Journal of Statistics: Advances in Theory and Applications
Volume 2, Number 1, 2009, Pages 11-44

THE CYCLICAL AUTO-LINEAR MODEL ON THE SPHERE

CHRYSOULA DIMITRIOU-FAKALOU

Department of Statistical Science
University College
London, U. K.
e-mail: chrysoula@stats.ucl.ac.uk

Abstract

For observations recorded regularly on a perfect circle or sphere, we present a new class of stationary models that clothe naturally the spatial second-order dependence. At first, the new model might be seen as the cyclical analogue of the standard auto-regressive model. Further and due to the spherical structure, a finite auto-regressive as well as finite moving-average dependence may be clothed simultaneously via the so-called CAL model. Some existing results from time series might be simplified and we may easily proceed with the estimation and statistical inference for the coefficients of best linear prediction, which are needed for kriging on the spherical surface.

1. Introduction

The analysis of stationary spatial processes has often become a generalization of the classical analysis of series on the time axis $\mathbb{Z} = \{0, \pm 1, \ldots\}$. Considering that a spatial process takes place in the plane $\mathbb{Z}^2$ is in fact an approximation due to the actual spherical shape of surfaces. Some previous attempts to present new mechanisms that

2000 Mathematics Subject Classification: 60G10, 62F10, 62F12.

Keywords and phrases: auto-correlation, auto-linear model, best linear prediction coefficient.

Received August 16, 2009

express solely the spherical spatial dependence, such as those of Jones [7] or McLeod [10], are spectral-density oriented and they lack the meaningfulness and simplicity of the parameters involved, like these are secured by a model-based approach.

Furthermore, these attempts have been hardly accompanied by definitions of relevant estimators and their asymptotic behavior; spatio-temporal processes have been considered instead, as increasing the number of timings can provide the necessary structure for inference over the spatial dimensions. For example, Cressie and Huang [4], and Ma [9] have focused on constructing valid covariance functions for spatio-temporal processes on flat surfaces, while Jun and Stein [8] have done the same on spheres. A general methodology regarding the estimation and statistical inference for the parameters expressing the spatial dependence using the temporal dynamics only, has been provided by Dimitriou-Fakalou [5].

For regularly spaced observations on a sphere, we have constructed meaningful stationary models, which allow for a methodology of identification, estimation, diagnostic checking, and kriging to be adopted; similarly to the stationary ARMA model for a time series, the CAL model may be seen as the main tool to proceed with the statistical analysis on a circle. A natural pattern has been followed that imitates Whittle [11] on $\mathbb{Z}$, and then on $\mathbb{Z}^2$, such that first the simple case of the circle is addressed, and then a discussion on the sphere follows.

The models we have built up respect the regular spacing of locations on a spherical surface, while they do not depend on the number of points observed. They use directly as parameters the coefficients needed for spatial prediction. For the new derivations, we most advantageously reproduce the methods of Besag [2]. A theorem establishes the asymptotic normality of the observed coefficients under a finite fourth moment condition; the result is distribution-free and copies the classical time series Bartlett's formulas in the variance matrix. How many 'neighbors' are needed for kriging? How can we estimate the relevant parameters and use them for prediction? Such questions will be answered next.

2. Notation

For a given circle and a positive integer N, we consider $s_1, \ldots, s_N$ to be N fixed and regularly spaced points on the circle, i.e., they form a polygon with N equal edges. Conventionally and starting from a first point s_1, we may label s_{j+1} to be the 'next' point from s_j, $j = 1, \ldots, N - 1$, and moving from left to right on the circle. We will then write $\{\varepsilon(s_j), j = 1, \ldots, N\}$ to be a set of uncorrelated random variables, each with zero mean and variance $0 < \sigma^2 < \infty$, taking place on the N points.

Next, we consider any positive integer p, such that we write $N = (p+1)M_p$ and $M_p = 1, 2, \ldots$, i.e., N is a multiple of $(p+1)$. The N original points on the circle might be now seen as M_p different sets of $(p+1)$ equally spaced points or M_p different polygons each one with $(p+1)$ equal edges. We will then define the conventional 'right-shift' operator R_p to be such that $R_p\varepsilon(s_j)$, $j = 1, \ldots, N$ is the ε on the 'next' point from left to right of the $(p+1)$- rather than the N-polygon.

As an example for the equally spaced $s_1, \ldots, s_{20}$ points, we may consider $p = 1$ with $M_1 = 10$ diameters, $p = 3$ with $M_3 = 5$ squares, $p = 4$ with $M_4 = 4$ pentagons and $p = 9$ with $M_9 = 2$ decagons; of course, the polygon with the most points we can make is just one and it is using all the 20 points. Then s_1 is on the same diameter as s_{11}, on the same square as s_6, s_{11}, and s_{16}, on the same pentagon as s_5, s_9, s_{13}, and s_{17} and, finally, on the same decagon as s_{2k+1}, $k = 1, \ldots, 9$. If we consider the random variables $\{\varepsilon(s_j), j = 1, \ldots, 20\}$, we may apply the different 'right-shift' operators on, say $\varepsilon(s_1)$, to derive

$$R_1\varepsilon(s_1) = \varepsilon(s_{11}),\ R_3\varepsilon(s_1) = \varepsilon(s_6),\quad R_4\varepsilon(s_1) = \varepsilon(s_5),\ R_9\varepsilon(s_1) = \varepsilon(s_3).$$

Similarly to the R_p operator, the 'left-shift' operator $L_p \equiv R_p^{-1}$ may be defined to move to the 'next' point of the $(p+1)$- polygon from right to left. It is not difficult to verify that $R_p^{(p+1)k+i} = R_p^i$ for any $i, k \in \mathbb{Z}$.

3. CAL Models on the Circle

For given p, we will write $l = [(p+1)/2]$. Based on $\{\varepsilon(s_j), j = 1, \dots, N\}$ and some real numbers $\theta_1, \dots, \theta_l$, we define the random variables

$$Y(s_j) = \varepsilon(s_j) + \sum_{i=1}^{l} \theta_i R_p^i \varepsilon(s_j), \; j = 1, \dots, N. \tag{1}$$

Then for some real numbers $0 < \nu < \infty$ and $\beta_1, \dots, \beta_l$ with $|\beta_i| \leq 1$, $i = 1, \dots, l$, we may write that

$$\text{Cov}(Y(s_j), Y(s_k)) \equiv \begin{cases} \nu, & \text{if } j = k, \\ -\beta_i \nu, & \text{if } |j-k| = M_p i, \; M_p(p+1-i), i = 1, \dots, l, \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

for $j, k = 1, \dots, N$. According to (2), all the random variables Y have the same variance ν. Further, two random variables $Y(s_j)$ and $Y(s_k)$ have a nonzero correlation only, if the locations s_j, s_k are on the same $(p+1)$- polygon and i of its 'steps' or edges away; this correlation $-\beta_i$ is a function of i and, in that sense, the N random variables have a (weakly) stationary dependence. The symmetry restrictions of the covariances imply that we do not use $(p+1)$ but l parameters β instead.

The numbers $\beta_1, \dots, \beta_l$ and ν might be directly derived from the numbers $\theta_1, \dots, \theta_l$ and σ^2, as (2) expresses the second-order properties of the random variables Y, while (1) models Y as linear functions of the ε. It should be highlighted though, that there are several ways to model linear functions of ε and derive covariance properties (2). One may wish to replace the one-sided representation (1) that uses $\theta_1, \dots, \theta_l$ by a two-sided filter

$$Y^*(s_j) = \varepsilon(s_j) + \sum_{i=1}^{p} \theta_i^* R_p^i \varepsilon(s_j), \; j = 1, \dots, N$$

with $\theta_1^*, \ldots, \theta_p^*$ instead, such that (2) applies for Y^*; Example 3.1 explains this further. For the ARMA (p, q) model on $\mathbb{Z}$, the same auto-covariance function or spectral density can be achieved by 2^{p+q} different equations based on the same uncorrelated random variables; one of these equations only defines a causal and invertible ARMA process and guarantees the identifiability of the auto-regressive and moving-average parameters from the auto-covariance function. The distinction between 'unilateral' and 'bilateral' pure auto-regressive processes as well as when these share the same spectral density, was first described by Whittle [11].

For any $j = 1, \ldots, N$, we will write the variance matrix

$$B = \frac{1}{\nu}\operatorname{Var}(Y(s_j)\ R_pY(s_j)\ldots R_p^pY(s_j))^{\top}.$$

It is not difficult to verify that it is

$$B \equiv \begin{bmatrix} 1 & -\beta_1 & -\beta_2 & \cdots & -\beta_l & -\beta_l & \cdots & -\beta_2 & -\beta_1 \\ -\beta_1 & 1 & -\beta_1 & & -\beta_{l-1} & -\beta_l & & -\beta_3 & -\beta_2 \\ & & & & & & & & \\ -\beta_1 & -\beta_2 & -\beta_3 & & -\beta_l & -\beta_{l-1} & \cdots & -\beta_1 & 1 \end{bmatrix}, \tag{3}$$

if p is even, or

$$B \equiv \begin{bmatrix} 1 & -\beta_1 & -\beta_2 & \cdots & -\beta_l & -\beta_{l-1} & \cdots & -\beta_2 & -\beta_1 \\ -\beta_1 & 1 & -\beta_1 & & -\beta_{l-1} & -\beta_l & & -\beta_3 & -\beta_2 \\ & & & & & & & & \\ -\beta_1 & -\beta_2 & -\beta_3 & & -\beta_{l-1} & -\beta_{l-2} & \cdots & -\beta_1 & 1 \end{bmatrix}, \tag{4}$$

if p is odd.

