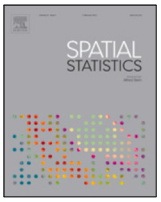
Spatial Statistics 50 (2022) 100596

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta



A selective view of climatological data and likelihood estimation

Federico Blasi [a], Christian Caamaño-Carrillo [b], Moreno Bevilacqua [c], Reinhard Furrer [d,*]

[a] Department of Mathematics, University of Zurich, Zurich, Switzerland
[b] Department of Statistics, Universidad del Bío-Bío, Concepción, Chile
[c] Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez, Viña del Mar, Chile
[d] Department of Mathematics and Department of Computational Science, University of Zurich, Zurich, Switzerland

ARTICLE INFO

Article history:
Received 1 November 2021
Received in revised form 26 December 2021
Accepted 7 January 2022
Available online 25 January 2022

Keywords:
Tapering
Composite likelihood
Sinh-arcsinh distribution
CMIP6 data
Random field
Spatial process

ABSTRACT

This article gives a narrative overview of what constitutes climatological data and their typical features, with a focus on aspects relevant to statistical modeling. We restrict the discussion to univariate spatial fields and focus on maximum likelihood estimation. To address the problem of enormous datasets, we study three common approximation schemes: tapering, direct misspecification, and composite likelihood for Gaussian and non-Gaussian distributions. We focus particularly on the so-called 'sinh-arcsinh distribution', obtained through a specific transformation of the Gaussian distribution. Because it has flexible marginal distributions – possibly skewed and/or heavy-tailed – it has a wide range of applications. One appealing property of the transformation involved is the existence of an explicit inverse transformation that makes likelihood-based methods straightforward. We describe a simulation study illustrating the effects of the different approximation schemes. To the best of our knowledge, a direct comparison of tapering, direct misspecification, and composite likelihood has never been made previously, and we show that direct misspecification is inferior. In some metrics, composite likelihood has a minor advantage over tapering. We use the estimation approaches to model a high-resolution global climate change field. All simulation code is available as a Docker container and is thus fully reproducible. Additionally, the present article describes where and how to get various climate datasets.

* Corresponding author.
E-mail addresses: federico.blasi@math.uzh.ch (F. Blasi), chcaaman@ubiobio.cl (C. Caamaño-Carrillo), moreno.bevilacqua@uai.cl (M. Bevilacqua), reinhard.furrer@math.uzh.ch (R. Furrer).

https://doi.org/10.1016/j.spasta.2022.100596

1. Climate data and their statistical analysis

"Weather is what you get, climate is what you expect". That is a colloquial explanation of the difference between climatological data and weather data. No mere unusual summer cold spell could be used to undercut the latest Summary for Policymakers (SPM) (Masson-Delmotte et al., 2021a) from the Sixth Assessment Report (AR6) of the Intergovernmental Panel on Climate Change (IPCC) (Masson-Delmotte et al., 2021b): "It is unequivocal that human influence has warmed the atmosphere, ocean and land. Widespread and rapid changes in the atmosphere, ocean, cryosphere and biosphere have occurred".

Broadly speaking, there are two categories of observed climate data: aggregated meteorological data (e.g., temperature and precipitation data) and climate proxy data, which is typically based on phenology (e.g., start of the cherry blossom, tree rings, ice cores, and sediments). However, the relationship between observed phenological variables and the climate is not well understood, even when time series are dense and contemporary (Güsewell et al., 2017, 2018). Satellite remote sensing systems can provide data from both categories (e.g., sea surface temperature or land cover) and at different levels of spatial and temporal aggregation (NASA, 2021). Many satellite data products are openly available from different repositories (see Appendix A.1).

A thorough study of climate, temperature, and precipitation is not enough, however, and many more variables, such as air pressure, winds, surface radiation, or humidity, are required (see https://gcos.wmo.int/en/essential-climate-variables/ for a complete list of the Essential Climate Variables, ECV).

Meteorological data is typically aggregated into 20- or 30-year averages to obtain climatological data. Changes are then assessed using moving averages or by comparing two time periods. The typically used reference periods for comparing future climate projections to previous observations are 1951–1980, 1850–1900, and (for variables with recent data only) 1981–2010 (Rhode, personal communication). Similarly, climate scientists often work with seasonal averages, such as the December to February (DJF) or June to August (JJA) averages that reflect climatological changes better than the spring and fall seasons.

Data from irregular monitoring networks can be processed to (global) grids or reduced to single value indices. Observed climate data is often processed to fine grids using statistical downscaling (Wilby et al., 2004) or so-called reanalysis approaches (Kalnay et al., 1996). Downscaling takes coarse-resolution information and constructs corresponding higher-resolution maps using statistical approaches (Fowler et al., 2007; Poggio and Gimona, 2015). A reanalysis consists of blending sparse, past weather observations with a numerical model to derive best guess fields. These fields can be processed further to derive summary statistics from these weather variables, such as extreme values (e.g., minimum winter temperature, maximum 24-hour precipitation). Somewhat counter intuitively, subsets of various gridded variables are aggregated to scalar indices, which are then often used to describe the state of the climate, for example, a heat wave index (Furrer et al., 2010), the Southern Oscillation Index (Nino3.4) (Trenberth, 1997), the North Atlantic Oscillation (NAO) (Hurrell et al., 2013), and many more.

Atmosphere–Ocean General Circulation Models (AOGCMs) or General Circulation Models (GCMs), for short, simulate the Earth's climate by representing its atmosphere–ocean–ice system numerically based on their simplified physical, chemical, and biological properties and interactions between the atmosphere, the ocean, and ice. An Earth System Model (ESM) expands on a GCM and resolves the carbon cycle and possibly other high resolution processes (see Glossary of Masson-Delmotte et al., 2021b). GCMs and ESMs can be run under different scenarios – so-called pathways – incorporating prescribed forcings (e.g., volcanic eruptions, changes in solar radiation, and, most relevantly, changes in anthropogenic emissions due to fossil fuel combustion and land-use conversion). Hence, pathways provide climate projections for different climate variables but also contribute to the knowledge of how the Earth's system responds to the different forcings.

Running a GCM numerically is very expensive and even high-resolution, state-of-the-art models cannot resolve local topographic features below ten-kilometer granularity. Obtaining high-resolution projections requires Regional Climate Models (RCMs), which function similarly to GCMs but trade off a much higher spatial resolution by only modeling a fraction of the Earth's surface.

Projections from a GCM provide overall boundary conditions, and an RCMs can be interpreted as a computationally expensive downscaling approach.

The first realistic climate model dates back to the late 1960s and, in the following decades, more and more models were developed and used for climate projections. The first AOGCMs incorporating realistic geography were run for multi-decadal simulations in the late 1980s. These early models differed in their precise implementation as well as in their projections. As a result, the World Climate Research Programme (WCRP) started a Coupled Model Intercomparison Project (CMIP) in the mid-1990s (Meehl et al., 1997). These and subsequent simulations made significant contributions to the IPCC Third Assessment Report (Houghton et al., 2001). CMIP Phase 3 (CMIP3) was the start of the modern era of open access to multi-model data via the internet (Meehl et al., 2007). The international effort continued and, at the time of writing this article, simulation output was still being added to the CMIP6 project and planning for CMIP7 was well under way. In the past decade, incredible amounts of simulation data have been archived and made publicly available. The CMIP3 repository comprises roughly 31 TB of data; CMIP5 requires 2 PB and CMIP6 will require over 5 PB (Meehl, 2019).

The availability of so much data needs to be complemented by appropriate, capable software – ideally open source – for analysis and modeling. AOGCM data are typically stored in Network Common Data Form (NetCDF). Satellite or remote sensing data that have been at least marginally processed are often stored in Hierarchical Data Format (HDF), GeoTIFF, or NetCDF. HDF is the de facto standard for products of NASA's Earth Observing System (which includes the Aqua, Terra, and Landsat 8 satellites) and GeoTIFF is a public domain metadata standard which allows georeferencing information to be embedded within a TIFF file. NetCDF has a large community of users and is broadly supported. It is often useful begin by browsing NetCDF files using, for example, `ncview` (http://meteora.ucsd.edu/~pierce/ncview_home_page.html), to get a quick overview of the available variables and their ranges. The free software environment, `R` (R. Core Team, 2021), supports these formats through the `rhdf5`, `hdf5r`, `raster`, `terra`, `ncdf4` and `RNetCDF` packages. See https://CRAN.R-project.org/view=Spatial for links and further packages. There are plenty of accessible resources about how to handle climatological spatial data in `R`. See, for example, Bivand et al. (2013), https://rspatial.org and its references, and also Hengl et al. (2015).

In summary, a plethora of climate and other environmental variables are available on regular grids or with global coverage. These options provide a perfect playground for spatial and spatio-temporal statistics. Typical tasks involving spatial climatological data include fitting parametric models, aggregating different data sources, downscaling interpolations, and, to a lesser extent, predicting the future. The statistical analysis of climatological data is interesting because of its societal relevance and challenging because of inherent modeling difficulties. However, it seems that statisticians have a tendency to use climate data to probe statistical models rather than provide profound climatological results—these have most often been published outside statistics journals, as evidenced by the reference lists in the last three Assessment Reports of the IPCC.

Working Group I's contribution to the IPCC's Sixth Assessment Report (Masson-Delmotte et al., 2021b) cites not a single article from *Spatial Statistics* (nor from any other well-known statistics journals, see Appendix A.2). Although the size of the assessment reports has increased over time (the current one has close to 4,000 pages), the number of statistical articles cited in them seems to have decreased. This decrease is probably because statistics journals almost exclusively provide new methodology without substantially contributing to other fields of study, even if that methodology was developed to meet the needs of a research question from another field. Scientists, on the other hand, are more likely to learn about a new statistical methodology if it is used as part of a substantial contribution to their specialty. Consequently, the original methodological article remains little cited. Visibility in high quality, applied science journals, such as *Spatial Statistics*, remains low, unfortunately, despite many excellent, relevant contributions.

References to climate or climatological data are made in about a quarter of the articles published in *Spatial Statistics* (123 out of 528, to end of 2021), but only about two dozen use the term `climat*` in the title or abstract. The only two articles to write explicitly about circulation models were those by Poggio and Gimona (2015) and Castruccio (2016). These numbers mask the substantial methodological advances revealed in *Spatial Statistics* that are proving beneficial for climate science.

These advances include classic interpolation methods (Franco-Villoria and Ignaccolo, 2017), machine learning based prediction (Li, 2021; Pathakoti et al., 2021), modeling or estimating with large or huge datasets (Kleiber and Nychka, 2015; Barbian and Assunção, 2017), models for bivariate fields (Salvaña and Genton, 2020; Bevilacqua et al., 2020b), covariance models (Cappello et al., 2021; Alegría et al., 2021), covariance approximations (Hong et al., 2021), multiresolution approximations (Nychka et al., 2018; Appel and Pebesma, 2020), Bayesian hierarchical models (Paciorek et al., 2015; Cameletti et al., 2019; Banerjee, 2020) to name just a few.

After this brief overview of climatological data, we now summarize the maximum likelihood approach for large spatial random fields. The summary is quite generic, and although it is presented for a single random field, most of the statements hold for space–time or multivariate data (pronounced differences occur when deriving asymptotic results). We refrain from summarizing other estimation and approximation methods and instead refer readers to Heaton et al. (2019), Hong et al. (2021). In Section 2, we address efficient estimation approaches focusing on the approximation methods of tapering, direct misspecification, and composite likelihood, all illustrated within the Gaussian process framework. There are several situations in which knowledge about the parametric distribution of observed or simulated climate fields is crucial. One example, is fingerprinting, a technique where patterns in models and observed changes are compared, typically based on a regression type approach (Allen and Tett, 1999). The patterns that best explain the observed data provide the strongest support for the causes of the change. Further examples relying on parametric models of the dependency structure in climate data are statistical downscaling or the aggregation of climate data down to single values. The hierarchical model structure developed by Furrer et al. (2007a) was strongly driven by extensive exploratory data analyses involving parametric model fitting.

