Pivotal discrepancy measures for Bayesian modelling of spatio-temporal data

Lindsay R. Morris¹ · Nokuthaba Sibanda¹

Received: 18 August 2020 / Revised: 11 January 2022 / Accepted: 26 January 2022 /

Published online: 12 February 2022

© The Author(s) 2022

Abstract

Within the field of geostatistics, Gaussian processes are a staple for modelling spatial and spatio-temporal data. Statistical literature is rich with estimation methods for the mean and covariance of such processes (in both frequentist and Bayesian contexts). Considerably less attention has been paid to developing goodness-of-fit tests for assessment of model adequacy. Jun et al. (Environmetrics 25(8):584–595, 2014) introduced a statistical test that uses pivotal discrepancy measures to assess goodness-of-fit in the Bayesian context. We present a modification and generalization of their statistical test. The initial method involves spatial partitioning of the data, followed by evaluation of a pivotal discrepancy measure at each posterior draw to obtain a posterior distribution of pivotal statistics. Order statistics from this distribution are used to obtain approximate p-values. Jun et al. (Environmetrics 25(8):584–595, 2014) use arbitrary partitions based on pre-existing spatial boundaries. The partitions are made to be of equal size. Our contribution is two-fold. We use K-means clustering to create the spatial partitions and we generalise Jun et al.'s approach to incorporate unequal partition sizes. Observations from a spatial or spatio-temporal process are partitioned using an appropriate feature vector that incorporates the geographic location of the observations into subsets (not necessarily of the same size). The method's viability is illustrated in a simulation study, and in an application to hoki (*Macruronus novaezelandiae*) catch data from a survey of the sub-Antarctic region.

Keywords Bayesian models · Goodness-of-fit · Pivotal discrepancy measure · Spatio-temporal models

Handling Editor: Pierre R. L. Dutilleul

✉ Nokuthaba Sibanda
nokuthaba.sibanda@vuw.ac.nz

¹ School of Mathematics and Statistics, Victoria University of Wellington, Wellington, New Zealand

1 Introduction

The literature that surrounds the subject of spatial and spatio-temporal statistics is predominantly concerned with parametric inference for the covariance structure. Within geostatistics (where data are observed at specific locations in time), a great deal of attention has been paid to proposing and describing new spatial and spatio-temporal models, studying their characteristics, and developing estimation methods from within both frequentist and Bayesian frameworks.

A variety of examples can be found in the literature. For instance, parametric spatial and spatio-temporal covariance models that assume stationarity have been developed and estimated in Cressie and Huang (1999), Gneiting (2002), and Stein (2005). Cressie and Huang (1999) derived theoretical results on positive definiteness of the stationary spatio-temporal covariance function. This led to the proposal of a class of non-separable parametric covariance functions, which they applied to the problem of mapping the east-west component of wind speed over a region in the tropical western Pacific Ocean. Gneiting (2002) extended the work of Cressie and Huang (1999) further and provided a more general class of covariance functions that do not depend on Fourier transform pairs. Their work was illustrated through an application to Irish wind data. Stein (2005) extended these ideas further to consider spherical functions, applied to the same Irish wind dataset used in Gneiting (2002). Non-stationary approaches have been proposed and investigated in Paciorek (2013), Ecker et al. (2013), and Fouedjio (2017). A Bayesian hierarchical approach to modeling spatial and spatio-temporal covariance functions was taken in Cameletti et al. (2011), Cameletti et al. (2013), Sahu and Bakar (2012), and Banerjee et al. (2014). Each of these papers makes use of the flexible Matérn class of covariance functions. White and Ghosh (2009) extended the conditional autoregressive model for areal data proposed in Besag (1974), to geostatistical data, and introduced the stochastic neighbourhood conditional autoregressive model. An overview of the state of Bayesian hierarchical spatio-temporal models is given in Gelfand and Banerjee (2017).

Markedly less attention has been paid to developing goodness-of-fit tests that identify model mis-specification or allow for selection of a “best” model. Model mis-specification in the context of parametric covariance models for spatio-temporal processes refers to models that have an incorrect mean function or covariance structure. At present, there is no generalized formal theory for assessing goodness-of-fit for spatio-temporal models that are defined using parametric covariance functions. Instead, there is a range of criteria and tests that have been used when fitting a spatio-temporal covariance model to data.

Literature has seen the use of Akaike information criterion, AIC (Akaike 1973) and Bayesian information criterion, BIC (Schwarz 1978), which are popular model selection tools for a wide range of frequentist analyses. Huang et al. (2007) proposed model comparison for space-time models using these criteria, and investigated their usefulness through simulation and an application to surface shortwave radiation budget analysis. Another criterion, deviance information criterion, DIC (Spiegelhalter et al. 2002), was used by Pollice (2011) to compare multivariate receptor models for identifying the spatial locations of major PM₁₀ pollution sources. To compare predictive capabilities, use of mean square and root mean square prediction errors at fixed times

were illustrated in Huang et al. (2007). Further, Sahu and Bakar (2012) applied the predictive model choice criterion (PMCC), which included a term for model complexity. In more recent times, we have seen the proposal and use of the widely applicable information criterion, WAIC, (Watanabe 2010). The WAIC is an information criterion constructed in the same vain as DIC, but is fully Bayesian as it uses the entire posterior distribution, unlike the DIC which is based on a point estimate. Vehtari and Gelman (2014) and Vehtari et al. (2017), adopted WAIC as a method for approximating leave-one-out cross validation for model goodness-of-fit.

These model selection/goodness-of-fit criterion are inappropriate for some spatio-temporal models. The WAIC relies on a partition of a data set into individual data points, and its computation involves a sum of the log-posterior evaluated over all data points and posterior draws. For spatial or spatio-temporal data, partition into individual data points is not valid due to the inherent structure in the data (Gelman et al. 2014). In addition, Jun et al. (2014) illustrated that global goodness-of-fit tests applied to the entire dataset provide limited power to detect model mis-specification when applied to spatial data.

