Spatially adaptive non-stationary covariance functions via spatially adaptive spectra

BY A. Pintore & C. Holmes

Department of Statistics, University of Oxford, Oxford, OX1 3TG, U.K.
cholmes@stats.ox.ac.uk

Summary

We describe a framework for constructing non-stationary covariance functions for use in modelling spatial Gaussian processes. The main aim of this paper is to highlight the use of the spectral domain as an efficient representation for building non-stationarity. We consider two possible decompositions of a stationary covariance function, namely the Fourier and Karhunen-Loève expansions. Non-stationarity is introduced by evolving the stationary spectrum over the space in such a way that the resulting non-stationary models are both valid and interpretable in terms of the original covariance function but now with local parameters. Moreover, we show that these models have an equivalent basis functions representation for which the variance of the coefficients varies over the space.

Some key words: Gaussian Processes, Fourier Analysis, Karhunen-Loève expansion, Kriging, Non-stationary covariance functions, Stationarity, Spatial Smoothing, Variance-varying coefficients models.

1 Introduction

Geostatistical modelling of Gaussian data is an important task in many branches of the Earth and Life Sciences. For instance, in areas as diverse as epidemiology, ecology, petroleum engineering and oceanography there is a need to develop statistical techniques that can model the dependence between observations of a process taken at different spatial locations. References can be found in Cressie (1993). Most current methods assume stationarity in the process which is equivalent to stating that the statistical association between two points is solely a function of the vector distance that separates them. Stationarity is usually adopted for reasons of mathematical convenience and is an idealisation that rarely holds in practise, non-stationarity being a more realistic assumption.
Indeed, in many applications, characteristics of the field under study such as roads, mountains, geological faults and rivers can induce local dependencies in the process that cause the dependence between points to vary with their location as well as their distance apart. However, although it is clear that non-stationarity is often present, modelling such non-stationarity is technically difficult, as we now discuss.

It is usual to consider a decomposition of the observable process as,

\[ Z(s) = \mu(s) + \epsilon(s), \]  

where \( Z(s) \) is the observed process at location \( s \in D \subseteq \mathbb{R}^2 \), \( \mu(s) \) is a deterministic trend component corresponding to “large-scale” variations in the process and \( \epsilon(s) \) is a zero-mean Gaussian stochastic component corresponding to “small-scale” variations. The process \( \epsilon(s) \) is then characterised by its second-order structure (i.e. its covariance function say). It is important to note that as underlined in Cressie (1989), the decomposition into trend and process is typically unidentifiable. These distinctions are further confounded in a Bayesian setting where \( \mu(s) \) is also random. It is clear from (1) that non-stationarity could be introduced via the mean, the stochastic component or both.

There are various ways to handle the modelling of a non-constant mean such as universal kriging, median-polish kriging, differencing or splines and we refer the interested reader to Cressie (1993) for a description of some of these. In this paper, we focus on modelling non-stationarity in the stochastic component of the process. The reason for this choice lies in the fact that in geostatistics, the interest is often in describing the dependencies between observations rather than the mean function per se.

We begin our outline with a brief description of the main methods used to tackle non-stationarity in the stochastic component of the process. An important feature of the various methods that have been developed up to now is that they build non-stationarity directly into the covariance function. This however is very challenging since a major hurdle is the necessity that any locally specified model must guarantee a globally valid positive-definite covariance structure for the data, and moreover from a modelling perspective we wish to ensure that the covariance models remain easily interpretable.

We will argue in this paper that building non-stationary covariance functions is much simpler and more explicit if one works in the spectral domain. Here, we consider the term “spectral” in
a broad sense, that is, in terms of the decomposition of the covariance function in an orthogonal basis. The spectral representation offers an equivalent representation for the covariance function in characterising a second-order stationary process. We hope to demonstrate the usefulness and effectiveness of this representation for tackling non-stationarity by underlying its two major strengths. Firstly, it allows one to introduce non-stationarity as a natural generalisation to the stationary case while ensuring valid models. Secondly, the spectral density function is explicit in defining the smoothing properties characterising the model as well as its non-stationary structure.

We introduce non-stationarity by evolving the spectrum over the space, thus allowing for spatial adaptiveness in the covariance structure of the process under study. Although the method we suggest is similar in spirit to the evolutionary spectrum of Priestley (1965), it nevertheless represents a generalisation in a different direction. Two cases are of interest.

- if the spectrum is of parametric form, $f(\omega, \theta)$, where $\theta$ is a parameter or set of parameters that characterises the properties of the spectrum, we evolve the spectrum in space by evolving $\theta$, thus defining a spatially adaptive spectrum $f(\omega, s; \theta(s))$ at each location $s$.

- if the spectrum is of non-parametric form, $f(\omega)$ say, we propose to evolve it over space by tempering, that is taking it to some power, $\{f(\omega)\}^{\eta(s)}$ at each location $s$, where $\eta(s)$ is some smooth function of $s$. This has an analogy with heating or cooling the process which in effect allows us to give more or less power to higher or lower frequencies. This idea is related to tempering of probability density functions (Geyer and Thompson 1995).

It is interesting to note that for common parametric forms for $f(\omega, s; \theta(s))$, such as the Matèrn or Gaussian for instance, the two approaches are similar in spirit although not quite equivalent. We will apply these ideas to different spectral representations of the covariance function and show that the resulting non-stationary models that we obtain are of the same form as their stationary counterparts. Our modelling methodology can be written as a four step procedure,

- Construct a best fit stationary model to the data.

- Perform a spectral decomposition of the resulting covariance matrix or function, using the Karhunen-Loève expansion or its Fourier representation.

- Allow the model to evolve in space by evolving the spectrum locally.
• Transfer back to the covariance function domain to obtain a valid non-stationary model of the same form as the stationary one but now with local parameters.

An important feature of our modelling procedure is that it has an equivalent basis functions representation

\[ Z(s) = \sum_{i=1}^{N} \phi_i(s) \beta_i, \ s \in D, \]  

where \( \phi_i(s), i = 1, ..., N \) is a set of basis functions, and where we allow the variance of the random coefficients to evolve over space, that is \( Var(\beta_i) = \sigma^2(s) \) under the condition that the correlation between \( \beta_i(s) \) and \( \beta_j(s') \) is equal to 1 for \( i = j \) and to 0 otherwise, for all \( s, s' \in D \) and all \( i, j = 1, ..., N \). We refer to this basis functions representation as \textit{variance-varying coefficients models}. These models induce a locally adaptive shrinkage of the basis coefficients and as such represent an alternative to varying-coefficients models, which were first introduced by O'Hagan (1978) and further developed by Hastie and Tibshirani (1993) and Gelfand et al. (2003). In this paper, we will limit ourselves to the basis functions induced by our spectral representations but the generalisation to any set of basis functions will however be clear from our derivation.

The rest of the paper is as follows. Section 2 gives a short overview of the current methods used in spatial non-stationary modelling of covariance functions. Section 3 presents two possible and widely used “spectral” representations of stationary covariance functions and Section 4 introduces our method for spatially evolving parametric spectra. We derive analytical expressions for the spatially localised generalisations of the most common used geostatistical covariance functions. In Section 5, we present the spectral tempering method which allows the non-parametric spectra to evolve overspace. Section 6 gives an equivalent basis functions representation with variance-varying coefficients of the spectral tempering approach used for non-parametric spectra. Section 7 shows some properties and simulations of the models induced by our method. Section 8 deals with the issue of prediction for representations of covariance functions using analytic and non-analytic basis functions. Section 9 gives detail on the modelling and estimation procedure for the latent spatial process used to evolve the spectrum overspace. We consider a two-dimensional real data application in Section 10 while Section 11 compares the relative merits of the two spectral representations used in this paper. We conclude with a Discussion raising some issues to be addressed in further research.
2 Non-stationary models

In this section, we give a brief description of the main methods used to tackle non-stationarity via the stochastic component $\epsilon(s)$ and how they compare to the method described in this paper. A good overview of some of these methods is also given in Sampson et al. (2001).

In the first instance it is useful to partition the methods into two categories corresponding to two different types of problems they address: firstly, those where only a single realisation of the process is available at each location (as is common in geostatistical applications) and secondly where more than one realisation is available (as is common for instance in many meteorological applications). This difference is of great importance since in the second case one can estimate the empirical (possibly non-stationary) covariance function, which provides the starting point for most methods in this class. The most common techniques using multiple realisations to model non-stationarity include the Empirical Orthogonal Functions approach (Cohen and Jones (1969), Creutin and Obled (1982), Obled and Creutin (1986)) and some extensions (Nychka and Saltzman (1998), Holland et al. (1999)), the Deformation approach initially presented by Sampson and Guttorp (1992) and further studied and developed by Guttorp et al. (1994), Smith (1996), Meiring et al. (1997) and Damian et al. (2001) and a multiple realisations kernel-based method due to Nott and Dunsmuir (2002). The underlying idea behind all these methods is to obtain a valid parametric model enabling one to retrieve the initial empirical covariance matrix (or the empirical dispersions between measurement points) obtained from the data, or some smooth version of it. One is then able to evaluate the covariance function between any two points outside of the measurement set. For instance, in the Empirical Orthogonal Functions approach one usually starts by computing the Karhunen-Loève expansion of the empirical (non-stationary) covariance matrix and, in the Deformation approach, one uses a mapping to fit a parametric model to the empirical dispersions. Other methods include the smoothing of the empirical covariance matrix due to Loader and Switzer (1992) and Oeflert (1993), and a fully Bayesian model initially described by Le and Zidek (1992). Note that for these latter methods, the mean is modelled via a parametric function as part of the model. A drawback of these methods is that they are usually unable to tackle applications where only single realisations of the process are available.

The second approach is sometimes termed “geostatistical” in the sense that it models non-stationarity when only a single realisation of the process is measured at each site location. Hence,
Unlike the previous methods, it can not make use of an initial empirical covariance matrix, which in a sense renders modelling more challenging. There are then essentially two philosophies here. The first type of methods are based on the assumption that locally some data points can be gathered together and can be assumed to follow a stationary process. The methods using this principle include the Moving window approach (Haas (1990), Haas (1995)), partition models approach (Denison et al. (2002)) and the mixture of Gaussian Processes approach (Rasmussen and Ghahramani (2002)). Note that these methods usually model the mean and the covariance structure of the stochastic part of the process simultaneously. These methods are natural as they appear to be a straightforward generalisation of localised kriging applied to non-stationary data. One advantage of this compared with standard kriging is that they are less expensive computationally. They can also be adjusted to model spatio-temporal data. However, they suffer from a major drawback in that they do not generally allow for the modelling of the covariance structure over the whole space. So that a full representation of the process in the form (1) is not possible and joint forecasts at arbitrary points are difficult to obtain.

Our approach is able to build a valid non-stationary covariance function for the whole space from a single realisation of the process. As such it should be compared with the existing methods of Higdon et al. (1999) and Fuentes and Smith (2002). The main idea behind these single realisation kernel-based methods is the estimation of some latent spatial stationary process(es) or parameters which are then used to introduce non-stationarity into the model. We now give further insight into these methods.

The method of Higdon et al. (1999) is based on a result, given in Priestley (1965), that states that if the correlation function between two points \( s_1, s_2 \) can be written,

\[
\rho(s_1, s_2) = \int_{\mathbb{R}^2} k_{s_1}(u)k_{s_2}(u)du,
\]

where \( k_s(u) \) denotes a kernel centred at location \( s \) whose shape is a function of \( s \), then the resulting (zero-mean) non-stationary process can be represented as a convolution,

\[
Z(s) = \int_{\mathbb{R}^2} k_s(u)x(u)du,
\]

where \( x(s) \) is a white noise process, provided \( k_s(u), u \in \mathbb{R}^2 \) is square-integrable for all \( s \). Higdon et al. (1999) model \( k_s(u) \) as an unknown function in terms of parameters which can then be estimated in a hierarchical Bayesian framework. These parameters are modelled using spatial stationary
processes in order to let the kernels evolve smoothly over space. Hence, one has to estimate the values of a number of latent spatial stationary processes in order to compute the moving average underlying the model. This is how non-stationarity is introduced into the model. This method has many nice features. First, in the case of Gaussian kernels one can obtain an analytical expression for the covariance function (3) over the whole space. Another advantage lies in the use of a hierarchical Bayes model which allows one to measure the departure from the stationary assumption. Finally, the method is able to capture spatially varying local anisotropy in the covariance structure of the process. However, the exposition in Higdon et al. (1999) relies fairly heavily on the Gaussian assumption for the form of the kernels and it is unsure how constraining this is for obtaining an exact analytical form, rather than a numerical approximation, for the covariance structure of the process.

Another single realisation kernel-based method is due to Fuentes and Smith (2002). In this method, which is a extension of the finite decomposition method of Fuentes (2001a), the spatial process is represented as a convolution of stationary processes, that is

$$Z(s) = \int_{D} k(s - x)Z_{\theta(x)}(s)d\mathbf{x}, \ s \in D$$

(5)

where $D$ is the domain over which the process is defined, $k$ is a stationary kernel function and \( \{Z_{\theta(x)}(s)\} \) is a family of independent stationary Gaussian processes indexed by a spatial field of parameter vectors $\theta(x)$, which is constrained to vary smoothly over the space. The parameters of the model are estimated using a Bayesian spatial estimation framework. This method provides a nice theoretical framework for modelling non-stationary geostatistical data. There are some open issues however. Firstly the non-stationary covariance function given by (5) can be difficult to evaluate analytically if not intractable even if simple assumptions are taken for the form of the kernels, such as of Gaussian type for instance. Also, this method appears to only model spatially varying locally isotropic models.

With our modelling approach, we are able to obtain analytical expressions for spatially varying locally anisotropic non-stationary covariance functions at the measurements sites. Moreover, although our method is “geostatistical” since it only requires one observation at each location, it is also able to make use of repeated measurements as the methods of the first type described previously. This is a feature of the possible use of the Karhunen-Loève representation. We give some
detail in Section 11. Finally, our model is interpretable leading to localised covariance functions, since we work with the generalised spectral decomposition of the process under study.

3 Two common spectral representations

In this paper, we shall consider the two most well known examples of spectral decomposition of covariance functions used in geostatistics, namely the Fourier representation (i.e. expansion in the basis formed by the complex exponential functions) and the Karhunen-Loève expansion. This last expansion is the main tool in the Empirical Orthogonal Functions methods and could be easily generalised to other choices of basic functions (e.g. wavelets). We now study these two expansions in greater detail.

3.1 the Fourier spectral representation of second-order stationary stochastic processes

From here on, we take \( D \) to denote equivalently \( \mathbb{R} \) and \( \mathbb{R}^2 \). Note that the numerical examples and corresponding plots will be given for \( D = \mathbb{R} \) only for ease of illustration, but all the results and conclusions remain true in higher dimensions.

It is well known from Khintchine’s theorem and Bochner’s characterisation of positive definite functions (Levy 1965) that the covariance function of any second order stationary stochastic process \( X(t), t \in D \) has a representation of the form

\[
C(s, t) = \int_{\Omega} \exp[i\omega(s - t)]d\mu(\omega), \quad s, t \in D, \quad \omega \in \Omega = \mathbb{R},
\]

where the measure \( \mu(\omega) \) has the properties of a possibly unnormalised probability distribution function. It can be shown (Loève 1955) that there is an equivalence between (6) and the representation of the process \( X(t), t \in \mathbb{R} \) as an expansion in terms of complex exponential via a Fourier-Stieltjes integral of the form

\[
X(t) = \int_{\Omega} \exp[i\omega t]dZ(\omega), \quad t \in \mathbb{R},
\]
with $E[dZ(\omega)\overline{dZ(\omega')}] = \delta(\omega, \omega') d\mu(\omega)$, where the function $\delta(\omega, \omega')$, $\omega, \omega' \in \Omega$ is defined as

$$
\delta(\omega, \omega') = \begin{cases} 
0 & \text{if } \omega \neq \omega' \\
1 & \text{o.w.}
\end{cases}
$$

The measure $\mu(\omega)$ is called the integrated spectrum. Moreover, if the measure $\mu(\omega)$ is absolutely continuous with respect to the Lebesgue measure then $d\mu(\omega) = f(\omega)d\omega$ and $f(\omega)$ is called the spectral density function or spectrum of the process $X(t)$, $t \in D$. We will assume that the measure is absolutely continuous from here on.

In geostatistics one usually adopts a second-order stationary model with a parametric form for the covariance function $C$, such as of the Gaussian or Matérn type (Cressie 1993). This reduces the modelling task to simply estimating the parameters of the covariance function. It is important to note that most common parametric forms or models for $C$ correspond to different parametric expressions for the spectral density function $d\mu(\omega)$ of the underlying process. Moreover, the smoothing induced by a particular covariance model is made explicit by the parameters parameterising the spectral density function which helps to describe the properties of the corresponding second-order spatial structure. To illustrate this, it is instructive to consider the parametric spectral density functions corresponding to some well-known covariance functions used throughout the spatial statistics literature, and more recently in machine learning, e.g. Neal (1999).

