

RESEARCH ARTICLE

# A Four Dimensional Spatio-Temporal Analysis of an Agricultural Dataset

Margaret R. Donald<sup>1\*</sup>, Kerrie L. Mengersen<sup>2</sup>, Rick R. Young<sup>3</sup>

**1** Mathematics and Statistics, University of New South Wales, Sydney, NSW, Australia, **2** Statistics Department, Queensland University of Technology, Brisbane, QLD, Australia, **3** Tamworth Agricultural Institute, NSW Department of Primary Industries, Calala, NSW, Australia

\* [m.r.donald@bigpond.com](mailto:m.r.donald@bigpond.com)



OPEN ACCESS

**Citation:** Donald MR, Mengersen KL, Young RR (2015) A Four Dimensional Spatio-Temporal Analysis of an Agricultural Dataset. PLoS ONE 10(10): e0141120. doi:10.1371/journal.pone.0141120

**Editor:** Manuel Reigosa, University of Vigo, SPAIN

**Received:** March 9, 2015

**Accepted:** October 5, 2015

**Published:** October 29, 2015

**Copyright:** © 2015 Donald et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The Supporting Information file contains the data analysed in this paper as analysis.csv. (Yval is the log(neutron count)).

**Funding:** This work was supported by Australian Research Council Discovery Project DP140103564. New South Wales Department of Agriculture and The Commonwealth Scientific and Industrial Research Organisation (CSIRO) gratefully acknowledge the funding support for the field experiment from the Grains Research and Development Corporation (GRDC projects DAN 159N and DAN 407), the New South Wales State Government via Salt Action (projects NW1094.94, NW1093.94, NW1120.96 and NA0778.99), the Land and Water Resources

## Abstract

While a variety of statistical models now exist for the spatio-temporal analysis of two-dimensional (surface) data collected over time, there are few published examples of analogous models for the spatial analysis of data taken over four dimensions: latitude, longitude, height or depth, and time. When taking account of the autocorrelation of data within and between dimensions, the notion of closeness often differs for each of the dimensions. Here, we consider a number of approaches to the analysis of such a dataset, which arises from an agricultural experiment exploring the impact of different cropping systems on soil moisture. The proposed models vary in their representation of the spatial correlation in the data, the assumed temporal pattern and choice of conditional autoregressive (CAR) and other priors. In terms of the substantive question, we find that response cropping is generally more effective than long fallow cropping in reducing soil moisture at the depths considered (100 cm to 220 cm). Thus, if we wish to reduce the possibility of deep drainage and increased groundwater salinity, the recommended cropping system is response cropping.

## Introduction

Where data are collected from a set of sites, at a series of time points, observations taken close to each other in either time or space may be autocorrelated. Highly autocorrelated observations reduce the number of effective observations, and statistical analyses and inferences which fail to take this autocorrelation into account are more prone to identification of erroneous significant relationships. In many applications, the spatial autocorrelations are the focus of interest, but in other applications, the aim is to account for them in order to obtain accurate and precise parameter estimates.

Spatio-temporal data are often analysed using models where spatial and temporal autocorrelation effects are separable, that is, with the assumption of no interaction between time and space. Cressie and Wikle [1] comment that separable covariance models “have very particular properties that are rarely seen in empirical studies of spatio-temporal dependence”.

A range of alternatives to address this drawback have been suggested in the more recent literature. Dynamical spatio-temporal models described in [1–3] utilise differential equations

Research and Development Corporation (LWRRDC project CDS16), and the Murray-Darling Basin Commission/LWRRDC (Natural Resource Management Strategy project AGS1) that enabled this research. The design and management of the field experiment and the acquisition, analysis, decision to publish, or preparation of the manuscript were not influenced in any way by the funds providers or the organisation management.

**Competing Interests:** The authors have declared that no competing interests exist.

describing physical processes together with hierarchical models [4] involving a data model, a process model and a parameter model. The dynamical aspect of these models implies the possibility of a physical description. However, as Wikle and Hooten [5] note, substantial simplifications in the dynamics must often be made, with the redistribution kernels necessary for analytical solution not necessarily being representative of the data, and the assumption of homogeneity of the kernels over space and time being possibly unrealistic.