Of course, in reality, the joint distribution of climate fields is hardly Gaussian, and there are several ways to address this. A selection of approaches are: working with differences, i.e., climate changes (Furrer and Sain, 2009); the transformation of responses (like log-transformations Damian et al., 2003; square-root transformations Jeong and Jun, 2015; fourth-root transformations Furrer and Sain, 2010), highly flexible mean parametrization (Furrer et al., 2007a), aggregation of several variables (Flury et al., 2021), multi-layer hierarchical Bayesian approaches (Sain et al., 2011), the use of non-Gaussian processes (Xu and Genton, 2016, 2017; Bevilacqua et al., 2021, 2020a), as well as further approaches and references in Schmidt and Guttorp (2020).

Section 3 extends the Gaussian framework to a particular transformation setting, thus yielding a so-called *sinh-arcsinh* process. This process incorporates flexible marginal distributions involving two additional parameters that model heavier or lighter tails than those induced by Gaussian processes and/or possible asymmetries. One advantage of the sinh-arcsinh process compared to other recent proposals (e.g., Xu and Genton, 2017) is that the transformation involved is explicitly invertible, which allows likelihood-based methods of estimation to be applied directly.

In Section 5 we use temperature projections from one particular climate model and emissions scenario and estimate the parameters of the distributions modeled using the approximation schemes mentioned above. Considerable emphasis is put on explaining how to obtain the data. The simulation and estimation were implemented in R software, and we provide the analysis via a Docker image (Boettiger, 2015, docker.com). Hence, the results of this article are fully reproducible.

We conclude our investigation in Section 6, and provide further pertinent information in several sections of Appendix A.

2. Maximum likelihood-type estimation approaches

One well-known assumption made when modeling climate data using spatial processes is to assume that the joint distribution of the spatial random vector studied is Gaussian. This unlocks some extensively developed statistical theory involving desirable estimator properties and efficient computational implementation.

Let $\{Z(\mathbf{s}), \mathbf{s} \in \mathcal{D} \subset \mathbb{R}^d, d \geq 1\}$ be a Gaussian random field, with $\mathbf{s}$ denoting the spatial location, and, typically, $d = 2$. We assume that the random field can be decomposed as

$$Z(\mathbf{s}) = \mu(\mathbf{s}) + \sigma Z^s(\mathbf{s}),$$

where the mean function $\mu(\mathbf{s})$ can be expressed as $\mu(\mathbf{s}) = \mathbf{x}(\mathbf{s})^\top \boldsymbol{\beta}$ with $\mathbf{x}(\mathbf{s})$ a vector of length q of the observable covariates at location $\mathbf{s}$; $\boldsymbol{\beta}$ a q-vector of unknown regression coefficients; $\sigma > 0$ is the scale parameter, and $Z^s(\cdot)$ is a standard Gaussian process. For simplicity, we will assume a parametric, isotropic setup, which means that the correlation $\mathrm{Corr}\big(Z^s(\mathbf{s}_i), Z^s(\mathbf{s}_j)\big)$ can be expressed by a function $\rho(h_{ij}; \boldsymbol{\vartheta})$ with $h_{ij} = \|\boldsymbol{s}_i - \boldsymbol{s}_j\|$ and $\rho(0; \boldsymbol{\vartheta}) = 1$, parametrized by a vector $\boldsymbol{\vartheta}$ of length m.

A wide variety of parametric correlation functions are available, see, e.g., Wackernagel (2006). The parametric family of Matérn functions has received a lot of attention over the past decade (Matérn, 1986; Stein, 1999). We include a white noise component and parametrize the correlation as

$$\rho(h; \boldsymbol{\vartheta}) = \frac{(1-\lambda)}{\Gamma(\nu)2^{\nu-1}}\left(\frac{h}{\gamma}\right)^{\nu} K_\nu\left(\frac{h}{\gamma}\right) + \lambda I_{\{h=0\}}, \tag{1}$$

where $\gamma > 0, \nu > 0, 0 \leq \lambda \leq 1$, $K_\nu(\cdot)$ is the modified Bessel function of the second kind of order ν (Abramowitz and Stegun, 1970), $I_{\{h=0\}}$ is the indicator function of event $h = 0$, $\Gamma(\cdot)$ is the Gamma function, and $\boldsymbol{\vartheta} = (\gamma, \nu, \lambda)^\top$, so $m = 3$. Another frequently used correlation function corresponds to the parametric family of Wendland functions (Wendland, 1995, 1998), which has the characteristic of being compactly supported

$$\rho(h; \boldsymbol{\vartheta}) = (1-\lambda) I_{\{h<\gamma\}}\left(1 - \frac{1}{\gamma}\right)^{2(w+1)} P_w\left(\frac{h}{\gamma}\right) + \lambda I_{\{h=0\}}, \quad w = 0, 1, 2, \ldots \tag{2}$$

where $\boldsymbol{\vartheta} = (\gamma, \lambda)^\top$ and $P_w(\cdot)$ is a polynomial of order w. See Appendix A.3 for explicit examples of $w = 1$ and $w = 2$, corresponding to Wendland$_1$ and Wendland$_2$, respectively. They are special cases of a more general family called Generalized Wendland functions (Bevilacqua et al., 2019).

Note that the particular parametrization regarding the noise effect ($\lambda I_{\{h=0\}}$ component of ρ) implies a more robust estimation of σ than a decomposition according to the spatial scales (Cressie, 1993, pp 112–113).

Now, considering $\mathbf{z}$ as one realization of the process $Z(\mathbf{s})$ at locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\} \subset \mathcal{D}$ (for simplicity of the exposition, mutually distinct) and the matrix $\mathbf{X}$ with rows $\mathbf{x}(\mathbf{s}_i)^\top$, the log-likelihood of the parameter vector $\boldsymbol{\psi} = (\boldsymbol{\beta}^\top, \boldsymbol{\vartheta}^\top, \sigma)^\top \in \mathbb{R}^{q+m+1}$ given $\mathbf{z}$ is

$$\begin{aligned} l(\boldsymbol{\psi}; \mathbf{z}) &= l(\boldsymbol{\beta}, \boldsymbol{\vartheta}, \sigma; \mathbf{z}) \\ &= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2}\log\det \boldsymbol{\Omega}(\boldsymbol{\vartheta}) - \frac{1}{2}(\sigma^{-1}(\mathbf{z} - \mathbf{X}\boldsymbol{\beta}))^\top \boldsymbol{\Omega}(\boldsymbol{\vartheta})^{-1}(\sigma^{-1}(\mathbf{z} - \mathbf{X}\boldsymbol{\beta})), \end{aligned} \tag{3}$$

where the correlation matrix $\boldsymbol{\Omega}(\boldsymbol{\vartheta}) = [\rho(h_{ij}; \boldsymbol{\vartheta})]_{i,j=1}^n$. A vector $\hat{\boldsymbol{\psi}}_{\mathrm{ML}}$ that maximizes $l(\cdot; \mathbf{z})$ is called a Maximum Likelihood (ML) estimate and is found via numerical optimizers.

One of the simplest implementations of a numerical optimizer evaluates $l(\cdot; \mathbf{z})$ for a feasible set of parameters $\{\boldsymbol{\psi}_j \in \boldsymbol{\Psi}, j = 1, 2, \ldots\}$, with $\boldsymbol{\Psi}$ being the parameter space of $\boldsymbol{\psi}$, and then selecting the one that achieves the highest likelihood. A more appealing subclass of numerical algorithms is made up of gradient-based algorithms, such as steepest-ascent, which iteratively updates its estimate based on the gradient of $l(\boldsymbol{\psi}; \mathbf{z})$. See Nash (2014) for a thorough overview.

The profile likelihood approach is based on gradient algorithms and has the advantage of splitting the optimization of $\boldsymbol{\psi}$ into two steps, first estimating $\boldsymbol{\vartheta}$ and then $\boldsymbol{\beta}$. To begin with, let $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$ for any plausible and thus admissible values. Next, a unique closed-form solution for $\hat{\boldsymbol{\beta}}$ that maximizes $l(\cdot; \mathbf{z})$ is

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\vartheta}_0) = \left(\mathbf{X}^\top \boldsymbol{\Omega}^{-1}(\boldsymbol{\vartheta}_0)\mathbf{X}\right)^{-1} \mathbf{X}^\top \boldsymbol{\Omega}^{-1}(\boldsymbol{\vartheta}_0)\mathbf{z}. \tag{4}$$

This estimate corresponds to the generalized least squares estimate of $\boldsymbol{\beta}$ when $\mathrm{Corr}(\mathbf{Z}) = \boldsymbol{\Omega}(\boldsymbol{\vartheta}_0)$. We now consider (4) as a function of $\boldsymbol{\vartheta}$ and rewrite (3) to

$$l_p(\boldsymbol{\vartheta}, \sigma; \mathbf{z}) = l(\hat{\boldsymbol{\beta}}(\boldsymbol{\vartheta}), \boldsymbol{\vartheta}, \sigma; \mathbf{z}) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2}\log\det \boldsymbol{\Omega}(\boldsymbol{\vartheta}) - \frac{1}{2\sigma^2}\mathbf{z}^\top \mathbf{P}(\boldsymbol{\vartheta})\mathbf{z}, \tag{5}$$

where $\mathbf{P}(\boldsymbol{\vartheta}) = \boldsymbol{\Omega}^{-1}(\boldsymbol{\vartheta}) - \boldsymbol{\Omega}^{-1}(\boldsymbol{\vartheta})\mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}(\boldsymbol{\vartheta})\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}(\boldsymbol{\vartheta})$. The function $l_p(\boldsymbol{\vartheta}, \sigma; \mathbf{z})$ is known as the profile log-likelihood function of $(\boldsymbol{\vartheta}, \sigma^2)$ (Waller and Carlin, 2010) and can be optimized by

numerical algorithms as well. Finally, $\hat{\boldsymbol{\beta}}_{\mathrm{ML}} = \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\vartheta}}_{\mathrm{ML}})$. As climate fields are typically large, the computation of $\mathbf{P}(\boldsymbol{\vartheta})$ is costly and thus, in practice, the spatial structure is often neglected when estimating fixed effects. That means that $\hat{\boldsymbol{\beta}}_{\mathrm{LS}} = \left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{z}$ is used instead of (4). Accordingly, the profile log-likelihood is based on the vector of residuals $\mathbf{e} = \mathbf{z} - \mathbf{X}\hat{\boldsymbol{\beta}}$ and is approximated by

$$\widetilde{l}_p(\boldsymbol{\vartheta}, \sigma; \mathbf{e}) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2}\log\det\boldsymbol{\Omega}(\boldsymbol{\vartheta}) - \frac{1}{2\sigma^2}\mathbf{e}^{\mathsf{T}}\boldsymbol{\Omega}^{-1}(\boldsymbol{\vartheta})\mathbf{e}. \tag{6}$$

This approach can be refined by adding a couple of iterations over Eqs. (4) and (6).

The major disadvantage of gradient algorithms is the computational burden of evaluating the (profile) log-likelihood at each step, which can be prohibitive even for medium-sized samples as we need to evaluate the determinant of $\boldsymbol{\Omega}(\boldsymbol{\vartheta})$, and solve linear systems involving $\boldsymbol{\Omega}(\boldsymbol{\vartheta})$. The number of flops required for solving those linear systems is up to $O(n^2)$, with $O(n^3)$ required for storage. We therefore present three techniques that aim to ease the computational burdens present when modeling spatial climatological processes using an ML approach.

2.1. Tapering

One method of mitigating the computational burden of gradient algorithms is the Tapering Approach (TA) (Furrer et al., 2006). This induces sparseness in the correlation matrix $\boldsymbol{\Omega}(\boldsymbol{\vartheta})$ by tapering the correlation function with a compactly-supported correlation function $\rho_\delta(h)$, i.e., with $\rho_\delta(h) = 0$ if $h > \delta$, and $\delta > 0$ being the taper range. The resulting tapered correlation function is then defined as

$$\rho_T(h; \boldsymbol{\vartheta}, \delta) = \rho(h; \boldsymbol{\vartheta})\rho_\delta(h), \tag{7}$$

which is also valid because the product of two valid covariance functions is also valid. The associated tapered covariance matrix is therefore defined by the element-wise product (or Schur product) of $\boldsymbol{\Omega}(\boldsymbol{\vartheta})$ and a matrix $\mathbf{T}(\delta) = [\rho_\delta(h_{ij})]_{i,j=1}^n$. There exist efficient Cholesky decomposition algorithms for sparse matrices (see Furrer and Sain, 2010, for example). Replacing $\boldsymbol{\Omega}(\boldsymbol{\vartheta})$ with $\boldsymbol{\Omega}(\boldsymbol{\vartheta}) \odot \mathbf{T}(\delta)$ in (6) results in a pseudo-likelihood whose maximization is computationally feasible for large datasets, but this comes with the shortcoming of biased estimates (Kaufman et al., 2008). An extension of the TA known as the double tapering approach can be used to overcome this drawback, since it yields unbiased estimations by tapering both the model and the log-likelihood's sample covariance matrix. However, this too comes with the downside of requiring the full inverse of the tapered covariance matrix, which compromises computational efficiency (Kaufman et al., 2008).