To begin, consider the well-known N-dimensional Gaussian correlation function

$$
C(s, t) = \exp[-\alpha||s - t||^2], \quad s, t \in \mathbb{R}^N, \quad N \geq 1, \quad \alpha > 0,
$$

where $\alpha$ is a global smoothing parameter to be estimated. The corresponding N-dimensional spectral distribution function is given by

$$
f(\omega) = (4\pi \alpha)^{-N/2} \exp[-\omega^2/4\alpha], \quad \omega \in \Omega = \mathbb{R}^N.
$$

In Figure 1 we plot $f(\omega)$, (9), in the one-dimensional case (i.e. N=1) at a 100 points uniformly spaced on [0, 10] for $\omega > 0$ and for different values of $\alpha$.

As can be seen in Figure 1, the Gaussian covariance function (8) gives decreasing weights to the higher frequency components in the process, thus smoothing the latter. This is clear from the analytical expression of $f(\omega)$ given in (9) which shows that the frequency weight tends to zero as
ω → ±∞. The smoothing parameter α in (8) governs the shape of the spectral density function in (9) and hence the behaviour of the process, such that the larger α is, the more weight is given to high frequency components relative to the low frequency ones. In this way the parameter α governs the amount of smoothing induced by the model and the spectral representation (9) makes precise this interpretation.

Another popular class of covariance functions is the Matérn class given by

\[ C(s,t) = \frac{\pi^{N/2}}{2^{\nu-1} \Gamma(\nu + N/2)} e^{\nu(s-t)\|s-t\|^\nu} K_\nu(c\|s-t\|_N), \quad s, t \in \mathbb{R}^N, N \geq 1, \]

where the function \( K_\nu \) is the modified Bessel function of the second kind of order \( \nu \) (see Stein (1999) or Abramowitz and Stegun (1965) for more details). The \( N \)-dimensional spectral distribution function corresponding to (10) is somewhat simpler,

\[ f(\omega) \propto (c^2 + \|\omega\|^2)^{-\nu-N/2}, \quad \omega \in \Omega = \mathbb{R}^N, N \geq 1, \quad c > 0, \quad \nu > 0, \]

where the parameter \( c^{-1} \) can be considered as the autocorrelation range and the parameter \( \nu \) governs the degree of smoothness of the process. From (11) it is clear that the greater the value of \( \nu \), the less weight is given to the higher frequency components of the spectrum, and thus the smoother the process is. This can also be seen in Figure 2, where we have plotted the one-dimensional spectral distribution function (i.e. \( N=1 \)) \( f(\omega) \), (11), at a 100 points uniformly taken on \([0,10]\) for \( \omega > 0 \) and for different values of \( \nu \).
Figure 2: Plot of the Matérn spectral density function $f(\omega)$ on 100 points taken uniformly on $[0,10]$ for $c = 1$, $\nu = 0.1$ (*), $\nu = 1$ (squares) and $\nu = 10$ (•), Section 3.1.

Note that the exponential correlation function is a Matérn covariance function with $\nu$ equal to 0.5. It is interesting to compare Figures 1 and 2 which make explicit the consequences of adopting one particular model over another.

3.2 the Karhunen-Loève expansion

We now consider another important expansion that has been used successfully throughout the geostatistics literature. As stated in Cohen and Jones (1969), for a process with a given covariance function there exists a unique orthogonal expansion of that process for which the coefficients are uncorrelated. This decomposition is known as the Karhunen-Loève expansion and the generalised spectral representation of the process is then:

$$Z(s) = \sum_{i=1}^{\infty} \alpha_i \phi_i(s)$$  \hspace{1cm} (12)

where the deterministic functions $\phi_i(s)$ and the uncorrelated random variables $\alpha_i$ satisfy

$$\int \phi_i(s)\phi_j(s)ds = \delta_{ij},$$  \hspace{1cm} (13)

$$E[\alpha_i] = 0$$  \hspace{1cm} (14)

$$E[\alpha_i \alpha_j] = \delta_{ij} \lambda_i,$$  \hspace{1cm} (15)

and where $\delta_{ij}$ is the usual Kronecker function $\delta$. Equation (12) is known as the Karhunen-Loève spectral representation of the process $Z(s)$. The functions $\phi_k(s)$, $k = 1, \ldots$ are the solutions of the
following homogeneous Fredholm integral equation of the second kind,
\[ \int C(s_i, s_j) \phi_k(s_j) ds_j = \lambda_k \phi_k(s_i). \] (16)

The corresponding value, \( \lambda_k \), specifies the variance of the random coefficient \( \alpha_k \). Hence, \( \lambda_k \) quantifies the power of the process at the \( k^{th} \) scale. From (16) it is seen that the functions \( \phi_k(s) \), \( k = 1, \ldots, n, \ldots \) are eigenfunctions and the values \( \lambda_k \), \( k = 1, \ldots, n, \ldots \) their corresponding eigenvalues. For distinct eigenvalues the representation is unique. It follows from (12)-(15) that the covariance function can be written in the following form
\[ C(s_i, s_j) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s_i) \phi_k(s_j). \] (17)

It is important to point out that the eigenvalues \( \lambda_k \) are usually a non-parametric function of \( k \), since no close expression is available for them, except in very rare cases. Moreover, this is always true in the corresponding discrete case which we now consider.

In the discrete case, with a finite set of data points, the Karhunen-Loève expansion is equivalent to a principal component analysis and the Fredholm equation, (16), is the analog of the matrix eigenvector equation. Hence, if one computes the single-value decomposition of the matrix \( C \) in (17), we find
\[ C = V D V', \] (18)

where \( D \) is a diagonal matrix of eigenvalues \( \lambda_i \) and \( V = (v_{i,j})_{i,j=1,\ldots,n} \) is the matrix whose columns form an orthonormal basis of the \( n \)-dimensional space spanned by our finite set of data points. We will assume from here on that the eigenvalues are ordered, so that: \( \lambda_i \geq \lambda_{i+1}, \ i = 1, \ldots, n. \) We also standardise \( V \) and \( D \) so that \( \lambda_1 = 1. \) We are now able by analogy with (12) to represent the spatial process \( Z(s) \) in basis function form,
\[ Z(s_i) = \sum_{j=1}^{n} \alpha_j v_{i,j} + \epsilon_i, \quad i = 1, \ldots, n \] (19)

where the random coefficients \( \alpha_j \) are normally distributed as,
\[ \alpha_j \sim N(0, \lambda_j). \] (20)

and where the \( \epsilon_i \), \( i = 1, \ldots, n \) are independent and identically distributed with
\[ \epsilon_i = \epsilon(s_i) \sim N(0, \gamma^2) \quad i = 1, \ldots, n. \] (21)
The expansion obtained for the covariance matrix is then given by:

\[
C_{i,j}^S = \sum_{k=1}^{n} \lambda_k v_{i,k} v_{j,k} + \gamma^2 \delta_{i,j}, \quad i, j = 1, \ldots, n
\]

(22)

where \( \sum_{k=1}^{n} \lambda_k v_{i,k} v_{j,k} \) is the expansion of the correlation matrix \( C_{i,j} \). The eigenvalues obtained from the numerical signalisation of the matrix \( C \) are estimates of the generalised spectrum (17) and the eigenvectors are estimates of the eigenfunctions \( \phi_i(s) \) at the locations given by the data points.

Note that unlike the Fourier representation of the covariance function given in the previous subsection, which is an expansion in terms of fixed periodic functions, the functions in the Karhunen-Loève expansion depend on the actual values taken by the covariance function and are not always periodic. We will discuss this issue in more detail when we compare these two methods. In order to gain insight into this expansion, it is illustrative to plot the eigenvectors obtained from a “typical” stationary correlation matrix. Here we consider the Gaussian correlation matrix of (8), that is,

\[
C_{i,j} = C(s_i, s_j) = \exp(-\alpha ||s_i - s_j||^2) \quad i, j = 1, \ldots, n
\]

(23)

where \( || \cdot || \) is the usual Euclidean norm. In Figure 3 we plot the 1\(^{st}\), 5\(^{th}\) and 20\(^{th}\) eigenfunctions (eigenfields) for a one-dimensional example, corresponding to (23) with the value of \( \alpha \) taken to be 400. The number of locations used to plot Figure 3 is \( n = 150 \) and they are sampled regularly on \( s \in [-1, 1] \) in order to obtain relatively “regular” shapes for the eigenvectors corresponding to the biggest eigenvalues. This choice was made for illustration purposes only. More general shapes corresponding to irregularly spaced data will be considered in Section 8. Nevertheless, it is interesting to note that in the case of regularly sampled data points, the shape of the eigenvectors is very close to that of the cosine and sine functions used in the Fourier representation. This feature however does not remain true in more general cases.

A corresponding two-dimensional example is shown in Figure 4 and Figure 5 for the same covariance structure on a regular lattice. The value of \( \alpha \) is the same as before. Similarly to Figure 3, Figure 4 and Figure 5 demonstrate that the eigenvectors appear to be very localised in the frequency domain. That is, each eigenvector has a very precise frequency. Indeed, eigenvectors corresponding to the biggest eigenvalues capture smooth variations of the spatial structure whereas the eigenvectors corresponding to the smallest eigenvalues capture more high-frequency variations. In other words,
the large scale variations of the process are captured by the slowest varying eigenfields (i.e. large eigenvalues) and the small-scale ones are made up of the faster varying eigenfields (i.e. small eigenvalues). It is interesting to note the self-similarity between the 1st and 20th eigenfunctions of the spatial fields in Figure 4.

In Figure 6 we have plotted the normalised eigenvalues of the covariance matrix used in the previous one-dimensional example shown in Figure 3.

From Figures 6 and 1 the analogy between the Fourier representation and the Karhunen-Loève expansion is clear. Indeed, the eigenvalues $\lambda_k$ have a similar interpretation to the spectrum in the Fourier representation since they determine the relative power of the process at the $k^{th}$ scale. Indeed the effect of the monotone spectrum is to damp down the contribution of the faster varying eigenvectors relative to the slower varying ones hence reducing the influence of the higher frequency components. This is perhaps more clearly seen if we consider the basis function representation of the process (19)-(21) induced by the Karhunen-Loève expansion, that is,

$$Z = V \cdot \alpha + \epsilon,$$

(24)
Figure 5: Surface plots of the 1st, 5th and 20th eigenvectors in the two-dimensional case, Section 3.2.

Figure 6: Plot of the normalised eigenvalues for the one-dimensional example with $\alpha = 400$, Section 3.2.
where $Z = \{Z(s_1), ..., Z(s_n)\}$ and $V$ is the $n \times n$ matrix defined in (17) and

$$\alpha \sim MVN(0, D), \quad (25)$$

$$\epsilon \sim MVN(0, \sigma^2 I_n) \quad (26)$$

where $D$ is the $n \times n$ matrix of eigenvalues defined by (18) and $\sigma^2$ is some noise parameter. Note that (24)-(26) is equivalent to a Bayesian basis function model or to a Linear Mixed Model (McCulloch and Searle 2001), with distributions on random coefficients given by (25). This representation can be linked with the spectral Fourier representation of the process given by (7). This representation provides further insight into the role of the eigenvalues corresponding to finite-dimensional kernel defined by the matrix $D$. Indeed, if we consider the model given by (24)-(26) then the best linear unbiased predictor of $\alpha$ (i.e. the posterior mean in a Bayesian framework) is given by:

$$\hat{\alpha} = (\sigma^2 D^{-1} + V'V)^{-1} V'V \hat{\alpha}_{mle} \quad (27)$$

where $\hat{\alpha}_{mle}$ is the least-squares or maximum likelihood estimate (mle) of $\alpha$, and $V$ and $D$ are given by (18). In our case, $V'V = I_n$ since $V$ is an orthonormal matrix and hence, we obtain

$$\alpha_i = \frac{\lambda_i}{\sigma^2 + \lambda_i} \hat{\alpha}_{mle}, \quad i = 1, ..., n \quad (28)$$

where $\hat{\alpha}_i$ is the $i^{th}$ coefficient of $\alpha$. We therefore observe that the eigenvalues $\lambda_i$ lead to a shrinking of the classical estimate $\alpha_{mle}$ and that the shrinking for the $i^{th}$ eigenfield is proportional to $\lambda_i/(\sigma^2 + \lambda_i)$. As a consequence, since $\lambda/(\sigma^2 + \lambda)$ diminishes as the value of $\lambda$ becomes smaller, we can again interpret the influence of the eigenvalues in terms of damping down (or shrinking) the eigenvectors relative to one another. Moreover, it can be seen that the high frequency eigenvectors (i.e. those with small eigenvalues) receive greater damping relative to the smooth eigenvectors as shown in Figure 6.

The basis function representation (19)-(21) of the process is important since it clearly provides a link with other possible expansions. Indeed, the basis functions could be taken to be orthogonal polynomials, Demmler-Reinsch splines (Eubank 1999) or wavelets, which are very popular across the statistics literature.

It is important to point out that, although the Karhunen-Loève expansion and the Fourier representation of the covariance function have a similar interpretation, a major difference between
them lies in the fact that the spectrum resulting from the Karhunen-Loève expansion is non-parametric, unlike the main spectral density functions obtained in Fourier analysis. Thus different methods will have to be used when evolving it over space. This is also true for other basis functions representations.

Another difference between both methods, although of less importance in this article, lies in the fact that the Fourier representation of the covariance function depends crucially on the stationarity assumption of \( C \) whereas it plays no role in the Karhunen-Loève expansion, which is estimated from any valid \( C \). This is an important distinction, which we will come back to later.

Now that we have given some insight into the “spectral” representation of the process and its covariance function, we describe how it forms a useful device to model non-stationarity.

4 Direct spatial adaptation of parametric spectra

Stationary covariance functions have proved extremely popular in geostatistics. However, a major drawback is that they assume a constant degree of dependence between points in the space. That is, the spectral characteristics of the process do not vary and hence these models are unable to adapt to local changes in the behaviour of the process underlying the data. In the case of the Gaussian or the Matérn spectral density functions, this is reflected in the fact that the parameters of the covariance function, or equivalently of the spectrum, are fixed and not functions of location, while in the Karhunen-Loève expansion stationarity implies that the coefficients \( \lambda_k \) are the same at all location. Another way to illustrate this is by highlighting the fact that, in such a case, the random coefficients of the exponential basis functions in (7) or the Karhunen-Loève expansion in (12) are the same for all \( s \in \mathbb{R} \).

In this paper we propose to capture non-stationarity via a local spectral representation of the process where the local spectrum is now a function of location. Indeed, we have shown in the previous section that by controlling the damping on the spectral components, the spectrum determines the smoothness properties of the spatial model used. As a consequence, by allowing it to evolve over space, we allow for our model to spatially adapt to the local characteristics in the process. It has been pointed out in the previous section that the Karhunen-Loève and the Fourier spectra are fundamentally different in the sense that the former is non-parametric while the latter
is parametric, at least for the covariance models we are interested in.

In this section, we propose a method for allowing a parametric spectrum to evolve over space. The extension to non-parametric spectra follows as a natural generalisation of this approach for the most common used geostatistical covariance functions. We describe it in the next section.

First, recall that we assume throughout the paper the integrated spectrum to be absolutely continuous with respect to the Lebesgue measure so that the spectral density function of the processes we consider are well-defined. Given a parametric model $f(\omega; \theta)$ for the spectrum, where $\theta$ is a parameter or set of parameters governing the properties of the spectrum, we define at each location $s$,

$$f_{NS}^s(\omega) \propto f(\omega, s; \theta(s)), \ \forall \ s \in D, \ \forall \ \omega \in \Omega,$$

where $f_{NS}^s$ is the localised non-stationary (NS) spectral density of the process at $s \in D$. The constant of proportionality can be chosen, for example, to ensure that the overall power (i.e. variance) in the process remains constant at any location $s$. That is,

$$\int_{\Omega} f_{NS}^s(\omega) d\omega = \int_{\Omega} f(\omega, s; \theta(s)) d\omega = K, \ \forall \ s \in D.$$  \hspace{1cm} (30)

for some constant $K$ usually taken to be one.

More generally, model (29) leads to the following joint spectral density at locations $s, t \in D$,

$$f_{NS}^{st}(\omega) = f_{NS}(\omega, s, t; \theta(s), \theta(t)) = f_{NS}^s(\omega)^{1/2} f_{NS}^t(\omega)^{1/2} = f(\omega, s; \theta(s))^{1/2} f(\omega, t; \theta(t))^{1/2}, \ \forall \ \omega \in \Omega,$$

so that

$$f_{NS}^s(\omega) \equiv f_{NS}^{st}(\omega) = f(\omega, s; \theta(s)), \ s \in D.$$  \hspace{1cm} (32)

Clearly, if $\theta$ governs the smoothing induced by the stationary spectrum and $\theta(s) > \theta$ at location $s$, then the covariance structure at $s$ is smoother than its initial stationary counterpart, that is, we obtain a local spectrum with relatively more power in low frequencies. Similarly, if $\theta(s) < \theta$, the resulting covariance exhibits more variation than the stationary model. Moreover, if we consider standard forms such as (9) and (11), we might suspect that the resulting local spectral densities $f_{NS}^{st}(\omega)$ will be of the same parametric form. It turns out that this is the case for these models as shown in the next subsection.
It is important to highlight that the modelling procedure we propose here represents a generalisation of the usual spectral theory in a different direction to that of the evolutionary spectrum of Priestley (1965). Indeed, Priestley (1965) considers models of the form

\[ f_{K_S}(\omega) = A_s(\omega)f(\omega; \theta) \] (33)

rather than (29). These models are obtained by considering correlation functions between \( s_1, s_2 \) that can be written as

\[ \rho(s_1, s_2) = \int_{\Omega} \phi_{s_1}(\omega) \phi^*_{s_2}(\omega) d\mu(\omega) \] (34)

where \( \mu(\omega) \) is a measure, usually taken to be absolutely continuous with respect to the Lebesgue measure with density \( f(\omega) \) and the functions \( \phi_u(\omega) \) are of oscillatory form, that is,

\[ \phi_u(\omega) = A_u(\omega) \exp\{iu\} \] (35)

where conditions are imposed on \( A_u(\omega) \) for the model to be identifiable.