Definition 1. The random variables $\{X(s_j),\ j = 1, \ldots, N\}$ follow the Cyclical Auto-Linear model of order p and we write $\{X(s_j)\} \sim CAL(p)$, if for the zero mean random variables $\{Y(s_j),\ j = 1, \ldots, N\}$ with covariances described by (2) and with positive-definite matrix B defined by (3) or (4), it holds for any $j = 1, \dots, N$ that

$$Y(s_j) = \begin{cases} X(s_j) - \sum_{i=1}^{l} \beta_i (R_p^i X(s_j) + L_p^i X(s_j)), & \text{if } p \text{ is even}, \\ X(s_j) - \sum_{i=1}^{l-1} \beta_i (R_p^i X(s_j) + L_p^i X(s_j)) - \beta_l R_p^l X(s_j), & \text{if } p \text{ is odd}. \end{cases} \tag{5}$$

The term 'auto-linear' stems directly from the definitions of Dimitriou-Fakalou [6] on $\mathbb{Z}^d$ and for any positive integer d. The auto-linear model has clothed the same second-order properties as an auto-regressive equation, only that it is natural for spatial processes and, unlike an auto-regression unless specified to be causal, it guarantees a one-to-one correspondence with the finite number of elements in the denominator of the spectral density. Furthermore, it is simultaneous, which distinguishes it from the conditional approaches of Besag [1]. For the sake of example, the process $\{X_t, t \in \mathbb{Z}\}$ is a (weakly) stationary auto-regression of order p, i.e., it satisfies

$$X_t - \varphi_1 X_{t-1} - \dots - \varphi_p X_{t-p} = \varepsilon_t, \tag{6}$$

where $\{\varepsilon_t\}$ are zero mean, uncorrelated random variables with a finite variance. Then the spectral density of $\{X_t\}$ is

$$g_X(\omega) \propto \{(1 - \varphi_1 z - \dots - \varphi_p z^p)(1 - \varphi_1 z^{-1} - \dots - \varphi_p z^{-p})\}^{-1},$$

$$z = e^{i\omega},\ i = \sqrt{-1},\ \omega \in (-\pi, \pi), \tag{7}$$

which can also be expressed as

$$g_X(\omega) \propto \{1 - \beta_1(z + z^{-1}) - \dots - \beta_p(z^p + z^{-p})\}^{-1},\ z = e^{i\omega},\ i = \sqrt{-1},\ \omega \in (-\pi, \pi), \tag{8}$$

and $\beta_1, \dots, \beta_p$ may be derived from $\varphi_1, \dots, \varphi_p$ in an obvious way. Nevertheless, there are 2^p different sets of φ that can generate the same β. The unique auto-linear representation relating to (8) is

$$X_t - \sum_{i=1}^{p} \beta_i (X_{t-i} + X_{t+i}) = Y_t, \tag{9}$$

and $\{Y_t\}$ is now a finite moving-average process with auto-correlations equal to $-\beta_{|i|}$ at lags $|i| = 1, \dots, p$ and zero anywhere further. While a causal representation (6) is useful to express the dependence over time due to the unidirectional flow of the time axis, the auto-linear model (9) might be more appropriate to express the spatial dependence on the line transect (see Whittle [11]) or for smoothing in time series. Once the definitions are generalized on $\mathbb{Z}^d$ with any positive integer d, the auto-linear model might be used for processes occuring on approximately flat surfaces $\mathbb{Z}^2$ or the three-dimensional space $\mathbb{Z}^3$.

Back to Definition 1 and following the same sequel as on $\mathbb{Z}$, the Cyclical Auto-Linear equation of order p on the circle models simultaneously N random variables $\{X(s_j), j = 1, \dots, N\}$ as linear combinations of some other N random variables $\{Y(s_j), j = 1, \dots, N\}$ with a finite 'moving-average' dependence (2); this is in the sense that each $Y(s_j), j = 1, \dots, N$ relates to only p nonzero correlations with other $Y(s_k), k = 1, \dots, N, k \neq j$. The model is natural for the spatial dependence, as it does not imply any directional preference on the circle. The variables on all the locations of the $(p+1)$-polygon have been taken into the equation but symmetry conditions on the coefficients β are also present. Why the new random variables $\{X(s_j), j = 1, \dots, N\}$ have been most advantageously defined from $\{Y(s_j), j = 1, \dots, N\}$ according to the CAL equation, follows in the next section.

3.1. The example of a CAL(2) model

For the three fixed points s_1, s_2, and s_3 on the circle, which form an equilateral triangle, we may consider the uncorrelated random variables $\{\varepsilon(s_j), j = 1, 2, 3\}$, as we described before. If we define the new random variables

$$Y(s_1) = \varepsilon(s_1) + \theta\varepsilon(s_2),\ Y(s_2) = \varepsilon(s_2) + \theta\varepsilon(s_3),\ Y(s_3) = \varepsilon(s_3) + \theta\varepsilon(s_1),$$

it is not difficult to verify that the correlation between the random variables is

$$\mathrm{Corr}(Y(s_1), Y(s_2)) = \mathrm{Corr}(Y(s_2), Y(s_3)) = \mathrm{Corr}(Y(s_3), Y(s_1)) = \frac{\theta}{1+\theta^2}.$$

On the other hand, we may also define the variables $\{Y^*(s_j), j = 1, 2, 3\}$ from $Y^*(s_1) = \varepsilon(s_1) + \theta_1\varepsilon(s_2) + \theta_2\varepsilon(s_3)$, and $Y^*(s_2) = \varepsilon(s_2) + \theta_1\varepsilon(s_3) + \theta_2\varepsilon(s_1)$, $Y^*(s_3) = \varepsilon(s_3) + \theta_1\varepsilon(s_1) + \theta_2\varepsilon(s_2)$, and use two parameters θ_1, θ_2 to reduce again to one correlation only

$$\mathrm{Corr}(Y^*(s_1), Y^*(s_2)) = \mathrm{Corr}(Y^*(s_2), Y^*(s_3))$$
$$= \mathrm{Corr}(Y^*(s_3), Y^*(s_1)) = \frac{\theta_1 + \theta_2 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2}.$$

There are many ways to define our Y variables as linear combinations of the three ε, which could use up to three parameters θ and they would all result in one identical correlation, say $-\beta$, between all the possible pairs of different Y. Provided that the matrix

$$B = \begin{bmatrix} 1 & -\beta & -\beta \\ -\beta & 1 & -\beta \\ -\beta & -\beta & 1 \end{bmatrix}$$

is positive-definite, we write $\{X(s_j), j = 1, 2, 3\} \sim \mathrm{CAL}(2)$ according to the equations

$$X(s_j) - \beta(R_2X(s_j) + L_2X(s_j)) = Y(s_j), \; j = 1, 2, 3.$$

Note that a positive-definite B secures, for example, that the three Y^* cannot be the same random variable, as when $\theta_1 = \theta_2 = 1$, or $|\beta| = 1$. Vice versa, the one-sided filter will not be able to ever produce a correlation higher than 0.5 in absolute value.

For more general cases, when $N = 3M_2$, $M_2 = 1, 2, \ldots$ and there are equally spaced points $s_1, \ldots, s_N$ on the circle, assuming that uncorrelated random variables $\{\varepsilon(s_j), j = 1, \ldots, N\}$ are taking place, we may define $\{X(s_j),\ j = 1, \ldots, N\} \sim \mathrm{CAL}(2)$ similarly, provided that B as above is positive-definite.

4. Covariance Structures

It holds that $\{X(s_j),\ j = 1, \ldots, N\} \sim \mathrm{CAL}(p)$, and we define the random vectors

$$\mathbf{X}_j = \left(X(s_j) \quad R_p X(s_j) \quad \ldots \quad R_p^p X(s_j)\right)^\top, \text{ and}$$

$$\mathbf{Y}_j = \left(Y(s_j) \quad R_p Y(s_j) \quad \ldots \quad R_p^p Y(s_j)\right)^\top.$$

The Cyclical Auto-Linear model implies that

$$B\,\mathbf{X}_j = \mathbf{Y}_j \quad \text{or} \quad \mathbf{X}_j = B^{-1}\mathbf{Y}_j,\ j = 1, \ldots, N. \tag{10}$$

For any $j = 1, \ldots, N$, we may write $\Sigma_X \equiv \mathrm{Var}(\mathbf{X}_j)$ and $\Sigma_Y \equiv \mathrm{Var}(\mathbf{Y}_j)$ and we know that it holds $\Sigma_Y = \nu B$. Straight from (10), we may also derive that

$$\Sigma_X = B^{-1}\,\Sigma_Y\,B^{-1} = \nu B^{-1} \quad \text{or} \quad \Sigma_X^{-1} = \nu^{-1} B. \tag{11}$$

Further to the covariances in Σ_X as in (11), it holds that

$$\mathrm{Cov}(X(s_j),\ X(s_k)) = 0, \text{ if } |j - k| \neq M_p i,\ i = 0, 1, \ldots, p, \tag{12}$$

when the two points s_j and s_k correspond to Y on different $(p+1)$-polygons, which are uncorrelated according to (2).