Although the TA is unable to deliver unbiased estimates compromising the interpretability of the resulting estimates, its most compelling characteristic is associated with prediction. This means that using the same tapered model for prediction, the resulting root mean squared errors are very close to those of the correct underlying model (Furrer et al., 2006), even in non-Gaussian or multivariate settings (Bachoc et al., 2020; Bevilacqua et al., 2019).

The selection of the taper range δ is crucial. One must strike a balance between a small value that implies faster calculation and a larger value that reduces the bias by being able to better capture the spatial dependency. As a rule of thumb, 50 to 100 locations should be within the taper range.

2.2. Deliberate misspecification

A simpler approach than TA considers directly a compactly-supported correlation structure for $\boldsymbol{\Omega}(\boldsymbol{\vartheta})$, regardless of the process's true underlying correlation structure, and fixing the range parameter to a convenient value. Due to sparse matrices, this Deliberate Misspecification (DM) approach leads to computational benefits similar to those of TA, but without honoring the underlying covariance structure of the process being studied.

2.3. Composite likelihood based on pairs

Another efficient method for easing the computational burden of the classic ML approach is the weighted Composite Likelihood (CL) method based on pairs. This is one of a general class of CL estimating methods (Lindsay, 1988; Varin et al., 2011) based on the likelihood of marginal or conditional events, and it has been successfully applied in recent years for estimating complex models and/or handling large datasets (see, e.g., Baddeley, 2017; Fronterrè et al., 2018; Lie and Eidsvik, 2021 and many more as indicated in Appendix A.2).

As outlined by Lindsay et al. (2011), the choice of a suitable CL function should be driven by the statistical and computational considerations of estimation problem at hand. For Gaussian random fields in particular, there is a clear computational advantage in using CL based on pairs of observations rather than using other types of CL methods (Eidsvik et al., 2013; Stein et al., 2004). Specifically, the weighted CL method based on pairs uses the log-likelihood functions $l_{ij}(\boldsymbol{\psi})$ and $l_{i|j}(\boldsymbol{\psi})$ associated with the bivariate random vector $(Z(\mathbf{s}_i), Z(\mathbf{s}_j))^\top$ and the random variable $Z(\mathbf{s}_i)|Z(\mathbf{s}_j) = z_j$, respectively. The weighted marginal pairwise log-likelihood (wpl_M) and the weighted conditional pairwise log-likelihood (wpl_C) functions are given by

$$wpl_M(\boldsymbol{\psi}; \mathbf{z}) = \sum_{i=1}^{n}\sum_{j\neq i}^{n} w_{ij} l_{ij}(\boldsymbol{\psi}), \qquad wpl_C(\boldsymbol{\psi}; \mathbf{z}) = \sum_{i=1}^{n}\sum_{j\neq i}^{n} w_{ij} l_{i|j}(\boldsymbol{\psi}), \tag{8}$$

respectively, where w_{ij} are suitable positive weights. The associate estimators are defined as $\hat{\boldsymbol{\psi}}_a = \text{argmax}_{\boldsymbol{\psi}}\, wpl_a(\boldsymbol{\psi})$, for approaches $a = M, C$.

In general, wpl_a, $a = M, C$ estimation is expected to be less statistically efficient than the ML estimation, and the role of weights w_{ij} is to minimize that loss. Using the theory of optimal estimating equations (Heyde, 1997), it can be seen (Bevilacqua et al., 2012) that determining the optimal weights requires the calculation of the inverse of an $n(n-1) \times n(n-1)$ matrix, which is computationally more demanding than what is required for an ML estimation. Some approximations of the optimal weights have been proposed in the literature (e.g., Li and Sang, 2018; Pace et al., 2019). However, calculating these kinds of weights remains computationally demanding for large datasets. Assuming symmetric weights, (8) simplifies to

$$wpl_M(\boldsymbol{\psi}; \mathbf{z}) = 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} w_{ij} l_{ij}(\boldsymbol{\psi}), \qquad wpl_C(\boldsymbol{\psi}; \mathbf{z}) = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} w_{ij}\big(2l_{ij}(\boldsymbol{\psi}) - l_i(\boldsymbol{\psi}) - l_j(\boldsymbol{\psi})\big), \tag{9}$$

where $l_i(\cdot)$, $i = 1, \ldots, n$ is the marginal log-likelihood.

One symmetric weight function based on distances that has been successfully applied by different authors (Bai et al., 2014; Feng et al., 2014; Heagerty and Lele, 1998) is given by

$$w_{ij}(k) = \begin{cases} 1, & \|\boldsymbol{s}_i - \boldsymbol{s}_j\| < k, \\ 0, & \text{otherwise}, \end{cases} \tag{10}$$

where $k \in \mathbb{R}^+$ is an arbitrary distance greater than the minimum distance between the locations. This kind of weight function allows a certain percentage (depending on k) of the total number of pairs to be eliminated and provides a clear computational advantage over a constant weight function. Additionally, it has been shown that these kind of weights improve the statistical efficiency of both the wpl_a, $a = M, C$ methods compared to the use of constant weights (see, for instance, Joe and Lee, 2009; Davis and Yau, 2011; Bevilacqua et al., 2012).

Via their extensive simulation study Bevilacqua and Gaetan (2015) showed that wpl_C slightly outperforms wpl_M, from a statistical efficiency viewpoint, when estimating the parameters of Gaussian random fields. For this reason, we will focus on the wpl_C estimation, which we refer to as CL in what follows.

2.4. Asymptotics

Two well-defined asymptotic frameworks exist with which to study the limiting distribution of estimators under a spatial framework. These are known as increasing-domain asymptotics and

infill-domain asymptotics (Cressie, 1993). Under increasing-domain asymptotics, we consider an expanding sampling region with increasing distances between locations; the ML and CL estimators are consistent and asymptotically normal under weak conditions (Mardia and Marshall, 1984; Cressie and Lahiri, 1996; Bevilacqua and Gaetan, 2015). On the other hand, under infill-domain asymptotics, the sampling region is fixed and sampling gets denser. The theoretical results here are more limited, stating, among other things, the existence of two types of parameters: those that are consistently estimable, known as microergodic parameters, and those that are not (Kaufman and Shaby, 2013; Bevilacqua et al., 2019). The microergodic parameters have similar limiting distributions to those under an increasing-domain asymptotic.

In an infill-domain asymptotic framework the taper function needs to be as differentiable at the origin as the covariance function and some other minor assumptions (Furrer et al., 2006; Kaufman et al., 2008; Stein, 2013). For a fixed range, a high smoothness also implies stronger tapering, and thus, we often opt for Matérn covariance functions with smoothness parameter $\nu \leq 0.5$, a Spherical covariance taper, with $0.5 < \nu \leq 1.5$, a Wendland$_1$ and with $1.5 < \nu \leq 2.5$, a Wendland$_2$. In an increasing-domain asymptotic framework, on the other hand, the taper range needs to increase, but we only have weak restrictions on the taper function (Bachoc et al., 2020).

The theoretical justification for the DM approach is discussed in Bevilacqua et al. (2019) for an infill-domain asymptotic framework.

In practice, we have a single set of observations, i.e., one finite "n". This type of setting does not dictate the asymptotic scheme and picking the scheme with the best properties in the particular setting is legitimate. For climate science, we believe that infill-domain asymptotics are more intuitive and mimic well the situation of increasing the resolution of climate models.

3. Sinh-arcsinh distribution

The sinh-arcsinh (SAS) distribution was introduced by Jones and Pewsey (2009) as a general means of generating classes of distributions containing symmetric and asymmetric cases with varying tail-weights, and thus it is well suited for modeling climatological data. Let us consider the continuous, strictly monotonic function $S_{a,b} : \mathbb{R} \to \mathbb{R}$ defined as

$$S_{a,b}(y) = \sinh\big(b \sinh^{-1}(y) - a\big), \tag{11}$$

with $b > 0$ and $a \in \mathbb{R}$. Given a symmetrical random variable U, a so-called *generating random variable*, with the probability density function (pdf) f_U and the cumulative distribution function (cdf) F_U, we define $Y^s_{\alpha,\kappa}$, an SAS random variable, as

$$Y^s_{\alpha,\kappa} := S_{-\frac{\alpha}{\kappa},\frac{1}{\kappa}}(U) = \sinh\big(\kappa^{-1}(\sinh^{-1}(U) + \alpha)\big) \tag{12}$$

with a closed form inverse

$$U = \sinh\big(\kappa \sinh^{-1}(Y^s_{\alpha,\kappa}) - \alpha\big) = S_{\alpha,\kappa}(Y^s_{\alpha,\kappa}). \tag{13}$$

The pdf and cdf of the SAS random variable are

$$f_{Y^s_{\alpha,\kappa}}(y) = \kappa \left(\frac{1 + S^2_{\alpha,\kappa}(y)}{1 + y^2}\right)^{1/2} f_U\big(S_{\alpha,\kappa}(y)\big) \qquad \text{and} \qquad F_{Y^s_{\alpha,\kappa}}(y) = F_U\big(S_{\alpha,\kappa}(y)\big).$$

One particularly appealing property of the SAS transformation is that its parameters are clearly interpretable. The parameters α and κ can be interpreted as skewness and kurtosis parameters, respectively. If they are studied separately, the parameter $\alpha \in \mathbb{R}$ controls the distribution's skewness. In particular, $\alpha > 0$ and $\alpha < 0$ yield a right-skewed and a left-skewed distribution, respectively. The parameter κ controls the tail-weights, where tail-weights decrease with increasing κ. Specifically, $\kappa < 1$ and $\kappa > 1$ yield heavier and lighter tails than the Gaussian distribution, respectively. Additionally, the special case where $\alpha = 0$ and $\kappa = 1$ leads to the identity transformation, i.e., yielding the generating random variable $Y^s_{0,1} = U$.

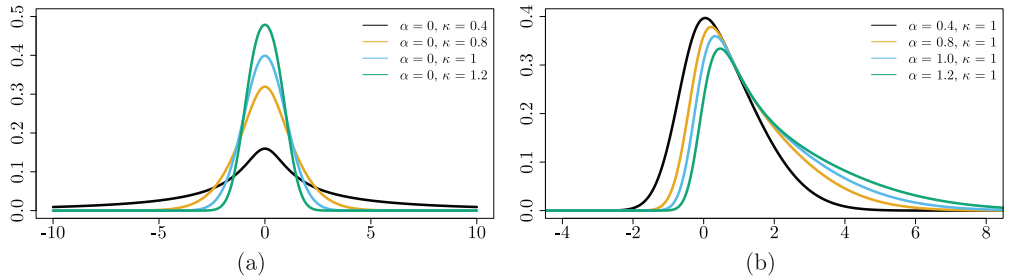


Fig. 1. SAS density for (a) $\alpha = 0$ and $\kappa = 0.4, 0.8, 1$, and 1.2, (b) $\kappa = 1$ and $\alpha = 0.4, 0.8, 1$, and 1.2.

One important case is when the generating random variable is standard Gaussian in which the pdf of the SAS random variable $Y^s_{\alpha,\kappa}$, is given by

$$f_{Y^s_{\alpha,\kappa}}(y) = \kappa\left(\frac{1+S^2_{\alpha,\kappa}(y)}{2\pi\left(1+y^2\right)}\right)^{1/2}\exp\left(-\frac{1}{2}S^2_{\alpha,\kappa}(y)\right). \tag{14}$$

Fig. 1 shows the flexibility of the SAS distribution for increasing values of both the tail parameter (with fixed $\alpha = 0$) and the skewness parameter (with fixed $\kappa = 0$) using a standard Gaussian generating distribution.