The aim of this paper is to evaluate a number of space-time modelling methods, which are described and compared in the context of a substantive case study of an agricultural field experiment designed to assess the impact of different cropping systems on soil moisture. The data arise from a balanced lattice design with different experimental treatments by plot. Earlier exploratory analyses of these data are given in [6, 7]. The dataset has a range of challenging features, including spatial autocorrelations that are not constant over time, nor over any of the three spatial dimensions. The model must therefore take account of both the spatial and temporal autocorrelations in the four dimensions of the data, the possible non-additive nature of the autocorrelation in time and space, and autocorrelations within the three-dimensional space ignoring time. The objectives of the data analysis were threefold: firstly, to estimate the contrast across time between the treatments of long fallow rotation and response cropping together with 95% credible intervals; secondly, to understand the time-varying nature of the contrasts; and thirdly, to find appropriate credible intervals for the contrasts when considered as time-series.

The primary spatial approach adopted in this paper is a conditional autoregressive (CAR) model [8, 9] with the proper priors of [10]. This model defines a local neighbourhood for the spatial domain, which improves computational efficiency since the precision matrix is sparse and repeated matrix inversions are not required. This is preferred to a kriging approach [11, 12] since kriging models are slow to converge in a Bayesian setting when datasets are large [13]. Moreover, [14] and [15] show an equivalence between the two types of model. This is supported by [16] and [17], who calibrate CAR models to kriging models. The primary temporal approaches include random walk and autoregressive models [9, 18–21] and penalised spline smoothers [22, 23].

Because of the complexity of the model developed to address the features of the data described above, the first modelling method fits a separate spatial model to each time period (date). The block updating Gibbs sampler of [7] is employed to provide estimates of the contrast of interest for each depth and day and their associated credible intervals. The second method takes these contrasts estimates and uses a time series model to describe their behaviour over time to gain insight into the time-varying process. In contrast to this two-stage approach, the third method fits a single structured additive model [8, 22] to the whole dataset. A feature of these approaches is that estimates can be smoothed across unequally spaced covariates and unequal time intervals.

## Data

The four dimensional data were soil water measurements from an agricultural field experiment. The field layout was a balanced lattice design comprising 4 replicate blocks of 9 treatments. Each plot was 15 m × 40 m, and measurements were taken from three neutron moisture meter access tubes (50 mm × 3 m deep installed 10 m apart along the centre of each plot). Thus, the four dimensional data consisted of soil water observations from 108 sites (defined across the surface), and taken at 15 soil depths from 20 cm to 300 cm, for 56 different dates spaced roughly monthly over a five-year period. The 108 measurement sites were arranged in 6 rows of 18 columns. Hence, data at each time point consisted of 1620 measurements at 108 sites (over 2 dimensions) at 15 depths, while the entire dataset consisted of 90,720

observations. The experiment was designed to determine a cropping system which would minimise water leakage below the crop root zone, and consisted of three cropping systems, running in different phases, and three perennial pastures, giving rise to 9 treatments. The first soil water measurements were made on June 26, 1995, and the last on April 27, 2000.

The treatments were

- Treatments 1–3: Phases of a long fallow wheat/sorghum rotation, where one wheat and one sorghum crop were grown in three years with an intervening 10–14 month fallow period. The 3 treatments were each of 3 phases of the long fallow 3 strip system.
- Treatment 4: Continuous cropping in winter with wheat and barley grown alternately.
- Treatments 5 and 6: Response cropping, where an appropriate crop (either a winter or a summer crop) was planted when the depth of moist soil exceeded a predetermined level.
- Treatments 7–9: Perennial pastures. The three treatments were lucerne (a deep rooted perennial forage legume with high water use potential), lucerne grown with a winter growing perennial grass, and a mixture of winter and summer growing perennial grasses.

The primary question for crop scientists was whether response cropping produced lower moisture values both at the intermediate and greater depths, in comparison with long fallow cropping, and whether this was sustained over different stages of the cropping cycle. This contrast was calculated as the average of treatments 1, 2 and 3 minus the average of treatments 5 and 6. The units of measurement for the contrast are log(neutron count ratio), a surrogate measure for moisture, with the higher neutron count ratios indicating wetter soils. Further details of the experimental design and aims are given in [24].