Concerning the validity of our model, we have the following result

**Lemma 4.1:** Let \( C(s,t), s, t \in D \) be a stationary covariance function whose spectrum is \( f(\omega) \). Then, the function on \( D \times D \) given by

\[ C_{KS}(s,t) = \int_{\Omega} \exp\{i\omega(s - t)\}f(\omega, s; \theta(s))^{1/2}f(\omega; t; \theta(t))^{1/2} d\omega \] (36)

is a valid covariance function if and only if

\[ \int_{\Omega} |f(\omega, s; \theta(s))| d\omega < \infty \] (37)

for all \( s \) in \( D \).

**Proof:** The proof of this result is given in Appendix A.1.

When modelling \( \theta(s) \) we demonstrate that the smoothness assumptions are made explicit and the resulting non-stationary model is interpretable. Moreover, it contains the stationary model as a special case and allows one to measure the departure from it at each location. Finally, it is clear from that our spectral approach for modelling non-stationarity is capable of capturing spatially varying local anisotropy features. For instance, consider a line (of direction \( h \)) passing through \( s \) such that \( \theta(h) \) is increasing along the line. Then, the covariance kernel centred at \( s \) will be asymmetric in the direction \( h \). We give some visualisations in Section 10. We now consider applications of this modelling procedure to some common used geostatistical covariance functions.
4.1 Non-stationary covariance functions obtained via spatial adaptation of the parametric Fourier spectrum

In this section, we give some examples of \( N \)-dimensional non-stationary correlation functions that are obtained from some common stationary correlation functions used throughout the spatial statistics literature. Moreover, we demonstrate that the non-stationary correlation functions thus \( C_{NS}(s, t), s, t \in T \) obtained have a similar parametric form to the corresponding stationary case.

4.1.1 The localised Gaussian correlation function

We consider the following form for the stationary Gaussian correlation function,

\[
C(s, t) = \exp \left[ -\|s - t\|^2 / \alpha \right], \quad s, t \in \mathbb{R}^N, \quad N \geq 1, \quad \alpha > 0,
\]

(38)

with smoothing parameter \( \alpha \) so that the corresponding spectral density function takes the form

\[
f(\omega) = h \exp \left[ -\alpha \omega^T \omega / 4 \right], \quad \omega \in \Omega = \mathbb{R}^N,
\]

(39)

where \( h = (4\pi / \alpha)^{-N/2} \) rather than (9). We favour working with (39) instead of (9) since the smoothness now increases with parameter \( \alpha \). Using our model (29)-(31) the resulting local spectrum at \( s \in D \) can be written in the form

\[
f^k_{NS}(\omega) = h_{NS}(s) \exp \left[ -\alpha(s) \omega^T \omega / 4 \right], \quad \omega \in \Omega = \mathbb{R}^N,
\]

(40)

with \( h_{NS}(s) = h(s)^2 (4\pi / \alpha(s))^{-N/2}, \quad s \in D \), where \( h(s) \) is a bounded function on \( D \), which is chosen to ensure constant power in the process (30), see Appendix A.2. for details. Then, the correlation function corresponding to (40) is

\[
C_{NS}(s, t) = h_{s,t} \exp \left[ -\|s - t\|^2 / \alpha_{s,t} \right], \quad \forall s, \ t \in D,
\]

(41)

where the local parameters \( h_{s,t} \) and \( \alpha_{s,t} \) are given by

\[
\alpha_{s,t} = \frac{[\alpha(s) + \alpha(t)]}{2},
\]

(42)

\[
h_{s,t} = 2^{N/2} \left[ \alpha(s) \alpha(t) \right]^{N/4} \left[ \alpha(s) + \alpha(t) \right]^{-N/2},
\]

(43)

Note that \( h_{s,t} \) depends on the choice of the function \( h(s), \ s \in D \). Here it is taken to ensure that (41) is a correlation function. We have the following result:
Theorem: If \( \alpha(t) > 0 \) for all \( t \in D \), then the function \( C_{NS}(s, t) \), \( s, t \in D \) given by (41) is a valid non-stationary correlation function obtained from localising the spectrum of the stationary Gaussian correlation function.

Proof: The proof of this result together with a full derivation of (41) is given in Appendix A.2.

4.1.2 The localised Matérn covariance function

Consider again the Matérn covariance function

\[
C(s, t) = \frac{\pi^{N/2}}{2^{\nu-1} \Gamma(\nu + N/2) \alpha^{2\nu}} (\alpha ||s - t||/N)^\nu K_\nu(\alpha ||s - t||/N), \quad s, t \in \mathbb{R}^N, N \geq 1,
\]

with corresponding spectral distribution function

\[
f(\omega) = (\alpha^2 + ||\omega||_N^2)^{-\nu - N/2}, \quad \omega \in \Omega = \mathbb{R}^N, N \geq 1, \quad \alpha > 0, \quad \nu > 0.
\]

Using our approach (29)-(29) the resulting local spectrum at \( s \in D \) is

\[
f_{NS}^s(\omega) = h(s)^2 (\alpha^2 + ||\omega||_N^2)^{-\nu(s) - N/2}, \quad \omega \in \Omega = \mathbb{R}^N,
\]

The corresponding covariance function is

\[
C_{NS}(s, t) = h_{s,t}(\alpha ||s - t||/N)^{\nu(s,t)} K_{\nu(s,t)}(\alpha ||s - t||/N),
\]

with local parameters \( \nu_{s,t} \) and \( h_{s,t} \) given by,

\[
\nu_{s,t} = \frac{[\nu(s) + \nu(t)]}{2} \quad (48)
\]

\[
h_{s,t} = \frac{h(t)h(s) \pi^{N/2}}{2^{\nu(s,t) - 1} \Gamma(\nu_{s,t} + N/2) \alpha^{2\nu_{s,t}}}, \quad (49)
\]

for all \( s, t \in D \), where \( h(t), t \in D \) is a bounded function chosen, if one so wishes, to impose the condition \( C_{NS}(t, t) = 1, \forall t \in D \), so that

\[
h(t) = \left( \frac{(\pi)^{N/2} \Gamma(\nu_{t,t})}{\alpha^{2\nu_{t,t}} \Gamma(\nu_{t,t} + N/2)} \right)^{-1/2}
\]

(50)

Note that we have \( \nu_s \equiv \nu_{s,s}, \forall s \in D \). By using the framework of the previous section we can state the following result.

Theorem: If \( \nu(t) > 0 \) for all \( t \in D \), then the function \( C_{NS}(s, t) \), \( s, t \in D \) given by (47) with (50) is
a valid non-stationary correlation function obtained from localising the spectrum of the stationary Matérn correlation function.

Proof: The proof of this result together with a full derivation of (47) is given in Appendix A.3.

Note that in Appendix D.1, we also provide the analytical form of the non-stationary correlation function obtained from a separable correlation function. When using these non-stationary covariance functions in practical applications, we will usually relate them to the corresponding stationary models by modelling $\theta(s)$ by $\eta(s)\theta$ where $\theta$ is the parameter obtained from the “best” stationary fit to the data. This allows for a straightforward characterisation of the non-stationarity in the data with respect to a “fixed” model (i.e. the stationary one).

It is important to emphasise that sometimes one may not be able to calculate the integrals given by (31) analytically, for other stationary models, as in the case for the separable correlation function for instance. One must then turn to numerical approximations in order to obtain the non-stationary covariance structure between two points.

It is interesting to note that in the case of the Gaussian and Matérn structures, spatial adaptation is similar to applying a particular transformation to the spectrum, that is, taking it to some power $\eta(s)$. This is generally the case for any stationary covariance function with a log-spectrum linear in its parameters. We refer to this transformation as spectral tempering. In Appendix C, we provide some results on the non-stationary Gaussian and Matérn covariance structures obtained when using spectral tempering. Note that the results are close to those obtained using direct spatial adaptation. More importantly, this analogy provides us with a natural extension of direct spatial adaptation to non-parametric spectra.

5 Spatial adaption of non-parametric spectra via tempering

It was shown in the previous section that tempering the spectrum is similar to spatial adaptation for the most common geostatistical covariance functions. More generally, when dealing with any non-parametric spectrum, tempering appears to be a simple method for evolving the spectrum over space by controlling the damping on the spectral components. Indeed, by conceptually heating or cooling the spectrum at each location, we are able to allow for spatially localised smoothing. Thus we transform the non-parametric spectrum at each location so that $f(\omega, s) = |f(\omega)|^{\eta(s)}$ at each
location $s$ for some latent spatial process $\eta(s)$, $s \in T$. As an example, we show in Figure 7 the damping effect induced on the Gaussian stationary model shown in Figure 6 if the spectrum is taken to the power $\eta = 0.2$ or $\eta = 5$, i.e. $f(\omega, s) = [f(\omega)\eta]$, $s \in T$. The normalised eigenvalues obtained in these two cases have been plotted along those for $\eta = 1$, the initial stationary case in Figure (23).

![Figure 7: Plots of the normalised eigenvalues for the one-dimensional example for $\eta = 1$ (squares), $\eta = 5$ (circles) and $\eta = 0.2$ (diamonds), Section 5.](image)

Clearly, the larger $\eta(s)$ is, the more damping is induced on the high frequency components. Hence, $\eta(s) < 1$ leads to a spatial process with higher frequency at location $s$ than in the stationary case. On the contrary, if $\eta(s) > 1$, only the slowest varying components have non-negligible power and thus, the process obtained is smoother than the initial stationary model. The effect of our modelling procedure as well as its interpretation are clear. Moreover, as for direct spatial adaptation, our modelling procedure possesses the critical property that the resulting covariance functions are analytical and their validity (i.e. positive definiteness) is easy to check.

It is important to note that spectral tempering is a general transformation that can be applied to both parametric and non-parametric spectra. However, when dealing with parametric spectra, the method described in the previous section has more appeal since the parameters that are evolved over space are directly interpretable as characterising the properties of the spatial dependence being modelled. However, it is interesting to point out that in some cases spectral tempering leads to tractable fully analytical forms for the resulting covariance structure when direct spatial adaptation doesn’t (i.e. numerical approximations must be used). An example is given in Appendix D.

Because spectral tempering can be applied to both parametric and non-parametric spectra, we first give some detail on how to implement it with respect to the general Fourier representation of a
stationary covariance function. Then, we apply the results to the particular case of the Karhunen-Loève representation which is a typical example of a non-parametric spectrum.

5.1 The general tempering procedure

We denote the spectrum by $f(\omega)$ to emphasise the fact that the spectrum need not be parametric. Then, in order to evolve the spectrum spatially, we consider a latent power or tempering process $\eta(s)$, a strictly positive function of location $s$, which we use to raise the stationary spectrum at location $s$ to the power,

$$f_{NS}(\omega, s) = f^s_{NS}(\omega) \propto [f(\omega)]^{\eta(s)}, \forall s \in D, \forall \omega \in \Omega,$$  \hspace{1cm} (51)

where $f^s_{NS}(\omega)$ is the non-stationary (NS) localised spectral density of the process at $s \in D$. Again, the constant of proportionality can be chosen to ensure (30). This leads to a joint spectral density at $s, t \in D$ of the form,

$$f^{s,t}_{NS}(\omega, s, t) = f^{s,t}_{NS}(\omega) = f^s(\omega)^{1/2} f^t(\omega)^{1/2} \propto [f(\omega)]^{\frac{\eta(s)+\eta(t)}{2}}, \forall s, t \in D, \forall \omega \in \Omega$$  \hspace{1cm} (52)

so that again,

$$f^s_{NS}(\cdot) \equiv f^s_{NS}(\cdot), s \in D.$$

Clearly for $\eta(s) > 1$, one obtains a local spectrum with relatively more power in the low frequencies, while $\eta(s) < 1$ induces more power in the high frequencies. Moreover, in the case of non-parametric spectra, the latter at each location are directly comparable with the corresponding original spectra they were derived from. Considering the model given by (51) and (52) we have the following result.

Lemma 5.1: Let $C(s,t)$ be a stationary covariance function whose spectrum is $f(\omega)$. Then the function given by

$$C_{NS}(s,t) = \int_{\Omega} \exp[i\omega(s-t)]f(\omega)^{\eta(s)/2} f^*(\omega)^{\eta(t)/2} d\omega$$  \hspace{1cm} (53)

is a valid covariance function if and only if

$$\int_{\Omega} [f(\omega)]^{\eta(s)} d\omega < +\infty.$$  \hspace{1cm} (54)
for all $s$ in $D$.

Proof: The necessary condition is clear by definition of $C_{NS}(\cdot, \cdot)$ so that only the sufficient part of the statement needs to be proved. This is done in Appendix B.

We have thus set out the condition under which our modelling approach for non-stationary processes given by (51)-(53) is valid. That is, provided the condition (54) is met the local covariance function $C_{NS}(s, t)$, $s, t \in D$ induced by evolving the spectrum using the latent process $\eta(s)$, $s \in D$ will be a valid covariance function. Examples of the procedure for the Gaussian and Matérn covariance functions are given in Appendix C. We now describe how to apply the tempering procedure in the case of non-parametric spectra by considering the particular case of the Karhunen-Loève expansion.

5.2 Tempering the non-parametric Karhunen-Loève spectrum

Although we consider the particular Karhunen-Loève spectrum here, it should be clear how the procedure generalises to other non-parametric spectra. When considering non-parametric spectra, one needs only adapt the general procedure described in the previous section. Thus, similarly to (51), we evolve the initial Karhunen-Loève spectrum spatially, that is,

$$\lambda_k^{NS} = \lambda_k^n, \quad k = 1, \ldots, n, \quad (55)$$

so that similarly to (53), we have

$$C_{NS}(s, t) = \sum_{i=1}^{\infty} \lambda_k^{n(s)/2} \lambda_k^{n(t)/2} \phi_k(s) \phi_k(t) \quad (56)$$

It is then straightforward to check that Lemma 5.1 of Subsection 5.1 remains valid here. Indeed, the proof is along the same lines. It is important to note however, that the condition given by Lemma 1 is not very useful here since one is rarely capable of obtaining an analytical expression for the eigenvectors $\phi_k(t)$, $k = 1, \ldots, \infty$. Nevertheless, it appears this is of little importance since for most applications we use the Karhunen-Loève expansion on a finite data set. In other words, one really considers (19) and (18) rather than (12) and (16) respectively.

It is clear in the finite-dimensional case that the matrix $C_{NS}^{NS}$ defined by

$$C_{NS}^{NS} = \sum_{k=1}^{n} \lambda_k^{n_i+n_j}/2 \, v_{i,k} v_{j,k} + \sigma^2 \delta_{ij}, \quad i, j = 1, \ldots, n. \quad (57)$$
where \( \{ v_{i,j} \}_{i=1}^n \) are defined as in Section 3.2, is positive definite. This is due to the fact that there are only a finite number eigenvalues, all of which are strictly positive. In the next section, we show how an equivalent basis functions representation leads to the same result (57).

Note that we sometimes favour working with the corresponding correlation matrix induced by the covariance matrix (76),

\[
C_{NS}^{i,j} = \left[ \frac{1}{\sqrt{\text{Cov}_{i,i}^{NS} \sqrt{\text{Cov}_{j,j}^{NS}}} \sum_{k=1}^{n} \lambda^{(n_i+n_j)/2} v_{i,k} v_{j,k}} \right] + \sigma^2 \delta_{ij}, \quad i, j = 1, \ldots, n.
\]

Spectral tempering for non-parametric spectra leads to models which retain similar properties to those obtained for parametric spectra using the method described in the previous section. Indeed, if we compare the non-stationary and stationary structures, that is (57) with (22), the behaviour and smoothing assumptions induced by our method are clear. The stationary model is again a special case corresponding to \( \eta(s) = 1, \forall s \in D \) and the resulting non-stationary models are again capable of capturing spatially varying local anisotropy features. Some visualisations will be given in Section 10.

### 6 An equivalent basis functions representation

In this section, we show how the spectral tempering modelling procedure for non-parametric spectra can alternatively be derived as a special type of regression model with spatially evolving coefficients, which we will refer to as *variance-varying coefficient models*. These models are appealing since they represent an alternative to varying-coefficients models. We consider the Karhunen-Loève expansion, though our procedure is equally valid for any basis.