Bastos and O'Hagan (2009) proposed numerical and graphical diagnostic tools for Bayesian model checking in the context of Gaussian processes. A Bayesian approach for assessing goodness-of-fit for Gaussian random fields (GRFs) based on pivotal discrepancy measures was introduced by Johnson (2007) and further discussed in Yuan and Johnson (2012) and Jun et al. (2014). The approach can be used for GRFs with stationary and nonstationary covariances and to data observed at regular or irregularly spaced locations. More recently, Lobo and Fonseca (2020) use a cross-validation approach to assess goodness-of-fit of spatial models. These approaches, however, consider only spatial data. In this paper, we extend the approach described in Jun et al. (2014) to assess goodness-of-fit of Bayesian spatio-temporal models using pivotal discrepancy measures. Jun et al. (2014) proposed the method of partitioning the data to increase the power to detect model mis-specification, but used equal partition sizes. The primary innovations in this article are to extend the Jun et al. (2014) approach to partitions of unequal sizes and the use of K-means partitioning as a method for inducing homogeneity within partitions, when there are no pre-set spatial boundaries.

The paper is divided in to the following sections. We introduce and define the general spatio-temporal model in Sect. 2. In Sect. 3, we present the pivotal discrepancy measure for a spatio-temporal model evaluated at a sample from the posterior distribution. Further, we present the pivotal discrepancy measure for subsets of data of unequal size. Section 4 is dedicated to evaluation of the method using a simulation study. This is followed in Sect. 5 by an application and assessment of spatio-temporal models to hoki catch weight data. We discuss the findings of the paper in Sect. 6

2 Spatio-temporal Gaussian process model

We are interested in modelling the covariance structure of an observed univariate spatio-temporal process $\{y(s, t) : (s, t) \in \mathbb{R}^d \times \mathbb{R}\}$. The Gaussian process model is a commonly used model due to its flexibility in modelling the effect of relevant covariates as well as time and space dependence and is defined in several articles. Following

the notation of Cameletti et al. (2011) and Sahu and Bakar (2012), we assume that an observation $y(s_i, t)$ measured at location s_i , where $i = 1, \dots, n$ and time $t = 1, \dots, T$, can be modelled by a Gaussian process with measurement equation,

$$y(s_i, t) = \mu(s_i, t) + Z(s_i, t) + \varepsilon(s_i, t),$$

where $\mu(s_i, t) = \mathbf{x}(s_i, t)\boldsymbol{\beta}$ and $\mathbf{x}(s_i, t) = (1, x_1(s_i, t), \dots, x_p(s_i, t))$ denotes the $(p+1)$ -dimensional vector of covariates for location s_i at time t , $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)'$ is the coefficient vector, and n and T are the numbers of spatial locations and time points respectively. The residual is partitioned into two components, one spatial ($Z(s_i, t)$) and one non-spatial ($\varepsilon(s_i, t)$). The measurement error (nugget effect), $\varepsilon(s_i, t)$, is modelled independently as a white noise process, $N(0, \sigma_\varepsilon^2)$. Lastly, $Z(s_i, t)$ is a realization of a latent spatio-temporal process that is modelled by a Gaussian process that changes in time with first order autoregressive dynamics, and coefficient ρ as follows,

$$Z(s_i, t) = \rho Z(s_i, t-1) + \omega(s_i, t), \quad t = 2, \dots, T,$$

where $|\rho| < 1$. $Z(s_i, 1)$ is such that,

$$Z(s_i, 1) \sim N\left(0, \sigma_\omega^2 \frac{R(s_i, s_j; \boldsymbol{\phi})}{1 - \rho^2}\right).$$

Furthermore, $\omega(s_i, t)$ is modelled by a zero-mean Gaussian distribution, in which we assume temporal independence. It is characterized fully by the spatio-temporal covariance function,

$$\text{Cov}(\omega(s_i, t), \omega(s_j, t^*)) = \begin{cases} 0 & \text{if } t \neq t^* \\ \sigma_\omega^2 R(s_i, s_j; \boldsymbol{\phi}) & \text{if } t = t^*, \end{cases} \quad (1)$$

where $i \neq j$, and t and t^* are two different time points. The parameter σ_ω^2 denotes the spatial variance and $R(\cdot)$ is a correlation function that depends on parameter vector $\boldsymbol{\phi}$, such that the resulting correlation matrix, $\mathbf{R}$ is positive definite. Note that $\mathbf{R}$ is an $n \times n$ matrix with elements $R(s_i, s_j; \boldsymbol{\phi})$. We implicitly make the assumption that the overall spatial covariance structure of the data is constant over time. Also note that we make no assumption of spatial stationarity or isotropy, as evidenced by the spatial covariance function in Eq. 1.

By collecting all the observations measured at time t in a vector denoted by $\mathbf{y}_t = (y(s_1, t), \dots, y(s_n, t))'$, we can write

$$\mathbf{y}_t = \boldsymbol{\mu}_t + \mathbf{Z}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \text{MVN}(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}_n),$$

for $t = 1, \dots, T$, where $\boldsymbol{\mu}_t = \mathbf{X}_t \boldsymbol{\beta}$, $\mathbf{X}_t = (\mathbf{x}(s_1, t)', \dots, \mathbf{x}(s_n, t)')'$, σ_ε^2 is the variance of the nugget effect, and $\mathbf{I}_n$ is the identity matrix with dimension n . As before, the spatio-temporal process is decomposed into spatial and temporal terms,

$$\mathbf{Z}_t = \rho \mathbf{Z}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim \text{MVN}(\mathbf{0}, \sigma_\omega^2 \mathbf{R})$$

for $t = 1, \dots, T$ and positive definite correlation matrix $\mathbf{R}$.

Also let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma_\varepsilon^2, \rho, \sigma_\omega^2, \boldsymbol{\phi})$ denote the vector of parameters. It is then implied that

$$\mathbf{y}_t | \mathbf{Z}_t, \boldsymbol{\theta} \sim \text{MVN}(\boldsymbol{\mu}_t + \mathbf{Z}_t, \sigma_\varepsilon^2 \mathbf{I}_n), \quad (2)$$

for $t = 1, \dots, T$, and

$$\mathbf{Z}_t | \mathbf{Z}_{t-1}, \boldsymbol{\theta} \sim \text{MVN}(\rho \mathbf{Z}_{t-1}, \sigma_\omega^2 \mathbf{R}), \quad (3)$$

for $t = 2, \dots, T$, and that $\mathbf{Z}_1$ comes from the stationary distribution of the AR(1) process,

$$\mathbf{Z}_1 | \boldsymbol{\theta} \sim \text{MVN}\left(\mathbf{0}, \frac{\sigma_\omega^2}{1 - \rho^2} \mathbf{R}\right). \quad (4)$$

From Eqs. 2–4 we can then write that the marginal distribution of $\mathbf{y}_t$ (given the parameters) is

$$\mathbf{y}_t | \boldsymbol{\theta} \sim \text{MVN}\left(\boldsymbol{\mu}_t, \frac{\sigma_\omega^2}{1 - \rho^2} \mathbf{R} + \sigma_\varepsilon^2 \mathbf{I}_n\right). \quad (5)$$

We now present a pivotal discrepancy measure for spatio-temporal data $\mathbf{y}_t$ based on a posterior sample of the parameter vector $\boldsymbol{\theta}$.