In modelling the key contrast evolving over time, the transformed variables log(rainfall+1), linear, quadratic and cubic time effects, together with interactions of year with sine and cosine terms with periods of a year and a half-year were considered as possible useful covariates.

## Ethics Statement

The cropping experiment was run on the private property ‘Hudson’, near Willow Tree, New South Wales, Australia, owned by Mr Robert and Mrs Edwina Duddy, with their full consent. No animals, endangered, protected or otherwise, were involved.

## Methods

### Subsection 1: Method 1

Let  $y_{tid}$  be the response variable measured on date  $t$ , at site  $i$  (of  $S$  horizontal plot sites), at depths from 20 cm–300 cm, indexed by  $d$  ( $d = 1, \dots, 15$ ). Let  $j$  be the treatment given at site  $i$ .

Method 1 fits a common spline model at each date,  $t$ , as follows:

$$\begin{aligned}
 y_{tid} &= f_{ij}(d) + \psi_{tid} + \epsilon_{tid}, \\
 \epsilon_{tid} &\sim N(0, \sigma_{id}^2), \\
 \psi_{tid} | \psi_{i'd}, i \neq i' &\sim N\left(\rho_t \sum_{i' \in \partial i} \frac{\psi_{i'd}}{n_i}, \frac{\tau_{id}^2}{n_i}\right), \\
 f_{ij}(d) &= \mathbf{X}\beta_{ij},
 \end{aligned} \tag{1}$$

where  $n_i$  is the number of sites adjacent to site  $i$ , and  $i' \in \partial i$  denotes that site  $i'$  is a neighbour of site  $i$ . Neighbours are defined as the first order spatially closest observations at the same

depth. This is the proper CAR model of [10], with  $\rho_t$  constrained to be common across all depths for a given date,  $t$ .  $\mathbf{X}$  is the (common) design matrix for a set of spline basis functions across the 15 depths, while  $\beta_{tj}$  is the vector of coefficients for each day,  $t$ , and treatment,  $j$ . See [7] for details. The estimates from these fitted equations allowed the calculation of the contrasts of interest. From the treatment effects, the contrast of long following versus response cropping was calculated at each depth as  $(e_1 + e_2 + e_3)/3 - (e_5 + e_6)/2$ , where  $e_j$  indicates the estimate for treatment  $j$  for a common depth,  $d$ . (Subscripts for depth have been omitted.) Thus, from Method 1, using [7], 56 treatment means at each depth ( $d$ ) and date ( $t$ ), i.e.,  $56 \times 9 \times 15 = 7560$ , and  $56 \times 15$  (840) contrast estimates are estimated, together with their credible intervals.

Use of this method satisfies the first objective of the Introduction of estimating the key contrast together with credible intervals over time. However, it provides no insight into their time-varying nature.

### Subsection 2: Method 2

Here, time-series models were fitted to the contrast estimates of Method 1, This gives 15 time-series (one for each depth). We considered the time-series for the depths from 100 cm to 220 cm (focussing on the deeper depths, and 7 series in all). The set of 7 time-series can be modelled in two ways. The first alternative is as a multivariate time-series. However, in this case study the correlation structure of the contrasts was quite variable over the four dimensions. In particular, although there is some evidence of continuity in the contrasts across depths, it was unclear how to adequately model the depth-varying component of the corresponding multivariate random walk, or of a more general dynamic linear model. The second alternative is as a set of univariate time series, one for each depth. Since the depth is common for each time period, the subscript  $d$  is omitted from the equations below.

Three classes of time-series models were considered: regression models with time-varying covariates (Eq 2), random walk and autoregressive time-series models ignoring regression covariates (Eqs 3–5), and a combination of an autoregressive model with regression components using time-varying covariates and penalised spline models over time (Eq 6). The regression models accounted for covariates but assumed that the contrasts were independent over time. The autoregressive models ignored covariates but allowed for temporal dependence, although it was assumed that observations were equally spaced over time. The combined model provided a way of including covariates and accommodating unequal time periods. Furthermore, since these data may exhibit seasonality and perhaps trends, it was anticipated that the spline models might suggest further explanatory variables for the regression.