Recall from Section 3.2, that the finite-dimensional Karhunen-Loève expansion of the standard kriging model can be written in an equivalent basis function representation,

\[
Z = V \cdot \alpha + \epsilon
\]

where \( Z = \{ Z(s_1), \ldots, Z(s_n) \} \), \( V \) is given by (17) and

\[
\alpha \sim MVN(0, D), \quad (60)
\]

\[
\epsilon \sim MVN(0, \sigma^2 I_n) \quad (61)
\]
where $D$ defined by (18) and $\sigma^2$ is some noise parameter. Our tempering modelling procedure is equivalent to letting the vector of coefficients $\alpha$ evolve over space. In order to do that we first re-write the model given by (59)-(61) in the following basis function form

$$Z(s_i) = (v_{i,1}, \ldots, v_{i,n})' \alpha_i + \epsilon_i, \quad i = 1, \ldots, n, \quad (62)$$

where we have, following (61),

$$\epsilon_i \sim N(0, \sigma^2) \quad (63)$$

and where $\alpha_i = \alpha(s_i)$ are a separate set of random coefficients associated with each location $s_i$ distributed as

$$\alpha_i \sim MVN(0, C_{\alpha_i}) \quad (64)$$

where initially we take $C_{\alpha_i} = D$ so that

$$C_{\alpha_i} = \begin{pmatrix} \lambda_1 & 0 & \cdots \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix}, \quad (65)$$

Hence, from the general regression model (59)-(61) we have constructed the following localised regression model:

$$Z = X\tilde{\alpha} + \epsilon \quad (66)$$

where $\epsilon$ is defined as previously, (61), and $X$ is a $n \times n^2$ matrix such that:

$$X\tilde{\alpha} = \begin{pmatrix} v_{1,1:n} & 0 & \cdots & 0 \\ 0 & v_{2,1:n} & \cdots & \vdots \\ \vdots & \cdots & \ddots & \vdots \\ 0 & \cdots & 0 & v_{n,1:n} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}, \quad (67)$$

where the matrix $V$ was defined in (18). The $n^2 \times 1$ vector $\tilde{\alpha}$ stores the localised $\alpha_i$'s as $\tilde{\alpha} = (\alpha_1, \ldots, \alpha_n)'$, and we assume,

$$E(\tilde{\alpha}) = 0_{n^2}, \quad (68)$$

$$Var(\tilde{\alpha}) = C_{\tilde{\alpha}} \quad (69)$$

27
with \( C_\alpha \) a \( n^2 \times n^2 \) matrix defined as

\[
C_\alpha = \begin{pmatrix}
C_{\alpha_1} & C_{\alpha_1 \alpha_2} & \cdots & C_{\alpha_1 \alpha_n} \\
C_{\alpha_1 \alpha_2} & C_{\alpha_2} & \cdots & C_{\alpha_2 \alpha_n} \\
\vdots & \vdots & \ddots & \vdots \\
C_{\alpha_1 \alpha_n} & \cdots & C_{\alpha_{n-1} \alpha_n} & C_{\alpha_n}
\end{pmatrix},
\]

(70)

where \( C_{\alpha_i \alpha_j} = \text{Cov}(\alpha_i, \alpha_j) \).

Our modelling procedure is now equivalent to setting

\[
C_{\alpha_i} = D_{i}(s_i),
\]

(71)

so that the eigenvalues at location \( s_i \) are taken to the power \( \eta(s_i) = \eta_i \) as in the previous subsection. From this equivalent point of view it is again clear that we allow for spatially localised smoothing via tempering of the spectrum. Indeed, if we compare the basis function representation given by (59)-(61) and the representation (62)-(65), we see that changing the eigenvalues from \( \lambda_j \) to \( \lambda_j^\eta \) at location \( s_i \), that is changing \( C_{\alpha_i} \) from \( D \) to \( D_{\eta_i} \), implies a localised shrinkage at \( s_i \) which, recalling (27) and (28) in Section 3.2, is given by

\[
\hat{\alpha}_{i,j} = \frac{\lambda_j^{\eta_i}}{\sigma^2 + \lambda_j^{\eta_i}} \hat{\alpha}_{\text{mle}}, \quad j = 1, \ldots, n
\]

rather than by

\[
\hat{\alpha}_{i,j} = \frac{\lambda_j}{\sigma^2 + \lambda_j} \hat{\alpha}_{\text{mle}}, \quad i = 1, \ldots, n
\]

as given by the initial stationary model, where \( \alpha_{\text{mle}} \) is the least squares or maximum-likelihood estimate (mle) of \( \alpha \) given in (60). Hence, as already stated previously, by changing the value of \( \eta \), the process can be made smoother or less smoother at location \( s_i \) than the initial stationary process. We now show how this representation leads back to (57).

From (71) we can now write (70) in covariance form,

\[
C_\alpha = \begin{pmatrix}
D^\eta & \rho_\theta^{12} D^{(\eta_1 + \eta_2)/2} & \cdots & \rho_\theta^{1n} D^{(\eta_1+\eta_n)/2} \\
\rho_\theta^{12} D^{(\eta_1 + \eta_2)/2} & D^\eta & \cdots & \rho_\theta^{2n} D^{(\eta_2+\eta_n)/2} \\
\vdots & \vdots & \ddots & \vdots \\
\rho_\theta^{1n} D^{(\eta_1+\eta_n)/2} & \cdots & \rho_\theta^{(n-1)n} D^{(\eta_{n-1}+\eta_n)/2} & D^\eta
\end{pmatrix},
\]

(72)
where $\rho_{ij}^{\theta} = \rho_{\theta}(s_i, s_j)$ defines the correlation between $\alpha_i$ and $\alpha_j$ for $i, j = 1, ..., n$. We denote $C_{\alpha}$ by $C_{\alpha_{\theta}}$

**Theorem:** The matrix $C_{\alpha_{\theta}}$ so defined is positive definite for $\rho_{\theta}(d)$ a monotonic decreasing function of $0 \geq d < \infty$, $\rho_{\theta}(0) = 1$.

**Proof:** see Appendix E.1.

Clearly working with (72), which is a $n^2 \times n^2$ matrix is unattractive. However, (66)-(72) defines a Gaussian process,

$$ Z \sim MVN(0, Cov^{NS}) $$

(73)

where $Cov^{NS}$ is a $n \times n$ matrix defined as:

$$ Cov^{NS} = X^\prime C_{\alpha_{\theta}} X + \sigma^2 I_n $$

(74)

with $C_{\alpha_{\theta}}$ defined by (72). The matrix $X$ was defined in (66). The resulting covariance matrix $Cov^{NS}$ has elements,

$$ Cov^{NS}_{i,j} = \sum_{k=1}^{n} \rho_{ij}^{\theta} \lambda_k^{(n_{i}+n_{j})/2} v_{i,k}v_{j,k} + \sigma^2 \delta_{ij}, \quad i, j = 1, ..., n. $$

(75)

The validity of $Cov^{NS}$ is proved in Appendix E.2.

We choose to take $\rho_{ij}^{\theta} = 1$ for $i, j = 1, ..., n$, in (75), which leads to a valid model allowing us to recover the initial stationary model if the spectral damping is one over the space (i.e. $C_{\alpha_{\theta}} = D$ for $i = 1, ..., n$ with $D$ given by (17)), which leads to

$$ Cov^{NS}_{i,j} = \sum_{k=1}^{n} \lambda_k^{(n_{i}+n_{j})/2} v_{i,k}v_{j,k} + \sigma^2 \delta_{ij}, \quad i, j = 1, ..., n. $$

(76)

Thus, we have shown that our spectral tempering procedure has an equivalent basis function representation with variance-varying coefficients.

**7 Properties**

We now illustrate the behaviour and the properties of the models induced by allowing the spectrum to evolve spatially. For illustration purposes, we take the initial (stationary) correlation matrix to be Gaussian,

$$ C_{ij} = \exp(-\alpha ||s_i - s_j||^2), \quad i, j = 1, ..., n $$

(77)
for some value of the parameter $\alpha$. We consider two models corresponding to the two spectral representations:

- The first model is built from spatially evolving the spectrum of the Fourier representation of (77), given by (9). This spectrum is parametric with parameter $\alpha$ and we take $\alpha(s) = \alpha \eta(s)$ for some function $\eta(\cdot)$ over $D$.

- The second model is obtained from the Karhunen-Loève expansion of (77). The spectrum is non-parametric so we use spectral tempering with parameter $\eta(\cdot)$ to evolve it over space.

For clarity, we first consider the one-dimensional case, $s \in [-1, 1]$, and three different modelling shapes for $\eta(s)$ which help to illustrate some of the properties and characteristics of our model.

- Model 1, piecewise constant,

$$
\eta_i = \begin{cases} 
0.1 & \text{if } -1 < s_i < 0 \\
10 & \text{otherwise}
\end{cases}
$$

- Model 2, cubic,

$$
\eta_i = (s_i + 1)^3 + 10^{-4}
$$

- Model 3, stationary,

$$
\eta_i = 1
$$

Each simulation uses $n = 150$ data points equally spaced on $(-1, 1)$ and we have taken the value of $\alpha$ to be 400 (for ease of illustration). Model 1 characterises a situation where the first half-set of locations have a high-frequency spatial structure and the second set a smooth one. Model 2 describes a structure which becomes smoother as the coordinates of the locations tend to 1. Note we have added a small term so that $\eta(s) > 0, s \in (-1, 1)$ Finally, Model 3 corresponds to the initial stationary process. The different models for the $\eta_i$ parameters have been plotted in Figure 8. Realisations of the process for Models 1 to 3 for both spectral representations are illustrated in Figures 9 to 11 respectively.

It is clear from Figures 9, 10 and 11 that $\eta(s)$ characterises the local frequency of the process under study. Indeed, the step function, first plot in Figure 8, induces a process, Figure 9, that
Figure 8: From the left, Plots of the $\eta_i$ parameters for Models 1, 2 and 3, Section 7.

Figure 9: Data simulated from Model 1 (78), see 1st plot in Figure 8. From the left, Plot 1 & 2: Fourier spectral representation. Plot 3 & 4: Karhunen-Loève representation, Section 7.

Figure 10: Data simulated from Model 2 (79), see 2nd plot in Figure 8. From the left, Plot 1 & 2: Fourier spectral representation. Plot 3 & 4: Karhunen-Loève representation, Section 7.

Figure 11: Data simulated from Model 3 (80), see 3rd plot in Figure 8. From the left, Plot 1 & 2: Fourier spectral representation. Plot 3 & 4: Karhunen-Loève representation, Section 7.
shows a distinct change in frequency at the mid-point \( s = 0 \). The smoothly increasing function, second plot in Figure 8, induces a process that becomes gradually smoother toward the right hand side of the space as shown in Figure 10. Finally, and as expected, the stationary process, third plot in Figure 8, exhibits a constant level of smoothness, Figure 11.

Figures 9, 10 and 11 also reveal that these properties are shared by both spectral representations. This is an important feature of our building strategy since it illustrates that the properties and interpretation of the \( \eta(s) \) process are the same for particular choices of spectra. Finally, one can see from Figures 9, 10 and 11 that, while the local frequency changes, the overall scale of the process, \( \text{var}(z(s_i)) \) at any location remains constant. This is a consequence of ensuring (30).

We now consider simulations in two dimensions. Locations are defined on a \( 8 \times 8 \) (i.e. \( n = 64 \)) lattice and the value of \( \alpha \) is taken to be 0.25 (again for ease of illustration). This choice is purely due to illustration purposes. We consider the following two shapes for \( \eta(s) \),

- Model 4, piecewise constant,
  \[
  \eta_i = \begin{cases} 
  0.1 & \text{if } s_i \in [1, 4] \times [1, 8], \\
  10 & \text{otherwise}
  \end{cases}
  \]  \hspace{1cm} (81)

- Model 5, cubic,
  \[
  \eta_i = 0.01 \times \left( \frac{||s_i||}{\sqrt{2}} \right)^3,
  \]  \hspace{1cm} (82)

In Model 2, the norm \( || \cdot || \) corresponds to the Euclidean norm. Simulations from the induced models for these two examples and for both spectral representations are given in Figure 12 and Figure 13.

Again the behaviour and interpretation of \( \eta(s) \) is clear and the conclusions drawn from the one-dimensional simulations remain valid in the two-dimensional case.

Further insight is offered by examining the non-stationary correlation functions \( C^{NS}(\cdot, s_j) \) obtained for both representations when the noise is taken to be zero. We consider Model 2, (79), as its non-stationary structure serves our illustration purposes. Figures 14 and 15 correspond to the values of the non-stationary correlation functions \( C^{NS}(\cdot, s_j) \), plotted with \( n = 150 \) data points, for \( s_i \in \{-0.95, -0.6, 0.6, 0.95\} \), for the Fourier and Karhunen-Loève representations respectively, plotted for \( n = 150 \) data points. The shape of \( \eta(s) \) is given in the 2nd Plot in Figure 8 and the
corresponding simulations are shown in Figure 10. Note that the values $s_i$ have been chosen to characterise regions of the space with different smoothness constraints.

From Figures 14 and 15 it is clear that the correlation function $C^{NS}(s_j, s_i)$ considered as a function of $s_i$ is well-behaved and interpretable. The widths of the kernels change with location indicating the changing levels in association over the space. This is in agreement with the change in the smoothness structure of the data.

8 Prediction

In this section, we consider the issue of predicting the resulting processes obtained from direct spatial adaptation of parametric spectra and spectral tempering of non-parametric spectra at any location in the space. If the basis functions have analytical expressions then prediction is straightforward as we show in the next section. On the other hand, if the basis functions are not analytical, then more care is needed for the prediction. This is the case with the Karhunen-Loève expansion.
Figure 14: From the left, Plots of the correlation functions $C^{NS}(\cdot, s_i)$ for $s_i = -0.95 (i = 5), -0.6 (i = 50), 0.6 (i = 100)$ and $0.95 (i = 145)$, for the Fourier representation with $\eta(s)$ as in (79), Section 7.

Figure 15: From the left, Plots of the correlation functions $C^{NS}(\cdot, s_i)$ for $s_i = -0.95 (i = 5), -0.6 (i = 50), 0.6 (i = 100)$ and $0.95(i = 145)$, for the Karhunen-Loève representation with $\eta(s)$ as in (79), Section 7.
for instance. We now consider each case successively.

8.1 Prediction with analytical basis functions

In order to make predictions when the expansion of the initial covariance function is with respect to some analytical basis (e.g. Fourier representation), one can simply use kriging as outlined in Cressie (1993) since the non-stationary localised covariance between any two points in the space can be determined, provided one has the value of the latent spatial process \( \eta(\cdot) \) used to evolve the spectrum at any location. This issue will be dealt with in the next section. Thus, the conditional predictive distribution density at any new location \( s^* \in T \) is given by

\[
p(y^*|Z, S, s^*, C_{NS}(\cdot, \cdot)) \sim N(\mu, \sigma^2),
\]

with \( \mu \) and \( \sigma^2 \) given by

\[
\mu = C_{NS}(s^*, S)'C_{NS}(S, S)^{-1}Z
\]

\[
\sigma^2 = C_{NS}(s^*, s^*) - C_{NS}(s^*, S)'C_{NS}(S, S)^{-1}C_{NS}(s^*, S),
\]

where \( Z = (z_1, \ldots, z_n) \) are the observations at locations \( S = (s_1, \ldots, s_n) \), \( C_{NS}(s^*, S) \) is a \( n \times 1 \) vector with \( i^{th} \) element given by \( C_{NS}(s^*, s_i) \) and \( C_{NS}(S, S) \) is a \( n \times n \) matrix with \( (i, j)^{th} \) element \( C_{NS}(s_i, s_j) \). We note that in this case, the computational cost of prediction is the same as in the stationary case.

8.2 Prediction with non-analytical basis functions

In this section we focus on the issue of evaluating the covariance between any two points in the space when one uses a representation of the covariance function in terms of non-analytical basis functions. Here, we consider the expression (76) induced by the finite-dimensional Karhunen-Loève expansion (i.e. the singular-valued decomposition of the covariance matrix) since the full analytical infinite-dimensional Karhunen-Loève expansion of a given covariance function can rarely be worked out. The method we present can be easily generalised to other non-analytical basis functions, such as wavelets or the Demmler-Reinsh basis.

To begin note that if \( C(s_i, s_j), i, j = 1, \ldots, n \) is a valid \( n \times n \) spatial covariance matrix, then the Karhunen-Loève expansion leads to a decomposition of \( C \) in terms of the eigenvalues \( \{\lambda_i\}_{i=1, \ldots, n} \)
of \( C \) and their corresponding eigenvectors \( \{ \hat{\phi}_i \}_{i=1}^{n} \) where \( \hat{\phi}_i = (\hat{\phi}_i(s_1), \ldots, \hat{\phi}_i(s_n)) \). The main issue for the purpose of prediction is the need to extend the eigenvectors \( \{ \phi_i \}_{i=1}^{n} \) to functions \( \{ \phi_i(s) \}_{i=1}^{n} \) defined for all \( s \in D \). This will enable one to calculate the quantity

\[
\sum_{i=1}^{n} \lambda_i \phi_i(s) \phi_i(u), \ s, u \in D
\]

for any locations \( s \) and \( u \) thus leading to the determination of \( C(s,u) \).