This paper applies the SAS transformation to a zero mean unit variance, isotropic Gaussian random field $Z^s(\mathbf{s}) = \left\{Z^s(\mathbf{s}), \mathbf{s} \in \mathcal{D} \subset \mathbb{R}^d\right\}$ with correlation function $\rho(\cdot;\boldsymbol{\vartheta})$. Specifically, we define $Y^s_{\alpha,\kappa}(\mathbf{s}) = \left\{Y^s_{\alpha,\kappa}(\mathbf{s}), \mathbf{s} \in \mathcal{D} \subset \mathbb{R}^d\right\}$ with $Y^s_{\alpha,\kappa}(\mathbf{s}) := S_{-\frac{\alpha}{\kappa},\frac{1}{\kappa}}\left(Z^s(\mathbf{s})\right)$, a standard SAS random field with the marginal density given in (14).

A location and scale transformation gives a non-standard SAS random field $Y_{\alpha,\kappa} = \left\{Y_{\alpha,\kappa}(\mathbf{s}), \mathbf{s} \in \mathcal{D} \subset \mathbb{R}^d\right\}$, defined as

$$Y_{\alpha,\kappa}(\mathbf{s}) = \mu(\mathbf{s}) + \sigma Y^s_{\alpha,\kappa}(\mathbf{s}), \tag{15}$$

where $\mu(\mathbf{s}) \in \mathbb{R}$ is a spatially varying location parameter that, as in Section 2, can be expressed as $\mu(\mathbf{s}) = \boldsymbol{x}(\mathbf{s})^\top\boldsymbol{\beta}$, and $\sigma > 0$ is a scale parameter. The mean and variance of this process are

$$\mathrm{E}\left(Y_{\alpha,\kappa}(\mathbf{s})\right) = \mu(\mathbf{s}) + \frac{\sigma\sinh(\alpha/\kappa)e^{1/4}}{\sqrt{8\pi}}\left(K_{\frac{\kappa+1}{2\kappa}}(1/4) + K_{\frac{1-\kappa}{2\kappa}}(1/4)\right),$$

$$\mathrm{Var}\left(Y_{\alpha,\kappa}(\mathbf{s})\right) = \frac{\sigma^2\cosh(2\alpha/\kappa)e^{1/4}}{\sqrt{32\pi}}\left(K_{\frac{\kappa+2}{2\kappa}}(1/4) + K_{\frac{2-\kappa}{2\kappa}}(1/4)\right) - \frac{1}{2} - \left(\mathrm{E}(Y_{\alpha,\kappa}(\mathbf{s}))\right)^2,$$

where K_ζ is the modified Bessel function of the second kind of order ζ (Jones and Pewsey, 2009).

Since the transformation $S_{-\frac{\alpha}{\kappa},\frac{1}{\kappa}}(\cdot)$ is monotonic, such that $\int_{-\infty}^{\infty} S^2_{-\frac{\alpha}{\kappa},\frac{1}{\kappa}}(t)\varphi(t)dt < \infty$ with $\varphi(\cdot)$, the standard Gaussian density, $Y^s_{\alpha,\kappa}(\mathbf{s})$ can be expressed as

$$Y^s_{\alpha,\kappa}(\mathbf{s}) = \sum_{j=0}^{\infty}\frac{\xi_j(\alpha,\kappa)H_j\left(Z^s(\mathbf{s})\right)}{j!},$$

where $H_j(\cdot)$, $j = 0, 1, 2, \ldots$ are the jth order (probabilistic) Hermite polynomials (Abramowitz and Stegun, 1970). A closed form expression for the coefficients $\xi_j(\alpha,\kappa)$ can be found in Appendix A.4. In addition, the correlation function of the SAS random field can be written as

$$\rho_{Y^s_{\alpha,\kappa}}(h;\boldsymbol{\vartheta}) = \sum_{j=1}^{\infty}\frac{\xi_j(\alpha,\kappa)^2}{j!}\rho(h;\boldsymbol{\vartheta})^j. \tag{16}$$

Using (16), it can easily be shown that mean-square continuity and degrees of mean-square differentiability can be inherited from the generating Gaussian random field $Z^s(\mathbf{s})$. In particular,

$Y^s_{\alpha,\kappa}(\mathbf{s})$ is m-times mean-square differentiable if $Z^s(\mathbf{s})$ is m-times mean-square differentiable. As a consequence, flexible correlation functions, like the Matérn model, can be used to parametrize the mean square differentiability of the SAS random field, as in the Gaussian case.

Given $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\} \in \mathcal{D}$ a set of distinct locations, let $\boldsymbol{\Omega} = [\rho(\|\mathbf{s}_i - \mathbf{s}_j\|; \boldsymbol{\vartheta})]_{i,j=1}^n$, the correlation matrix associated with a parametric correlation model of the process $Z^s(\mathbf{s})$. Let $S_{\alpha,\kappa}(\boldsymbol{Z}^S)$ be the componentwise SAS transformation of the elements of the vector $\boldsymbol{Z}^s$, where $\boldsymbol{Z}^s = \left(Z^s(\mathbf{s}_1), \ldots, Z^s(\mathbf{s}_n)\right)^\top$ is a Gaussian random vector, where $Z^s(\mathbf{s}_i) = S_{\alpha,\kappa}\left(\sigma^{-1}(y_i - \mu_i)\right)$ for $i = 1, \ldots, n$. Thus, the log-likelihood function associated to the random vector $\boldsymbol{Y} = \left(Y_{\alpha,\kappa}(\mathbf{s}_1), \ldots, Y_{\alpha,\kappa}(\mathbf{s}_n)\right)^\top$ is given by

$$l(\boldsymbol{\psi}; \mathbf{y}) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2}\log\det\boldsymbol{\Omega}(\boldsymbol{\vartheta}) - \frac{1}{2}S_{\alpha,\kappa}\left(\sigma^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right)^\top\boldsymbol{\Omega}(\boldsymbol{\vartheta})^{-1}S_{\alpha,\kappa}\left(\sigma^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right) + n\log(\kappa) - \frac{1}{2}\sum_{i=1}^{n}\left(\log\left(1 + S^2_{\kappa,\alpha}\left(\sigma^{-1}(y_i - \mathbf{x}_i^\top\boldsymbol{\beta})\right)\right) - \log\left(1 + \left(\sigma^{-1}(y_i - \mathbf{x}_i^\top\boldsymbol{\beta})\right)^2\right)\right), \tag{17}$$

where $S_{\alpha,\kappa}(\mathbf{y}) = [S_{\alpha,\kappa}(y_i)]_{i=1}^n$ and $\boldsymbol{\psi} = (\boldsymbol{\beta}^\top, \boldsymbol{\vartheta}^\top, \sigma, \kappa, \alpha)^\top \in \mathbb{R}^{q+m+3}$. If $\kappa = 1$ and $\alpha = 0$, (17) simplifies to the Gaussian log-likelihood (3).

4. Simulation

We ran a simulation study to illustrate the behavior of the four estimation methods presented here using sixteen synthetic climatological settings. We present the distribution of biases and the computational efficiency of the four discussed estimation approaches.

4.1. Simulation framework

We consider a common baseline framework of isotropic SAS random fields, with the underlying Matérn correlation model defined in Eq. (1) with $\gamma = 0.05$, scale parameter $\sigma = 1$, range parameter $\gamma = 0.05$, and mean function $\mu(\mathbf{s}) = 1 + 0.1\cos(s_x) + 0.2\cos(s_y)$, $\mathbf{s} \in [0, 1]^2$. A full factorial design including four factors is considered on this common framework, defining the sixteen synthetic frameworks shown in Table 1 and from which 200 replications are drawn. The *Smoothness* factor relates to the process's mean square differentiability. The 'high' setting sets $\nu = 2.3$ and $\lambda = 0$, mimicking temperature fields. The 'low' setting sets $\nu = 0.5$ and adds a small white noise component, with $\lambda = 0.1$, mimicking precipitation fields (Furrer et al., 2007a). The *Gridded* factor relates to the locations' spatial configuration, in gridded or non-gridded fields, where gridded fields are given by one realization of a simple inhibition process with minimum distance equal to 0.01 (Cressie, 1994), mimicking the minimum separation between the stations of a monitoring network. A *Size* factor relates to the sample size, with small and large size levels of 784 and 3025, respectively. The fourth and final *Distribution* factor relates to the random field's underlying distribution, which is an SAS distribution with skewness and kurtosis parameters of $\alpha = 0$ and $\kappa = 1$ (i.e., Gaussian) or $\alpha = 0.5$ and $\kappa = 0.8$, respectively.

Each field is then estimated using the four approaches described above: ML, TA, DM, and CL. We assume a Matérn correlation structure with known, fixed smoothness parameters for ML, TA, and CL, i.e., $\boldsymbol{\vartheta} = (\gamma, \lambda)^\top$. For TA, the tapering function is selected using the infill-domain asymptotics framework described in Section 2.4, with the taper range set to half of the true effective range, i.e., $\delta = 0.075$ and $\delta = 0.143$ for low and high smoothness, respectively. The same correlation model is used for DM.

A general optimization task is done by tuning memory allocation for TA and DM, and by selecting the parameter k from Eq. (10) for the CL approach. For the Gaussian frameworks, the parameters are estimated using the profile-likelihood (6) with the initial $\boldsymbol{\vartheta}_0$ set to the true value. All the parameters for the SAS random fields are estimated jointly based on (17). A box constraints optimizer method is used (L-BFGS-B), with the same initialization for all the estimation methods.

Table 1
Specifications of the simulation frameworks.

Framework	Smoothness	Gridded	Size	Distribution
1	Low	Yes	784	Gaussian
2	High	Yes	784	Gaussian
3	Low	No	784	Gaussian
4	High	No	784	Gaussian
5	Low	Yes	3025	Gaussian
6	High	Yes	3025	Gaussian
7	Low	No	3025	Gaussian
8	High	No	3025	Gaussian
9	Low	Yes	784	SAS
10	High	Yes	784	SAS
11	Low	No	784	SAS
12	High	No	784	SAS
13	Low	Yes	3025	SAS
14	High	Yes	3025	SAS
15	Low	No	3025	SAS
16	High	No	3025	SAS

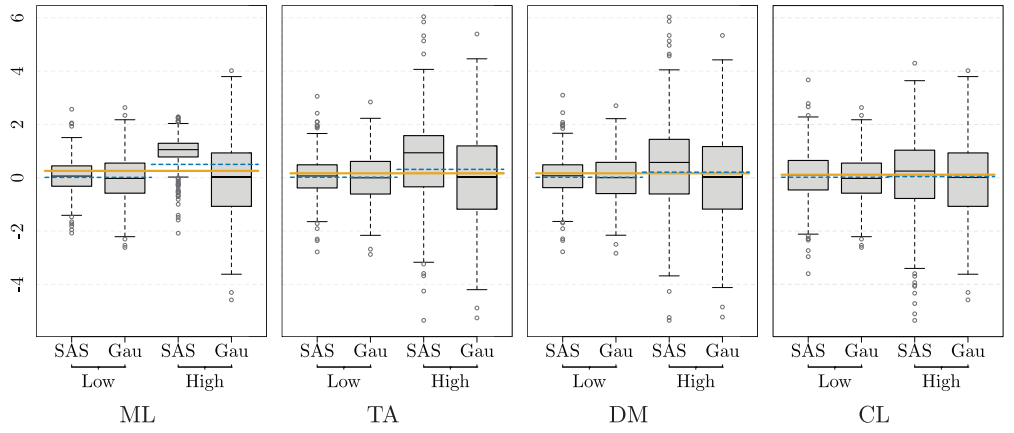


Fig. 2. Bias of $\hat{\beta}_0$. The solid orange solid lines represent the mean value calculated using each estimation method. Blue dashed lines represent the mean value for each estimation method according to their smoothness setting. Single boxplots are by underlying distribution. The remaining factors (Gridded and Size) were collapsed since no relevant differences were found.

All computing was performed using `R` software (R. Core Team, 2021), version 4.1.1, on an AMD Ryzen 9 5900X × 24 threads computer, with 128 GB RAM. The code employed was written according to the principles of research transparency and reproducibility (see Appendix A.5) and relied heavily on the `spam` (Furrer et al., 2021), `geoModels` (Bevilacqua et al., 2018), and `optimParallel` (Gerber and Furrer, 2019) `R` software packages.