3 Pivotal discrepancy measures

Let $\tilde{\boldsymbol{\theta}}^{(m)}$ represent the m th draw of the parameter vector $\boldsymbol{\theta}$ from the posterior distribution $\pi(\boldsymbol{\theta} | \mathbf{y})$, where $\mathbf{y} = \{\mathbf{y}_1, \dots, \mathbf{y}_T\}$. We can construct a pivotal quantity,

$$S(\mathbf{y}_t, \tilde{\boldsymbol{\theta}}^{(m)}) = (\mathbf{y}_t - \boldsymbol{\mu}_t^{(m)})' \left(\frac{\sigma_\omega^{2(m)}}{1 - \rho^{(m)2}} \mathbf{R}^{(m)} + \sigma_\varepsilon^{2(m)} \mathbf{I}_n \right)^{-1} (\mathbf{y}_t - \boldsymbol{\mu}_t^{(m)}) \sim \chi_n^2, \quad (6)$$

for $t = 1, \dots, T$, $m = 1, \dots, M$, where M is the total number of posterior draws. Then $S(\mathbf{y}_t, \tilde{\boldsymbol{\theta}}^{(m)})$ is χ^2 -distributed on n degrees of freedom (Johnson 2007).

Jun et al. (2014) highlighted two complications that arise when Eq. 6 is used in a Bayesian goodness-of-fit test for spatial models, which also arise for the spatio-temporal case. The first complication is that a test based on the test statistic in Eq. 6 typically provides little power to detect model misspecification when it is applied globally to the entire data vector. Jun et al. (2014) illustrated this through an example where a simple Bayesian linear regression is fitted to a fictional dataset that exhibited larger variability at the extreme values of a covariate, and smaller variability around the mean of the covariate. Their test, based on the spatial equivalent of the test statistic in Eq. 6, was unable to detect departure of data from the simple linear model. This was

due to the cancellation of the large and small contributions from the residuals when the statistic $S(\mathbf{y}_t, \tilde{\boldsymbol{\theta}}^{(m)})$ was applied to the entire data set. Jun et al. (2014) proposed a partitioning strategy, where the chi-squared diagnostic was constructed using residuals from distinct regions of the spatial domain. Use of the partitioning strategy allowed lack of fit of the model to the data in each partition to be correctly detected and the goodness-of-fit test failed. Partitioning of the data was further motivated in a simulation test and applications to Colorado precipitation data and total column ozone data. We propose an extension of their strategy in Sect. 3.1.

The second complication is how to combine the pivotal discrepancy measures based on many posterior draws and a partitioned dataset when conducting a goodness-of-fit test. A single posterior draw, $\tilde{\boldsymbol{\theta}}^{(m)}$, from the posterior distribution based on a non-partitioned dataset gives the statistic $S(\mathbf{y}_t, \tilde{\boldsymbol{\theta}}^{(m)}) \sim \chi_n^2$. Each posterior draw gives a different value of the test statistic and these values will be correlated. Jun et al. (2014) proposed diagnostics based on bounds on the distribution of order statistics (Caraux and Gascuel 1992; Rychlik 1992) to carry out a goodness-of-fit test that makes use of multiple correlated statistics obtained from the posterior draws. We adopt that approach in this article.

3.1 Partitioning the observed locations into C subsets (not necessarily of equal size)

Jun et al. (2014) proposed partitioning the set of observed locations into C subsets of equal size w and showed that partitioning the observation vector into regions of high and low variability allowed the test to detect model mis-specification. They suggest partitioning based on either prior knowledge regarding regions of likely homogeneity, or according to well defined spatial boundaries. We extend their approach to consider partitions of spatio-temporal data, with partitions of the spatial domain of unequal sizes, w_j , $j = 1, \dots, C$. Applying Eq. 6 to the partitioned spatio-temporal data gives:

$$S_j(\mathbf{y}_{tj}, \tilde{\boldsymbol{\theta}}^{(m)}) = (\mathbf{y}_{tj} - \boldsymbol{\mu}_{tj}^{(m)})' \left(\frac{\sigma_\omega^{2(m)}}{1 - \rho^{(m)2}} \mathbf{R}_j^{(m)} + \sigma_\varepsilon^{2(m)} \mathbf{I}_{w_j} \right)^{-1} (\mathbf{y}_{tj} - \boldsymbol{\mu}_{tj}^{(m)}) \sim \chi_{w_j}^2, \quad (7)$$

for $t = 1, \dots, T$, $j = 1, \dots, C$, and $m = 1, \dots, M$, where $\mathbf{y}_{tj}$, $\boldsymbol{\mu}_{tj}$, and $\mathbf{R}_j$ denote the parts of Eq. 6 corresponding to subset j , and w_j is the number of observed locations in subset j .

In addition to allowing the subsets to vary in size, we propose use of the K-means clustering algorithm (Alsabti et al. 1997) to partition the spatial domain in cases where there are no a priori well-defined spatial boundaries. We design the algorithm (see Algorithm 1 below) to minimise intra-cluster distances from the cluster centroid, thus creating subsets of locations that are likely to have more heterogeneity than the entire domain as a whole. Other clustering algorithms can be used and we consider these alternatives briefly in the Discussion section.

Algorithm 1 K-Means clustering

- 1: Choose an appropriate number of subsets, C , to partition the data into.
- 2: Randomly select the initial cluster centroids, $\boldsymbol{y}_1, \dots, \boldsymbol{y}_C \in \mathbb{R}^d$, where $\boldsymbol{y}_j$ is a d -dimensional location vector that will result in subsets that each exhibit more homogeneity than the full dataset. For example, $\boldsymbol{y}_j = \boldsymbol{s}'_j$, where $\boldsymbol{s}'_j$ is the location vector for the centroid of the j^{th} subset.
- 3: Repeat until convergence:

- for $i = 1, \dots, n$, set,

$$c_i := \arg \min_j (||\boldsymbol{u}_i - \boldsymbol{y}_j||)^2,$$

where $\boldsymbol{u}_i$ is the location vector for observation i .