Apart from the covariances of $\{X(s_j)\}$ and $\{Y(s_j)\}$ studied separately, it would be interesting to understand the cross-covariance structure that relates these two sets of random variables. It holds that

$$\mathrm{Cov}(\mathbf{X}_j,\ \mathbf{Y}_j) = \mathrm{Cov}(\mathbf{X}_j,\ B\,\mathbf{X}_j) = \Sigma_X\,B = \nu B^{-1} B = \nu I_{p+1},\ j = 1, \ldots, N, \tag{13}$$

where we write I_r for the identity matrix with r rows. For the same reasons as in (12), it holds that

$$\text{Cov}(\boldsymbol{X}_j, \boldsymbol{Y}_k) = O_{(p+1)\times(p+1)}, \text{ if } |j-k| \neq M_p i, \ i = 0, 1, \ldots, p, \tag{14}$$

where we write $O_{r\times c}$ for the zero matrix with r rows, c columns. By combining Equations (13) and (14), we may conclude for $j, k = 1, \ldots, N$, that

$$\text{Cov}(X(s_j), Y(s_k)) = \begin{cases} \nu, & j = k, \\ 0, & j \neq k. \end{cases} \tag{15}$$

Let us now interpret what this means. Suppose that we are interested in finding the best linear predictor of any $X(s_j)$, $j = 1, \ldots, N$ based on the information provided from $R_p X(s_j), \ldots, R_p^p X(s_j)$. We write the linear function $\hat{X}(s_j) = \sum_{i=1}^{p} c_i R_p^i X(s_j)$ and this should be the best one, in the sense that the prediction variance $E(X(s_j) - \hat{X}(s_j))^2$ is the minimum achieved by any such linear predictor. Then, the least squares property implies that these constants $c_1, \ldots, c_p$ are the unique solutions of the linear system

$$\text{Cov}(X(s_j) - \hat{X}(s_j), R_p^i X(s_j)) = 0, \ i = 1, \ldots, p.$$

Indeed, if we re-write the system above as

$$\text{Cov}(X(s_j), R_p^i X(s_j)) - \sum_{k=1}^{p} c_k \text{Cov}(R_p^k X(s_j), R_p^i X(s_j)) = 0, \ i = 1, \ldots, p,$$

or in the vector form

$$\text{Var}\begin{pmatrix} R_p X(s_j) \\ \vdots \\ R_p^p X(s_j) \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_p \end{pmatrix} = \begin{pmatrix} \text{Cov}(X(s_j), R_p X(s_j)) \\ \vdots \\ \text{Cov}(X(s_j), R_p^p X(s_j)) \end{pmatrix},$$

then the uniqueness of the coefficients $c_1, \ldots, c_p$ is guaranteed from the fact that B, B^{-1} are positive-definite and that it holds

$$\mathbf{X}_j^{\top} = (X(s_j)\ (R_p X(s_j)\ \ldots\ R_p^p X(s_j)))\ \text{and}\ \mathrm{Var}(\mathbf{X}_j) = \nu B^{-1}.$$

As a result, $\mathrm{Var}((R_p X(s_j)\ \ldots\ R_p^p X(s_j))^{\top})$ is positive-definite and the linear system has a unique solution.

If we set the predictor

$$\hat{X}(s_j) \equiv \begin{cases} \sum_{k=1}^{l} \beta_k (R_p^k X(s_j) + L_p^k X(s_j)), & \text{if } p \text{ is even}, \\ \sum_{k=1}^{l-1} \beta_k (R_p^k X(s_j) + L_p^k X(s_j)) + \beta_l\, R_p^l X(s_j), & \text{if } p \text{ is odd}, \end{cases} \tag{16}$$

then the prediction error $Y(s_j) \equiv X(s_j) - \hat{X}(s_j)$ is uncorrelated with all $R_p^i X(s_j),\ i = 1, \ldots, p$ indeed. In fact, as $Y(s_j)$ is according to (15), uncorrelated with any $X(s_k),\ k = 1, \ldots, N,\ k \neq j$, it holds that $\hat{X}(s_j)$ in (16) is the best linear predictor of $X(s_j)$ based on all $R_p^i X(s_j),\ i = 1, \ldots, p$, plus any other information from X on the N points. The (minus) correlations $\beta_1, \ldots, \beta_l$ for the Y now take the form of best linear prediction coefficients for the X. The variance of the best linear prediction error is

$$\mathrm{Var}(X(s_j) - \hat{X}(s_j)) \equiv \mathrm{Var}(Y(s_j)) = \nu.$$

4.1. The circle versus the line transect

For a causal auto-regression $\{X_t,\ t \in \mathbb{Z}\}$ of order p defined by (6), it holds that the best linear predictor of X_t based on $X_{t-i},\ i = 1, \ldots, p$ is $\varphi_1 X_{t-1} + \ldots + \varphi_p X_{t-p}$. Further, this is also the best linear predictor of X_t based on $X_{t-i},\ i = 1, \ldots, p$, plus any other information $X_{t-i},\ i > 0$, i.e., from the 'past'. On the other hand, using the auto-linear representation (9) of $\{X_t\}$, Dimitriou-Fakalou [6] explained that the best linear predictor of X_t based on $X_{t-i},\ X_{t+i},\ i = 1, \ldots, p$ is

$$\beta_1(X_{t-1} + X_{t+1}) + \ldots + \beta_p(X_{t-p} + X_{t+p}).$$

This is also the best linear predictor of X_t based on X_{t-i}, X_{t+i}, $i = 1, \ldots, p$, plus any other information X_{t-i}, $i \in \mathbb{Z}$, i.e., anywhere from the 'past' or 'future'. Since for Gaussian random variables the best linear predictors are conditional expectations, the auto-normal processes of Besag [1] are a special case of the auto-linear processes of Dimitriou-Fakalou [6].

Moving from the line transect to $\{X(s_j), j = 1, \ldots, N\} \sim$ CAL(p) on the circle, the methods of Besag [2] might be used then; and (16) is the best linear predictor for $X(s_j)$ adding the extra element that it is also a conditional expectation when the random variables are Gaussian. Besag [2] provided the decomposition of an inverse variance matrix for Gaussian random variables, as a product of a square matrix with elements the coefficients of the conditional expectations and a diagonal matrix with the reciprocals of the relevant conditional variances. On the circle, Besag's decomposition simplifies to $\Sigma_X^{-1} = \nu^{-1}B$, using ν to be the same variance, $(p+1)$ times on the same polygon and the diagonal matrix reduces to $\nu^{-1}I_{p+1}$; B as in (3) or (4) is the square matrix with elements the best linear prediction coefficients β.

Next to the similarities of the methods of Besag [1], [2], and the cross-covariance structure of the X and Y, that is, essential for the derivation of the natural best linear predictors over space, there are also differences that need to be acknowledged. We write the auto-linear representation of the random variables $\{X_t, t \in \mathbb{Z}\}$ as functions of the random variables $\{Y_t, t \in \mathbb{Z}\}$ with auto-correlations generated by the spectral density

$$g_Y(\omega) \propto \{1 - \beta_1(z + z^{-1}) - \ldots - \beta_p(z^p + z^{-p})\},\ z = e^{i\omega},\ i = \sqrt{-1},\ \omega \in (-\pi, \pi). \tag{17}$$

While the process $\{X_t\}$ achieves its best linear predictors to be functions of a finite number of 'neighbors', i.e., it has a finite auto-regressive or auto-linear behavior, the process $\{Y_t\}$ has the auto-correlation function cutting off to zero after a finite number of 'lags', i.e., it exhibits a finite moving-average dependence. For stationary processes on $\mathbb{Z}$, both a finite auto-regressive as well as finite moving-average second-order dependence cannot take place at the same time.

Nevertheless, $\{X(s_j), j = 1, \dots, N\} \sim \text{CAL}(p)$ on the circle combine both these finite auto-linear and finite moving-average properties; firstly, the best linear predictor of $X(s_j)$ based on all $X(s_k), k \neq j$ is a function of p 'neighbors' only, no matter how large N might be. Secondly and according to (12), the covariances of $X(s_j)$ with $X(s_k), k \neq j$ are nonzero for p locations s_k only. Consequently for some real numbers $0 < \mu < \infty$ and $\gamma_1, \dots, \gamma_l$ with $|\gamma_i| \leq 1, i = 1, \dots, l,$ we will be able to write for any $j, k = 1, \dots, N$ that

$$\text{Cov}(X(s_j), X(s_k)) \equiv \begin{cases} \mu, & \text{if } j = k \\ \gamma_i \mu, & \text{if } |j - k| = M_p i,\ M_p(p + 1 - i),\ i = 1, \dots, l. \\ 0, & \text{otherwise} \end{cases} \tag{18}$$

In order to compute $\mu, \gamma_i, i = 1, \dots, l$ from $\nu, \beta_i, i = 1, \dots, l,$ we express

$$\Sigma_X = \mu \Gamma \equiv \nu B^{-1},$$

where Γ is a matrix with elements unity in the main diagonal and γ_i, $i = 1, \dots, l,$ in the same sense that B is a matrix with elements unity in the main diagonal and $-\beta_i, i = 1, \dots, l,$ in (3) or (4). Then according to Definition 1, $\{Y(s_j)\} \sim$ CAL(p) since it holds that

$$X^*(s_j) = \begin{cases} Y(s_j) + \sum_{i=1}^{l} \gamma_i (R_p^i Y(s_j) + L_p^i Y(s_j)), & \text{if } p \text{ is even}, \\ Y(s_j) + \sum_{i=1}^{l-1} \gamma_i (R_p^i Y(s_j) + L_p^i Y(s_j)) + \gamma_l R_p^l Y(s_j), & \text{if } p \text{ is odd}, \end{cases}$$

and $\{X^*(s_j) = (\nu / \mu) X(s_j)\}$ has a finite moving-average correlation structure as in (18). The two properties are clearly dual.

The CAL(p) model restricts all dependence to appear via the information from p 'neighbors' only, but these are evenly placed on the circle. As a result, no matter how large N might be, the same coefficients $\beta_1, \ldots, \beta_l$ are to be estimated, and used for spatial prediction on a given circle and asymptotic results for the unknown coefficients might be made available. With a proper selection of order, the CAL(p) model can adopt the (weakly) stationary dependence taking place between the N random variables. When prediction is based on the p nearest neighbors available, rather than the recordings of the $(p+1)$-polygon, this is still a special case of the $\text{CAL}(N-1)$ equation. However, unlike the CAL(p) model that manages different cliques of Y or X uncorrelated random vectors, the $\text{CAL}(N-1)$ model that uses the l nearest neighbors from right and left, corresponds to an $(N \times N)$ matrix B and to $(N-1)$ rather than p nonzero correlations; it puts all N random variables into one clique only.

Since, the CAL(p) model combines a finite auto-linear and moving-average dependence, it should be considered the cyclical analogue of the ARMA model expressing any (weakly) stationary dependence. Is the CAL model useful? Is it reasonable to consider stationary dependence on the circle? Exactly like the ARMA model, it might be used as the main tool to proceed with the statistical inference for a fixed number of parameters that are needed for spatial prediction. Thus, after properly extracting any non-stationary features from the recordings available, a methodology of identification, estimation, diagnostic checking and kriging on the circle, as we introduce it next, should be followed using the CAL model.