4.2. Results

The results of our simulation study are shown in Figs. 2–6. For each parameter, we identified the two relevant factors that best described the differences between the estimation methods. The remaining factors were collapsed to make visualization easier.

Fig. 2 shows the bias for $\hat{\beta}_0$. Overall, the estimation methods led to similar biases. Regarding the dispersion of the biases, there was a clear pattern according to smoothness, with a smaller spread at low level of smoothness than at high smoothness. The high smoothness SAS frameworks were the most challenging ones: ML, TA, and DM frameworks demonstrated a considerable positive shift of

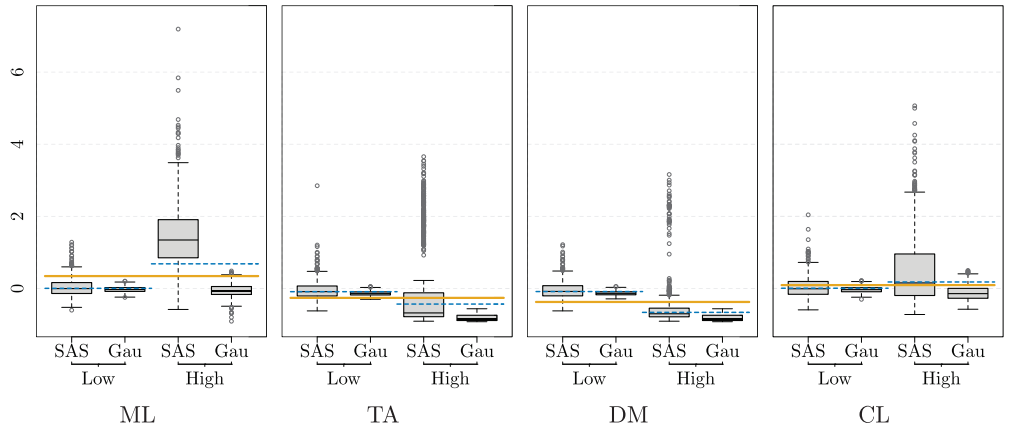


Fig. 3. Bias of $\hat{\sigma}$. The solid orange lines represent the mean value calculated using each estimation method. Blue dashed lines represent the mean value for each sample size by estimation method. Single boxplots are by underlying distribution. The remaining factors (Gridded and Size) were collapsed since no relevant differences were found.

their medians. In this framework, CL stood out as the most robust approach. The results for β_1 and β_2 biases were similar, but with a slight negative median bias for ML and TA in the high smoothness SAS frameworks.

Fig. 3 shows the bias of $\hat{\sigma}$ with similar features to those seen in Fig. 2, e.g., similar overall behavior of their medians, with CL having the lowest one; higher spreads for the high smoothness frameworks; and the high smoothness SAS frameworks representing the most challenging scenarios. However, for all four methods the distribution of bias is strongly right-skewed.

Fig. 4 presents the square root of the absolute bias for the range $\hat{\gamma}$ standardized by the true parameter γ. The absolute biases do not differ greatly from the raw ones because approximately 96% of them are positive. DM was excluded since we deliberately fixed the range parameter. In this figure, all three estimation methods present markedly different behavior. ML exhibits the best overall results for medians and dispersion, without any clearly observable difference between the frameworks presented. CL shows markedly larger estimates, especially for the smaller sample size. Finally, TA was expected to return biased estimates, which were mild in the previous results, but have now deteriorated and led to extremely wide ranges. An increase in the sample size mitigates the most severe biases.

Fig. 5 shows the bias of $\hat{\lambda}$. A clear difference can be seen between the low and high smoothness frameworks across the four estimation methods, with the latter being better estimated, in general. Within those high smoothness frameworks, CL exhibits considerably worse behavior for low sample sizes, presenting a raw bias of 0.1 when the true parameter is 0. This flaw is overcome as the sample size increases. For the low smoothness frameworks, CL again exhibits particularly bad behavior for low sample sizes, which is again corrected as the sample size increases. Next, there is an expected improvement associated with the sample size increase in the low smoothness scenario with ML, in addition to having both boxplots almost centered in 0. An interesting pattern can be seen for TA and DM in the low smoothness frameworks: an increase in sample size shifts the boxplots off 0, exhibiting narrower but more biased results.

All estimation methods provided mostly consistent estimates of the SAS fields' kurtosis and skewness. The distribution of the skewness parameter is stronger right-skewed for ML and CL compared to TA and DM.

Fig. 6 summarizes the relative consumption of time (a) and peak memory (b) required by the different approximate methods in comparison to ML, and split by sample size (small to the left and large to the right of the central line). Most importantly, the three other estimation methods were substantially quicker than ML. CL presents the best relative reduction in time needed, especially

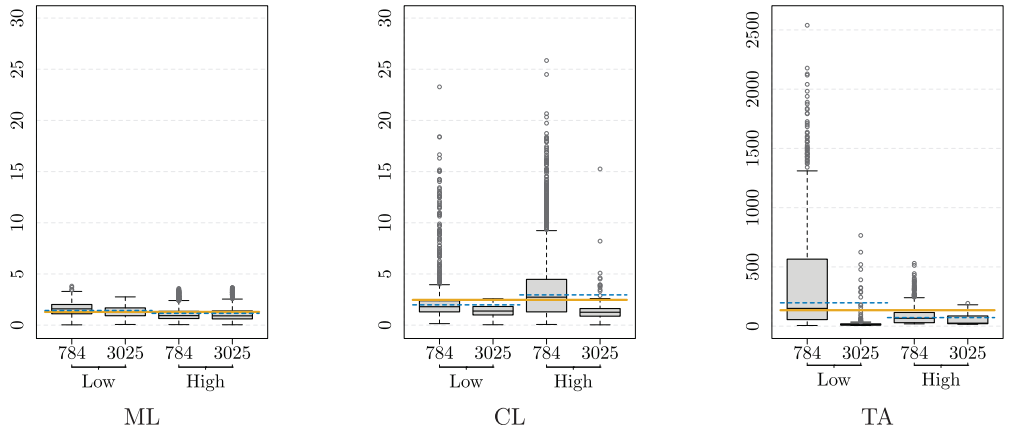


Fig. 4. Square root of the absolute bias of $\hat{\gamma}$ standardized by the true parameter γ. The solid orange solid lines represent the mean value calculated using each estimation method. The blue dashed lines represent the mean value for each estimation method by smoothness setting. Single boxplots are by sample size. The remaining factors (Gridded and Distribution) were collapsed since no relevant differences were found. Note the different scales.

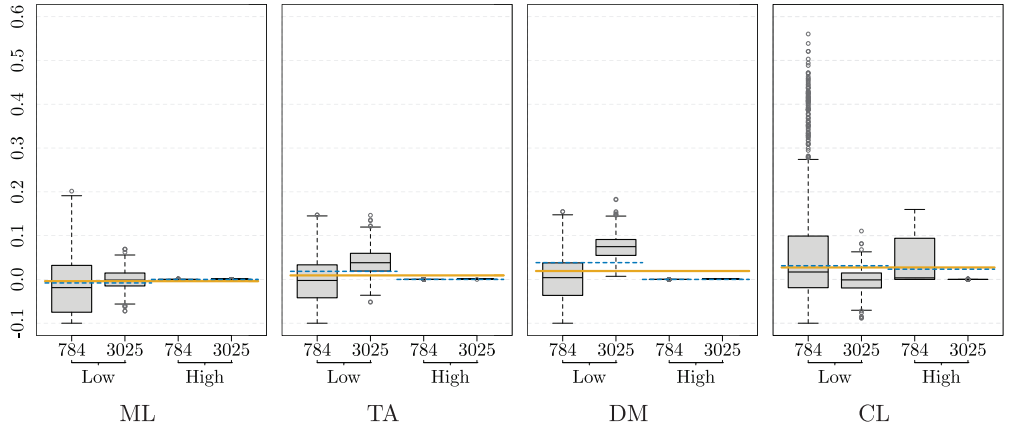


Fig. 5. Bias of $\hat{\lambda}$. Solid orange solid lines represent the mean value calculated using estimation method. Blue dashed lines represent the mean value by parameter setting for each estimation method. Single boxplots are by sample size. The remaining factors were collapsed since no relevant differences were found.

for the large sample size frameworks. For the low sample size, the joint median of the other three distributions is around 0.2, and it is 0.05 for the large sample size, meaning that they required just 20% and 5% of the time required by ML. Only a small proportion of estimates required more time when using TA and DM (3.6% and 0.4%, respectively). Each of the approximate estimation methods in panel (b) of Fig. 6 exhibits a markedly different pattern for the given sample sizes. For small sample sizes, the methods generally consumed more peak memory than ML, whereas they were substantially lower than ML when using larger sample sizes. TA and DM showed the largest reductions, with a joint median of approximately 0.4; the median was 0.81 for CL.

4.3. Discussion

We have presented an overview of the statistical estimation approaches' behaviors via a simulation study of sixteen synthetic frameworks mimicking different climatological datasets. We

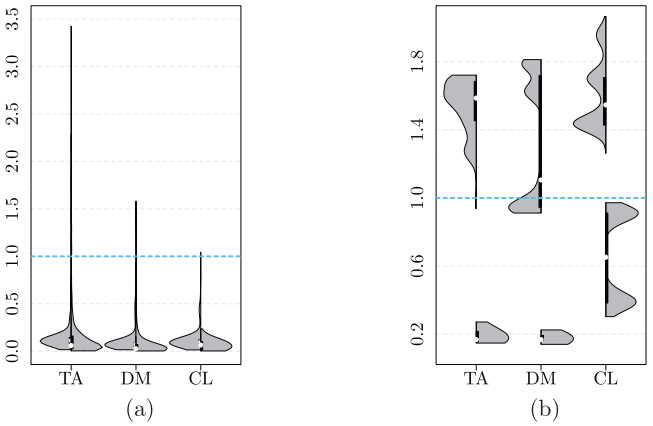


Fig. 6. Violin plots showing the relative consumption of (a) time and (b) peak memory by the approximate estimation methods in comparison to ML, split by sample size (small left and large right).

addressed the interpretability of their resulting estimates by analyzing their bias distributions; we addressed their computational performance by looking at their consumption of time and peak memory. One limitation to this simulation study was its inability to differentiate between effects due to the underlying distribution and those due to the fitting procedure of our full factorial design. Both effects are aligned on their levels and therefore the effect can be only estimated jointly. Regarding the factors studied, one that never stood out as a main component to explain the differences between the estimates was Gridded. The better noise estimation capabilities associated with non-gridded fields seems to vanish when limiting the minimum distance between locations (at least with the chosen distance).

Regarding the interpretability of our estimates, none of the approximate methods stood out as a potential overall preferred alternative to ML, but some approximate methods could be used for some specific frameworks. CL performed well for the fixed effect and scale parameter estimations, even outperforming ML, and delivered less biased estimates at the cost of a small increase of variability. However, CL performed worse for the range and noise estimations over noisy frameworks when dealing with a low sample size. TA slightly outperformed CL within the low smoothness frameworks, delivering less variable estimates for the noise parameter with a slight increase in their bias, with the exception, of course, of the range estimates. DM's performance can be summarized as slightly worse than TA: in general, it showed marginally more dispersed and biased estimates, with the extra limitation of not being able to provide range estimates. Finally, ML was the best approach for range and noise estimations over low sample-size frameworks, but it showed subpar performance for the fixed effects and scale parameters in high smoothness SAS frameworks.

Regarding computational performance, there was a clear pattern due to sample size. For small sample sizes, the approximate methods and mechanisms ended up requiring more memory and time than ML (up to 1.4 times more than ML). This is because approximation methods are tuned to work well with large and massive sample sizes, which was also seen in their substantially lower peak memory consumption (41% of the time required by ML). Since CL is mainly used as part of the `GeoFit` function in `GeoModels`, we can expect to observe higher memory consumption. In general, the approximate methods required less computation time, with greater relative reductions for the larger sample size (as low as 5% of the time required by ML).