From the theory of the numerical treatment of integral equations (Baker 1977), we know that there are two classes of methods for extending the eigenvectors \( \{ \hat{\phi}_i \}_{i=1}^{n} \). The first class corresponds to methods based on integration formulae and the second is based on expansion methods. The methods based on integration formulae are more appealing in the sense that they are easier to implement and thus computationally faster. In the expansion methods, an approximate eigenfunction \( \phi_i(s) \), \( s \in D \) is derived as a linear combination of linearly independent \( a \) priori chosen functions \( \psi_1(s), \ldots, \psi_n(s) \). These methods are computationally expensive and the \( a \) priori choice of the generating functions is often not straightforward. Obled and Creutin (1986) used a method of similar philosophy on the empirical (non-stationary) covariance matrix \( C \) obtained from spatio-temporal data.

Our modelling approach allows us to use the first class of methods as, unlike Obled and Creutin (1986), our method does not compute the Karhunen-Loève expansion on the empirical (non-stationary) covariance matrix but on a stationary parametric covariance function fitted to the data. Thus, we can obtain the stationary covariance between any two points in the space and this is important information which we make use of. The methodology we present essentially follows that given in Baker (1977) and used by Williams and Seeger (2001) in a different context.

Returning to the general case, recall that the eigenvalues and eigenfunctions of the covariance function (or \textit{kernel}) \( C(s,u) \), \( s, u \in D \) are the solutions of the following Fredholm homogeneous equation of the second kind,

\[
\int_D C(s,u)f(u)du = \kappa f(s), s \in D, \tag{83}
\]

where the eigenfunctions \( f(u) \) are orthonormalised. That is, if \( f_m(u) \) and \( f_n(u) \) are the eigenfunctions corresponding to the eigenvalues \( \kappa_m \) and \( \kappa_n \) respectively, then we must have

\[
\int_D f_m(u)f_n(u)du = \delta_{mn}.
\]
This suggests that in the finite dimensional case, one may obtain approximate eigenvalues $\bar{\kappa}$ and their corresponding approximate eigenfunctions $\bar{f}(\cdot)$ by solving the equation

$$\sum_{j=1}^{n} \omega_j C(s, s_j) \bar{f}(s_j) = \bar{\kappa} \bar{f}(s), \ s \in D. \quad (84)$$

where the weights ($\omega_1, ..., \omega_n$) are strictly positive real numbers that sum to one. In order to solve (84) we consider the system of equations obtained when replacing successively $s$ with $s_1, ..., s_n$ in (84). The system has a unique solution since the matrix $C(s_i, s_j), \ i, j = 1, ..., n$ is positive definite and its solutions are the eigenvalues $\{\lambda_i\}_{i=1,...,n}$ and the corresponding eigenvectors $\{\bar{\phi}_i\}_{i=1,...,n}$ obtained from the single value decomposition of $C$. The eigenvectors must then be rescaled so that they satisfy

$$\sum_{j=1}^{n} \omega_j \bar{f}_n(s_j) \bar{f}_m(s_j) \approx \delta_{mn}. \quad (85)$$

We denote these vectors by $\bar{\phi}_i, \ i = 1, ..., n$. Note that these vectors together with the eigenvalues $\bar{\kappa}_i, \ i = 1, ..., n$ are not usually those that one uses in the Karhunen-Loève expansion in the finite dimensional case. Indeed, one usually uses the eigenvectors and eigenvalues obtained from the single value decomposition of the matrix $C$. That is because it can be shown that both representations are equivalent so that one generally chooses to work directly on the matrix $C$. Also the quadrature rule we have chosen to compute approximations to integrals is very simple here (i.e., a diagonal matrix with weights given by the $\omega_i$'s). More complex methods could be used. We refer the reader to Baker (1977) for more detail.

The eigenvalues $\bar{\kappa}_i, \ i = 1, ..., n$ obtained are strictly positive since the matrix $C$ is positive definite and thus, by substituting in (84), the vectors $\{\bar{\phi}_i\}_{i=1,...,n}$ may be extended to functions $\{\phi_i(s)\}_{i=1,...,n}, \ s \in D$ using the relation

$$\phi_i(s) = (1/\bar{\kappa}_i) \sum_{j=1}^{n} \omega_j C(s, s_j) \bar{\phi}_i(s_j), \ i = 1, ..., n, \ s \in D \quad (86)$$

If one uses the single-value decomposition of $C$ to compute the Karhunen-Loève expansion, then one would have to use the formula given by

$$\phi_i(s) = (1/\lambda_i) \sum_{j=1}^{n} C(s, s_j) \bar{\phi}_i(s_j), \ i = 1, ..., n, \ s \in D \quad (87)$$
to extend the eigenvectors $\{\hat{\phi}_i\}_{i=1,...,n}$ where $\lambda_i$, $i = 1, ..., n$ are the eigenvalues obtained from the single value decomposition. Note that approximations satisfying (86) are sometimes known as Nyström’s method and $\phi_i(s)$, $i = 1, ..., n$ is the Nyström extension of the eigenvector $\hat{\phi}_i$, $i = 1, ..., n$. It is straightforward to check that (86) and (87) leads to a valid model for the covariance structure over the space. Some examples of the predictions obtained for unequally spaced data are given in Figures 16 and 17. These figures were drawn from $n = 50$ randomly chosen data points in $[-5,5]$ and for four different types of covariance functions usually used in spatial applications: a Matérn covariance matrix with $\nu = 0.25$ (Matérn 1 here after), $\nu = 0.5$ (i.e. Exponential covariance function), $\nu = 1$ and for the Gaussian covariance function (i.e. $\nu \to \infty$). We expand the first and fourth eigenvector in Figures 16 and 17 respectively. Again, we focus on the one-dimensional case for illustration purposes only.

![Eigenvector Comparison](image.png)

Figure 16: Numerical estimates of the 1st eigenfunction at 1000 equally spaced points in $[-8,8]$ shown by the solid line, using 50 unequally spaced points (circles) for the Matérn 1 (top left), the Exponential (top right), the Matérn 2 (bottom left) and the Gaussian (bottom right) covariance function, Section 8.2.

It is interesting to note that the shape of the eigenvectors are not as smooth as those shown in Section 3.2. That is because we have unequally spaced data here. In other words, the basis functions used in the Karhunen-Loève expansion depend on the location of the data points and
Figure 17: Numerical estimates of the 4th eigenfunction at 1000 equally spaced points in $[-8, 8]$ shown by the solid line, using 50 unequally spaced points (circles) for the Matèrn 1 (top left), the Exponential (top right), the Matèrn 2 (bottom left) and the Gaussian (bottom right) covariance function, Section 8.2.

thus the form of the eigenvectors may vary from one application to another. This is a fundamental difference between the Karhunen-Loève expansion and the Fourier representation for which the basis functions (i.e. complex exponentials) remain the same.

Note also that the shape of the eigenvectors within Figures 16 and 17 are quite similar. This is not surprising since all the covariance functions considered are special cases of the Matèrn covariance function whose smoothness is controlled by the parameter $\nu > 0$. This suggests therefore that given the same locations, the eigenvectors obtained from a Karhunen-Loève expansion on the four types of covariance functions considered yields similar shapes for the eigenvectors and that the smoothness is controlled by the generalised spectrum for each type (i.e. the eigenvalues which give more or less weight to higher frequency components depending on the value of $\nu$). We will reflect further on the issues we have mentioned here in the Discussion. It is clear from Figures 16 and 17 that the predicted expansion of the eigenvectors works well and we have found it to be the case on numerous applications and simulations. They also exhibit an important feature which is that they tend to zero far (in distance) from the support of the data points. This will result in predicting a
covariance of zero between a data point and a point far from any other data point, which is what one would obtain with kriging.

As already stated, the expression (87) is only useful in the case when one can compute the value of \( C(s, u) \) between any points in the domain. This is an advantage of performing the Karhunen-Loève expansion on the best stationary parametric model as an initial stage. Note also that (87) has intuitive appeal since it corresponds to the projection of the vector \( C_\beta = (C(s, s_1), ..., C(s, s_n)) \) on the space spanned by the eigenvectors \( \{\phi_i\}_{i=1}^n \).

Using the method described above, we are now capable to obtain an analytical expression (and hence to compute) the covariance function (76) between any two points \( s \) and \( u \) in the plane \( D \). Indeed, using (87) it only remains to predict \( \eta(s) \) at a new location which in our case can easily be done with as we show in the next section.

9 Modelling of the latent spatial process \( \eta(s_i), i = 1, ..., n \)

Recall that we model the spatially evolving parameter \( \theta(s) \) of a parametric spectrum \( f(\omega, \theta) \) by \( \theta(s) = \eta(s) \theta \) where \( \eta(\cdot) \) is a spatial latent process over \( D \). A similar process is also used to temper the non-parametric Karhunen-Loève spectrum. In this section, we propose a method for modelling \( \eta(s) \) which we have found to work well across a range of data sets. However, the reader is free to use any modelling procedure so long as \( \eta(s) \) remains strictly positive. When modelling \( \eta(s) \), we wish to dictate that nearby locations have similar spectral representations so that \( \eta(s) \) should be a smooth function of location \( s \). Another important feature is that of computational efficiency when estimating the non-stationary covariance structure. Finally, we want to be able to obtain uncertainty intervals for our predictions. We therefore suggest modelling \( \eta(s) \) using a Bayesian regression spline model on \( \log(\eta(s)) \) of the form:

\[
\log \eta(s) = \beta_0 + s \beta_1 + \sum_{j=1}^k \psi(s, u_j) \beta_{j+2}
\]

(88)

where \( \{\beta_0, \beta_1\} \) capture linear trends in \( \log \eta(s) \) and the regression splines \( \psi(s, u_j) \) with knot points \( u_j \) allow for spatial variation in \( \eta(s) \). We refer the reader to Denison et al. (2002) for more detail. In two dimensions, we use the thin-plate regression spline basis, \( \psi(s, u_j) = ||s - u_j||^2 \log ||s - u_j|| \), as a natural choice. In the one dimension examples, we have chosen to work with the basis,
\(\psi(s, u_j) = ||s - u_j||\), which appeared to give us good practical results. For a particular data set we place the knots using a \(k\)-means clustering algorithm. In our examples and simulations, we have found that using 5 knots for one-dimensional data sets of size \(n \approx 50\) and 10 knots for two-dimensional ones of similar sizes gave good results in practice.

Using a regression spline approach has a number of advantages: first, we reduce the dimension of the space over which the search is undertaken to that of the number of parameters to estimate, which improves the computational cost of our method. Secondly, predictions of the \(\eta(s)\) at new locations are straightforward. Finally, we have put a prior on the parameters,

\[
\beta = (\beta_0, \beta_1, \beta_3, ..., \beta_{k+2})' \sim N(0, b^2 I),
\]

and \(b^2\) is fixed to some reasonable value. We then optimise the parameters of our regression model as well as the new nugget effect, using the following marginal scale-invariant log-likelihood (i.e. REML), e.g. McCulloch and Searle (2001), given by

\[
l_m(C^{NS}; Z) \propto -\frac{1}{2} \log |C^{NS}| - \frac{n}{2} \log (Z' C^{NS-1} Z) - (1/b^2)\beta' \beta \tag{90}
\]

In our applications, we used the Nelder-Mead optimisation procedure to maximise the scale-invariant marginal log-likelihood \(l_m(C^{NS}; Z)\). This leads to a set of optimal \(\hat{\beta}_i\)'s for the spline surface. Uncertainty in \(\eta(s)\) is then expressed by uncertainty in these coefficients. We suggest using normal approximations to the posterior density of \(\beta\), evaluated at \(\hat{\beta}\) to quantify this uncertainty. Alternatively we could use stochastic simulation such as Markov Chain Monte Carlo to draw inferences. We do not pursue this issue here as this is beyond the scope of this paper. Finally, note that prediction of \(\eta(s)\) at new locations is straightforward.

As a straightforward example of the flexibility of our method, consider the Doppler function,

\[
f(s) = (s(1 - s))^{1/2} \sin((2\pi)(1 + a)/(s + a)), \quad s \in [0, 1],
\]

where \(a\) is a constant. We generate \(n = 128\) data points regularly on \([0,1]\) with \(a = 0.05\). Then, following Donoho and Johnstone (1995), the sample points \(f(s)\) are rescaled so that their variance is equal to 7. Some zero-mean Gaussian white-noise with variance one is added. The “true” and “noisy” data are plotted in Plot 1 in Figure 18.

We fit a Gaussian stationary model to the data and compare its behaviour to that of the corresponding Gaussian non-stationary model (73) obtained using a Fourier representation. The
\( \eta(\cdot) \) model is given by (88) and we use 10 basis functions. Results are shown in Plots 2, 3 and 4 in Figure 18.

![Plots 2, 3, 4 and 5](image)

Figure 18: Plot 1: true function (- -) and noisy data (*). Plot 2: true function (- -) and “best” Gaussian stationary fit. Plot 3: true function (- -) and “best” Gaussian non-stationary fit. Plot 4: Estimated values of \( \log \eta(s) \) for the stationary (- -) and non-stationary (\( \cdot \)) models, Section 9.

It is clear from Plots 2 and 3 in Figure 18 that the non-stationary model can adapt to the change of smoothness in the signal. Plot 4 shows that the values of \( \eta(s) \) for the non-stationary model increases with \( s \) as one would expect since the smoothness in the data increases with \( s \). This one-dimensional example highlights the flexibility allowed by our modelling procedure for \( \eta(\cdot) \) and once again, underlines the interpretability of the values of \( \eta(s) \) over the space. Results are similar if we consider the Karhunen-Loève expansion.

10 A two-dimensional real data example

In this section, we give an application of our method on a real data two-dimensional environmental monitoring example, that was initially studied in Zirschky and Harris (1986). The data are dioxin concentrations on the Piazza Road pilot study area, which is part of an EPA Superfund site in Missouri. A log-transformation of the initial data was taken so that the resulting data could be assumed to follow a Gaussian distribution. We are interested in this data set as the study by Higdon et al. (1999) showed that the spatial dependence between points was expected to vary with location. That is, not only did the spatial structure in the data reveal anisotropy but this anisotropy was also shown to vary spatially.

The actual pilot data contain over 1000 concentration measurements, but following Higdon et al. (1999) we only consider a subset of \( n = 60 \) data points here, which are chosen so that the whole
space is covered. We model the mean trend $\mu(s)$ as

$$\mu(s) = m_0 + m_1 s_1 + m_2 s_2 + m_3 s_1 s_2,$$

where $s = (s_1, s_2)$ and $(m_0, m_1, m_2, m_3)$ are a set of parameters to be estimated. We estimate these parameters jointly with the parameters of the Gaussian covariance function in the first step of the procedure, that is when fitting the stationary model to the data. We then apply our model to the data, using the Karhunen-Loève expansion and the true Fourier representation. When fitting these two models, we choose to take the mean trend to be equal to that obtained from the first stationary step. Indeed, we are solely interested here in the covariance structures obtained for both models and in comparing them one with another and also with the stationary case. These comparisons are made easier to interpret when all models are assumed to have the same trend. Note however that our model could be easily adapted to account for a different mean trend.

Many methods exist for obtaining an initial stationary covariance matrix. One can refer to Cressie (1993) for a full description of various methods and the reader is free to adopt any of these procedures. Here, we propose to use the following parametric form for the stationary covariance structure,

$$C(s_i, s_j) = \rho^2 \left( \gamma^2 \delta_{ij} + \exp(-\chi^2 ||s_i - s_j||^2) \right), \quad i, j = 1, ..., n. \quad (91)$$

where $\rho$ is the scale parameter and $\gamma^2$ the nugget effect. To estimate the parameters we integrate over $\rho^2$ to obtain a marginal scale-invariant log-likelihood for $\gamma$ and $\chi$, equivalent to the REstricted Maximum-Likelihood (REML) described for instance by McCulloch and Searle (2001),

$$l_m(C; Z) \propto -\frac{1}{2} \log |C| - \frac{n}{2} \log (Z' C^{-1} Z). \quad (92)$$

To maximise the marginal log-likelihood, we can use any practical method of optimisation, see Fletcher (1987) for examples. Here, as in Section 9, we choose to use a Nelder-Mead optimisation algorithm.

The data are given in in Figure 19. The bubble patches are centred at the locations.

We model the spatial latent process $\eta(s)$ for each model as described in Section 9, using 10 splines. The estimated values of $\log \eta_i$'s using the Karhunen-Loève and the Fourier representations are plotted in Figure 20, in the right and left bubble patch plots respectively. Sites with values
Figure 19: Plot 1: Bubble patch of the measurements of concentrations at the $n = 60$ data locations, Section 10.