5. Methodology Using the CAL Equation

We consider the evenly spaced points $s_1, \ldots, s_N$ on the given circle. We make the following assumptions:

(C1) It holds that (i) the zero mean random variables $\{\varepsilon(s_j),\ j = 1, \ldots, N\}$ are independent and identically distributed with variance $0 < \sigma^2 < \infty$, and (ii) $E(\varepsilon(s_j)^4) = \eta\sigma^4 < \infty,\ j = 1, \ldots, N$.

For fixed positive integer p and $l = [(p+1)/2]$, such that $N \equiv M_p(p+1),\ M_p = 1, 2, \ldots,$ and $s_1, \ldots, s_N$ are equally spaced on the circle, we define the random variables $\{Y(s_j),\ j = 1, \ldots, N\}$ from (1) with covariances (2); we will write the vector $\boldsymbol{\beta}_0 = (\beta_1, \ldots, \beta_l)^\top$ and the matrix B as in (3) or (4). Provided that B is positive-definite, it holds that $\{Y(s_j)\} \sim \mathrm{CAL}(p)$. Next, we present Theorem 1, which will be needed for the new methodology we will propose. We will write $N \to \infty$, in the sense that $M_p \to \infty$, i.e., that N is large in relation to the fixed order p.

Theorem 1. *We write*

$$b_i = -\frac{\sum_{j=1}^{N} Y(s_j)\, R_p^i Y(s_j)}{\sum_{j=1}^{N} Y(s_j)^2}, \; i = 1, \ldots, l$$

and $\mathbf{b} = (b_1, \ldots, b_l)^\top$. *Under condition* (C1)(i), *it holds as* $N \to \infty$ *that*

(i) $\mathbf{b} \xrightarrow{P} \boldsymbol{\beta}_0$,

(ii) $N^{1/2}(\mathbf{b} - \boldsymbol{\beta}_0) \xrightarrow{D} N(\mathbf{0}, W)$ *under* (C1)(ii), *where we write* $W = [w_{n,m}]_{n,m=1}^{l}$, *and the elements*

$$w_{n,m} = \sum_{i=0}^{p} \{\rho_{i+n}\rho_{i+m} + \rho_{i-n}\rho_{i+m} + 2\rho_n\rho_m\rho_i^2 - 2\rho_n\rho_i\rho_{i+m} - 2\rho_m\rho_i\rho_{i+n}\} \tag{19}$$

are given from Bartlett's formula, with

$$\rho_0 = 1, \quad \rho_i = -\beta_i, \; i = 1, \ldots, l, \quad \rho_{-i} = \rho_{p+1-i} = \rho_{p+1+i} = \rho_i, \; i \in \mathbb{Z}.$$

5.1. Identification

Suppose we have the recordings $\{x(s_j), \; j = 1, \ldots, N\}$ on the circle and a (weakly) stationary pattern $\{X(s_j)\} \sim$ CAL(p) is taking place. First, we will need to identify the correct order p, where N must be a multiple of $(p + 1)$.

The 'saturated' CAL model allows for a number of $[N/2]$ different auto-correlations. For this model, we may construct the cyclical analogue of a time series sample auto-correlogram, by first computing

$$\tilde{\gamma}_i = \frac{\sum_{j=1}^{N} X(s_j)\, R_{N-1}^{i} X(s_j)}{\sum_{j=1}^{N} X(s_j)^2},\; i = 1, \ldots, [N/2], \tag{20}$$

and then plotting $\tilde{\gamma}_i$ versus the lag i.

If $\tilde{\gamma}$ is very large in absolute value, then we have reasons to believe that the true auto-correlation on the relevant lag is not equal to zero. Similarly to time series and using Theorem 1, we may consider the formal bounds $\pm 1.96 / \sqrt{N}$, such that, if the true auto correlations are zero, they include each sample auto-correlation with probability 95%; otherwise, Bartlett's formula (19) has to be used with the variances $\tilde{w}_{i,i} / N$, $i = 1, \ldots, [N/2]$ using the sample plug-ins $\tilde{\gamma}$ instead of the true but unknown correlations.

Evidence in favor of a specific order p should be that all the relevant $\tilde{\gamma}$ on the $(p+1)$-polygon are significant. However, we would not expect all the remaining auto-correlations, especially those on the nearest lags $|i| = 1, 2, \ldots, M_p - 1$ of the N-polygon, to easily give up any sign of existing dependence and be close to zero. Thus, more formal selection criteria should be used and further investigation is required to assess them. As an example, we refer to the Akaike Information Criterion (Brockwell and Davis [3]), which will properly combine the Gaussian likelihood of $\{X(s_j),\, j = 1, \ldots, N\}$, assuming that they follow the CAL(p) equations, together with the number of parameters $(l+1)$. How to compute the Gaussian likelihood of $\{X(s_j)\} \sim$ CAL(p) follows next. The AIC will take into account that the more the edges of the polygon preferred the more the variability likely to be explained and, in order to serve the principle of parsimony, it will penalize each candidate model accordingly, such that the order p relating to the minimum number is preferred.

5.2. Estimation

From the observations $\{X(s_j),\, j = 1, \ldots, N\} \sim$ CAL(p) with fixed order p, we now aim at proceeding with the statistical inference for the true best linear prediction coefficients $\boldsymbol{\beta}_0 = (\beta_{1,0}, \dots, \beta_{l,0})^{\top}$. We write the following condition:

(C2) The parameter space $\mathcal{B}$ is such that, for any $\boldsymbol{\beta} = (\beta_1, \dots, \beta_l)^{\top}$, it holds that the matrix B, as defined by (3) or (4), is positive-definite. Further, it holds that $\boldsymbol{\beta}_0 \in \mathcal{B}$.

Using $\mathbf{X}_j$, $j = 1, \dots, N$ as defined previously the observations may be written as M_p vectors, i.e., $\mathbf{X}_j$, $j = 1, \dots, M_p$; which are independent under condition (C1). Further, the Gaussian likelihood will fully describe the second-order properties of interest, which are parameterized by ν and by the matrix B, with elements that change according to $\boldsymbol{\beta} \in \mathcal{B}$. The Gaussian likelihood of the N observations will take the form

$$L(\beta, \nu) \propto \nu^{-N/2} |B|^{M_p/2} \exp\{ -\sum_{j=1}^{M_p} \mathbf{X}_j^{\top} B \mathbf{X}_j / 2\nu\}, \ \beta \in \mathcal{B}, \ 0 < \nu < \infty. \quad (21)$$

The random part $\sum_{j=1}^{M_p} \mathbf{X}_j^{\top} B \mathbf{X}_j$ in (21) is linear with respect to the parameters β. This leaves the deterministic part to have a maximal effect on the computation of the likelihood and, thus, the properties of the maximum likelihood estimators are difficult to establish. To approximate the estimates, a search over $\mathcal{B}$ under (C2) is required. Similar limitations were discussed by Besag [1] for the parameters of auto-normal processes on $\mathbb{Z}^2$.

However using Theorem 1, we may easily proceed with the statistical inference for the parameters β by defining estimators, which may be computed speedily as solutions of equations. For the first step, some consistent estimators, say $\tilde{\beta}$, will be needed. For this, the best linear prediction equations might be imitated to define $\tilde{\beta}_1, \dots, \tilde{\beta}_l$, such that

$$\sum_{j=1}^{N} \tilde{Y}(s_j) [R_p^k X(s_j) + L_p^k X(s_j)] = 0, \ k = 1, \dots, l - 1,$$

and either

$$\sum_{j=1}^{N} \tilde{Y}(s_j)[R_p^l X(s_j) + L_p^l X(s_j)] = 0,$$

if p is even, or

$$\sum_{j=1}^{N} \tilde{Y}(s_j)\ R_p^l X(s_j) = 0,$$

if p is odd, where $\{\tilde{Y}(s_j),\ j = 1, \dots, N\}$ are computed according to (5) using the $\{X(s_j),\ j = 1, \dots, N\}$ and the $\tilde{\beta}$-s. The $\tilde{\beta}_1, \dots, \tilde{\beta}_l$ are solutions of an $(l \times l)$ linear system and might be derived speedily. Their consistency under (C1) (i) stems directly from the fact that they imitate the rules of best linear prediction.

Finally, we may define our estimators

$$\hat{\beta}_i = -\frac{\sum_{j=1}^{N} \tilde{Y}(s_j) R_p^i \tilde{Y}(s_j)}{\sum_{j=1}^{N} \tilde{Y}(s_j)^2},\ i = 1, \dots, l, \tag{22}$$

which follow the properties of Theorem 1, due to the consistency of the plug-ins $\tilde{\beta}_1, \dots, \tilde{\beta}_l$.

5.3. Diagnostic checking

After deriving the estimates $\hat{\beta}$, we may use these and $\{X(s_j),\ j = 1, \dots, N\}$ to compute $\{\hat{Y}(s_j),\ j = 1, \dots, N\}$ according to (5). The random variables $\{\hat{Y}(s_j)\}$ must expose similar behavior to $\{Y(s_j)\} \sim$ CAL(p), i.e., nonzero auto-correlations on the relevant lags only. These might be verified either graphically by plotting the sample auto-correlogram of $\{\hat{Y}(s_j)\}$, or with formal tests using Theorem 1.

5.4. Spatial prediction

Once we have estimated the parameters β rather than γ from $\{X(s_j), j = 1, \dots, N\}$, we may use the estimated coefficients directly to proceed with spatial prediction. More specifically, for another fixed point of the circle, say s, such that

$$\{X(s_j),\ j = 1, \dots, N, R_p^i X(s),\ i = 0, 1, \dots, p\} \sim \text{CAL}(p),$$

not only will we be able to predict the value of $X(s)$ based on given information $R_p^i X(s), i = 1, \dots, p,$ but we will also construct 'prediction intervals'.

