Fuglstad, Geir-Arne; **Castruccio, Stefano**

Compression of climate simulations with a nonstationary global spatiotemporal SPDE model. (English) Zbl 1446.62139

Ann. Appl. Stat. 14, No. 2, 542-559 (2020).

Summary: Modern climate models pose an ever-increasing storage burden to computational facilities, and the upcoming generation of global simulations from the next Intergovernmental Panel on Climate Change will require a substantial share of the budget of research centers worldwide to be allocated just for this task. A statistical model can be used as a means to mitigate the storage burden by providing a stochastic approximation of the climate simulations. Indeed, if a suitably validated statistical model can be formulated to draw realizations whose spatiotemporal structure is similar to that of the original computer simulations, then the estimated parameters are effectively all the information that needs to be stored. In this work we propose a new statistical model defined via a stochastic partial differential equation (SPDE) on the sphere and in evolving time. The model is able to capture nonstationarities across latitudes, longitudes and land/ocean domains for more than 300 million data points while also overcoming the fundamental limitations of current global statistical models available for compression. Once the model is trained, surrogate runs can be instantaneously generated on a laptop by storing just 20 Megabytes of parameters as opposed to more than six Gigabytes of the original ensemble.

MSC:

62H11 Directional data; spatial statistics
60H15 Stochastic partial differential equations (aspects of stochastic analysis)
62-08 Computational methods for problems pertaining to statistics
62P12 Applications of statistics to environmental and related topics
86A08 Climate science and climate modeling
62R07 Statistical aspects of big data and data science

Keywords:

stochastic partial differential equation (SPDE); space-time model; global model; nonstationary; climate model

Software:

GMRFLib

Full Text: DOI Euclid

References:

[1] Baker, A. H., Xu, H., Dennis, J. M., Levy, M. N., Nychka, D., Mickelson, S. A., Edwards, J., Vertenstein, M. and Wegener, A. (2014). A methodology for evaluating the impact of data compression on climate simulation data. In Proceedings of the 23rd International Symposium on High-Performance Parallel and Distributed Computing, ACM HPDC '14 203-214.

[2] Baker, A. H., Hammerling, D. M., Mickelson, S. A., Xu, H., Stolpe, M. B., Naveau, P., Sanderson, B., Ebert-Uphoff, I., Samarasinghe, S. et al. (2016). Evaluating lossy data compression on climate simulation data within a large ensemble. Geosci. Model Dev. 9 4381-4403.

[3] Baker, A. H., Xu, H., Hammerling, D. M., Li, S.-M. and Clyne, J. P. (2017). Toward a multi-method approach: Lossy data compression for climate simulation data. In High Performance Computing: ISC High Performance 2017 (J. Kunkel, R. Yokota, M. Taufer and J. Shalf, eds.) 30-42. Springer, Berlin.

[4] Bakka, H., Vanhatalo, J., Illian, J. B., Simpson, D. and Rue, H. (2019). Non-stationary Gaussian models with physical barriers. Spat. Stat. 29 268-288.

[5] Bolin, D. and Lindgren, F. (2011). Spatial models generated by nested stochastic partial differential equations, with an application to global ozone mapping. Ann. Appl. Stat. 5 523-550. · Zbl 1235.60075 · doi:10.1214/10-AOAS383

[6] Castruccio, S. and Genton, M. G. (2016). Compressing an ensemble with statistical models: An algorithm for global 3D spatio-temporal temperature. Technometrics 58 319-328.

[7] Castruccio, S. and Guinness, J. (2017). An evolutionary spectrum approach to incorporate large-scale geographical descriptors on global processes. J. R. Stat. Soc. Ser. C. Appl. Stat. 66 329-344.

[8] Castruccio, S. and Stein, M. L. (2013). Global space-time models for climate ensembles. Ann. Appl. Stat. 7 1593-1611. · Zbl

1454.62436 · doi:10.1214/13-AOAS656

[9] Castruccio, S., Hu, Z., Sanderson, B., Karspeck, A. and Hammerling, D. (2019). Reproducing internal variability with few ensemble runs. J. Climate 32 8511-8522.

[10] Edwards, M., Hammerling, D. and Castruccio, S. (2020). Marginally parametrized spatio-temporal models and stepwise maximum likelihood estimation. Available at arXiv:1806.11388.

[11] Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J. and Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. Geosci. Model Dev. 9 1937-1958.

[12] Fuentes, M. (2007). Approximate likelihood for large irregularly spaced spatial data. J. Amer. Statist. Assoc. 102 321-331. · Zbl 1284.62589 · doi:10.1198/016214506000000852

[13] Fuglstad, G.-A. and Castruccio, S. (2020). Supplement to "Compression of climate simulations with a nonstationary global SpatioTemporal SPDE model." https://doi.org/10.1214/20-AOAS1340SUPPA, https://doi.org/10.1214/20-AOAS1340SUPPB

[14] Fuglstad, G.-A., Lindgren, F., Simpson, D. and Rue, H. (2015a). Exploring a new class of non-stationary spatial Gaussian random fields with varying local anisotropy. Statist. Sinica 25 115-133. · Zbl 06497337

[15] Fuglstad, G.-A., Simpson, D., Lindgren, F. and Rue, H. (2015b). Does non-stationary spatial data always require non-stationary random fields? Spat. Stat. 14 505-531.