- for $j = 1, \dots, C$, set

$$\boldsymbol{y}_j = \frac{\sum_{i=1}^n \mathbf{1}_{\{c_i=j\}} \boldsymbol{u}_i}{\sum_{i=1}^n \mathbf{1}_{\{c_i=j\}}}.$$

3.2 Nominal distribution of the ordered pivotal statistics

The screening diagnostics we use are based on bounds of order statistics given in Proposition 3 in Caraux and Gascuel (1992). These bounds are applied to non-identically distributed dependent variables, thereby allowing us to generalise the diagnostics proposed by Jun et al. (2014).

Let $X_1, \dots, X_N$ denote a sample from a dependent set of N random variables with non-identical distribution functions, $F_{X_1}, \dots, F_{X_N}$. We define a set of order statistics $X_{(1)}, \dots, X_{(N)}$ for these random variables. Also, let $F_{X_{r:N}}$ denote the distribution function for the r th-order statistic out of a sample of N dependent draws from $F_{X_1}, \dots, F_{X_N}$. Then,

$$\sup \left(0, 1 - \frac{\sum_{j=1}^N (1 - F_{X_j}(x))}{N - r + 1} \right) \leq F_{X_{r:N}}(x) \leq \inf \left(\frac{\sum_{j=1}^N F_{X_j}(x)}{r}, 1 \right). \quad (8)$$

We partition the spatial domain into C groups of potentially unequal size, w_j , $j = 1, \dots, C$. The pivotal statistic $S_j(\mathbf{y}_{tj}, \tilde{\boldsymbol{\theta}}^{(m)})$ is calculated for each partition, j , $j = 1, \dots, C$, time point t , $t = 1, \dots, T$ and posterior draw m , $m = 1, \dots, M$. This results in a total of CTM dependent test statistics $\{S_j(\mathbf{y}_{tj}, \tilde{\boldsymbol{\theta}}^{(m)}) : j = 1, \dots, C; t = 1, \dots, T, m = 1, \dots, M\}$, with respective density function $\chi_{w_j}^2$. We denote the r th-order statistic from this set by $S_{(r)}$, where $r = 1, \dots, CTM$, and let F_r denote the distribution function of the $\chi_{m_j}^2$ distribution. It follows from above that,

$$P(S_{(r)} < t) \leq \inf \left(1, \frac{\sum_{r=1}^{CTM} F_r(t)}{r} \right),$$

$$P(S_{(r)} > t) \leq \inf \left(1, \frac{\sum_{r=1}^{CTM} (1 - F_r(t))}{CTM - r + 1} \right).$$

3.3 Pivotal discrepancy measure goodness-of-fit test for Bayesian inference

We propose the following procedure for testing goodness-of-fit for Gaussian spatio-temporal models:

1. Partition the set of observed locations into C subsets, Q_j of size w_j , where $j = 1, \dots, C$, using K-means clustering. For each $t = 1, \dots, T$, let y_{tj} , X_{tj} , and R_j denote the parts of Eq. 5 that correspond to subset Q_j .
2. Generate posterior samples for θ , $\theta^{(1)}, \dots, \theta^{(M)}$, based on the entire observed dataset $(y_1, \dots, y_T)$.
3. For every sampled parameter vector $\theta^{(i)}$, and each data subset y_{tj} , for every t , calculate the pivotal statistic in Equation 7.
4. Collect all CTM statistics in an ordered set $\{S_j(y_{tj}, \theta^{(i)}) : j = 1, \dots, C, t = 1, \dots, T, i = 1, \dots, M\}$, and denote the k th-order statistic from this set by $S_{(k)}^*$.
5. Perform a two-sided goodness-of-fit test at significance level α by first specifying integers l and u such that $1 \leq l < u \leq CTM$. Then determine t_l and t_u such that

$$\left(\left[\frac{\sum_{k=1}^{CTM} F_k(t_l)}{l} \right] - \frac{\alpha}{2} \right)^2, \quad (9)$$

and

$$\left(\left[1 - \frac{\sum_{k=1}^{CTM} (1 - F_k(t_u))}{CTM - u + 1} \right] - \frac{\alpha}{2} \right)^2 \quad (10)$$

are minimized. If either $S_{(l)}^* < t_l$ or $S_{(u)}^* > t_u$, then the assumed model can be rejected in a two-sided test of size α .

Jun et al. (2014) recommend that l and u be selected such that $l = r_l CTM$ and $u = r_u CTM$, where $0 < r_l < r_u < 1$, for example $r_l = 0.1$ and $r_u = 0.9$.

4 Simulation

A simulation experiment was performed to assess the ability of the goodness-of-fit test to detect misspecification of the covariance structure of a model. A total of 30 pairs of longitude and latitude values, $s_i = (s_{i1}, s_{i2})$, $i = 1, \dots, n = 30$, were sampled randomly from one of three subsets within the unit square. Within the first subset, Q_1 , 5 locations were generated uniformly from the lower left $[0, 0.2] \times [0, 0.2]$ portion of the unit square. In the second subset, Q_2 , 10 locations were uniformly sampled from the lower right $[0.8, 1] \times [0, 0.2]$ portion of the unit square. Finally, in subset Q_3 , 15 locations were uniformly sampled from the entire unit square. The motivation is that the fit of a spatial model can be best tested by comparing its fit in distinct regions (where its local smoothness properties can be evaluated), with its fit to points distributed throughout the domain (where its global features can be evaluated) as mentioned in Jun et al. (2014). Subsets Q_1 and Q_2 provided clusters of locations that allow for the assessment of local model fit, whereas subset Q_3 provides

Fig. 1 Domain and locations of the simulated data, colour-coded by subset

motivation for assessing global model fit. Figure 1 shows the simulated locations and the corresponding subsets.

Three datasets were simulated using the spatio-temporal process defined in Sect. 2. The mean process, μ_t , was set to zero, to allow for detection of model misspecification through the covariance structure only. Observed data $\{y_t\}$ were simulated for $t = 1, \dots, 5$ time points.

Three variants of the general Matérn correlation function

$$R(s_i, s_j; \nu, \phi) = \frac{1}{\Gamma(\nu)2^{\nu-1}} \left(\frac{\sqrt{8\nu}}{\phi} \|s_i - s_j\| \right)^\nu K_\nu \left(\frac{\sqrt{8\nu}}{\phi} \|s_i - s_j\| \right), \quad (11)$$

with closed form expressions were used to construct the covariance matrix $\sigma_\omega^2 \mathbf{R}$. In the equation above, $\nu > 0$ controls the smoothness of the realised random field, ϕ is a spatial scale parameter, K_ν is a modified Bessel function of order ν and $\|s_i - s_j\|$ is the Euclidean distance between the locations (Banerjee et al. 2014).