Assuming that $\{\varepsilon(s_j), j = 1, \dots, N\}$ are independent from $\{R_p^i \varepsilon(s),$ $i = 0, 1, \dots, p\}$, the estimators $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \dots, \hat{\beta}_l)^\top$ are also independent from $\{R_p^i X(s), i = 0, 1, \dots, p\}$. For the predictor, say $\hat{X}^*(s)$, we may imitate $\hat{X}(s)$ in (16) by using the $\hat{\beta}$-s. For the variance of the prediction error, and since $Y(s)$ is uncorrelated with $R_p^i X(s), i = 1, \dots, p,$ it holds that

$$\text{Var}(X(s) - \hat{X}^*(s)) = \text{Var}(Y(s) + \hat{X}(s) - \hat{X}^*(s)) = \nu + \text{Var}(\hat{X}(s) - \hat{X}^*(s)). \tag{23}$$

Due to the aforementioned independence, we can write that $\text{Var}(\hat{X}(s) - \hat{X}^*(s)) = E\{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top \Gamma^* (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\}$, where the $(l \times l)$ matrix Γ^* is such that

$$\Gamma^* = \begin{cases} \text{Var}(R_p X(s) + L_p X(s), \dots, R_p^l X(s) + L_p^l X(s))^\top, & \text{if } p \text{ is even}, \\ \text{Var}(R_p X(s) + L_p X(s), \dots, R_p^{l-1} X(s) + L_p^{l-1} X(s), R_p^l X(s))^\top, & \text{if } p \text{ is odd}. \end{cases}$$

For our result above, we have used an identical argument like Brockwell and Davis [3] for the final prediction error, i.e., the independence between the estimators and the variables used for prediction, which is needed when taking a conditional expectation. We write the Cholesky decomposition $\Gamma^* \equiv C^\top C$ and, thus,

$$E\{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top \Gamma^* (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\} = E\{(C(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0))^\top (C(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0))\}$$

$$= E\{\mathrm{tr}[(C(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0))(C(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0))^\top]\} = \mathrm{tr}[E\{(C(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0))(C(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0))^\top\}]$$
$$= \mathrm{tr}[\mathrm{Var}(C(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0))] = \mathrm{tr}[CWC^\top]/N,$$

where W is the matrix as given in Theorem 1 with elements from Bartlett's formulas.

If we combine all these results, we may approximate (23) by

$$\tilde{\nu} + \mathrm{tr}\ [\tilde{C}\,\tilde{W}\,\tilde{C}^\top]/N,$$

where all the estimates will be computed from $\{x(s_j),\ j = 1, \ldots, N\}$. The consistent plug-ins $\tilde{\gamma}$ will be used to construct $\tilde{\Gamma}^* \equiv \tilde{C}^\top\tilde{C}$, the matrix $\tilde{W}$ will use the $\tilde{\beta}$-s, and finally, we will compute

$$\tilde{\nu} = \sum_{j=1}^{N} \tilde{Y}(s_j)^2 / N.$$

Note that a missing observation might be always filled in this way, if recordings from all its neighbors on the $(p+1)$- polygon are available.

Table 1. Original point, applying the operator and moving to another point, for all the members of the CAL(3×3) clique

Point	Operator R_1	Operator R_2	Operator L_2	Operator L_1
'H'	'E'	'N'	'W'	'S'
'T'	'W'	'S'	'E'	'N'
'W'	'N'	'H'	'S'	'T'
'E'	'S'	'T'	'N'	'H'
'N'	'T'	'E'	'H'	'W'
'S'	'H'	'W'	'T'	'E'

6. CAL Models on the Sphere

First, we will present a basic model on the sphere, i.e., the CAL (3×3) equation. To construct a clique of neighbors for this model, we need to have a starting point, say 'S', the South Pole of the sphere. We next consider four equally spaced arcs on the sphere, all starting from 'S', and with length equal to 1/4th of a full circle going through the North and South Poles. The four new points of the clique lie on the end of each arc; conventionally, we label them in order as 'H', 'W', 'T', and 'E' (and then 'H' again), all equally placed on the equator of the sphere.

Of course, for all the four new points considered in the clique and excluding the arc that returns them to 'S', another three arcs must be considered. Indeed, if we include a final sixth point 'N' to be the North Pole of the sphere, then for all the six points considered, four arcs originate from each one of them and they all terminate to another point already considered; the six points have been equally placed on the sphere and they will form a clique.

For the given $p = 3$, we consider the two different operators R_1 and R_2 to move the random variable of interest from 'S' to two consecutive points, out of the four considered on the equator, say the points 'H' and 'W', respectively. Then the operators $L_2 \equiv R_2^{-1}$ and $L_1 \equiv R_1^{-1}$ must move from 'S' to 'T' and 'E', respectively. After arranging the four paths for the first point 'S', we can construct Table 1 that demonstrates applying all the operators on any of the six points of the clique. By looking at Table 1, it is not difficult to verify that it holds

$$R_1^{3+i} \equiv R_1^i,\ i \in \mathbb{Z} \quad \text{and} \quad R_2^2 \equiv R_1. \tag{24}$$

We write $\mathbf{s}$ for any of the six points of interest, and we consider the zero mean random variables $\{Y(\mathbf{s})\}$ all with variance $0 < \nu < \infty$, such that

$$\text{Corr}(Y(\mathbf{s}),\ R_1^i R_2^j Y(\mathbf{s})) \equiv \begin{cases} 1, & \text{if } (i,\ j) = \ (0,\ 0) \\ -\beta_{(1,0)}, & \text{if } (i,\ j) = \pm(1,\ 0) \\ -\beta_{(0,1)}, & \text{if } (i,\ j) = \pm(0,\ 1) \\ -\beta_{(1,1)}, & \text{if } (i,\ j) = \pm(1,\ 1) \end{cases}, \tag{25}$$

for some real numbers $|\beta_{(1,0)}|, |\beta_{(0,1)}|, |\beta_{(1,1)}| \leq 1$. For the last branch of (25), note that it holds anyway that

$$R_1 R_2 Y(\mathbf{s}) = L_1 L_2 Y(\mathbf{s}) \equiv R_1^{-1} R_2^{-1} Y(\mathbf{s}),$$

and both paths lead to the same random variable. Further to (25) and following the relevant paths according to Table 1, it can be verified for other correlations that

$$\mathrm{Corr}(Y(\mathbf{s}), R_1^i R_2^j Y(\mathbf{s})) \equiv \begin{cases} 1, & \text{if } (i, j) = \pm(1, -2), (2, 2), \\ -\beta_{(1,0)}, & \text{if } (i, j) = \pm(0, 2), (1, 2), (2, 0), (2, -2), \\ -\beta_{(0,1)}, & \text{if } (i, j) = \pm(1, -1), (2, 1), \\ -\beta_{(1,1)}, & \text{if } (i, j) = \pm(2, -1). \end{cases}$$

Using those last equations as well as (24) and (25), all the correlations $\mathrm{Corr}(Y(\mathbf{s}), R_1^i R_2^j Y(\mathbf{s}))$, $i, j \in \mathbb{Z}$ are functions of $\beta_{(1,0)}, \beta_{(0,1)}$, and $\beta_{(1,1)}$ only.

The order with which we apply the operators does not have any effect on the final point that the random variable is taking place, i.e., we have defined an 'objective' two-dimensional system. For example, all the permutations $R_1 R_2 R_1$, $R_1^2 R_2$, and $R_2 R_1^2$ applied to the random variable taking place on, say 'H', move the variable to 'W'. In general, it can be checked that for all different permutations of i identical operators R_1 and j identical operators R_2, the operator $R_1^i R_2^j$ might be used instead. This would not have been possible, if instead of the cyclical order of operators $R_1, R_2, L_2 \equiv R_2^{-1}, L_1 \equiv R_1^{-1}$ (and then R_1 again) as we used it on the four branches, we had preferred a 'right-left' and 'up-down' sense of order as it is on $\mathbb{Z}^2$ expressed via the operators, say $R, U, L \equiv R^{-1}, D \equiv U^{-1}$ and then R again; we would not be able to demonstrate Table 1 with all the paths properly completed then. However, thanks to our current definitions and given that we consider (weak) stationary such that $\mathrm{Corr}(Y(\mathbf{s}), R_1^i R_2^j Y(\mathbf{s})) = \mathrm{Corr}(Y(\mathbf{s}^*), R_1^i R_2^j Y(\mathbf{s}^*))$, $i, j \in \mathbb{Z}$ for any two points of the clique $\mathbf{s}, \mathbf{s}^*$, we can write that

$$R_1^i R_2^j Y(\mathbf{s}) = R_1^{i-i^*} R_2^{j-j^*} (R_1^{i^*} R_2^{j^*} Y(\mathbf{s})),\ i,\ i^*,\ j,\ j^* \in \mathbb{Z},$$

and secure a standard result as for stationary processes on $\mathbb{Z}^2$, i.e.,

$$\mathrm{Corr}(R_1^i R_2^j Y(\mathbf{s}),\ R_1^{i^*} R_2^{j^*} Y(\mathbf{s})) \equiv \mathrm{Corr}(Y(\mathbf{s}),\ R_1^{i-i^*} R_2^{j-j^*} Y(\mathbf{s}));$$

the symmetry condition

$$\mathrm{Corr}(Y(\mathbf{s}),\ R_1^i R_2^j Y(\mathbf{s})) = \mathrm{Corr}(Y(\mathbf{s}), L_1^i L_2^j Y(\mathbf{s}))$$

is a direct consequence.