[16] Fuglstad, G.-A., Simpson, D., Lindgren, F. and Rue, H. (2019). Constructing priors that penalize the complexity of Gaussian random fields. J. Amer. Statist. Assoc. 114 445-452. · Zbl 07095889 · doi:10.1080/01621459.2017.1415907

[17] Guinness, J. and Hammerling, D. (2018). Compression and conditional emulation of climate model output. J. Amer. Statist. Assoc. 113 56-67. · Zbl 1398.68157 · doi:10.1080/01621459.2017.1395339

[18] Ingebrigtsen, R., Lindgren, F. and Steinsland, I. (2014). Spatial models with explanatory variables in the dependence structure. Spat. Stat. 8 20-38.

[19] Ingebrigtsen, R., Lindgren, F., Steinsland, I. and Martino, S. (2015). Estimation of a non-stationary model for annual precipitation in southern Norway using replicates of the spatial field. Spat. Stat. 14 338-364.

[20] Jeong, J., Castruccio, S., Crippa, P. and Genton, M. G. (2018). Reducing storage of global wind ensembles with stochastic generators. Ann. Appl. Stat. 12 490-509. · Zbl 1393.62123 · doi:10.1214/17-AOAS1105

[21] Jeong, J., Yan, Y., Castruccio, S. and Genton, M. G. (2019). A stochastic generator of global monthly wind energy with Tukey \(g\)-and-\(h\) autoregressive processes. Statist. Sinica 29 1105-1126. · Zbl 1421.62158

[22] Jun, M. and Stein, M. L. (2007). An approach to producing space-time covariance functions on spheres. Technometrics 49 468-479.

[23] Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., Arblaster, J. M., Bates, S. C., Danabasoglu, G. et al. (2015). The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. Bull. Am. Meteorol. Soc. 96 1333-1349.

[24] Lindgren, F., Rue, H. and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. J. R. Stat. Soc. Ser. B. Stat. Methodol. 73 423-498. · Zbl 1274.62360 · doi:10.1111/j.1467-9868.2011.00777.x

[25] Meehl, G. A., Moss, R., Taylor, K. E., Eyring, V., Stouffer, R. J., Bony, S. and Stevens, B. (2014). Climate model intercomparisons: Preparing for the next phase. Eos Trans. AGU 95 77-78.

[26] Nikiema, O. and Laprise, R. (2011). Budget study of the internal variability in ensemble simulations of the Canadian Regional Climate Model at the seasonal scale. J. Geophys. Res., Atmos. 116.

[27] Rozanov, Ju. A. (1977). Markovian random fields, and stochastic partial differential equations. Math. USSR, Sb. 32 515-534. · Zbl 0396.60057 · doi:10.1070/SM1977v032n04ABEH002404

[28] Rue, H. and Held, L. (2005). Gaussian Markov Random Fields: Theory and Applications. Monographs on Statistics and Applied Probability 104. CRC Press/CRC, Boca Raton, FL. · Zbl 1093.60003

[29] Sampson, P. D. and Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. J. Amer. Statist. Assoc. 87 108-119.

[30] Small, R. J., Bacmeister, J., Bailey, D., Baker, A., Bishop, S., Bryan, F., Caron, J., Dennis, J., Gent, P. et al. (2014). A new synoptic scale resolving global climate simulation using the Community Earth System Model. J. Adv. Model. Earth Syst. 6 1065-1094.

[31] Stein, M. L. (2005). Statistical methods for regular monitoring data. J. R. Stat. Soc. Ser. B. Stat. Methodol. 67 667-687. · Zbl 1101.62115 · doi:10.1111/j.1467-9868.2005.00520.x

[32] Sun, Y. and Genton, M. G. (2011). Functional boxplots. J. Comput. Graph. Statist. 20 316-334.

[33] Taylor, K., Stouffer, R. and Meehl, G. (2012). An overview of CMIP5 and the experiment design. Bull. Am. Meteorol. Soc. 93 485-498.

[34] Torn, R. D. (2016). Evaluation of atmosphere and ocean initial condition uncertainty and stochastic exchange coefficients on ensemble tropical cyclone intensity forecasts. Mon. Weather Rev. 144 3487-3506.

[35] van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V. et al. (2011). The representative concentration pathways: An overview. Clim. Change 109 5-31.

[36] Whittle, P. (1954). On stationary processes in the plane. Biometrika 41 434-449. · Zbl 0058.35601 · doi:10.1093/biomet/41.3-4.434

[37] Yoshimori, M.