The first variant of $R(s_i, s_j; \nu, \phi)$,

$$R(s_i, s_j; \phi) = \exp \left(\frac{-\|s_i - s_j\|}{\phi} \right), \quad (12)$$

is the closed form of the Matérn correlation function, where the smoothness parameter, ν , is set to 0.5 and is also known as the exponential correlation function. The second

variant,

$$R(s_i, s_j; \phi) = \exp \left[- \left(\frac{\|s_i - s_j\|}{\phi} \right)^2 \right], \quad (13)$$

is the closed form of the Matérn correlation function, where the smoothness parameter, $\nu \rightarrow \infty$, and is known as the Gaussian correlation function. The third variant,

$$R(s_i, s_j; \phi) = s_{i2}s_{j2} \exp \left(\frac{-\|s_i - s_j\|}{\phi} \right), \quad (14)$$

is a non-stationary form of the exponential correlation function given by Eq. 12, that allows the correlation between observations separated by distance $\|s_i - s_j\|$ to scale by their latitudes, s_2 .

The following parameters were chosen to simulate the data $\mathbf{y}_t$. The measurement variance (nugget variance) was set to $\sigma_\varepsilon^2 = 0.0001$. A small value was chosen to focus on identifying an incorrect spatio-temporal covariance structure. Further, we set $\rho = 0.7$ and $\sigma_\omega^2 = 1$. Finally, we set $\phi = 0.2$ in Eqs. 12 and 14, and $\phi = 0.8$ in Eq. 13.

We fitted the spatio-temporal model given in Sect. 2 with the covariance function in Eq. 12 (exponential correlation function) to each of the three datasets. We excluded covariates, with only a single intercept term included in the mean function, such that $\mu_t = \mathbf{1}_{30}\beta$, where $\mathbf{1}_{30}$ is a vector of 1's. The parameters $\theta = (\beta, \sigma_\varepsilon^2, \sigma_\omega^2, \phi, \rho)'$ were assumed *a priori* independent, and were assigned the non-informative prior distributions,

$$\beta \sim N(0, 100), \quad \sigma_\varepsilon^2 \sim \text{IG}(2, 1), \quad \sigma_\omega^2 \sim \text{IG}(2, 1), \quad \phi \sim U(0.001, 2), \quad \rho \sim U(-1, 1).$$

Markov chain Monte Carlo (MCMC) was used to fit the model to the data and this was done in **R** using the package NIMBLE (NIMBLE Development Team 2017). Two chains of 100,000 iterations each were generated for the parameter vector $\theta = (\beta, \rho, \phi, \sigma_\omega^2, \sigma_\varepsilon^2)^T$ for each dataset. The first 90,000 iterations from each chain were discarded as warm-up, and the remaining draws were combined, resulting in a posterior sample of size $M = 20,000$. Convergence of the Markov chain was assessed using traceplots (not provided) and potential scale reduction factor ($\hat{R}$) values. We took a value of $\hat{R} > 1.1$ to indicate lack of convergence. For each fitted model, pivotal quantities for every posterior sample were calculated inline with Eq. 7. We considered three cases of partitioning to assess the impact it has on testing goodness-of-fit. In the first case, the locations were not partitioned into subsets. Pivotal quantities for each fitted model $S(\mathbf{y}_t, \tilde{\theta}^{(m)})$ for $t = 1, \dots, 5$ and $m = 1, \dots, 20,000$ were calculated, combined and ordered. For the second case, the locations were partitioned into $C = 3$ subsets of $w = 10$. Pivotal quantities for each fitted model $S(\mathbf{y}_{tj}, \tilde{\theta}^{(m)})$ for $t = 1, \dots, 5$, $j = 1, 2, 3$, and $m = 1, \dots, 20,000$ were calculated, combined and ordered. Finally, the locations were partitioned into the subsets S_1 , S_2 , and S_3 , that were used to simulate the locations. Pivotal quantities for each fitted model $S(\mathbf{y}_{tj}, \tilde{\theta}^{(m)})$ for

$t = 1, \dots, 5$, $j = 1, 2, 3$, and $m = 1, \dots, 20,000$ were calculated, combined and ordered.

Table 1 displays the true parameter values that were used in the simulation of the data, as well as the summary statistics obtained from the posterior draws when the model was applied to the three data sets. $\hat{R}$ values for assessing convergence are also shown. The results show that the summary statistics for the model applied to data set 1 are the closest to the true values. The 95% credible intervals include the true values for β , ϕ , and σ_ω^2 .

Table 2 gives the 10th and 90th percentiles of the aggregated (over time and subset) ordered pivotal discrepancy measures for the model applied to each of the three data sets, in each of the three cases of subsetting. In order to confirm that the model provided a good fit, the quantities need to be within the interval given by the critical values that are calculated from the nominal χ^2 distributions assumed when they were calculated.

In the first case, when the locations were not partitioned for calculation of the pivotal quantities, it was found that the 10th and 90th percentiles of ordered pivotal discrepancy quantities were within the corresponding nominal percentiles of (12.76, 56.33). This suggests that the model provided a good fit to each of the three simulated data sets. In the second case, when the locations were partitioned into 3 even subsets to calculate the pivotal quantities, it was found that the 10th and 90th percentiles of the aggregated ordered pivotal discrepancy quantities were within the corresponding nominal percentiles of (1.827 and 27.11) when the model was applied to data set 1. The 10th and 90th percentiles of the aggregated ordered pivotal discrepancy quantities were, however, outside the nominal percentiles when the model was applied to data sets 2 and 3. This suggests that the model provides a good fit only to data set 1. A similar result was observed for the final case, with the locations being partitioned into three unequal subsets. It was found that the 10th and 90th percentiles of the aggregated ordered pivotal discrepancy quantities were within the corresponding nominal percentiles of (0.4894 and 31.71) when the model was applied to data set 1. The 10th and 90th percentiles of the aggregated ordered pivotal discrepancy quantities were, however, outside the nominal percentiles when the model was applied to data sets 2 and 3. This suggests that the model provides a good fit only to data set 1.