Figure 20: From the left, Plot 1: Bubble patch of log $\eta_i$'s for the non-stationary model using the Karhunen-Loève expansion. Plot 2: Bubble patch of log $\eta_i$'s for the non-stationary model using the Fourier representation., Section 10.
of log \( \eta \) greater than 0 correspond to a spatial structure exhibiting greater smoothness than in the stationary case. Conversely, those with values smaller than 0 reveal less smoothness than in the initial stationary case.

It is clear from both plots in Figure 20 that our model suggests non-stationarity. Indeed, on inspection of Plots 1 and 2 in Figure 20 two regions stand out. The first region (region 1, say) is roughly defined as \( \{(x, y) \in [0, 40] \times [0, 60]\} \) where the small values of log \( \eta(s) \) compared to the other region suggest that this area has much higher frequency than the rest of the field and thus different properties in its spatial structure. On the other hand, the region (region 2, say) defined as \( \{(x, y) \in [70, 100] \times [0, 100]\} \) exhibits greater smoothness than in the stationary case. The rest of the field appears to model the transition between these two regions, the values of log \( \eta(s) \) increasing from region 1 to region 2. These conclusions appear to be supported by the data in Figure 19.

It is also important to point out that both models support the same conclusions with respect to the covariance structure of the residuals, although the values of log \( \eta(s) \) are different from one model to another as one would expect since the basis functions used are different for both models. It is also interesting to note that the conclusions drawn by our method are quite similar to that of Higdon et al. (1999) regarding this data set.

Finally, we show that our method is able to capture spatially varying anisotropy through three examples, corresponding to the correlation functions taken around points A, B, C and D shown in Figure 19. Points A and B correspond to measurement sites while C and D were taken at locations different from the data points. For the Karhunen-Loève model, correlation functions are obtained by predicting the eigenfields, Section 8.2, while their derivation is straightforward using the Fourier representation. The corresponding kernels for both representations are given in Figure 21.

Plots A, B, C and D in Figure 21 for both representations show our model is able to capture asymmetric second-order structures at different locations, thus exhibiting spatial adaptiveness. The correlation kernel around B is the most smooth, while around D the correlation drops of rapidly, as expected from examining Figure 20. Moreover, when comparing Plots A, B and C, it is clear that the covariance structures obtained from either representation exhibit the same features. That is, at any given location, the spatial dependence is stronger in directions corresponding to larger values of log \( \eta(s) \) since these directions correspond to increasing smoothness. Conversely, the spatial dependence decreases rapidly in directions corresponding to small values of \( \eta(s) \), thus revealing high
Figure 21: From the left, Top row plots: Correlation structure around points A, B, C and D using the Karhunen-Loève representation. From the left, Bottom row plots: Correlation structure around points A, B, C and D using the Fourier representation, Section 10.

frequency. These characteristics explain how our model is able to capture locally varying anisotropy features in the spatial dependence, as well as how to interpret the values and the behaviour of the \( \eta(s) \) process. The differences in the covariance function for both representations around Plot D appear somewhat more marked than for points A, B and C, although they both exhibit the same spatial dependence structure (high frequencies to the left and lower ones to the right, as expected).

It is important to note that we do not impose any particular a priori direction on the anisotropy structure but rather let the data dictate in which direction such anisotropy occurs. With respect to the computational time, the whole study, that is obtaining a “best fit” stationary matrix and estimating the coefficients of the \( \eta(s) \) process and the nugget effect, took us less than 10 minutes on a standard Unix machine for each representation.

11 Comparisons between the two methods

In this section we provide the reader with a discussion of the relative merits between the two proposed expansions of the covariance function. That is, we wish to determine the advantages and the drawbacks of each method with respect to the other.
11.1 The advantages of the Fourier approach over the Karhunen-Loève expansion

The Fourier representation differs from the Karhunen-Loève expansion in a number of key respects.

- Firstly, as already pointed out in Section 8.2, a major drawback of the Karhunen-Loève expansion is that the eigenvectors depend not only on the structure of the covariance function but also on the locations of the measurement sites. As a consequence, the method depends on the particular application at hand. In light of these difficulties, the Fourier approach offers two important advantages in that the basis functions are the same in all applications and the Fourier basis has stronger interpretation in terms of frequencies. This allows comparisons across applications and stationary counterparts, which is not the case with the Karhunen-Loève expansion.

- Unlike the Karhunen-Loève basis eigenvectors, the Fourier basis is analytic. Thus, predictions are straightforward and do not require integral approximation methods.

- The resulting non-stationary covariance functions have a complete analytical form similar to their stationary counterparts but with localised parameters. This provides an important feature as it allows one to work directly with the covariance functions without needing to consider the Fourier domain used to derive them. This is of particular importance for non-technical users.

11.2 The advantages of the Karhunen-Loève expansion over the Fourier representation

In some situations, it may be more useful for one to use a Karhunen-Loève representation. For instance:

- In the case when one has repeated measurements (e.g. meteorological applications, see Section 2) and can thus obtain a sample non-stationary covariance matrix for the data, then the Karhunen-Loève representation and the corresponding tempering methodology still apply, which is not the case with the Fourier representation. Indeed, in the latter one needs an initial stationary covariance function in order to determine the initial spectrum and to apply
the method to it. On the other hand, the Karhunen-Loève expansion can be applied to any valid covariance matrix, whether stationary or not. Of course, it must be pointed out that if repeated measurements are available then one may wish to adapt the Fourier representation method to tackle non-stationary spatial-temporal structures. This is the subject of further work.

- One might prefer to use the Karhunen-Loève expansion when only a numerical approximation of the Fourier spectrum or of the integral (53) is available, e.g. with the spherical covariance function for instance.

- The Karhunen-Loève expansion can be applied even if the Gaussian assumption is dropped. This ensures the method can be generalised to other fields of applications.

- The Karhunen-Loève expansion can be modified to allow for rank-reduced methods (Williams and Seeger 2001), as already mentioned in Section 8. This is an important feature from a practical point of view since one of the main drawbacks of most geostatistical methods for handling non-stationary or even stationary data is their computational cost, which often limits quite severely the number of applications one is able to tackle and their complexity.

- Finally, although we have used the Karhunen-Loève expansion, the method could be modified to account for other possible basis functions, a natural choice being wavelets, Demmler-Reinsh splines or orthogonal polynomials for instance. Thus the method is quite general and although perhaps less appealing than the Fourier representation from a mathematical point of view, provides useful practical properties.

Finally, it is perhaps interesting to note that in the very specialised case of a covariance matrix which is (block-)Toeplitz circulant, see for example Davis (1998), then the approach via the Karhunen-Loève expansion and the Fourier basis are equivalent. The reason for this lies in the fact that the Karhunen-Loève expansion performed on Toeplitz circulant matrices leads to the (discrete) Fourier basis and hence the eigenvalues are the (discrete) Fourier spectrum. These matrices arise when one has equally spaced data on a torus (i.e. the data is equally spaced and can be assumed periodic). This analogy is of minor practical interest in geostatistics however since periodicity is a condition that rarely holds in spatial applications.
12 Discussion

In this paper, we have presented a framework for constructing single-parameter non-stationary Gaussian spatial models. The non-stationarity is introduced via a latent spatial process $\eta(s)$ which is used to evolve the stationary spectrum over space. Two methods are described, which correspond to parametric and non-parametric spectra respectively. We show that the resulting non-stationary models have analytic covariance structures of the same parametric form as the stationary case but now with interpretable localised parameters. We have also shown that our method could provide an interesting device for implementing regressions with coefficients evolving spatially through their variance components, thus leading to variance-varying coefficients models which show adaptive shrinkage.

Future work will consider alternative models for $\eta(s)$, possible rank-reduced methods for improvement of the computational cost, and the application of the method to non-stationary spatial-temporal processes.

References


Supporting Information

A Proofs of the Results in Section 4.

A.1 Proof of Lemma 4.1

The necessary condition is clear from the definition of a positive-definite function. With respect to the sufficient condition, we must show that the function $C_{NS}(s, t)$ defined on $T \times T$ by (36) is positive-definite and valid for all $s, t \in D$. The positive-definiteness is clear since we have, for all integer $n$, all real numbers $c_1, \ldots, c_n$ and all $s_1, \ldots, s_n \in D$,

$$
\sum_{i,j=1}^{n} c_i c_j C_{NS}(s_i, s_j) = \int_{\Omega} \sum_{i,j=1}^{n} c_i c_j \exp\{i \omega (s_i - s_j)\} f(\omega, s_i; \theta(s_i))^1/2 f(\omega, s_j; \theta(s_j))^1/2 d\omega \quad (1)
$$

$$
= \int_{\Omega} \sum_{i=1}^{n} c_i \exp\{i \omega s_i\} f(\omega, s_i; \theta(s_i))^1/2 \sum_{j=1}^{n} c_j \exp\{-i \omega s_j\} f(\omega, s_j; \theta(s_j))^1/2 d\omega 
$$

$$
= \int_{\Omega} \left| \sum_{i=1}^{n} c_i \exp\{i \omega s_i\} f(\omega, s_i; \theta(s_i))^1/2 \right|^2 d\omega 
$$

$$
\geq 0
$$

Condition (37) follows for the validity of $C_{NS}(\cdot, \cdot)$ from the inequality of Cauchy-Schwartz, that is,

$$
|C_{NS}| = \left| \int_{\Omega} \exp\{i \omega (s - t)\} f(\omega, s; \theta(s))^1/2 f(\omega, t; \theta(t))^1/2 d\omega \right| \quad (2)
$$

$$
\leq \sqrt{\int_{\Omega} |\exp\{i \omega s\} f(\omega, s; \theta(s))^1/2|^2 d\omega} \sqrt{\int_{\Omega} |\exp\{-i \omega t\} f(\omega, t; \theta(t))^1/2|^2 d\omega}
$$

$$
= \sqrt{\int_{\Omega} f(\omega, s; \theta(s)) d\omega} \sqrt{\int_{\Omega} f(\omega, t; \theta(t)) d\omega}
$$

since we consider real spectra here. This result can however be easily generalised to complex functions.

A.2 Proof of the Theorem on the Gaussian correlation function, Section 4.2.

Proof: We give the proof for the one-dimensional case (i.e. $N=1$), the generalisation to $N > 1$ being straightforward. Consider the usual Gaussian correlation function,

$$
C(s, t) = \exp[-(s-t)^2/\alpha], \forall s, t \in \mathbb{R}, \alpha > 0.
$$

1
Then the corresponding spectral density function is given by
\[ f(\omega; \alpha) = (4\pi/\alpha)^{-1/2} \exp[-\alpha\omega^2/4], \quad \forall \omega \in \Omega = \mathbb{R}, \]  
(4)

Now, following the description of Section 4.1, we consider the function
\[ f(\omega, s, t; \alpha(s), \alpha(t)) = h(s) h(t) f(\omega, s; \alpha(s))^{1/2} f(\omega, t; \alpha(t))^{1/2} \]
\[ = h_{s,t} \exp\{-\omega^2/(4\beta_{s,t})\} \]
(5)

where \( \beta_{s,t} \) and \( h_{s,t} \) are defined by
\[ \beta_{s,t} = 2/(\alpha(s) + \alpha(t)) \]
(6)
\[ h_{s,t} = (16\pi^2/(\alpha(s)\alpha(t)))^{-1/4} \]
(7)

where \( \alpha(t) \) is assumed strictly positive for all \( t \in D \). Now, following Lemma 4.1, in order for the Theorem to be true one must show that \( C_{NS}(\cdot, \cdot) \) given by
\[ C_{NS}(s, t) = \int_\Omega \exp\{i\omega(s-t)\} f(\omega, s, t; \alpha(s), \alpha(t)) d\omega \]
(8)
is a valid covariance function if \( \alpha(t) > 0 \) for all \( t \in \mathbb{R} \). That is, if the function \( \omega \in \Omega \rightarrow f(\omega, t; \alpha(t)) = \exp\{-\omega^2/(4\alpha(t))\} \) is integrable with respect to the Lebesgue measure for all \( \alpha(t) > 0 \). This is true since
\[ u \in \mathbb{R} \rightarrow \exp[-\beta u^2], \]
is integrable with respect to the Lebesgue measure \( \forall \beta > 0 \).

It remains only to be shown that an analytical expression can be obtained for (8) and that it is given by (41). We have \( \forall \ s, \ t \in \mathbb{R}, \)
\[ C_{NS}(s, t) \propto \int_\Omega \exp\{i\omega(s-t)\} \frac{h(s)h(t)}{(16\pi^2/(\alpha(s)\alpha(t)))^{1/4}} \exp\{-\omega^2/(4\beta_{s,t})\} d\omega \]
(9)
\[ \propto \frac{(4\pi\beta_{s,t})^{1/2}}{(16\pi^2/(\alpha(s)\alpha(t)))^{1/4}} \int_\Omega \exp\{i\omega(s-t)\}(4\pi\beta_{s,t})^{-1/2} \exp\{-\omega^2/(4\beta_{s,t})\} d\omega \]
(10)
\[ \propto \frac{(4\pi\beta_{s,t})^{1/2}}{(16\pi^2/(\alpha(s)\alpha(t)))^{1/4}} \exp[-\beta_{s,t}(s-t)^2] \]
(11)
\[ \propto \left[\frac{\alpha(s)\alpha(t)}{\alpha(s) + \alpha(t)}\right]^{1/4} \exp[-\beta_{s,t}(s-t)^2] \]
(12)
\[ \propto \left[\frac{\alpha(s)\alpha(t)}{\alpha(s) + \alpha(t)}\right]^{1/2} \exp[-\beta_{s,t}(s-t)^2] \]
(13)
where $\beta_{s,t}$ is given by (6). The passage from (10) to (11) is due to the fact that we know, from the stationary case, that
\[
\int_{\Omega} \exp[-i\omega(s-t)](16\pi^2\omega)^{-1/2}\exp[-\alpha\omega^2/4]d\omega = \exp[-(s-t)^2/\alpha].
\]
Finally, we wish to enforce the covariance function to have constant variance one at all locations. We thus choose the constant of proportionality accordingly which yields
\[
C_{NS}(s,t) = 2^{1/2} \frac{[\alpha(s)\alpha(t)]^{1/4}}{[\alpha(s) + \alpha(t)]^{1/2}} \exp[-\beta_{s,t}(s-t)^2]
\]
and the result is proved for $N = 1$.

The generalisation to the N-dimensional case for $N > 1$ is straightforward since the N-dimensional spectral density function of the correlation function
\[
C(s,t) = \exp[-||x - y||^2_N/\alpha], \forall s, t \in \mathbb{R}^N,
\]
\[
(14)
\]
is given by
\[
f(\omega) = (4\pi/\alpha)^{-N/2} \exp[-\alpha^N/4], \omega \in \Omega = \mathbb{R}^N.
\]
\[
(15)
\]
It can be seen from (15) that it is possible to do the calculation in each direction separately. And, by applying the one-dimensional case result described above to each $\omega_i, s_i, t_i, i = 1, ..., N$, we obtain (41), which completes the proof.

A.3 Proof of the Theorem on the stationary Matérn covariance function, Section 4.2.

Consider the following N-dimensional function
\[
f_{NS}^{s,t}(\omega) = h(t)h(s)(\alpha^2 + ||\omega||_N^2)^{-\nu_{s,t}-N/2}, \omega \in \Omega = \mathbb{R}^N,
\]
\[
(16)
\]
where the function $h(t)$ is bounded on $T$ and where we define $\nu_{s,t}$ to be
\[
\nu_{s,t} = (\nu_s + \nu_t)/2
\]
\[
(17)
\]
We know that $\forall s \in \mathbb{R}^N$, the function
\[
\omega \in \Omega = \mathbb{R}^N \rightarrow \int_{\Omega} \exp[i\omega s](\alpha^2 + ||\omega||_N^2)^{-b}d\omega,
\]
\[
(18)
\]
is integrable on $\Omega$ with respect to the Lebesgue measure if and only if $b > N/2$. Hence, if $\nu(t) > 0$ for all $t \in D$, then $C_{NS}(\cdot, \cdot)$ given by

$$C_{NS}(s, t) = \int_{\Omega} \exp [i\omega'(s - t)] f_{NS}(\omega) d\omega$$

(19)

is a valid covariance function. We now determine the analytical form of $C_{NS}(s, t)$.

We know that the covariance function corresponding to the spectral density function

$$f(\omega) = (\alpha^2 + |\omega|)^{\nu - N/2}$$

(20)

is given by

$$C(s, t) = \frac{\pi^{N/2}}{2^{\nu-1} \Gamma(\nu + N/2) \alpha^{2\nu}} (|s - t|)^\nu K_\nu(\alpha |s - t|), \forall \, s, \, t \in \mathbb{R}^N.$$  

(21)

Therefore, recalling (16) and using the fact that $\nu_{s,t}$, given by (17) is such that $\nu_{s,t} > 0$ since $\nu(t) > 0$ for all $t \in D$, we have

$$C_{NS}(s, t) = \frac{h(s)h(t)\pi^{N/2}}{2^{\nu_{s,t}-1} \Gamma(\nu_{s,t} + N/2) \alpha^{2\nu_{s,t}}} (|s - t|)^{\nu_{s,t}} K_{\nu_{s,t}}(\alpha |s - t|), \forall \, s, \, t \in \mathbb{R}^N.$$  

(22)

Therefore, $C_{NS}(s, t)$ given by (19) is a valid non-stationary covariance function if $\nu(t) > 0$ for all $t \in D$, and its analytical form is given by (22). Thus, the proof is completed.