In summary, based on this simulation study, any recommendations on whether or not to use approximate methods must reference sample size and the smoothness of the process being studied. If the sample size is considerably large, then the CL approach stands out as a rapid, robust alternative to ML, delivering interpretable estimates. With a medium sample size, any restrictions on computing time or peak memory consumption, which would be exceeded using ML, point to the TA approach

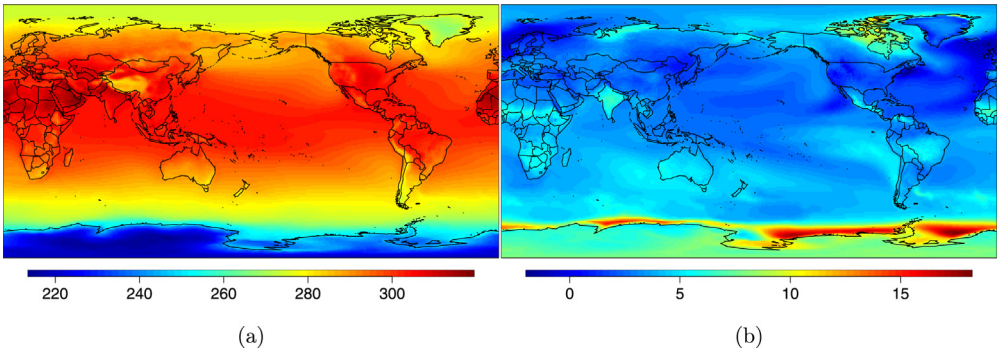


Fig. 7. (a) JJA average of near-surface air temperature (in degrees Kelvin) (years 1–20), (b) the projected change in near-surface air temperature (JJA, years 131–150 minus 1–20) (in degrees Kelvin). (Data: CESM2 model with experimental abrupt quadrupling of CO_2).

that provides relatively acceptable interpretability, except for the range parameter. TA offers a direct parameter trading-off bias and computational gains. Aside the range estimates, our experience shows that the choice of δ has much less effect on the estimates compared to k of CL.

5. Illustration

This section applies the three approximate estimation approaches to a gridded, high resolution, global climate field.

We use data from the CMIP6 repository, which can be downloaded for free via the web interface https://esgf-node.llnl.gov/search/cmip6/. Details on the downloading process are available in Appendix A.6. Climate projections have been calculated based on the Community Earth System Model Version 2 (CESM2) (Danabasoglu et al., 2020), made available to the public by the USA's National Center for Atmospheric Research (NCAR). We chose near-surface air temperature (denoted as TAS) as our climatological variable, based on a GCM experiment constrained as follows. The simulation started with a carbon dioxide (CO_2) concentration set to the level from the global annual mean 1850 value. Then an instantaneous quadrupling of the atmospheric CO_2 concentration was imposed with no further changes over the 150 years of simulation. Climate scientists use a setup like this to evaluate the climate model's true climate sensitivity and to diagnose the strength of various feedbacks. Furthermore, this experiment characterizes the radiative forcing that arises from an increase in atmospheric CO_2 as well as changes that arise indirectly due to warming (Pascoe et al., 2020).

Data is provided in its native 0.9×1.25 finite volume grid, resulting in 192×288 latitude–longitude grids. To avoid numerical instabilities, both boundary latitude rows have been eliminated. The left panel of Fig. 7 shows the summer seasonal climate average at the beginning of the simulation (average of June, July, and August, JJA, over years 1–20). The right panel shows the projected change in TAS towards the end of a 150-year simulation (average JJA, over the years 130–150). We then fit an SAS random field (15) to this dataset.

As we have data on the entire globe, we use a great circle distance (Furrer et al., 2007b) with a Wendland$_2$ covariance function that is valid on the sphere (Guinness and Fuentes, 2016). We fit an SAS model to the data with a taper, deliberate misspecification and a pairwise composite likelihood approach, denoted TA, DM and CL, respectively. A profile likelihood approach was adopted by first setting $\boldsymbol{\Omega}$ equal to the identity matrix to estimate mean, which was modeled by an additive model (cubic smoothing splines of sin and cosine of the latitude with a total of 18.74 degrees of freedom).

Table 2 summarizes the parameter estimates. The uncertainty estimates were approximated using the inverse of the Hessian matrix at the maximizer. To ensure the numerical stability of

Table 2
Parameter estimates for the three different approximate methods, where available, standard error estimates are given in parentheses.

Method	γ		σ		λ		α		κ	
TA	24.86	(11.213)	0.17	(0.002)	0.00	(0.167)	0.02	(0.002)	0.53	(0.008)
DM	6	(fixed)	0.17	(0.003)	0.00	(0.033)	0.02	(0.002)	0.53	(0.002)
CL	22.78		0.18		0.00		0.02		0.53	

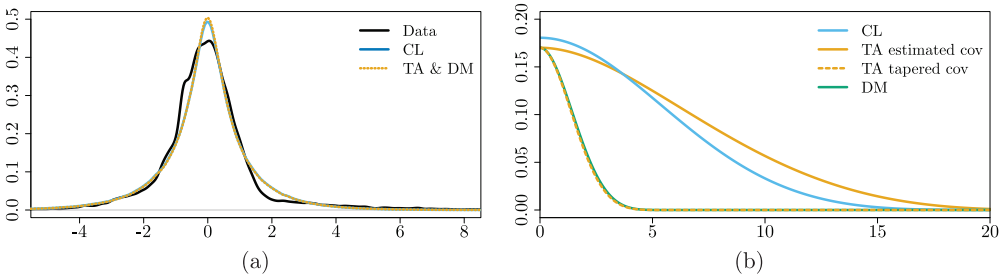


Fig. 8. (a) Marginal densities of the fitted models. The black line is a kernel density estimate of the spatial residuals. (b) Covariance functions based on parameter estimates.

this task, we down-sampled the data to 40,000 locations. For CL, no direct uncertainty measure is available for the model considered.

The left panel of Fig. 8 shows the marginal densities resulting from the parameter estimates given in Table 2. The resulting SAS density is a good match for the spatial residuals summarized by a kernel density estimate. The right panel of Fig. 8 gives the resulting covariance functions based on the parameter estimates. Note that although the resulting covariances of TA and DM are virtually identical, a range estimate is available for TA.

In summary, the three different estimation approaches yield similar fits and confirm that the choice of the estimation approach may be driven by the user aptness of the methodology and by the availability of the implemented code of the methodology.

6. Concluding remarks

The present article illustrated how to bridge the gap between statistical methodology and inference with regards to climate data. Within this contribution we would like to encourage readers to take advantage of the huge amounts of openly available climate data or to collaborate more often with climate scientists. We hope that this paper will also provide some guidance and starting points for spatial statisticians at all levels.

CRediT authorship contribution statement

Federico Blasi: Methodology, Software, Visualization, Formal analysis, Writing – original draft, Writing – review & editing, Visualization, Computing resources. **Christian Caamaño-Carrillo:** Methodology, Software. **Moreno Bevilacqua:** Methodology, Software, Supervision, Writing – original draft, Writing – review & editing. **Reinhard Furrer:** Methodology, Software, Supervision, Writing – original draft, Writing – review & editing, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank the reviewer and the editors for their support and comments. This work is supported by the Swiss National Science Foundation SNSF-175529 (RF and FB). Partial support was provided by FONDECYT grant 1200068 of Chile, by regional MATH-AmSud program, grant number 20-MATH-03, by grant ANID/PIA/ANILLOS ACT210096 for MB and by Proyecto Regular DIUBB 2120538 IF/R de la Universidad del Bío-Bío and by FONDECYT grant 11220066 of Chile for CC.

Appendix A

A.1. Links to data repositories

The table below provides a selective list of repositories of climate, meteorological, and environmental datasets. All links were accessed on January 5th 2022.

NCAR Climate Data (Satellite Data Products): https://climatedataguide.ucar.edu/data-type/satellite-data-products
Satellite imagery and other satellite-derived datasets linked to Southern Ocean: https://www.soos.aq/index.php?option=com_sppagebuilder&view=page&id=57
Free Satellite Imagery Sources: https://eos.com/blog/free-satellite-imagery-sources/
CMIP data browser (different phases): https://esgf-node.llnl.gov/search/cmip3, https://esgf-node.llnl.gov/search/cmip5 and https://esgf-node.llnl.gov/search/cmip6 (several mirrors exist).
Data provided by Google Earth Engine: https://developers.google.com/earth-engine/datasets/ with download described at https://developers.google.com/earth-engine/guides/exporting

See also the generic search link https://datasetsearch.research.google.com/.

A.2. Paper and citation counts

The table below provides the search strings used at https://www.scopus.com (using the University of Zurich's institutional subscription) for some of the statements in the text. The number corresponds to the number of articles found (Jan. 5th, 2022).

Query string	Number of articles found
SRCTITLE (''spatial statistics'') PUBYEAR < 2022	528
SRCTITLE (''spatial statistics'') PUBYEAR < 2022 ALL (climat*)	123
SRCTITLE (''spatial statistics'') PUBYEAR < 2022 TITLE-ABS-KEY(climat*)	26
SRCTITLE (''spatial statistics'') PUBYEAR < 2022 ALL (''Climate model'')	21
SRCTITLE (''spatial statistics'') PUBYEAR < 2022 ALL (''circulation model'')	2
SRCTITLE (''spatial statistics'') PUBYEAR < 2022 ALL (*GCM)	2
SRCTITLE (''spatial statistics'') PUBYEAR < 2022 TITLE-ABS-KEY (''composite *likelihood'')	16
SRCTITLE (''spatial statistics'') PUBYEAR < 2022 TITLE-ABS-KEY (''taper*'')	6

To count the number of citations in Masson-Delmotte et al. (2021b) for articles published in Spatial Statistics, the Journal of Statistical Software, the Journal of the Royal Statistical Society: Series B, the Journal of the American Statistical Association, Statistical Science, Annals of Applied Statistics, and Environmental and Ecological Statistics, respectively, we used the following approach (Linux tcsh 6.21.00):

```
> wget https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Full_Report.pdf
> pdftotext IPCC_AR6_WGI_Full_Report.pdf
> grep -c -e 10.1016/j.spasta -e 10.18637/jss -e 10.1111/rssb -e 10.1080/01621459 -e
   '10.1214/*-STS' -e '10.1214/*-AOAS' -e 10.1007/s10651 ar4_wg1.txt
```

Because DOIs were not consistently used for (Solomon et al., 2007) we used:

```
> wget https://www.ipcc.ch/site/assets/uploads/2018/05/ar4_wg1_full_report-1.pdf
> pdftotext ar4_wg1_full_report-1.pdf ar4_wg1.txt
> grep -w 'Stat\.' ar4_wg1.txt
```

Manual checks were performed for both approaches.

A.3. Correlation functions with compact support

Spherical: $\rho(h;\boldsymbol{\vartheta}) = (1-\lambda)I_{\{h<\gamma\}}\left(1-\frac{3}{2}\frac{h}{\gamma}+\frac{1}{2}\frac{h^3}{\gamma^3}\right)+\lambda I_{\{h=0\}}$,

Wendland$_1$: $\rho(h;\boldsymbol{\vartheta}) = (1-\lambda)I_{\{h<\gamma\}}\left(1-\frac{h}{\gamma}\right)^4\left(\frac{4h}{\gamma}+1\right)+\lambda I_{\{h=0\}}$,

Wendland$_2$: $\rho(h;\boldsymbol{\vartheta}) = (1-\lambda)I_{\{h<\gamma\}}\left(1-\frac{h}{\gamma}\right)^6\left(\frac{35}{3}\frac{h^2}{\gamma^2}+6\frac{h}{\gamma}+1\right)+\lambda I_{\{h=0\}}$, where $\boldsymbol{\vartheta}=(\gamma,\lambda)^{\mathsf{T}}$, $\gamma>0$, $0\le\lambda\le 1$.

