independent replicates. For example, suppose we are interested in an annual data of last year that can be either temperature over North America, or the county level Midwest crop yield or zip code level human immunodeficiency virus (HIV) new diagnoses in Philadelphia; all these spatial data likely have only one observation at each location. KK22 proposed their method based on Vecchia (1988), who developed a procedure with a spatial process not necessarily with replicates. We conjecture the requirement for replicates is mainly to attain the nice posterior-contraction property and reduce uncertainty in the parameter estimation. Is it possible to find a way to relax the replicates requirement (i.e., $N = 1$) but still approximately obtain the posterior-contraction? The authors investigated how the parameters decay with i, but is there any other spatial structure not in the order of i but on the distance between different i's that can be exploited to reduce the dependence on replicates? More specifically, can d_i and $\mathbf{u}_i$ be also modeled as spatially dependent processes in addition to their dependence on i? Of course, modeling additional spatially dependent processes can increase computation, so some special techniques such as those modeling dependence only on the nearest neighbors (Datta et al., 2016) may be considered. The maximin ordering makes the spatial dependence between i's unclear, but it may be worth a deliberation.

2.2 Choice of m

The choice of m is a trade-off between estimation accuracy and computation. KK22 suggested to set m as the largest j such that $\exp(-\theta_3 j) > 0.001$, where j denotes the

neighbor number. We wonder whether it is better to set m as adaptive for different i. With the maximin ordering, the distance between the conditioning-set to $\mathbf{s}_i$ decreases as i increases, but the size of the pool, $\mathbf{y}_{1:i-1}^{(l)}$, is always increasing. So it seems reasonable to have m as an increasing function of i to better approximate the full conditional distribution with the conditional distribution given a few neighbors.

2.3 Spatiotemporal data

Spatial data often have temporally correlated "replicates", i.e., spatiotemporal data. The temperatures of North America, the Midwest crop yield, and the disease data in our early examples can all become spatiotemporal if we now collect the annual data for the last 10 years. For spatiotemporal data, the temporal correlation and its interaction with spatial correlation need to be considered. This brings an additional challenge as the observations in the temporal dimension rapidly inflate the size of the covariance matrix $\mathbf{\Sigma}$ unless some simplified assumption such as space-time separability is assumed. If the KK22 method can be extended to this wealth of data, it would certainly expand its applicability. There is more impetus to relax the replication requirement in this case though, as independent replicates for spatiotemporal data are rarely available.

Similar to KK22, the nonstationary covariance modeling method in Bornn et al. (2012) also requires independent replicates of spatial data. Shand and Li (2017) extends Bornn et al.'s idea to model nonstationary covariance in both space and time for spatiotemporal data, and discusses the scalability of the method by taking random samples for latent dimension estimation. In Shand and Li (2017), observations in space and time are treated somewhat as independent replicates when estimating temporal and spatial correlation, respectively. Those intermediate results are then taken as inputs when dependency in all different forms is considered holistically. This strategy helps us estimate both space-time separable and nonseparable covariance structures while eliminating the dependence on replicates.

In the context of KK22, the ordering of space-time observations can be a challenge because both spatial and temporal distances are involved and the maximin ordering cannot be directly applicable. However, KK22 also mentioned other ordering strategies such as the ordering based on correlation distance which would more naturally extend to space-time data. Once the space-time observations are ordered, the KK22 method can readily apply to such data. Regarding the inflated size of $\mathbf{\Sigma}$, Section 2.2 in KK22 already discussed how to deal with a very large covariance matrix. If we can assume space-time separability, the precision matrix will be a Kronecker product of the precision matrices in space and time. In that case, we wonder if the ordering can be calculated for each dimension separately, and then $\mathbf{u}_i$ in space and time can be estimated separately as well. On the other hand, if we assume a simple temporal correlation structure such as an autoregressive model of order 1 (AR(1)), we wonder whether the estimates of $\mathbf{u}_i$ in the spatial dimension can approximately attain the posterior-contraction property, because the observations in time may act as dependent replicates for spatial correlation estimation. All these discussions for spatiotemporal data can be generalized to multivariate spatial data modeling.

3 Miscellaneous discussion

The smoothness of random fields is difficult to capture for nonparametric modeling. Im et al. (2007) proposes a semiparametric method that models the spectral density as a linear combination of B-splines up to a certain frequency threshold ω_0 and then an algebraic power function with a smoothness parameter similar to Matérn model for high frequencies beyond ω_0. However, many other nonparametric models, including Choi et al. (2013) that constructs the spatial or space-time covariance function using completely monotone functions do not directly consider the smoothness of covariance models. It is inspiring that KK22 has a smoothness parameter in their priors. We are curious how θ_3 explicitly relates to the smoothness of random fields.

There are different measures to evaluate covariance structure estimation. For example, mean squared prediction error is common for comparing different covariance estimates as prediction is a typical task for spatial data analysis. We think it would be informative if the authors could briefly comment on whether KL divergence relates to the prediction performance measure in general.