Provided that the matrix

$$B \equiv \begin{bmatrix} 1 & -\beta_{(1,0)} & -\beta_{(0,1)} & -\beta_{(0,1)} & -\beta_{(1,0)} & -\beta_{(1,1)} \\ -\beta_{(1,0)} & 1 & -\beta_{(0,1)} & -\beta_{(1,1)} & -\beta_{(1,0)} & -\beta_{(0,1)} \\ -\beta_{(0,1)} & -\beta_{(0,1)} & 1 & -\beta_{(1,0)} & -\beta_{(1,1)} & -\beta_{(1,0)} \\ -\beta_{(0,1)} & -\beta_{(1,1)} & -\beta_{(1,0)} & 1 & -\beta_{(0,1)} & -\beta_{(1,0)} \\ -\beta_{(1,0)} & -\beta_{(1,0)} & -\beta_{(1,1)} & -\beta_{(0,1)} & 1 & -\beta_{(0,1)} \\ -\beta_{(1,1)} & -\beta_{(0,1)} & -\beta_{(1,0)} & -\beta_{(1,0)} & -\beta_{(0,1)} & 1 \end{bmatrix} \tag{26}$$

is strictly positive-definite, we write $\{X(\mathbf{s})\} \sim \mathrm{CAL}\ (3 \times 3)$, if it holds that

$$X(\mathbf{s}) - \beta_{(1,0)}(R_1 X(\mathbf{s}) + L_1 X(\mathbf{s})) - \beta_{(0,1)}(R_2 X(\mathbf{s}) + L_2 X(\mathbf{s})) - \beta_{(1,1)} R_1 R_2 X(\mathbf{s}) = Y(\mathbf{s}), \tag{27}$$

for any of the six points $\mathbf{s}$. In general, for $M_{(3\times3)} = 1, 2, \dots$ and $N \equiv 6 \times M_{(3\times3)}$ equally spaced points $\mathbf{s}_k$, $k = 1, \dots, N$ on the sphere, if we consider the zero mean random variables $\{Y(\mathbf{s}_k)\}$ with correlations (25) within the same clique and zero otherwise, then we define $\{X(\mathbf{s}_k)\} \sim \mathrm{CAL}(3 \times 3)$ according to (27). As we explained for the one-dimensional case, it will also hold that $\{Y(\mathbf{s}_k)\} \sim \mathrm{CAL}(3 \times 3)$. We write

$$b_{(1,0)} = -\frac{\sum_{k=1}^{N} Y(\mathbf{s}_k) R_1 Y(\mathbf{s}_k)}{\sum_{k=1}^{N} Y(\mathbf{s}_k)^2}, \quad b_{(0,1)} = -\frac{\sum_{k=1}^{N} Y(\mathbf{s}_k) R_2 Y(\mathbf{s}_k)}{\sum_{j=1}^{N} Y(\mathbf{s}_k)^2},$$

$$b_{(1,1)} = -\frac{\sum_{k=1}^{N} Y(\mathbf{s}_k) R_1 R_2 Y(\mathbf{s}_k)}{\sum_{j=1}^{N} Y(\mathbf{s}_k)^2},$$

and the vectors

$$\mathbf{b} = (b_{(1,0)}, b_{(0,1)}, b_{(1,1)})^{\tau} \text{ and } \boldsymbol{\beta}_0 = (\beta_{(1,0)}, \beta_{(0,1)}, \beta_{(1,1)})^{\tau}.$$

The consistency of $\mathbf{b}$ to $\boldsymbol{\beta}_0$ as well as the asymptotic normality may be established as in Theorem 1, provided that $\{Y(\mathbf{s}_k)\}$ are linear combinations of some $\{\varepsilon(\mathbf{s}_k)\}$, which must satisfy (C1).

As for the variance matrix replacing W, after we track down the proof of Theorem 1 on the circle, we may easily imitate it by defining the random vectors

$$\mathbf{M}(\mathbf{s}_k) = \begin{bmatrix} Y(\mathbf{s}_k) \\ R_1 Y(\mathbf{s}_k) \\ R_2 Y(\mathbf{s}_k) \\ R_1 R_2 Y(\mathbf{s}_k) \end{bmatrix} Y(\mathbf{s}_k), \; k = 1, \ldots, N,$$

as well as the set

$$\mathcal{F} = \{(0, 0), \pm (0, 1), \pm (1, 0), (1, 1)\}.$$

We will write for any $(i, j) \in \mathcal{F}$ the vectors

$$R_1^i R_2^j \mathbf{M}(\mathbf{s}_k) \equiv \begin{bmatrix} R_1^i \; R_2^j \; Y(\mathbf{s}_k) \\ R_1^{i+1} \; R_2^j \; Y(\mathbf{s}_k) \\ R_1^i \; R_2^{j+1} \; Y(\mathbf{s}_k) \\ R_1^{i+1} \; R_2^{j+1} \; Y(\mathbf{s}_k) \end{bmatrix} R_1^i \; R_2^j \; Y(\mathbf{s}_k), \; k = 1, \ldots, N,$$

and will need to compute all the elements of the matrix

$$\sum_{(i, j) \in \mathcal{F}} \mathrm{Cov}(\mathbf{M}(\mathbf{s}_k), R_1^i \; R_2^j \; \mathbf{M}(\mathbf{s}_k)).$$

Using identical arguments like in the proof of Theorem 1, if we write

$$\mathbf{i}_1 = (1, 0) \quad \mathbf{i}_2 = (0, 1) \quad \mathbf{i}_3 = (1, 1),$$

then the (n, m)-th element of the matrix W relating to the covariance of $b_{\mathbf{i}_n}$ with $b_{\mathbf{i}_m}$ $n, m = 1, 2, 3,$ might be computed from

$$w_{n,m} = \sum_{\mathbf{i} \in \mathcal{F}} \{\rho_{\mathbf{i}+\mathbf{i}_n}\rho_{\mathbf{i}+\mathbf{i}_m} + \rho_{\mathbf{i}-\mathbf{i}_n}\rho_{\mathbf{i}+\mathbf{i}_m} + 2\,\rho_{\mathbf{i}_n}\rho_{\mathbf{i}_m}\rho_{\mathbf{i}}^2$$

$$-2\,\rho_{\mathbf{i}_n}\rho_{\mathbf{i}}\,\rho_{\mathbf{i}+\mathbf{i}_m} - 2\,\rho_{\mathbf{i}_m}\rho_{\mathbf{i}}\,\rho_{\mathbf{i}+\mathbf{i}_n}\},$$

where we write $\rho_{\mathbf{i}} \equiv \mathrm{Corr}(Y(\mathbf{s}_k),\ R_1^i R_2^j Y(\mathbf{s}_k)),\ \mathbf{i} = (i, j),\ k = 1, \ldots, N;$ as we explained already, all $\rho_{\mathbf{i}}$ in the formulas above will reduce to $\rho_{\mathbf{i}}, \mathbf{i} \in \mathcal{F}$ only.

The model (27) reduces all covariance dependence to be expressed using three parameters only. Nevertheless, a further reduction of the parameters might be achieved by writing a natural ‘isotropic’ model, such that $\beta_{(1,0)} = \beta_{(0,1)}$. Similarly to the isotropic model on $\mathbb{Z}^2$, such a model on the sphere refuses to distinguish between the different operators R_1 and R_2, and it imposes meaningful equality conditions on the parameters expressing the covariance dependence.

A vague description on how to construct a clique of neighbors for the CAL $(p \times p)$ model follows next. Again, we consider a starting point of the clique, say the South Pole ‘S’. Next, $(p+1)$ equally spaced arcs originate from ‘S’ and they must all cover a $(p+1)$-th of a full circle going through the South and North Poles. On the $(p+1)$ other ends of the arcs, there are the next points of the clique. Excluding the arc already drawn, another p arcs must be considered from each one of these new points and so on. Ideally, the sphere will be filled with points equally scattered and the number of neighbors to form a clique will be counted.

We will write $l = [(p+1)/2]$ and the operators we consider next are subject to p. The $(p+1)$ points added into the clique following ‘S’ must be ‘ordered’ conventionally and must be matched to the operators; if $(p+1)$ is even, we will consider the l operators in order $R_1, \ldots, R_l,$ as well as

$$L_1 \equiv R_1^{-1}, \ldots L_l \equiv R_l^{-1}$$

in the opposite order, i.e.,

$$R_1 \ \dots \ R_l \quad L_l \ \dots \ L_1$$

altogether. If $(p+1)$ is odd, on the other hand, we will consider the extra operator $R_0 = L_0 \equiv R_0^{-1}$, then the operators $R_1, \dots, R_l$, and finally, $L_1, \dots, L_l$ based on $R_1, \dots, R_l$ as before; these will take place in the cyclical order

$$R_0 \ R_1 \ \dots \ R_l \quad L_l \ \dots \ L_1$$

on consecutive paths.