A.4. Coefficients of Hermite polynomials $\xi_j(\alpha,\kappa)$

A standard SAS random field $Y^s_{\alpha,\kappa}(\mathbf{s})$ can be expressed by an infinite sum of Hermite polynomials

$$Y^s_{\alpha,\kappa}(\mathbf{s}) = S_{-\frac{\alpha}{\kappa},\frac{1}{\kappa}}\left(Z^s(\mathbf{s})\right) = \sum_{j=0}^{\infty}\frac{\xi_j(\alpha,\kappa)H_j\left(Z^s(\mathbf{s})\right)}{j!},$$

where the coefficients $\xi_j(\alpha,\kappa)$ depend on $Y^s_{\alpha,\kappa}(\mathbf{s})$. According to Jones and Pewsey (2009), we have

$$Y^s_{\alpha,\kappa}(\mathbf{s}) = S_{-\frac{\alpha}{\kappa},\frac{1}{\kappa}}\left(Z^s(\mathbf{s})\right) = \frac{1}{2}\left(e^{\alpha/\kappa}(Z^s(\mathbf{s})+(Z^s(\mathbf{s})^2+1)^{1/2})^{1/\kappa}-e^{-\alpha/\kappa}(Z^s(\mathbf{s})+(Z^s(\mathbf{s})^2+1)^{1/2})^{-1/\kappa}\right).$$

Let $H_j(\cdot)$, where $j=0,1,2,\ldots$ is the jth order (probabilistic) Hermite polynomial, defined as

$$H_j(z) = e^{z^2/2}\left(\frac{-d}{dz}\right)^j e^{-z^2/2}.$$

$H_j(\cdot)$ can be also represented by the so-called Hermite–Kampé de Fériet series

$$H_j(z) = j!\sum_{r=0}^{[j/2]}\frac{z^{j-2r}(-1)^r}{2^r(j-2r)!r!}.$$

Therefore, the coefficients $\xi_j(\alpha,\kappa)$ are given by,

$$\xi_j(\alpha,\kappa) = \mathrm{E}\left(Y^s_{\alpha,\kappa}(\mathbf{s})H_j(Z^s(\mathbf{s}))\right) = j!\sum_{r=0}^{[j/2]}\frac{(-1)^r I(\alpha,\kappa,j,r)}{2^{r+1}\sqrt{2\pi}(j-2r)!r!},$$

where, using Wolfram Mathematica, we obtain

$$\begin{aligned} I(\alpha,\kappa,j,r) &= \int_{-\infty}^{\infty} e^{-z^s(\mathbf{s})^2/2} z^s(\mathbf{s})^{j-2r} \\ &\quad \times \left(e^{\alpha/\kappa}(z^s(\mathbf{s}) + (z^s(\mathbf{s})^2+1)^{1/2})^{1/\kappa} - e^{-\alpha/\kappa}(z^s(\mathbf{s}) + (z^s(\mathbf{s})^2+1)^{1/2})^{-1/\kappa}\right) dz^s(\mathbf{s}) \\ &= (-1)^{3+j-2r} c_{\alpha,\kappa,1} - c_{\alpha,\kappa,2} + c_{\alpha,\kappa,1} a_{\alpha,\kappa,1} \\ &\quad + (-1)^{2+j-2r} c_{\alpha,\kappa,2} a_{\alpha,\kappa,1} + 2^{-2-j+2r} c_{\alpha,\kappa,3} a_{\alpha,\kappa,2} \\ &\quad - (-0.5)^{2+j-2r} c_{\alpha,\kappa,3} a_{\alpha,\kappa,3} \end{aligned}$$

with

$$\begin{aligned} a_{\alpha,\kappa,1} &= \cosh\left(\frac{2\alpha}{\kappa}\right) + \sinh\left(\frac{2\alpha}{\kappa}\right) \\ a_{\alpha,\kappa,2} &= \sec\left(\frac{\pi}{2\kappa} - \frac{j\pi}{2} + \pi r\right) + \sec\left(\frac{\pi}{2\kappa} + \frac{j\pi}{2} - \pi r\right) a_{\alpha,\kappa,1} \\ a_{\alpha,\kappa,3} &= \sec\left(\frac{\pi}{2\kappa} + \frac{j\pi}{2} - \pi r\right) + \sec\left(\frac{\pi}{2\kappa} - \frac{j\pi}{2} + \pi r\right) a_{\alpha,\kappa,1} \\ c_{\alpha,\kappa,1} &= e^{-\alpha/\kappa} 2^{-0.5+1.5/\kappa+j/2-r} \Gamma\left(\frac{1+j}{2} + \frac{1}{2d} - r\right) \\ &\quad \times {}_2F_2\left(\frac{1}{2} - \frac{1}{2d}, -\frac{1}{2d}; 1 - \frac{1}{d}, \frac{1-j}{2} - \frac{1}{2d} + r; \frac{1}{2}\right) \\ c_{\alpha,\kappa,2} &= e^{-\alpha/\kappa} 2^{-0.5-1.5/\kappa+j/2-r} \Gamma\left(\frac{1+j}{2} - \frac{1}{2d} - r\right) \\ &\quad \times {}_2F_2\left(\frac{1}{2} + \frac{1}{2d}, \frac{1}{2d}; 1 + \frac{1}{d}, \frac{1-j}{2} + \frac{1}{2d} + r; \frac{1}{2}\right) \\ c_{\alpha,\kappa,3} &= \frac{e^{-\alpha/\kappa}\pi\,\Gamma(1+j-2r)\,{}_2F_2\left(\frac{1+j}{2} - r, 1 + \frac{j}{2} - r; \frac{3+j}{2} - \frac{1}{2d} - r, \frac{3+j}{2} + \frac{1}{2d} - r; \frac{1}{2}\right)}{\kappa\,\Gamma\left(\frac{3+j}{2} - \frac{1}{2d} - r\right)\Gamma\left(\frac{3+j}{2} + \frac{1}{2d} - r\right)}, \end{aligned}$$

where ${}_2F_2(\cdot)$ is a special case of the generalized hypergeometric function ${}_pF_q(\cdot)$ (Abramowitz and Stegun, 1970).

A.5. Source files

R source files are available in the git repository https://git.math.uzh.ch/reinhard.furrer/j.spasta.2022.100596.

The README.txt file gives an overview of the available files as well on how to run them. The repository also provides a 'Dockerfile' to run the simulation in a fully reproducible environment.

A.6. Details for accessing the dataset in Section 5

The CMIP6 data are publicly available and can be downloaded for free via the web interface https://esgf-node.llnl.gov/search/cmip6/. The menu on the left of that page efficiently helps to restrict the search parameters. For example, we chose our dataset using the following tags (top to bottom): Activity: CMIP; Source ID: CESM2; Experiment ID: abrupt-4xCO2; Table ID: Amon; Variable: tas. Depending on the selection, the aggregation can be changed using either the Frequency or Table ID. For projections over many decades, the output may be provided in several files. There may also be more than one experiment possible for a specific selection, as indicated by the variant label. For a detailed description of the file naming convention, see Taylor et al. (2018), Brunner et al. (2020), or (Pascoe et al., 2020).

References

Abramowitz, M., Stegun, I.A. (Eds.), 1970. Handbook of Mathematical Functions. Dover.

Alegría, A., Cuevas-Pacheco, F., Diggle, P., Porcu, E., 2021. The F-family of covariance functions: A Matérn analogue for modeling random fields on spheres. Spatial Stat. 43, 100512.

Allen, M.R., Tett, S.F.B., 1999. Checking for model consistency in optimal fingerprinting. Clim. Dynam. 15, 419–434.

Appel, M., Pebesma, E., 2020. Spatiotemporal multi-resolution approximations for analyzing global environmental data. Spatial Stat. 38, 100465.

Bachoc, F., Betancourt, J., Furrer, R., Klein, T., 2020. Asymptotic properties of the maximum likelihood and cross validation estimators for transformed Gaussian processes. Electron. J. Stat. 14, 1962–2008.

Baddeley, A., 2017. Local composite likelihood for spatial point processes. Spatial Stat. 22, 261–295.

Bai, Y., Kang, J., Song, P., 2014. Efficient pairwise composite likelihood estimation for spatial-clustered data. Biometrics 7, 661–670.

Banerjee, S., 2020. Modeling massive spatial datasets using a conjugate Bayesian linear modeling framework. Spatial Stat. 37, 100417.

Barbian, M.H., Assunção, R.M., 2017. Spatial subsemble estimator for large geostatistical data. Spatial Stat. 22, 68–88.

Bevilacqua, M., Caamaño Carrillo, C., Arellano-Valle, R.B., Morales-Oñate, V., 2021. Non-gaussian geostatistical modeling using (skew) t processes. Scand. J. Stat. 48, 212–245.

Bevilacqua, M., Caamaño Carrillo, C., Gaetan, C., 2020a. On modelling positive continuous data with spatio-temporal dependence. Environmetrics 31, e2632.

Bevilacqua, M., Diggle, P.J., Porcu, E., 2020b. Families of covariance functions for bivariate random fields on spheres. Spatial Stat. 40, 100448.

Bevilacqua, M., Faouzi, T., Furrer, R., Porcu, E., 2019. Estimation and prediction using generalized wendland covariance functions under fixed domain asymptotics. Ann. Statist. 47, 828–856.

Bevilacqua, M., Gaetan, C., 2015. Comparing composite likelihood methods based on pairs for spatial Gaussian random fields. Stat. Comput. 25, 877–892.

Bevilacqua, M., Gaetan, C., Mateu, J., Porcu, E., 2012. Estimating space and space–time covariance functions for large data sets: a weighted composite likelihood approach. J. Amer. Statist. Assoc. 107, 268–280.

Bevilacqua, M., Morales-Oñate, V., Caamaño Carrillo, C., 2018. GeoModels: A package for geostatistical Gaussian and non Gaussian data analysis. R package version 1.0.3-4. https://vmoprojs.github.io/GeoModels-page/.

Bivand, R.S., Pebesma, E., Gomez-Rubio, V., 2013. Applied Spatial Data Analysis with R, Second ed. Springer, NY.

Boettiger, C., 2015. An introduction to Docker for reproducible research. Oper. Syst. Rev. 49, 71–79.

Brunner, L., Hauser, M., Lorenz, R., Beyerle, U., 2020. The ETH Zurich CMIP6 Next Generation Archive: Technical Documentation. Technical report, http://dx.doi.org/10.5281/zenodo.3734128.

Cameletti, M., Gómez-Rubio, V., Blangiardo, M., 2019. Bayeslan modelling for spatially misaligned health and air pollution data through the inla-spde approach. Spatial Stat. 31, 100353.

Cappello, C., De Iaco, S., Maggio, S., Posa, D., 2021. Time varying complex covariance functions for oceanographic data. Spatial Stat. 42, 100426.

Castruccio, S., 2016. Assessing the spatio-temporal structure of annual and seasonal surface temperature for CMIP5 and reanalysis. Spatial Stat. 18, 179–193.

Cressie, N.A.C., 1993. Statistics for Spatial Data, revised ed. Wiley.

Cressie, N., 1994. 4 - models for spatial processes. In: Stanford, J.L., Vardeman, S.B. (Eds.), Statistical Methods for Physical Science, Volume 28 of Methods in Experimental Physics. Academic Press, pp. 93–124.

Cressie, N., Lahiri, S.N., 1996. Asymptotics for REML estimation of spatial covariance parameters. J. Statist. Plann. Inference 50, 327–341.

Damian, D., Sampson, P.D., Guttorp, P., 2003. Variance modeling for nonstationary spatial processes with temporal replications. J. Geophys. Res.: Atmos. 108 (8778).

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D.A., DuVivier, A.K., Edwards, J., Emmons, L.K., Fasullo, J., Garcia, R., Gettelman, A., Hannay, C., Holland, M.M., Large, W.G., Lauritzen, P.H., Lawrence, D.M., Lenaerts, J.T.M., Lindsay, K., Lipscomb, W.H., Mills, M.J., Neale, R., Oleson, K.W., Otto-Bliesner, B., Phillips, A.S., Sacks, W., Tilmes, S., Kampenhout, L., Vertenstein, M., Bertini, A., Dennis, J., Deser, C., Fischer, C., Fox-Kemper, B., Kay, J.E., Kinnison, D., Kushner, P.J., Larson, V.E., Long, M.C., Mickelson, S., Moore, J.K., Nienhouse, E., Polvani, L., Rasch, P.J., Strand, W.G., 2020. The community earth system model version 2 (CESM2). J. Adv. Modelling Earth Syst. 12, e2019MS001916.

Davis, R., Yau, C.-Y., 2011. Comments on pairwise likelihood in time series models. Statist. Sinica 21, 255–277.