Spatial random effects and multi-resolution models (e.g. Nychka et al., 2014) are very popular for capturing either stationary or nonstationary spatial structures of massive datasets. KK22 also included the resolution adaptive fixed rank kriging approach by Tzeng and Huang (2018) (autoFRK), as one of the competitive methods. The author Katzfuss published a very high-impact paper on multi-resolution approximation (Katzfuss, 2017). We would like to learn how the authors view the connection, the discrepancies and the comparison between the KK22 and multi-resolution models.

4 Summary

We congratulate the authors on developing a very useful method for large spatial data. This method would find many applications for modeling the complex dependency structure of global climate data, e.g., the teleconnection of climate variables (Choi et al., 2015). Studies for spatial extremes often approximate the block maxima as independent replicates (e.g. Cao and Li, 2018), so the KK22 method can be naturally used to model dependence in spatial extremes by combining with the copula technique that takes care of the marginal extreme value distribution, among many other exciting applications.

References

Bornn, L., Shaddick, G., and Zidek, J. V. (2012). "Modeling nonstationary processes through dimension expansion." *Journal of the American Statistical Association*, 107(497): 281–289. MR2949359. doi: https://doi.org/10.1080/01621459.2011.646919. 313, 315

Cao, Y. and Li, B. (2018). "Assessing models for spatial extremes and methods for uncertainty quantification on return level estimation." *Environmetrics*. MR3919899. doi: https://doi.org/10.1002/env.2508. 316

Choi, I., Li, B., and Wang, X. (2013). "Nonparametric estimation of spatial and space-time covariance function." *Journal of Agricultural, Biological, and Environmental Statistics*, 18(4): 611–630. MR3142603. doi: https://doi.org/10.1007/s13253-013-0152-z. 316

Choi, I., Li, B., Zhang, H., and Li, Y. (2015). "Modelling space–time varying ENSO teleconnections to droughts in North America." *Stat*, 4(1): 140–156. MR3405397. doi: https://doi.org/10.1002/sta4.85. 316

Datta, A., Banerjee, S., Finley, A. O., and Gelfand, A. E. (2016). "Hierarchical Nearest-Neighbor Gaussian Process Models for Large Geostatistical Datasets." *Journal of the American Statistical Association*, 111(514): 800–812. PMID: 29720777. MR3538706. doi: https://doi.org/10.1080/01621459.2015.1044091. 314

Huang, J. Z., Liu, N., Pourahmadi, M., and Liu, L. (2006). "Covariance matrix selection and estimation via penalised normal likelihood." *Biometrika*, 93(1): 85–98. MR2277742. doi: https://doi.org/10.1093/biomet/93.1.85. 313

Im, H. K., Stein, M. L., and Zhu, Z. (2007). "Semiparametric estimation of spectral density with irregular observations." *Journal of the American Statistical Association*, 102(478): 726–735. MR2381049. doi: https://doi.org/10.1198/016214507000000220. 316

Katzfuss, M. (2017). "A multi-resolution approximation for massive spatial datasets." *Journal of the American Statistical Association*, 112(517): 201–214. MR3646566. doi: https://doi.org/10.1080/01621459.2015.1123632. 316

Kidd, B. and Katzfuss, M. (2022). "Bayesian nonstationary and nonparametric covariance estimation for large spatial data." *Bayesian Analysis*, 1(1): 1–22. 313

Nychka, D., Bandyopadhyay, S., Hammerling, D., Lindgren, F., and Sain, S. (2014). "A Multiresolution Gaussian Process Model for the Analysis of Large Spatial Datasets." *Journal of Computational and Graphical Statistics*, 24. MR3357396. doi: https://doi.org/10.1080/10618600.2014.914946. 316

Sháfer, F., Katzfuss, M., and Owhadi, H. (2021a). "Sparse Cholesky Factorization by Kullback-Leibler Minimization." *SIAM Journal on Scientific Computing*, 43(3): A2019–A2046. MR4267493. doi: https://doi.org/10.1137/20M1336254. 314

Sháfer, F., Sullivan, T. J., and Owhadi, H. (2021b). "Compression, inversion, and approximate PCA of dense kernel matrices at near-linear computational complexity." *Multiscale modeling & Simulation*, 19(2): 688–730. MR4243658. doi: https://doi.org/10.1137/19M129526X. 314

Shand, L. and Li, B. (2017). "Modeling nonstationarity in space and time." *Biometrics*, 73(3): 759–768. MR3713110. doi: https://doi.org/10.1111/biom.12656. 313, 315

Tzeng, S. and Huang, H. (2018). "Resolution Adaptive Fixed Rank Kriging." *Technometrics*, 60: 1–11. doi: https://doi.org/10.1080/00401706.2017.1345701. 316

Vecchia, A. V. (1988). "Estimation and model identification for continuous spatial processes." *Journal of the Royal Statistical Society: Series B (Methodological)*, 50(2): 297–312. MR0964183. 313, 314