If one wishes to impose that $C_{NS}(s, t)$ must have constant variance one, then one must choose a particular form for the function $h(t), \, t \in \mathbb{R}^N$. Calculations show that one must choose $h(t)$ to be equal to

$$h(t) = \left( \frac{\pi^{N/2} \Gamma(\nu_{s,t})}{\alpha^{2\nu_{s,t}}} \right)^{-1/2}$$

(23)

which leads to a non-stationary correlation matrix derived from the Matérn class of stationary covariance functions.

B Proofs of the Results in Section 5.

B.1 Proof of Lemma 5.1

To begin, it will help in what follows to note that if we assume that the spectral density function $f(\omega), \, \omega \in \Omega = \mathbb{R}$ is real, then the expansion of the stationary process $X(t)$ in (7) can be re-written
in the form
\[ X(t) = \int_\Omega \exp [i\omega t] f(\omega)^{1/2} dW^X(\omega), \quad t \in T, \]
with \( E[dW^X(\omega)dW^X(\omega')] = \delta(\omega, \omega') d\omega \). Using our suggested method, we are able to build a non-stationary process \( X_{NS}(t) \), whose corresponding expansion is
\[ X_{NS}(t) = \int_\Omega \exp [i\omega t] h(t)f(\omega)^{\eta(t)/2} dW^X(\omega), \quad t \in T, \]
where \( h(t), \ t \in T \) is a bounded function, so that it is clear that the variance of the coefficients of the expansion are now varying with \( t \in \mathbb{R} \).

More generally, we are lead to consider models of the following form. Let \( f(\omega), \omega \in \Omega \) be the spectral density function of a stationary process. We propose a non-stationary generalisation of the process as
\[ X_{NS}(t) = \int_\Omega \exp [i\omega t] dZ^X_t(\omega), \quad t \in T, \quad (24) \]
with
\[ E[dZ^X_t(\omega)dZ^X_s(\omega')] = h(t) h(s)f(\omega)^{\eta(t)+\eta(s)/2} \delta(\omega, \omega') d\omega, \ \omega, \omega' \in \Omega, \quad (25) \]
where again \( h(t) \) is some bounded function on \( T \) and \( \eta(t), \ t \in T \) is a strictly positive stochastic process. Note that \( h(t) \) can be set in order to ensure constant power in the process.

In order to prove that under certain stated conditions, models given by (24)-(25) lead to valid non-stationary covariance functions, we first consider the class \( C \) of second-order stochastic processes with covariance functions \( C(s, t), \ s, t \in T \) having a representation of the form:
\[ C(s, t) = \int_\Omega \phi(s, \omega) \hat{\phi}(t, \omega) d\mu(\omega), \quad s, t \in T \quad (26) \]
for some \( \mu \), a \( \sigma \)-finite measure on a \( \sigma \)-algebra in \( \Omega \), a measurable space (Loève 1955). Note that since the processes are assumed to be of second order, the function \( \omega \in \Omega \rightarrow \phi(t, \omega) \in \mathbb{C} \) must be square-integrable with respect to the measure \( \mu \) for all \( t \in T \) in order for (26) to be valid.

A result given by Grenander and Rosenblatt (1957) shows that there is an equivalence between the class \( C \) and the class \( \mathcal{I} \) of processes which have a Fourier-Stieltjes integral representation of the
form:

\[ X(t) = \int_{\Omega} \phi(t, \omega) dZ(\omega), \ t \in T, \] \hspace{1cm} (27)

\[ E[dZ(\omega) dZ(\omega')] = \delta(\omega, \omega') d\mu(\omega), \ \omega, \omega' \in \Omega, \]

This result was first stated by Karhunen (1947) and the proof given by Grenander and Rosenblatt (1957). The proof is essentially that given by Cramér (1951). The class \( \mathcal{C} \) contains all second-order stationary processes. Clearly, if one considers the representation (26) with \( \phi(t, \omega) = \exp[i\omega t], \ \forall t \in T, \ \forall \omega \in \Omega \) then one obtains the Fourier representation in (6). Priestley’s oscillatory processes (Priestley 1965) are also a special case of (26) in which \( \phi(t, \omega) = A_r(\omega) \exp[i\omega t], \ \forall t \in T, \ \forall \omega \in \Omega \) with \( A_r(\omega) \) a smooth function in \( t \in T \). Finally, Higdon et al. (1999) also use (26) to model non-stationary spatial processes although their approach is somewhat different since they do not consider the decomposition of the covariance function in the frequency domain but rather in the original space. Thus, in order to prove the validity of models of the form (24)-(25), we simply need to show under what conditions the model (24)-(25) belongs to the class \( \mathcal{C} \) and thus leads to a valid non-stationary second-order stochastic process.

To achieve this we first take

\[ \phi_r(\omega) = \exp[i\omega t] h(t) f(\omega) \eta(t)^l, \ \omega \in \Omega, \ t \in T. \] \hspace{1cm} (28)

Then, in order for the covariance function

\[ C(s, t) = \int_{\Omega} \phi_r(\omega) \overline{\phi_r}(\omega) d\omega \] \hspace{1cm} (29)

to be valid, the following conditions must be satisfied (Grenander and Rosenblatt 1957):

- \( C(s, t), \ s, t \in T \), must be a positive definite function.

- \( |C(s, t)|, \ s, t \in T \) must be finite. A sufficient condition is thus that \( \phi_r(\omega) \) must be quadratically integrable with respect to the Lebesgue measure \( \forall t \in T \).

- the measure \( d\omega \) must be \( \sigma \)-finite.
Because the measure $d\omega$ is the Lebesgue measure, it is clearly $\sigma$-finite and hence the third condition is satisfied. The first condition is also satisfied since:

$$\forall c_i, c_j \in \mathbb{R}, \ i, j \in [1, n], \sum_i \sum_j c_i c_j C(s_i, s_j) = \int_\Omega \sum_i c_i \phi_{s_i}(\omega) \sum_j c_j \phi_{s_j}(\omega) d\omega$$

$$= \int_\Omega \left| \sum_i c_i \phi_{s_i}(\omega) \right|^2 d\omega$$

$$\geq 0.$$ 

Therefore, it only remains to show that the second condition is satisfied, that is that

$$\int_\Omega |\phi(\omega)|^2 d\omega < +\infty$$

(31)

And, from (24)-(25) we have

$$\int_\Omega |\phi(t)|^2 d\omega = \int_\Omega |f(\omega)|^2(t) d\omega, \ t \in T$$

(32)

and thus the result follows for the sufficient condition. The necessary condition is straightforward if $C_{NS}(\cdot, \cdot)$ is a valid covariance function. And thus the result is proved.

C Spectral tempering applied to some common used geostatistical stationary covariance functions

In this section, we derive the non-stationary analytical covariance models obtained when spectral tempering is applied to the parametric spectra of some common geostatistical stationary covariance functions. Although, as was pointed out in Section 4, spectral tempering and spatial adaptation of the spectrum are closely linked when the log-spectrum is linear in the parameters, it is interesting to note the differences between the non-stationary Gaussian and Matèrn covariance functions obtained via spectral tempering with their counterparts given in Section 4 and derived in Appendix A.

C.1 The non-stationary Gaussian correlation function obtained via spectral tempering

Theorem: The function $C_{NS}(s, t), \ s, t \in D$ given by

$$C_{NS}(s, t) = D_{s, t} \exp\{-\beta_{s, t}||s - t||^2_N\}, \ \forall \ s, t \in D$$

(33)
where the local parameters $\beta_{s,t}$ and $D_{s,t}$ are given by

$$\beta_{s,t} = 2\alpha/\eta(s) + \eta(t),$$

$$D_{s,t} = 2^{N/2} \frac{[\eta(s) + \eta(t)]^{N/4}}{\eta(s) + \eta(t)}$$

is a valid non-stationary spatial correlation function obtained by spectral tempering the stationary Gaussian correlation function with parameter $\alpha$ for all $t \in D$.

**Proof:** We are first going to give the proof in the one-dimensional case (i.e. $N=1$), the generalisation to $N > 1$ being straightforward.

Consider the usual Gaussian correlation function,

$$C(s, t) = \exp \left[ -\alpha (s - t)^2 \right], \forall s, t \in \mathbb{R}, \alpha > 0. \quad (36)$$

Then the corresponding spectral density function is given by:

$$f(\omega) = (4\pi\alpha)^{-1/2} \exp \left[ -\omega^2/4\alpha \right], \forall \omega \in \Omega = \mathbb{R}, \quad (37)$$

Now, following the description of spectral tempering given in Section 5, we consider the function:

$$f_{NS}^{s,t}(\omega) = h(s)h(t)f(\omega)^{(\eta(s) + \eta(t))/2}, \quad (38)$$

where $h : \mathbb{R} \to \mathbb{R}$ is a bounded function on $\mathbb{R}$. It is easy to see that the function

$$C_{NS}(s, t) = \int_{\Omega} \exp \left[ i\omega (s - t) \right] f_{NS}^{s,t} d\omega \quad (39)$$

is a valid covariance function. Indeed following the lemma given in Section 5.1, it suffices to show that $\forall t \in \mathbb{R}$ such that $\eta(t) < 1$, the function $\omega \in \Omega \to f(\omega)^{\eta(t)}$ is integrable with respect to the Lebesgue measure, which results from the fact that

$$u \in \mathbb{R} \to \exp \left[ -\beta u^2 \right],$$

is integrable with respect to the Lebesgue measure $\forall \beta > 0$.

We need only show now that an analytical expression can be obtained for (39) and that it is
given by (33). We have \( \forall s, \ t \in \mathbb{R} \),

\[
C_{NS}(s, t) = \int_\Omega \exp[i\omega(s - t)] \frac{h(s)h(t)}{(4\pi\alpha)^{\eta(s) + \eta(t)/2}} \exp[-\omega^2/(4\beta_{s, t})]d\omega \tag{40}
\]

\[
= \frac{h(s)h(t)(4\pi\beta_{s, t})^{1/2}}{(4\pi\alpha)^{\eta(s) + \eta(t)/2}} \int_\Omega \exp[i\omega(s - t)](4\pi\beta_{s, t})^{-1/2} \exp[-\omega^2/(4\beta_{s, t})]d\omega \tag{41}
\]

\[
= \frac{h(s)h(t)(4\pi\beta_{s, t})^{1/2}}{(4\pi\alpha)^{\eta(s) + \eta(t)/2}} \exp[-\beta_{s, t}(s - t)^2] \tag{42}
\]

where \( \beta_{s, t} \) is given by

\[
\beta_{s, t} = 2\alpha/(\eta(s) + \eta(t)). \tag{43}
\]

The passage from (41) to (42) is due to the fact that we know, from the stationary case, that

\[
\int_\Omega \exp[-i\omega(s - t)](4\pi\omega)^{-1/2} \exp[-\omega^2/4\alpha]d\omega = \exp[-\alpha(s - t)^2].
\]

Finally, we wish to enforce \( C_{NS}(s, t) \) to have constant power at all locations. We therefore choose

\[
h(t) = (C_{NS}(t, t))^{-1/2} = \frac{(4\pi\alpha)^{\eta(t)/4}}{(4\pi\alpha/\eta(t))^{1/4}}, \ \forall t \in \mathbb{R}.
\]

And, replacing \( h(t) \) and \( h(s) \) in (42) yields (33) for \( N = 1 \).

The generalisation to the \( N \)-dimensional case for \( N > 1 \) is straightforward since the \( N \)-dimensional spectral density function of the correlation function

\[
C(s, t) = \exp[-\alpha||x - y||^2_N], \ \forall s, \ t \in \mathbb{R}^N, \tag{44}
\]

is given by

\[
f(\omega) = (4\pi\alpha)^{-N/2} \exp[-\omega'/(4\alpha)], \ \omega \in \Omega = \mathbb{R}^N. \tag{45}
\]

It can be seen from (45) that it is possible to do the calculation in each direction separately. And, by applying the one-dimensional case result described above to each \( \omega_i, s_i, t_i, \ i = 1, \ldots, N \), we obtain (33), which completes the proof.
C.2  The non-stationary Matern correlation function obtained via spectral tempering

Theorem: The function $C_{NS}(s, t), \ s, t \in T$ given by

$$C_{NS}(s, t) = D_{s, t}(\alpha||s - t||_N)^{\nu_{s, t}K_{\nu_{s, t}}(\alpha||s - t||_N), \ s, t \in D}$$  \hspace{1cm} (46)

with local parameters $\nu_{s, t}$ and $D_{s, t}$ given by

$$\nu_{s, t} = \frac{[\eta(s) + \eta(t)]}{2}(\nu + N/2) - N/2, \hspace{1cm} (47)$$

$$D_{s, t} = \frac{h(t)h(s)\pi^{N/2}}{2^{\nu_{s, t}}\Gamma(\nu_{s, t} + N/2)\alpha^{2\nu_{s, t}}} \hspace{1cm} (48)$$

where $h(\cdot)$ is a bounded function on $D$, is a valid covariance function obtained by spectral tempering of the stationary Matern covariance function with parameter $\nu$, if and only if,

$$\eta(t) > N/(2\nu + N) \hspace{1cm} (49)$$

for all $t \in D$.

Proof: Similarly to the Gaussian case, spectral tempering of the spectrum of the stationary Matern covariance function with parameters $\nu$ and $\alpha$ leads to the following $N$-dimensional spectrum for the second-order structure between $s$ and $t$,

$$f_{NS}^{s,t}(\omega) = h(t)h(s)(\alpha^2 + ||\omega||_N^2)^{-\nu_{s, t} - N/2}, \ \omega \in \Omega = \mathbb{R}^N, \hspace{1cm} (50)$$

where the function $h(t)$ is bounded on $D$ and where we define $\nu_{s, t}$ to be

$$\nu_{s, t} = \frac{[\eta(s) + \eta(t)]}{2}(\nu + N/2) - N/2, \hspace{1cm} (51)$$

for $\nu > 0$. We know that $\forall \ s \in \mathbb{R}^N$, the function

$$\omega \in \Omega = \mathbb{R}^N \rightarrow \int_{\Omega} \exp[i\omega \cdot s](\alpha^2 + ||\omega||_N^2)^{-b}d\omega, \hspace{1cm} (52)$$

is integrable on $\Omega$ with respect to the Lebesgue measure if and only if $b > N/2$. Hence, $\forall t \in \mathbb{R}^N, \ \forall \nu > 0$, the function

$$\omega \in \Omega = \mathbb{R}^N \rightarrow \exp[i\omega \cdot t](\alpha^2 + ||\omega||_N^2)^{\eta(t)(\nu + N/2)} \hspace{1cm} (53)$$
is integrable with respect to the Lebesgue measure on $\Omega$ if and only if

$$
\eta(t)(\nu + N/2) > N/2 \iff \eta(t) > N/(2\nu + N), \quad \forall \, t \in \mathbb{R}^N.
$$

(54)

As a result, by using the Lemma in Section 5.1, we can state that the function

$$
C_{NS}(s, t) = \int_{\Omega} \exp[\imath \omega^T (s - t)] f_{NS}^t(\omega) d\omega
$$

is a valid covariance function if and only if $\eta(t) > N/(2\nu + N)$, $\forall \, t \in \mathbb{R}^N$. We now determine the analytical form of $C_{NS}(s, t)$.

We know that the covariance function corresponding to the spectral density function

$$
f(\omega) = (\alpha^2 + ||\omega||^2)\nu^\nu (\nu - N/2)
$$

is given by

$$
C(s, t) = \frac{(\pi)^{N/2}}{2^{\nu - 1} \Gamma(\nu + N/2) \alpha^{2\nu}} (\alpha|s - t|)^\nu K_\nu(\alpha|s - t|), \quad \forall \, s, \, t \in \mathbb{R}^N.
$$

(57)

Therefore, recalling (46) and using the fact that $\nu_{s, t}$, given by (47) is such that $\nu_{s, t} > 0$ since $\nu > 0$ and condition (49) is satisfied, we have

$$
C_{NS}(s, t) = \frac{h(s) h(t) (\pi)^{N/2}}{2^{\nu_{s, t} - 1} \Gamma(\nu_{s, t} + N/2) \alpha^{2\nu_{s, t}}} (\alpha|s - t|)^{\nu_{s, t}} K_{\nu_{s, t}}(\alpha|s - t|), \quad \forall \, s, \, t \in \mathbb{R}^N.
$$

(58)

Therefore, $C_{NS}(s, t)$ is a valid non-stationary covariance matrix, if and only if condition (54) is satisfied and its analytical form is given by (58). The proof is completed.