We would have expected that the model only provide a good fit to data set 1, because the model used to generate that data set is the same as the one being fitted. This is correctly executed in the two cases of partitioning. In the first case, the model provided a good fit to each data set, because a lack of partitioning caused a decrease in power to detect the differences. This is highlighted in Figs. 2, 3 and 4. In those figures, the pivotal discrepancy quantities from each model applied to each data set in each case of partitioning are plotted as a density, and are overlaid with the nominal densities. We see for each data set that when no partitioning occurs, there is sufficient overlap of the pivotal quantities observed and the nominal densities to suggest the model provides a good fit. This is also the case for the partitioning scenarios for data set 1, but not the case for data sets 2 and 3.

For comparison we calculated the WAIC values when fitting the model with an exponential correlation function to all three datasets. The results in Table 2 show the smallest WAIC value for Dataset 3. This result is counterintuitive, however. Compared

Table 1 True values and posterior summary statistics of parameters for simulated data

	Dataset 1				Dataset 2				Dataset 3			
	True	Median	(95% CI)	$\hat{R}$	True	Median	(95% CI)	$\hat{R}$	True	Median	(95% CI)	$\hat{R}$
β	0	-0.02	(-0.20, 0.18)	1	0	-0.006	(-0.21, 0.20)	1	0	-0.01	(-0.20, 0.17)	1.02
ϕ	0.2	0.20	(0.11, 0.36)	1	0.8	1.86	(1.52, 2.00)	1	0.2	0.39	(0.15, 0.89)	1
ρ	0.7	0.84	(0.75, 0.91)	1.01	0.7	0.65	(0.50, 0.78)	1	0.7	0.60	(0.40, 0.79)	1
σ_ε^2	0.0001	0.14	(0.07, 0.22)	1.1	0.0001	0.03	(0.02, 0.04)	1.01	0.0001	0.07	(0.04, 0.10)	1
σ_ω^2	1	0.65	(0.41, 1.02)	1.03	1	0.49	(0.34, 0.68)	1.02	1	0.25	(0.12, 0.47)	1.01

Table 2 The 10th and 90th percentiles of ordered pivotal discrepancy measures for the model applied to each of the three datasets. These are compared to the nominal 10th and 90th percentiles: 12.76 and 56.33 for the non-subset data; 1.827 and 27.11 for the even subset data; 0.4894 and 31.71 for the uneven subset data

	Nominal percentiles	Dataset 1	Dataset 2	Dataset 3
No subset	(12.76, 56.33)	(18.77, 42.21)	(15.33, 27.77)	(15.65, 35.04)
Even subset	(1.827, 27.11)	(3.945, 22.71)	(5.222, 40.04)	(0.4609, 12.54)
Uneven subset	(0.4894, 31.71)	(0.9661, 31.50)	(2.038, 60.47)	(0.1575, 19.65)
WAIC		415	456	158

Fig. 2 PDM density dataset 1

to the other two datasets, Dataset 3 has the worst model mis-specification and we would therefore expect it to have the highest WAIC value.

5 Application to hoki catch data

5.1 Hoki catch data from New Zealand sub-Antarctic survey

Research trawl surveys of the New Zealand sub-Antarctic region were carried out by the National Institute of Water and Atmospheric Research (NIWA) for the Ministry

Fig. 3 PDM density dataset 2

of Primary Industries, New Zealand (MPI). The survey design used was a stratified 2-phase adaptive design optimised to reduce variance in biomass estimates (Francis 1984). The primary focus of the survey design was abundance estimation of a particular fish species, *Macruronus novaezelandiae*, commonly known as hoki. The hoki survey has been run on an annual or biennial basis since 1991 (Bagley et al. 2013; Fisheries New Zealand 2019). For our application, we focused on the trawls that occurred between 2000 to 2008 in order to have a continuous annual time series. Catch weights of all species caught during the survey were recorded. As the survey design was based on hoki, we focus our application on modelling spatio-temporal hoki catch weights. Figure 5 illustrates the stratification used and shows that the largest catch weights for hoki were recorded for trawls near Puysegur Bank.

To obtain repeated measurements at the same locations annually, catch weight locations within a stratum were gridded. The strata were gridded in such a way that within each grid there was at least one catch weight observation per year. The median longitude and latitude of all observations within a grid was taken as the grid centroid, g . The 38 grid centroids are shown in Fig. 5. It should be noted that not all strata were used in the grid construction. For strata 25–28, there were years during which trawls did not occur, and those strata were excluded from the final dataset. To obtain a single observation per grid per year, a weighted mean of all observations within a grid in a given year was calculated. Observations were weighted by distance from the grid center, with observations closer to the grid center given more weight. This method of

Fig. 4 PDM density dataset 3

weighting allowed observations located closer to the grid center to contribute more to the grid mean than those further away. In addition, the weighted mean depth of each trawl within a grid was assigned as the depth for the entire grid in the same fashion. Depth was used as a covariate for a selection of the models (shown below).

Three models were fitted to the gridded hoki catch weight data of the form given in Eqs. 2–4. We let $\mathbf{y}_t = (y(\mathbf{g}_1, t), \dots, y(\mathbf{g}_{38}, t))'$ where $y(\mathbf{g}_i, t)$ denoted the log-transformed weighted mean catch weight of hoki in grid $\mathbf{g}_i$ for year t , and $n = 38$. The marginal distribution of $\mathbf{y}_t$ given the parameters is,

$$\mathbf{y}_t | \boldsymbol{\theta} \sim \text{MVN}\left(\boldsymbol{\mu}_t, \frac{\sigma_w^2}{1 - \rho^2} \mathbf{R} + \sigma_\varepsilon^2 \mathbf{I}_n\right),$$

where $\boldsymbol{\mu}_t = \mathbf{1}_{38}\beta$ for each model. The three models were distinguished by the correlation structure assumed for $\mathbf{R}$. For model M1 we used an exponential correlation function, given by Eq. 12. For model M2 we used a Gaussian correlation function, given by Eq. 13. For model M3, we used the general Matérn correlation function given in Eq. 11. For each model, the parameters were assumed a priori independent, and were assigned the non-informative prior distributions,

$$\beta \sim N(0, 100), \sigma_\varepsilon^2 \sim \text{IG}(2, 1), \sigma_w^2 \sim \text{IG}(2, 1), \phi \sim U(0.01, 1000), \rho \sim U(-1, 1).$$

Fig. 5 (Left) Map of 1255 summer trawls that show the locations where hoki were caught in the sub-Antarctic region and the observed catch weight in kilograms. The boundaries of the strata are shown along with their identification number. The size of a point is indicative of the catch weight in kilograms, and the colour gradient is indicative of the year the trawl occurred as shown on the legend. (Right) Map of 38 grid centres showing the median longitude and latitude of observations within each grid for the years 2000 to 2008

Further, for model M3, the smoothness parameter ν was assumed a priori independent of the other parameters and was assigned the non-informative prior distribution, $\nu \sim U(0.01, 10)$.