When a clique of neighbors has been formed and a table similar to Table 1 has been constructed using the new operators, the CAL $(p \times p)$ equation will take the form

$$\begin{aligned} X(\mathbf{s}) &- \sum_{i_1>0} \beta_{(i_1,0,\dots,0)}(R_1^{i_1}X(\mathbf{s}) + L_1^{i_1}X(\mathbf{s})) - \dots - \sum_{i_l>0} \beta_{(0,\dots,0,i_l)}(R_l^{i_l}X(\mathbf{s}) + L_l^{i_l}X(\mathbf{s})) \\ &- \sum_{i_1>0,\, i_2\neq 0} \beta_{(i_1,i_2,0,\dots,0)}(R_1^{i_1}R_2^{i_2}X(\mathbf{s}) + L_1^{i_1}L_2^{i_2}X(\mathbf{s})) - \dots \qquad (28) \\ &- \sum_{i_{l-1}>0,\, i_l\neq 0} \beta_{(0,\dots,0,i_{l-1},i_l)}(R_{l-1}^{i_{l-1}}R_l^{i_l}X(\mathbf{s}) + L_{l-1}^{i_{l-1}}L_l^{i_l}X(\mathbf{s})) - \dots \\ &- \sum_{i_1>0,\, i_2,\dots,i_l\neq 0} \beta_{(i_1,i_2,\dots,i_l)}(R_1^{i_1}R_2^{i_2}\dots R_l^{i_l}X(\mathbf{s}) + L_1^{i_1}L_2^{i_2}\dots L_l^{i_l}X(\mathbf{s})) = Y(\mathbf{s}), \end{aligned}$$

if $(p+1)$ is even, and

$$\begin{aligned} X(\mathbf{s}) &- \sum_{i_0>0} \beta_{(i_0,0,\dots,0)}R_0^{i_0}X(\mathbf{s}) - \sum_{i_1>0} \beta_{(0,i_1,0,\dots,0)}(R_1^{i_1}X(\mathbf{s}) + L_1^{i_1}X(\mathbf{s})) - \dots \\ &- \sum_{i_l>0} \beta_{(0,0,\dots,0,i_l)}(R_l^{i_l}X(\mathbf{s}) + L_l^{i_l}X(\mathbf{s})) \\ &- \sum_{i_0,i_1>0} \beta_{(i_0,i_1,0,\dots,0)}(R_0^{i_0}R_1^{i_1}X(\mathbf{s}) + R_0^{i_0}L_1^{i_1}X(\mathbf{s})) - \dots \end{aligned}$$

$$- \sum_{i_0, i_l > 0} \beta_{(i_0, 0, \ldots, 0, i_l)} (R_0^{i_0} R_l^{i_l} X(\mathbf{s}) + R_0^{i_0} L_l^{i_l} X(\mathbf{s}))$$

$$- \sum_{i_1 > 0, i_2 \neq 0} \beta_{(0, i_1, i_2, \ldots, 0)} (R_1^{i_1} R_2^{i_2} X(\mathbf{s}) + L_1^{i_1} L_2^{i_2} X(\mathbf{s})) - \ldots$$

$$- \sum_{i_{l-1} > 0, i_l \neq 0} \beta_{(0, \ldots, 0, i_{l-1}, i_l)} (R_{l-1}^{i_{l-1}} R_l^{i_l} X(\mathbf{s}) + L_{l-1}^{i_{l-1}} L_l^{i_l} X(\mathbf{s})) - \ldots$$

$$- \sum_{i_0, i_1 > 0, i_2, \ldots, i_l \neq 0} \beta_{(i_0, i_1, i_2, \ldots i_l)} (R_0^{i_0} R_1^{i_1} R_2^{i_2} \ldots R_l^{i_l} X(\mathbf{s})$$

$$+ R_0^{i_0} L_1^{i_1} L_2^{i_2} \ldots L_l^{i_l} X(\mathbf{s})) = Y(\mathbf{s}), \tag{29}$$

if $(p+1)$ is odd, provided that the same neighbors of $\mathbf{s}$ have not been included more than once in either equation. Once the order of the model is known, all the restrictions on the auto-correlations ρ or $-\beta$ must be studied with care, such that a minimal number of parameters is used in the CAL equation.

Contrary to the two-dimensional stationary models in the plane or Equation (27) on the sphere, the Equations (28) and (29) use l or $(l+1)$ operators, respectively, though studying a process that takes place on a surface remains the target. This is a consequence of the fact that every point of interest, now connects directly to $(p+1)$ rather than four other points. Nevertheless, for the case that $(p+1)$ is an even number, the points of the clique will seem placed on $(p+1)/2$ full circles and the problem could become two-dimensional by properly re-defining two operators only; one moving the variable between different circles and the other within the same circle. It would be worth deriving more specific results for these natural models that generalize Equation (27).

Back to our definitions with $l(\text{or}(l+1))$ operators, a form of partial ordering in the indexes $i_1, \ldots, i_l$ of (28) or $i_0, i_1, \ldots, i_l$ of (29) (for example, $i_1 > 0, i_2, \ldots, i_l \neq 0$ and $i_1 < 0, i_2, \ldots, i_l \neq 0$ instead of $i_2 > 0, i_1, i_3, \ldots, i_l \neq 0$ and $i_2 < 0, i_1, i_3, \ldots, i_l \neq 0$) is of no importance, as though it separates the neighbors into two groups it takes both groups into account with equal parameters. As in the case of Equation (27), when all neighbors in the clique could be linked to the point of interest with either all R-positive steps, i.e., R_1, R_2 and R_1R_2, or all R-negative steps, i.e., L_1, L_2 and L_1L_2, further investigation is also encouraged in (28) and (29) whether all the relevant summations may reduce to only positive and only negative indexes $i_1, \ldots, i_l$; especially for (28), it seems that applying either $R_1 \ldots R_l$ or $L_1 \ldots L_l$, will move the point of interest to the other end of the diagonal that goes through the center of the sphere.

An hierarchy in the 'dimensions' or in the operators has taken place, when we choose which of the $(p+1)$ branches will relate to which operator and we order the operators cyclically. Again, it might be more meaningful to consider isotropic solutions of the Equation (28) or (29), such that any partiality in the operators does not count at all. Whether a preference of an isotropic model on the sphere is just meaningful, or whether it is in fact necessary, remains a question of interest; either way, it is a special case of the CAL equation presented in this paper. Similarly to the case of the circle, the CAL model may become the main tool to proceed with identification of the order, estimation of the parameters, diagnostic checking of the preferred model and prediction over the sphere, given that a set of regular recordings has been made available.

7. Note on the Asymptotic Results

A stochastic process on the circle or sphere could be considered anywhere over the continuous space. In that sense, a random variable ε, Y, or X might be on any point of the spherical surface and the definitions over a continuous dimension of time should be imitated over space.

Nevertheless, in order to avoid the problems attached to defining uncorrelated ε random variables over the continuous space, such as not being able to assume that they have a finite variance, we have instead taken the route of the fixed and equally spaced points on the cyclical surface. We have established all our asymptotic results as these points are more and more compared to a fixed order of the preferred model, but we have not specified any mechanism producing them. This differs from a time series on $\mathbb{Z}$, when given the N points where the series was observed, we know where the next recording is taking place; similarly for spatial processes on $\mathbb{Z}^2$.

In order to proceed with inference regarding the spatial parameters, we have presented asymptotic results that are valid without using the extra dimension of time; as a result, the asymptotic normality of the estimators for the natural parameters expressing pure spatial dependence on the circle or the sphere has been established.

Appendix: Proof of Theorem 1.

(i) Since, it holds $Y(s_j) = \varepsilon(s_j) + \sum_{k=1}^{l} \theta_k \; R_p^k \varepsilon(s_j),\ j = 1, \dots, N,$ we may combine for fixed $i = 1, \dots, l$

$$Y(s_j) R_p^i Y(s_j) = [\varepsilon(s_j) + \sum_{k=1}^{l} \theta_k \; R_p^k \varepsilon(s_j)][R_p^i \varepsilon(s_j) + \sum_{w=1}^{l} \theta_w R_p^{w+i} \varepsilon(s_j)]. \quad (30)$$

The random variable (30) is a linear combination of the i.i.d. random variables $(R_p^k \varepsilon(s_j))^2$, $k = 0, 1, \dots, p$ with expected value σ^2, and the random variables $R_p^k \varepsilon(s_j) R_p^{k+w} \varepsilon(s_j)$, $k = 0, 1, \dots, p$, $w = 1, \dots, p$ with zero expected value. We know that the expected value of (30) is

$$E\{Y(s_j) \; R_p^i Y(s_j)\} = -\beta_i \nu.$$

Following exactly the same thinking, we may consider the sum of all such random variables on the same $(p+1)$- polygon, i.e., the random variable

$$\sum_{k=0}^{p} R_p^k Y(s_j) \; R_p^{i+k} Y(s_j),$$

which is again a linear combination of the i.i.d. random variables $(R_p^k \varepsilon(s_j))^2$, $k = 0, 1, \dots, p$ with expected value σ^2, and the random variables $R_p^k \varepsilon(s_j) R_p^{k+w} \varepsilon(s_j)$, $k = 0, 1, \ldots, p$, $w = 1, \ldots, p$ with zero expected value. We know that its expected value is now

$$E\left\{\sum_{k=0}^{p} R_p^k Y(s_j) R_p^{i+k} Y(s_j)\right\} = -\beta_i \nu\,(p+1). \tag{31}$$

Now, there are M_p such random variables and we can write that

$$\sum_{j=1}^{N} Y(s_j) R_p^i Y(s_j) = \sum_{j=1}^{M_p} \sum_{k=0}^{p} R_p^k Y(s_j) R_p^{i+k} Y(s_j),$$

and we have a sum of M_p Independent and Identically Distributed random variables, each with mean (31). From the Weak Law of Large Numbers, we may write as $M_p \to \infty$ that

$$\frac{\sum_{j=1}^{N} Y(s_j) R_p^i Y(s_j)}{N} = \frac{1}{p+1} \frac{\sum_{j=1}^{M_p} \sum_{k=0}^{p} R_p^k Y(s_j) R_p^{i+k} Y(s_j)}{M_p}$$

$$\xrightarrow{\;p\;} \frac{1}{p+1} E\left\{\sum_{k=0}^{p} R_p^k Y(s_1) R_p^{i+k} Y(s_1)\right\} = -\beta_i \nu.$$

Of course, in a similar way, it can be shown that

$$\frac{\sum_{j=1}^{N} Y(s_j)^2}{N} \xrightarrow{\;p\;} \nu.$$

Thus, as $M_p \to \infty$

$$b_i = -\frac{\sum_{j=1}^{N} Y(s_j) R_p^i Y(s_j) / N}{\sum_{j=1}^{N} Y(s_j)^2 / N} \xrightarrow{\;p\;} \beta_i,\ i = 1, \ldots, l.$$

(ii) To prove the asymptotic normality and the form of the variance matrix, we will mimic Theorem 7.2.1 of Brockwell and Davis [3]. We will first write the random vectors

$$\mathbf{M}(s_j) = [Y(s_j)\ R_pY(s_j)\ldots R_p^lY(s_j)]^\top Y(s_j),\ j = 1, \ldots, N.$$