Eidsvik, J., Shaby, B., Reich, B., Wheeler, M., Niemi, J., 2013. Estimation and prediction in spatial models with block composite likelihoods. J. Comput. Graph. Statist. 23, 295–315.

Feng, X., Zhu, J., Lin, P., Steen-Adams, M., 2014. Composite likelihood estimation for models of spatial ordinal data and spatial proportional data with zero/one values. Environmetrics 25 (8), 571–583.

Flury, R., Gerber, F., Schmid, B., Furrer, R., 2021. Identification of dominant features in spatial data. Spatial Stat. 41, 100483.

Fowler, H.J., Blenkinsop, S., Tebaldi, C., 2007. Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. Int. J. Climatol. 27, 1547–1578.

Franco-Villoria, M., Ignaccolo, R., 2017. Bootstrap based uncertainty bands for prediction in functional kriging. Spatial Stat. 21, 130–148.

Fronterrè, C., Giorgi, E., Diggle, P., 2018. Geostatistical inference in the presence of geomasking: A composite-likelihood approach. Spatial Stat. 28, 319–330.

Furrer, R., Flury, R., Gerber, F., 2021. Spam: Sparse matrix. R package version 2.7-0. https://CRAN.R-project.org/package=spam.
Furrer, R., Genton, M.G., Nychka, D., 2006. Covariance tapering for interpolation of large spatial datasets. J. Comput. Graph. Statist. 15, 502–523.
Furrer, E.M., Katz, R.W., Walter, M.D., Furrer, R., 2010. Statistical modeling of hot spells and heat waves. Clim. Res. 43, 191–205.
Furrer, R., Knutti, R., Sain, S.R., Nychka, D.W., Meehl, G.A., 2007a. Spatial patterns of probabilistic temperature change projections from a multivariate Bayesian analysis. Geophys. Res. Lett. 34 (L06711).
Furrer, R., Sain, S.R., 2009. Spatial model fitting for large datasets with applications to climate and microarray problems. Stat. Comput. 19, 113–128.
Furrer, R., Sain, S.R., 2010. Spam: A sparse matrix R package with emphasis on MCMC methods for Gaussian Markov random fields. J. Stat. Softw. 36, 1–25.
Furrer, R., Sain, S.R., Nychka, D.W., Meehl, G.A., 2007b. Multivariate Bayesian analysis of atmosphere-ocean general circulation models. Environ. Ecol. Stat. 14, 249–266.
Gerber, F., Furrer, R., 2019. Optimparallel: An R package providing a parallel version of the L-BFGS-B optimization method. R J. 11, 352–358.
Guinness, J., Fuentes, M., 2016. Isotropic covariance functions on spheres: Some properties and modeling considerations. J. Multivariate Anal. 143, 143–152.
Güsewell, S., Furrer, R., Gehrig, R., Pietragalla, B., 2017. Changes in temperature sensitivity of spring phenology with recent climate warming in Switzerland are related to shifts of the preseason. Global Change Biol. 23, 5189–5202.
Güsewell, S., Pietragalla, B., Gehrig, R., Furrer, R., 2018. Representativeness of stations and reliability of data in the Swiss Phenology Network. Technical report, MeteoSwiss, nr, p. 267.
Heagerty, P., Lele, S., 1998. A composite likelihood approach to binary spatial data. J. Amer. Statist. Assoc. 93, 1099–1111.
Heaton, M.J., Datta, A., Finley, A.O., Furrer, R., Guinness, J., Guhaniyogi, R., Gerber, F., Gramacy, R.B., Hammerling, D., Katzfuss, M., Lindgren, F., Nychka, D.W., Sun, F., Zammit-Mangion, A., 2019. A case study competition among methods for analyzing large spatial data. J. Agric. Biol. Environ. Stat. 24, 398–425.
Hengl, T.T., Pebesma, E.E., Hijmans, R.R.J., 2015. Spatial and spatio-temporal modeling of meteorological and climatic variables using open source software. Spatial Stat. 14, 1–3.
Heyde, C., 1997. Quasi-Likelihood and Its Application: A General Approach To Optimal Parameter Estimation. Springer, New York.
Hong, Y., Abdulah, S., Genton, M.G., Sun, Y., 2021. Efficiency assessment of approximated spatial predictions for large datasets. Spatial Stat. 43, 100517.
Houghton, J.T., Ding, Y., Griggs, D.J., Noguer, M., van der Linden, .P.J., Dai, X., Maskell, K., Johnson, C.A., 2001. Climate Change 2001: The Scientific Basis. Cambridge University Press.
Hurrell, J.W., Kushnir, Y., Ottersen, G., Visbeck, M., 2013. An Overview of the North Atlantic Oscillation. American Geophysical Union, pp. 1–35.
Jeong, J., Jun, M., 2015. A class of Matérn-like covariance functions for smooth processes on a sphere. Spatial Stat. 11, 1–18.
Joe, H., Lee, Y., 2009. On weighting of bivariate margins in pairwise likelihood. J. Multivariate Anal. 100, 670–685.
Jones, M.C., Pewsey, A., 2009. Sinh-arcsinh distributions. Biometrika 96 (4), 761–780.
Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woolen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K.C., Ropelewski, C., Wang, J., Leetma, A., Reynolds, R., Jenne, R., Joseph, D., 1996. The NCEP/NCAR 40-year reanalysis project. Am. Meteorol. Soc. Bull. 77, 437–471.
Kaufman, C.G., Schervish, M.J., Nychka, D.W., 2008. Covariance tapering for likelihood-based estimation in large spatial data sets. J. Amer. Statist. Assoc. 103, 1545–1555.
Kaufman, C.G., Shaby, B.A., 2013. The role of the range parameter for estimation and prediction in geostatistics. Biometrika 100, 473–484.
Kleiber, W., Nychka, D.W., 2015. Equivalent kriging. Spatial Stat. 12, 31–49.
Li, L., 2021. Geospatial constrained optimization to simulate and predict spatiotemporal trends of air pollutants. Spatial Stat. 45, 100533.
Li, F., Sang, H., 2018. On approximating optimal weighted composite likelihood method for spatial models. Stat 7, e194.
Lie, H.S., Eidsvik, J., 2021. Inference in cylindrical models having latent markovian classes with an application to ocean current data. Spatial Stat. 41, 100497.
Lindsay, B., 1988. Composite likelihood methods. Contemp. Math. 80, 221–239.
Lindsay, B.G., Yi, G.Y., Sun, J., 2011. Issues and strategies in the selection of composite likelihoods. Statist. Sinica 21, 71–105.
Mardia, K.V., Marshall, R.J., 1984. Maximum likelihood estimation of models for residual covariance in spatial regression. Biometrika 71, 135–146.
Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J., Maycock, T., Waterfield, T., Yelekçi, O., Yu, R., Zhou, B. (Eds.), 2021a. Climate Change 2021: The Physical Science Basis. Contribution of Working Group I To the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Chapter IPCC, 2021: Summary for Policymakers. Cambridge University Press, (in press).
Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J., Maycock, T., Waterfield, T., Yelekçi, O., Yu, R., Zhou, B. (Eds.), 2021b. IPCC, 2021: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I To the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, (in press).

Matérn, B., 1986. Spatial Variation, second ed. Springer-Verlag.
Meehl, G.A., 2019. The Coupled Model Intercomparison Project (CMIP) and interface with IPCC. In: AGU Fall Meeting, WCRP40, San Francisco.
Meehl, G.A., Boer, G.J., Covey, C., Latif, M., Stouffer, R.J., 1997. Intercomparison makes for a better climate model. Eos 78 (445).
Meehl, G.A., Covey, C., Delworth, T., Latif, M., McAvaney, B., Mitchell, J.F.B., Stouffer, R.J., Taylor, K.E., 2007. THE WCRP CMIP3 multimodel dataset: A new era in climate change research. Bull. Am. Meteorol. Soc. 88, 1383–1394.
NASA, 2021. Data processing levels. https://earthdata.nasa.gov/collaborate/open-data-services-and-software/data-information-policy/data-levels, (Accessed 31 October 2021).
Nash, J.C., 2014. Nonlinear Parameter Optimization Using R Tools. Wiley.
Nychka, D., Hammerling, D., Krock, M., Wiens, A., 2018. Modeling and emulation of nonstationary Gaussian fields. Spatial Stat. 28, 21–38.
Pace, L., Salvan, A., Sartori, N., 2019. Efficient composite likelihood for a scalar parameter of interest. Stat 8, e222.
Paciorek, C., Lipshitz, B., Zhuo, W., Prabhat, C.G., Thomas, R., 2015. Parallelizing Gaussian process calculations in R. J. Stat. Softw. 63, 1–23.
Pascoe, C., Lawrence, B.N., Guilyardi, E., Juckes, M., Taylor, K.E., 2020. Documenting numerical experiments in support of the Coupled Model Intercomparison Project phase 6 CMIP6. Geosci. Model Dev. 13, 2149–2167.
Pathakoti, M., Santhoshi, T., Aarathi, M., Mahalakshmi, D.V., Kanchana, A.L., Srinivasulu, J., Raja Shekhar, S.S., Soni, Vijay Kumar, Sesha Sai, M.V.R., Raja, P., 2021. Assessment of spatio-temporal climatological trends of ozone over the Indian region using machine learning. Spatial Stat. 43, 100513.
Poggio, L., Gimona, A., 2015. Downscaling and correction of regional climate models outputs with a hybrid geostatistical approach. Spatial Stat. 14, 4–21.
R. Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org.
Sain, S.R., Furrer, R., Cressie, N., 2011. A spatial analysis of multivariate output from regional climate models. Ann. Appl. Stat. 5, 150–175.
Salvaña, M.L.O., Genton, M.G., 2020. Nonstationary cross-covariance functions for multivariate spatio-temporal random fields. Spatial Stat. 37, 100411.
Schmidt, A.M., Guttorp, P., 2020. Flexible spatial covariance functions. Spatial Stat. 37, 100416.
Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M., Miller, H.L. (Eds.), 2007. Climate Change 2007 – the Physical Science Basis: Working Group I Contribution To the Fourth Assessment Report of the IPCC. Cambridge University Press, Cambridge.
Stein, M.L., 1999. Interpolation of Spatial Data. Springer-Verlag.
Stein, M.L., 2013. Statistical properties of covariance tapers. J. Comput. Graph. Statist. 22, 866–885.
Stein, M., Chi, Z., Welty, L., 2004. Approximating likelihoods for large spatial data sets. J. Royal Stat. Soc. B 66, 275–296.
Taylor, K.E., Juckes, M., Balaji, V., Cinquini, L., Denvil, S., Durack, P.J., Elkington, M., Guilyardi, E., Kharin, S., Lautenschlager, M., Lawrence, B., Nadeau, D., Stockhause, M., 2018. CMIP6 Global Attributes, DRS, Filenames, Directory Structure, and CV's. Technical report, 10 September 2018 (v6.2.7), https://goo.gl/v1drZl.
Trenberth, K.E., 1997. The definition of El Niño. Bull. Am. Meteorol. Soc. 78, 2771–2778.
Varin, C., Reid, N., Firth, D., 2011. An overview of composite likelihood methods. Statist. Sinica 21, 5–42.
Wackernagel, H., 2006. Multivariate Geostatistics, third ed. Springer-Verlag, New York.
Waller, L., Carlin, B., 2010. Disease mapping. In: Gelfand, A.E., Diggle, P.J., Fuentes, M., Guttorp, P. (Eds.), Handbook of Spatial Stat.. Taylor & Francis., Chapter 14.
Wendland, H., 1995. Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. Adv. Comput. Math. 4, 389–396.
Wendland, H., 1998. Error estimates for interpolation by compactly supported radial basis functions of minimal degree. J. Approx. Theory 93, 258–272.
Wilby, R.L., Charles, S.P., Zorita, E., Timbal, B., Whetton, P., Mearns, L.O., 2004. Guidelines for Use of Climate Scenarios Developed from Statistical Downscaling Methods. IPCC Data Distribution Centre.
Xu, G., Genton, M.G., 2016. Tukey max-stable processes for spatial extremes. Spatial Stat. 18, 431–443.
Xu, G., Genton, M.G., 2017. Tukey g-and-h random fields. J. Amer. Statist. Assoc. 112, 1236–1249.