MCMC sampling of the joint posterior density was used to fit the three models to the gridded hoki data and this was done through **R** using the package **NIMBLE** (NIMBLE Development Team 2017). Two chains, each with 1,000,000 iterations, were generated for the parameter vector $\theta = (\beta, \rho, \phi, \sigma_\omega^2, \sigma_\varepsilon^2)^T$ for models M1 and M2, and $\theta = (\beta, \rho, \phi, \sigma_\omega^2, \sigma_\varepsilon^2, \nu)^T$ for model M3. The first 900,000 iterations from each chain were discarded as warm-up, and the remaining draws were combined, resulting in a posterior sample of size $M = 200,000$. Convergence of the Markov chain was assessed using traceplots (not provided) and potential scale reduction factor ($\hat{R}$) values. Table 3 gives summary statistics for the posterior distributions and $\hat{R}$ values computed for each model. There appears to be agreement between the three models on the values of most of the parameters. For model M3, the smoothness parameter ν needed for the Matérn correlation structure has a posterior median of 0.5156, which means it is close to being estimated as an exponential correlation structure.

For each fitted model, pivotal quantities for every posterior sample were calculated using Eq. 7. The grid locations were partitioned into five subsets using the K-means clustering algorithm. The optimal number of subsets to use in the K-means algorithm was found using the elbow method (Kodinariya and Makwana 2013). Pivotal quantities for each fitted model $S(y_{tj}, \tilde{\theta}^{(m)})$ for $t = 1, \dots, 5$, $j = 1, \dots, 5$, and $m = 1, \dots, 200,000$ were calculated, combined and ordered.

The 10th and 90th percentiles of the nominal chi-square distribution were calculated according to Eqs. 9 and 10. They were found to be 0.133 and 28.6 respectively. The 10th and 90th percentiles of the ordered pivotal discrepancy measures for models M1, M2 and M3 are given in Table 4. It was found that for each model, the 10th

Table 3 Posterior summary statistics and convergence diagnostics for hoki data

	Model 1			Model 2			Model 3		
	Median	(95% CI)	$\hat{R}$	Median	(95% CI)	$\hat{R}$	Median	(95% CI)	$\hat{R}$
β	0.017	(− 0.18, 0.21)	1.01	0.018	(− 0.18, 0.21)	1.01	0.017	(− 0.19, 0.21)	1
ν	849.1	(584.3, 1000)	1.04	261.8	(209.4, 312.9)	1.01	0.52	(0.29, 0.82)	1.04
ϕ	0.97	(0.94, 0.99)	1.02	0.99	(0.98, 1.00)	1.04	828.9	(494.3, 1000)	1
ρ	0.79	(0.66, 0.94)	1	0.88	(0.74, 1.03)	1	0.97	(0.94, 0.99)	1.03
σ^2_ε	0.31	(0.15, 0.55)	1.01	0.24	(0.13, 0.39)	1.02	0.79	(0.66, 0.94)	1.01
σ^2_ω							0.31	(0.14, 0.54)	1.04

Fig. 6 Posterior densities for each parameter for each of the three models (top), density of observed ordered PDMs and their nominal distributions for each of the three models (bottom)

Table 4 The 10th and 90th percentiles of ordered pivotal discrepancy measures for each model applied to the hoki catch weight data

Nominal percentiles	Model 1	Model 2	Model 3
(0.133, 28.6)	(5.30, 15.8)	(2.90, 13.3)	(5.10, 16.0)
WAIC	951	967	949

These are compared to the nominal 10th and 90th percentiles, 0.133 and 28.6 respectively

percentile was higher than that of the nominal distribution and the 90th percentile was lower than that of the nominal distribution. As a result, it can be said that each model provided a good fit to the gridded hoki data. This is reflected in Fig. 6 which shows the posterior densities for each parameter for each of the three models, as well as the density of the ordered pivotal discrepancy measures. The latter are overlaid with the nominal distribution density. The posterior densities for each parameter are similar, with model M2 having the most different posterior parameter densities. It can be seen that the posterior densities for β , ρ , σ_ϵ^2 , and σ_ω^2 are similar in shape, and overlap between models. For the parameter ϕ , there is a difference between models M1 and M3, and M2. We conclude that Models M1 and M3 are very similar models, and that all three models provide a similar fit to the data, with the only differences being due to the correlation structure parameters. Looking at the densities of the observed pivotal discrepancy measures for each model, the conclusion that each model provides

a good fit is motivated. There is sufficient overlap of the pivotal discrepancy measure densities and their nominal densities such that the test cannot detect a difference. As was done for the simulated data, we provide values of WAIC for the three models for comparison. The overall conclusions based on WAIC are similar to those reached using the pivotal discrepancy measures in that models M1 and M3 provide a similar fit, while model M2 differs from the other two models.

6 Discussion

From the simulation study, it is clear that partitioning is necessary, and furthermore, that the number of observations within a partition need not be constant. We found that the goodness-of-fit test based on pivotal discrepancy measures was unable to identify an incorrect model mis-specification when there was no partitioning. Further, when the number of observations within a partition was not constant, the distribution of the ordered pivotal discrepancy measures were wider, making it more difficult to reject the null hypothesis that the model provided a good fit. This indicates an increase in power.

The choice of how to partition the data should be considered carefully. In the simulation study in this paper, the subsets were chosen sensibly, in that we partitioned the observations according to the subsets that were used to generate the data. In practice, this will not necessarily be known, and an objective method should be developed. In the case study, we showed that using the K-means clustering algorithm is a suitable objective approach to partitioning. Jun et al. (2014) suggested that partitions of the dataset need neither be disjoint nor represent a complete partition. Thus other suitable clustering methods could be used. Density-based or fuzzy clustering methods are two of the alternatives to K-means clustering. Density-based algorithms create clusters based on the density of data points, with regions that have larger numbers of points considered as clusters. This would partition the spatial domain into regions of higher and lower spatial autocorrelation. Fuzzy clustering creates clusters that overlap, thus allowing points to belong to more than one cluster. Such an approach may be suitable where there is an even spread of points across the spatial domain. We made the assumption that the covariance matrix is separable over time and space, that is the overall spatial covariance structure is constant over time. Future research could investigate use of different clustering algorithms and use of our approach for a covariance matrix that is non-separable in space and time.