If one wishes to impose that $C_{NS}(s, t)$ be a correlation matrix, then one must choose a particular form for the function $h(t)$, $t \in \mathbb{R}^N$. Calculations show that one must choose $h(t)$ to be equal to

$$
h(t) = \left( \frac{(\pi)^{N/2} \Gamma(\mu_{t})}{\alpha^{2\mu_{t}} \Gamma(\mu_{t} + N/2)} \right)^{-1/2}
$$

(59)

which leads to a non-stationary correlation matrix derived from the Matérn class of stationary covariance functions.

D The separable correlation function

In this section, we apply both methods (i.e. spatial adaptation and spectral tempering) to the parametric spectrum obtained from a stationary separable covariance function. This example is
instructive since it corresponds to a case where the parametric log spectrum is not a linear function of its parameters. Moreover, we show using this example that although less appealing than direct spatial adaptation at first sight, spectral tempering can lead to fully analytical tractable models in situations where direct spatial adaptation can’t. This highlights the flexibility of spectral tempering as a general method for evolving spectra spatially, whether parametric or not.

D.1 The non-stationary separable correlation function obtained via direct spectral adaptation

Consider the stationary separable exponential correlation function given by

\[ C(s, t) = \exp[-\alpha_1 |s_1 - t_1| - \cdots - \alpha_N |s_N - t_N|], \quad s, t \in \mathbb{R}^N, \quad N \geq 1. \]  

(60)

Its spectrum has the form

\[ f(\omega) = \pi^{-N} \prod_{i=1}^{N} \left[ \frac{\alpha_i}{(\omega_i^2 + \alpha_i^2)^2} \right], \quad \omega \in \Omega = \mathbb{R}^N, \]  

(61)

where \( \omega = (\omega_1, \ldots, \omega_N) \). Applying spatial adaptation to (61) by evolving the parameter \( \alpha_i \), \( i = 1, \ldots, N \) over the space, i.e. \( \alpha_i = \alpha_i(s), \quad \forall i = 1, \ldots, N \) leads to

\[ f(\omega, s; t; \alpha(s), \alpha(t)) = \pi^{-N} \prod_{i=1}^{N} \left[ \frac{\alpha_i(s)\alpha_i(t)}{(\omega_i^2 + \alpha_i(s)^2)(\omega_i^2 + \alpha_i(t)^2)} \right]^{1/2}, \quad \omega \in \Omega \]  

(62)

From the form of (62), it is straightforward to check that the resulting covariance function \( C_{NS}(s, t), \quad s, t \in D \) is a valid positive definite function if \( \alpha_i(s) > 0 \) for all \( s \in D \) and is of the form,

\[ C_{NS}(s, t) = \int_\Omega \exp\{i\omega(s - t)\} \pi^{-N} \prod_{i=1}^{N} \left[ \frac{\alpha_i(s)\alpha_i(t)}{(\omega_i^2 + \alpha_i(s)^2)(\omega_i^2 + \alpha_i(t)^2)} \right]^{1/2} d\omega \]  

(63)

The expression given by (63) is hard to evaluate and is best approximated by numerical techniques. Unlike this case, one can obtain a fully analytical expression when spectral tempering rather than direct spatial adaptation is used, as is shown in the next section.

D.2 The non-stationary separable correlation function obtained via spectral tempering

Consider again the stationary separable exponential correlation function given by (60) and recall that its spectrum is of the form (61). Using spectral tempering of the spectrum as described in
Section 5, the localised spectral density function obtained is given by

\[ f_{NS}^{s,t}(\omega) = Cst_1(s, t) \prod_{i=1}^{N} \left( \frac{\alpha_i}{(\omega_i^2 + \alpha_i^2)} \right)^{\beta_{s,t}} \]  

(64)

where we have

\[ \beta_{s,t} = (\eta(s) + \eta(t))/2 \]

and where the function \( Cst_1(s, t) \), \( s, t \in T \) is given by

\[ Cst_1(s, t) = h(s)h(t)(\pi)^{-\frac{N}{2}(\eta(s) + \eta(t))} \]

(65)

where \( h(t) \), \( t \in T \) is a bounded function. Expression (61) leads to the following non-stationary correlation function:

\[ C_{NS}(s, t) = K(s, t) \prod_{i=1}^{N} \left[ (\alpha_i|s_i - t_i|)^{\frac{1}{2}\beta_{s,t}^i} K_{\frac{1}{2}\beta_{s,t}^i}(\alpha_i|s_i - t_i|) \right], \quad \forall s, t \in T, \]

(66)

where \( s = (s_1, ..., s_N), t = (t_1, ..., t_N), \alpha_i > 0, i = 1, ..., N \). The function \( K(s, t) \) is given by

\[ K(s, t) = \frac{Cst_1(s, t)2^{N(\frac{1}{2}-\beta_{s,t})}(\sqrt{\pi}^N(\alpha_1...\alpha_N)^{\beta_{s,t}})}{\Gamma(\beta_{s,t}^i)(\prod_{j=1}^{N} \alpha_j)^{2\beta_{s,t}^i-1}} \]

(67)

and the function \( h(t), t \in T \) is chosen so that \( C_{NS}(s, s) = 1 \) \( \forall t \in T \), that is we choose

\[ h(t) = \left( \pi^{-N/2} \left( \frac{\Gamma(\eta(t) - \frac{1}{2})}{\Gamma(\eta(t))} \right)^{N} \left( \prod_{j=1}^{N} \alpha_j \right)^{-2\eta(t)} \right)^{-\frac{1}{2}}. \]

(68)

Recall from (44) that the function \( K_{\nu} \) in (66) is the modified Bessel function of the second kind of order \( \nu \), which is symmetric about 0 in the \( \nu \) parameter. The reader should refer to Stein (1999) or Abramowitz and Stegun (1965) for more detail. We must still determine from (54) the conditions under which the function \( C_{NS}(s, t) \) is a valid correlation function. We then have the following result.

**Theorem:** The function \( C_{NS}(s, s), s, t \in T \) given by (66) is a valid correlation matrix if

\[ \forall t \in \mathbb{R}^N, \eta(t) > \frac{1}{2}. \]

(69)

**Proof:** Consider the following function in an N-dimensional space:

\[ f_{NS}^{s,t}(\omega) = Cst(s, t) \left[ \prod_{i=1}^{N} \left( \frac{1}{(\omega_i^2 + \alpha_i^2)} \right) \cdot \frac{1}{(\omega_N^2 + \alpha_N^2)} \right]^{\beta_{s,t}} \]

(70)
where \( Cst(s, t) \) is given by

\[
Cst(s, t) = Cst_1(s, t)(\alpha_1 ... \alpha_N)^{\beta_{s,t}}
\]

and \( Cst_1(s, t) \) is given by (65). The function (70) is obtained by the method described in Section 5, from the separable exponential correlation function. It is clear from (70) that \( f_{NS}^{st}(\omega) \) is separable in \( \omega \). That is, it can be written

\[
f_{NS}^{st}(\omega) = Cst(s, t) \prod_{i=1}^{N} g_i^{st}(\omega_i) \tag{71}
\]

where \( \omega = (\omega_1, ..., \omega_N) \) and

\[
g_i^{st}(\omega_i) = (\alpha_i^2 + \omega_i^2)^{\beta_{s,t}}, \forall i = 1, ..., N. \tag{72}
\]

Therefore, following the Lemma in section 5.1, in order to show that the function

\[
C_{NS}(s, t) = \int_{\Omega} \exp[i\omega(s-t)]f_{NS}^{st}(\omega)d\omega \tag{73}
\]

is a well defined covariance function, we need only show that the functions

\[
\omega_i \in \mathbb{R} \rightarrow \exp[i\omega t]g_i^{st}(\omega_i)
\]

are integrable on \( \mathbb{R} \) with respect to the Lebesgue measure \( \forall \, t \in \mathbb{R}^N \). It can be shown that the functions

\[
u \in \mathbb{R} \rightarrow (\alpha^2 + u^2)^{-b}, \quad \alpha > 0, \quad b \in \mathbb{R}, \tag{74}
\]

is integrable with respect to the Lebesgue measure on \( \mathbb{R} \) if and only if \( b > 1/2 \). Therefore, \( C_{NS}(s, t) \) is well defined \( \forall \, s, \, t \in \mathbb{R}^N \) if and only if \( \beta_{s,t} > \frac{1}{2} \), that is

\[
\forall \, t \in \mathbb{R}^N, \quad \eta(t) > \frac{1}{2}. \tag{75}
\]

We now determine the analytical expression of \( C_{NS}(s, t) \), under condition (75). Using the fact that for \( j = 1, ..., N, \)

\[
\int_{\mathbb{R}} \exp[i\omega_j(s_j - t_j)](\alpha_j^2 + \omega_j^2)^{\beta_{s,t}}d\omega_j = \frac{2^{\beta_{s,t}/2} - \beta_{s,t} \sqrt{\pi}}{\Gamma(\beta_{s,t}) \alpha_j^{2\beta_{s,t} - 1}}(\alpha_j|s_j - t_j|)^{\beta_{s,t} - 1/2}K_{\beta_{s,t} - 1/2}(\alpha_j|s_j - t_j|), \tag{76}
\]

```latex
\]
```
where \( t = (t_1, \ldots, t_N), \forall t \in \mathbb{R}^N \), we have

\[
C_{NS}(s, t) = \frac{Cst(s, t) 2^{N(3/2-\beta_s,t)}(\sqrt{\pi})^N}{(\Gamma(\beta_s,t))^{N}(\prod_{j=1}^{N} \alpha_j)^{\beta_s,t-1/2}} \prod_{i=1}^{N} [(\alpha_i|s_i - t_i|)^{\beta_s,t-1/2} K_{\beta_s,t-1/2}(\alpha_i|s_j - t_j|)]
\]  

(77)

Therefore, \( C_{NS}(s, t) \) is a valid non-stationary covariance matrix, whose analytical form is given by (77) if and only if condition (75) is satisfied.

If one wishes to impose that \( C_{NS}(s, t) \) be a correlation matrix, then one must choose a particular form for the function \( h(t), t \in \mathbb{R}^N \) in \( Cst(s, t) \). Calculations show that one must choose for all \( t \in \mathbb{R}^N, \)

\[
h(t) = \left( \pi^{N/2} \left( \frac{\Gamma(\eta(t)) - \frac{1}{\eta(t)}}{\Gamma(\eta(t))} \right)^N \left( \prod_{j=1}^{N} \alpha_j \right)^{1-2\eta(t)} \right)^{-\frac{1}{2}},
\]

which is equal to (68).

Note that when \( \forall \ i, \alpha_i < 1 \), an equivalent expression for (60) is given by

\[
C_{S}(s, t) = \alpha_1^{[s_1 - t_1]} \cdots \alpha_N^{[s_N - t_N]}, s, t \in \mathbb{R}^N, N > 1.
\]  

(78)

Therefore, our results for (66) continue to hold for (78). We have thus shown that one can obtain an analytic expression for \( C_{NS}(\cdot, \cdot) \) which was not the case for spatial adaptation, where the method lead to intractable calculations and the need for numerical approximations.

**E  Proof of the Results in Section 6**

**E.1  Proof of the Theorem**

*Theorem:* The matrix \( C_{\beta_n} \) so defined is positive definite for \( \rho_\theta(d) \) a monotonic decreasing function of \( 0 \geq d < \infty, \rho_\theta(0) = 1. \)

*Proof:* Recall that the matrix \( C_{\beta_n} \) is defined as:

\[
C_{\beta_n} = \begin{pmatrix}
D^{\eta_1} & \rho_\theta^{1,2} D^{(\eta_1+\eta_2)/2} & \cdots & \rho_\theta^{1,n} D^{(\eta_1+\eta_n)/2} \\
\rho_\theta^{2,1} D^{(\eta_1+\eta_2)/2} & D^{\eta_2} & \cdots & \rho_\theta^{2,n} D^{(\eta_2+\eta_n)/2} \\
\vdots & \ddots & \ddots & \vdots \\
\rho_\theta^{n,1} D^{(\eta_1+\eta_n)/2} & \cdots & \rho_\theta^{(n-1),n} D^{(\eta_{n-1}+\eta_n)/2} & D^{\eta_n}
\end{pmatrix},
\]

(79)
where

\[ \eta_i = \eta(s_i), \quad i = 1, \ldots, n, \]
\[ \rho_{ij}^{(s)} = \rho_{ij}(\|s_i - s_j\|^2), \quad i, j = 1, \ldots, n, \]

Recall also that \( C_{\beta_i\beta_j} = C_{\beta_i} = D \forall i, j = 1, \ldots, n \) where \( D \) is defined by (12). Now, note that we can write \( C_{\beta_n} \) in the following form:

\[ C_{\beta_n} = D_1 \cdot I_1 \cdot D_1, \quad (80) \]

where \( D_1 \) is defined as:

\[
D_1 = \begin{pmatrix}
D^{m/2} & 0 & \cdots & 0 \\
0 & D^{n_2/2} & 0 & \cdots \\
\vdots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & D^{m/2}
\end{pmatrix}
\quad (81)
\]

and \( I_1 \) is defined as:

\[
I_1 = \begin{pmatrix}
I_n & \rho_{ij}^{12} I_n & \cdots & \rho_{ij}^{1n} I_n \\
\rho_{ij}^{12} I_n & I_n & \cdots & \rho_{ij}^{2n} I_n \\
\vdots & \ddots & \ddots & \vdots \\
\rho_{ij}^{1n} I_n & \cdots & \rho_{ij}^{(n-1)n} I_n & \rho_{ij}^{nn} I_n
\end{pmatrix}
\quad (82)
\]

Now, if \( \rho(d) \) is a valid correlation function (i.e. monotonic decreasing function of \( 0 \geq d < \infty \), and \( \rho(0) = 1 \)) and \( \rho_{ij}^{(s)} \neq 1, \forall i, j, i \neq j, = 1, \ldots, n \), then the matrix \( I_1 \) is positive-definite (O’Hagan 1978). Hence, we can write \( I_1 \) in the form:

\[ I_1 = V_1 \cdot D_2 \cdot V_1', \quad (83) \]

where \( V_1 \) is an orthogonal matrix and \( D_1 \) is a diagonal matrix with all its elements strictly greater than zero and different. Now, because the matrix \( D_1 \) is a diagonal matrix with all its elements strictly greater than zero, it can be written in the form:

\[ D_2 = D_{21} \cdot D_{21}', \quad (84) \]

and hence,

\[ I_1 = (V_1 D_{21}) \cdot (V_1 D_{21})', \quad (85) \]
so that

\[ C_{\beta_n} = D_1 \cdot (V_1 D_{21}) \cdot (V_1 D_{21})' \cdot D_1 \quad (86) \]

and since \( D_1 \) is symmetric, we have

\[ C_{\beta_n} = (D_1 V_1 D_{21}) \cdot (D_1 V_1 D_{21})' \quad (87) \]

Now, one can immediately see that:

- \( \text{rank}(D_1) = n^2 \) since \( D_1 \) is a diagonal matrix with all its elements different from zero.
- \( \text{rank}(V_1) = n^2 \) since \( V_1 \) is an orthogonal matrix.
- \( \text{rank}(D_{21}) = n^2 \) since \( D_{21} \) is a diagonal matrix with all its elements different from zero.

Hence, \((D_1 V_1 D_{21})\) is a \( n^2 \times n^2 \) matrix of rank \( n^2 \). So if we call \( B = (D_1 V_1 D_{21})' \), then \( C_{\beta_n} = B' \cdot B' \)

where \( B \) is a \( n^2 \times n^2 \) matrix of rank \( n^2 \). We then use the following corollary:

Corollary: If \( B \) is a \( p \times n \) matrix of rank \( r \), then

1. \( B' B \) and \( B B' \) are non-negative matrices.
2. \( B' B \) is a positive semidefinite matrix if \( r < n \).
3. \( B' B \) is a positive definite matrix if \( r = n \).

As a consequence of part (3) of the Corollary, \( C_{\beta_n} \) is positive-definite, hence the theorem is proved.

E.2 Proof that the matrix \( Cov^{NS} \) in Section 6. is valid

Theorem: The matrix \( Cov^{NS} \) is positive-definite.

Proof: Now, \( \forall \eta_1, \eta_2, \ldots, \eta_n \), the matrix \( C_{\alpha_n} \) can be written in the form:

\[ C_{\alpha_n} = V_{\alpha_n} \cdot V'_{\alpha_n} \quad (88) \]

where \( V_{\alpha_n} \) is a \( n^2 \times n \) matrix defined as

\[ V_{\alpha_n} = \begin{pmatrix} D_{\eta_1} \\ D_{\eta_2} \\ \vdots \\ D_{\eta_n} \end{pmatrix} \quad (89) \]
where $D$ was defined previously, [11]. So now,

$$
\text{Cov}^{NS} = B'_{\alpha_n} \cdot B_{\alpha_n}
$$

(90)

where $B_{\alpha_n} = V_{\alpha_n}^{'}X^{'}$ is a $n \times n$ matrix, which can be seen to be of rank $n$. Hence, using the previous Corollary again (i.e. given in the proof of the theorem), we can conclude that $\text{Cov}^{NS}$ is a positive-definite matrix.

**References**