Suppose that we also introduce the notation

$$R_p^i\mathbf{M}(s_j) \equiv [R_p^iY(s_j)\ R_p^{i+1}Y(s_j)\ldots R_p^{i+l}Y(s_j)]^\top R_p^iY(s_j),\ j = 1, \ldots, N,\ i \in \mathbb{Z}.$$

Then, we write for $j = 1, \ldots, N$

$$\mathrm{Cov}(\mathbf{M}(s_j),\ R_p^i\mathbf{M}(s_j)) \equiv V_i,\ i = 0, \pm 1, \ldots, \pm p, \tag{32}$$

and further, it holds that

$$V_{-i} = \mathrm{Cov}(\mathbf{M}(s_j),\ R_p^{-i}\mathbf{M}(s_j)) = \mathrm{Cov}(R_p^{-i}\mathbf{M}(s_j),\ \mathbf{M}(s_j))^\top = V_i^\top.$$

Also note that due to the cyclical structure, it holds that

$$V_i^\top = V_{p+1-i}.$$

The (n, m)-th element of $V_i,\ i = 0, 1, \ldots, p$ is

$$E\{Y(s_j)R_p^nY(s_j)R_p^iY(s_j)R_p^{i+m}Y(s_j)\} - \nu^2\rho_n\,\rho_m,\ n, m = 0, 1, \ldots, l,$$

for any $j = 1, \ldots, N$, and $\nu^2\rho_n\,\rho_m = E\{Y(s_j)R_p^nY(s_j))E(R_p^iY(s_j)R_p^{i+m}Y(s_j)\}$ when we take into account the relationship between the ρ-s and the β-s in Theorem 1. Next to (32), it also holds that

$$\mathrm{Cov}(\mathbf{M}(s_j),\ \mathbf{M}(s_k)) = O_{(l+1)\times(l+1)},\ \text{if } |j-k| \neq M_pi,\ j, k = 1, \ldots, N,\ i = 0, \ldots p. \tag{33}$$

Using the same thinking as for the consistency, we may write for any $\boldsymbol{\lambda} \in \mathbb{R}^{l+1}$ the random variable

$$N^{-1/2}\boldsymbol{\lambda}^\top\sum_{j=1}^{N}\mathbf{M}(s_j) = (p+1)^{-1/2}M_p^{-1/2}\sum_{j=1}^{M_p}\boldsymbol{\lambda}^\top\sum_{i=0}^{p}R_p^i\mathbf{M}(s_j). \tag{34}$$

Writing the vector of covariances $\boldsymbol{\beta}_0^* = \nu(1\ \boldsymbol{\beta}_0^\top)^\top$, it holds that

$$E\{\sum_{i=0}^{p} R_p^i \mathbf{M}(s_j)\} = \sum_{i=0}^{p} E\{R_p^i \mathbf{M}(s_j)\} = (p+1)\boldsymbol{\beta}^*_0,\ j = 1, \ldots, N, \tag{35}$$

and that

$$\mathrm{Var}\{\sum_{i=0}^{p} R_p^i \mathbf{M}(s_j),\ \sum_{k=0}^{p} R_p^k \mathbf{M}(s_j)\} = \sum_{i=0}^{p}\sum_{k=0}^{p} \mathrm{Cov}(R_p^i \mathbf{M}(s_j),\ R_p^k \mathbf{M}(s_j))$$

$$= \sum_{i=0}^{p}\sum_{k=0}^{p} \mathrm{Cov}(R_p^i \mathbf{M}(s_j),\ R_p^{i+k-i} \mathbf{M}(s_j)) = \sum_{i=0}^{p}\sum_{k=-i}^{p-i} \mathrm{Cov}(R_p^i \mathbf{M}(s_j),\ R_p^{i+k} \mathbf{M}(s_j))$$

$$= \sum_{i=0}^{p}\sum_{k=-i}^{p-i} V_k = (\sum_{k=0}^{p} V_k) + (\sum_{k=-1}^{p-1} V_k) + \ldots + (\sum_{k=-(p-1)}^{1} V_k) + (\sum_{k=-p}^{0} V_k)$$

$$= (\sum_{k=0}^{p} V_k) + (V_1^{\tau} + \sum_{k=0}^{p-1} V_k) + \ldots + (\sum_{k=0}^{p-1} V_k^{\tau} + V_1) + (\sum_{k=0}^{p} V_k^{\tau}) = (p+1)\sum_{k=0}^{p} V_k, \tag{36}$$

thanks to restrictions of the cyclical nature that we explained before. We will write $V \equiv \sum_{i=0}^{p} V_i$.

The random variables $\{\boldsymbol{\lambda}^{\tau}\sum_{i=0}^{p} R_p^i \boldsymbol{M}(s_j),\ j = 1, \ldots, M_p\}$ are Independent and Identically Distributed, as they take place on different $(p+1)$ - polygons. According to the Central Limit Theorem, and using (35) and (36), it holds as $M_p \to \infty$ that

$$M_p^{-1/2}\sum_{j=1}^{M_p} \boldsymbol{\lambda}^{\tau} \sum_{i=0}^{p} R_p^i \boldsymbol{M}(s_j) \xrightarrow{D} N((p+1)\boldsymbol{\lambda}^{\tau}\boldsymbol{\beta}_0^*,\ (p+1)\boldsymbol{\lambda}^{\tau} V \boldsymbol{\lambda}), \tag{37}$$

which combined with (34) implies that it holds

$$N^{-1/2}\boldsymbol{\lambda}^{\tau}\sum_{j=1}^{N} \mathbf{M}(s_j) \xrightarrow{D} N(\boldsymbol{\lambda}^{\tau}\boldsymbol{\beta}_0^*,\ \boldsymbol{\lambda}^{\tau} V \boldsymbol{\lambda}), \tag{38}$$

as $M_p \to \infty$. Then from the Cramer-Wold device (see Brockwell and Davis [3]), it holds

$$N^{-1/2}\sum_{j=1}^{N}\mathbf{M}(s_j)\xrightarrow{\quad D\quad} N(\boldsymbol{\beta}_0^*,\, V), \tag{39}$$

as $M_p \to \infty$.

Regarding the variance matrix V, if we consider its $(n,\, m)$-th, where $n,\, m = 0, 1, \ldots, l$, then this is equal to

$$\sum_{i=0}^{p}\{E(Y(s_j)R_p^nY(s_j)R_p^iY(s_j)\, R_p^{i+m}Y(s_j)) - \nu^2\rho_n\,\rho_m\} \equiv Q - (p+1)\nu^2\rho_n\rho_m, \tag{40}$$

for any $j = 1, \ldots, N$. Using identical arguments like in Proposition 7.3.1 of Brockwell and Davis [3], we may derive

$$Q = (\eta - 3)(\nu\,\rho_n)(\nu\,\rho_m) + \sum_{i=0}^{p}E(Y(s_j)R_p^nY(s_j))\; E(R_p^iY(s_j)R_p^{i+m}Y(s_j))$$

$$+\sum_{i=0}^{p}E(Y(s_j)R_p^iY(s_j))\; E(R_p^nY(s_j)\; R_p^{i+m}Y(s_j))$$

$$+\sum_{i=0}^{p}E(Y(s_j)\; R_p^{i+m}Y(s_j))\; E(R_p^nY(s_j)\; R_p^iY(s_j))$$

$$= (\eta - 3)\nu^2\,\rho_n\,\rho_m + \nu^2\sum_{i=0}^{p}\{\rho_n\,\rho_m + \rho_i\,\rho_{i-n+m} + \rho_{i+m}\,\rho_{i-n}\}. \tag{41}$$

If we combine (40) and (41), we come up with the $(n,\, m)$-th element of V to be equal to $(\eta - 3)\nu^2\,\rho_n\,\rho_m + \nu^2\sum_{i=0}^{p}\{\rho_i\,\rho_{i-n+m} + \rho_{i+m}\,\rho_{i-n}\}$, which resurrects Proposition 7.3.1 of Brockwell and Davis [3].

Finally, for the coefficients defined in Theorem 1 as

$$\mathbf{b}^{\top} = g(N^{-1/2}\sum_{j=1}^{N}\mathbf{M}^{\top}(s_j)),$$

where we use the function $g(\cdot)$ from $\mathbb{R}^{l+1}$ into $\mathbb{R}^{l}$, such that

$$g(x_0 \ x_1 \ \ldots \ x_l) = (-x_1 / x_0 \ \ldots \ -x_l / x_0), \ x_0 \neq 0,$$

Bartlett's formula is derived with identical arguments like in Theorem 7.2.1 of Brockwell and Davis [3].

References

[1] J. Besag, Spatial interaction and the statistical analysis of lattice systems (with discussion), J. R. Statist. Soc. B 36 (1974), 192-236.

[2] J. Besag, Statistical analysis of non-lattice data, Statistician 24 (1975), 179-195.

[3] P. J. Brockwell and R. A. Davis, Time Series: Theory and Methods, 2nd ed., Springer-Verlag, New-York, (1991).

[4] N. A. Cressie and H. C. Huang, Classes of nonseparable, spatiotemporal stationary covariance functions, J. Amer. Stat. Assoc. 94 (1999), 1330-1340.

[5] C. Dimitriou-Fakalou, Modelling data observed irregularly over space and regularly in time, Statistical Methodology 6 (2009), 120-132.

[6] C. Dimitriou-Fakalou, Statistical inference for spatial auto-linear processes, (submitted), (2009).

[7] R. H. Jones, Stochastic processes on a sphere, The Annals of Mathematical Statistics 34 (1963), 213-218.

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

[9] C. Ma, Families of spatio-temporal stationary covariance models, Journal of Statistical Planning and Inference 116 (2003), 489-501.

[10] M. G. McLeod, Stochastic processes on a sphere, Physics of the Earth and Planetary Interiors 43 (1986), 283-299.

[11] P. Whittle, On stationary processes in the plane, Biometrika 41 (1954), 434-449.

■