A final consideration is that of how to select the best model from competing models fit to the same data. The goodness-of-fit test based on pivotal discrepancy measures currently offers no way to select the best model, instead opting for a decision based test only. Jun et al. (2014) and Johnson (2007) talk briefly on calculating bounds on Bayesian p-values that may offer an appropriate route to model selection.

In conclusion, we have developed a general goodness-of-fit test for Bayesian spatio-temporal models using partitioning and pivotal discrepancy measures. The test was successful in simulation as well as application to New Zealand hoki data.

Acknowledgements We thank Dr Matt Dunn and the National Institute for Water and Atmospheric Research (NIWA) for providing the Ministry of Primary Industries hoki trawl survey data.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Proceedings of the second international symposium on information theory. Akadémiai Kiadó, pp 267–281
- Alsabti K, Ranka S, Singh V (1997) An efficient k-means clustering algorithm. *Electr Eng Comput Sci* 43:1–10
- Bagley NW, Ballara SL, O'Driscoll RL, Fu D, Lyon WS (2013) A review of hoki and middle-depth summer trawl surveys of the sub-Antarctic, November–December 1991–1993 and 2000–2009. Ministry for Primary Industries, Wellington
- Banerjee S, Carlin BP, Gelfand AE (2014) Hierarchical modeling and analysis for spatial data. CRC Press, New York
- Bastos LS, O'Hagan A (2009) Diagnostics for Gaussian process emulators. *Technometrics* 51(4):425–438
- Besag J (1974) Spatial interaction and the statistical analysis of lattice systems. *J R Stat Soc B* 1:192–236
- Cameletti M, Ignaccolo R, Bande S (2011) Comparing spatio-temporal models for particulate matter in Piemonte. *Environmetrics* 22(8):985–996
- Cameletti M, Lindgren F, Simpson D, Rue H (2013) Spatio-temporal modeling of particulate matter concentration through the SPDE approach. *ASta Adv Stat Anal* 97(2):109–131
- Caraux G, Gascuel O (1992) Bounds on distribution functions of order statistics for dependent variates. *Stat Probab Lett* 14(2):103–105
- Cressie N, Huang H-C (1999) Classes of nonseparable, spatio-temporal stationary covariance functions. *J Am Stat Assoc* 94(448):1330–1339
- Ecker MD, De Oliveira V, Isakson H (2013) A note on a non-stationary point source spatial model. *Environ Ecol Stat* 20(1):59–67
- Fisheries New Zealand (2019) Fisheries Assessment Plenary May 2019: stock assessments and stock status. Compiled by the Fisheries Science and Information Group, Fisheries New Zealand, Wellington
- Fouedjio F (2017) Second-order non-stationary modeling approaches for univariate geostatistical data. *Stoch Environ Res Risk Assess* 31(8):1887–1906
- Francis RICC (1984) An adaptive strategy for stratified random trawl surveys. *N Zeal J Mar Freshw Res* 18(1):59–71
- Gelfand AE, Banerjee S (2017) Bayesian modeling and analysis of geostatistical data. *Annu Rev Stat Appl* 4:245–266
- Gelman A, Hwang J, Vehtari A (2014) Understanding predictive information criteria for Bayesian models. *Stat Comput* 24(6):997–1016
- Gneiting T (2002) Nonseparable, stationary covariance functions for space-time data. *J Am Stat Assoc* 97(458):590–600
- Huang H-C, Martinez F, Mateu J, Montes F (2007) Model comparison and selection for stationary space-time models. *Comput Stat Data Anal* 51(9):4577–4596
- Johnson VE (2007) Bayesian model assessment using pivotal quantities. *Bayesian Anal* 2(4):719–733
- Jun M, Katzfuss M, Hu J, Johnson VE (2014) Assessing fit in Bayesian models for spatial processes. *Environmetrics* 25(8):584–595

- Kodinariya TM, Makwana PR (2013) Review on determining number of clusters in K-means clustering. *Int J Adv Res Comput Sci Manage Stud* 1(6):90–95
- Lobo VGR, Fonseca TCO (2020) Bayesian residual analysis for spatially correlated data. *Stat Model* 20(2):171–194. <https://doi.org/10.1177/1471082X18811529>
- NIMBLE Development Team (2017) NIMBLE: an R package for programming with BUGS models, version 0.6-6. <https://www.R-nimble.org/>
- Paciorek CJ (2013) Spatial models for point and areal data using Markov random fields on a fine grid. *Electron J Stat* 7:946–972
- Pollice A (2011) Recent statistical issues in multivariate receptor models. *Environmetrics* 22(1):35–41
- Rychlik T (1992) Stochastically extremal distributions of order statistics for dependent samples. *Stat Probab Lett* 13(5):337–341
- Sahu SK, Bakar KS (2012) Hierarchical Bayesian autoregressive models for large space-time data with applications to ozone concentration modelling. *Appl Stoch Models Bus Ind* 28(5):395–415
- Schwarz G (1978) Estimating the dimension of a model. *Ann Stat* 6(2):461–464
- Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A (2002) Bayesian measures of model complexity and fit. *J R Stat Soc B* 64(4):583–639
- Stein ML (2005) Statistical methods for regular monitoring data. *J R Stat Soc B* 67(5):667–687
- Vehtari A, Gelman A (2014) WAIC and cross-validation in Stan
- Vehtari A, Gelman A, Gabry J (2017) Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput* 27(5):1413–1432
- Watanabe S (2010) Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J Mach Learn Res* 11:3571–3594
- White G, Ghosh SK (2009) A stochastic neighborhood conditional autoregressive model for spatial data. *Comput Stat Data Anal* 53(8):3033–3046
- Yuan Y, Johnson VE (2012) Goodness-of-fit diagnostics for Bayesian hierarchical models. *Biometrics* 68(1):156–164

Lindsay Morris is a Ph.D. student at the School of Mathematics and Statistics, Victoria University of Wellington, New Zealand.

Nokuthaba Sibanda is Senior Lecturer in Statistics at the School of Mathematics and Statistics, Victoria University of Wellington, New Zealand.