These classes of model are now described in more detail. Let  $e_t$  represent the contrast estimate at time,  $t$ .

The regression models are given by:

$$e_t \sim N(\mathbf{X} \beta, V), \quad 1/V \sim \text{Gamma}(10^{-6}, 10^{-6}), \tag{2}$$

where  $\mathbf{X}$  is a design matrix of time-varying covariates, namely  $\log(\text{rainfall}+1)$  and interactions of year (as a factor) by sine and cosine terms with periods of a year and a half year.

The local level state-space model (random walk) of e.g., [18, 20, 21] was fitted.

$$\begin{aligned} e_t &= \mu_t + v_t, & v_t &\sim N(0, V), \\ \mu_t &= \mu_{t-1} + \omega_t, & \omega_t &\sim N(0, W), \\ 1/V &\sim \text{Gamma}(10^{-6}, 10^{-6}), & 1/W &\sim \text{Gamma}(10^{-6}, 10^{-6}). \end{aligned} \tag{3}$$

In a further version of this model, t-distributions with 10 and 4 degrees of freedom were

substituted for the normal distributions for the observation and state errors, to allow for fatter tailed errors.

An alternative formulation of the random walk model which uses CAR neighbourhood models was also used. This formulation permits the neighbours to be weighted, and thus allows a correction for the unequal time intervals. Hence, weighted random walk models of order 1 (RW1) and order 2 (RW2) were fitted using [25].

$$\begin{aligned}
 e_t &= \mu_t + \omega_t + \psi_t, \omega_t \sim N(0, V), \\
 \psi_t | \psi_{t'}, t \neq t' &\sim N\left(\sum_{t' \in \partial t} \frac{w_{t'} \psi_{t'}}{w_+}, \frac{W}{w_+}\right), \text{ where} \\
 w_{t'} &= 1/|t - t'|, \text{ and } w_+ = \sum_{t' \in \partial t} w_{t'},
 \end{aligned} \tag{4}$$

with  $V, W$  defined as in Eq 3. The weight used is the reciprocal of the time difference between neighbours on the time scale (that is, the date difference).

The autoregressive models were fitted as follows [26, 27]:

$$\begin{aligned}
 e_t &\sim N(\mu_t, V), & 1/V &\sim \text{Gamma}(10^{-6}, 10^{-6}), \\
 \mu_t &= \alpha_0 + \alpha_1 Y_{t-1}, & \text{or,} \\
 \mu_t &= \alpha_0 + \alpha_1 Y_{t-1} + \mathbf{X}\beta, & \text{or,} \\
 \mu_t &= \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2}, & \text{or,} \\
 \mu_t &= \alpha_0 + \alpha_1 Y_{t-1} + \alpha_{12} Y_{t-12}.
 \end{aligned} \tag{5}$$

Note that the simple regression models using time-varying covariates (Eq 2), and the autoregressive models with time-varying covariates (Eq 5) were fitted to deal with the issue of a non-equally spaced time-series. The assumption of equally spaced contrasts across time in the models of Eqs 5 and 3 motivated the weighted random walk models of Eq 4. However, an alternative was to fit missing data models. This was done for the random walk of order 1 model (Eq 3) only. For these data, the highest common factor of the time intervals was 1, which gives a time-series of largely missing data (56 observed of 1768 observations or 3.2% non-missing observations). The missing data model was fitted for one depth only (140 cm).

The combined method involved fitting generalised additive models using penalised spline smooths over time [22, 28] using the software, BayesX (<http://www.stat.uni-muenchen.de/bayesx/bayesx.html>). Let the contrast,  $e_t$ , at date,  $t$ , and depth,  $d$ , be defined as

$$e_t = f(t) + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2), \tag{6}$$

for each depth,  $d$ , with  $f(t)$  being fitted as a penalised spline over time with a random walk penalty of order 2. These models, like the regression model of Eq 2, do not account for autocorrelation over the time dimension, but they use the unequally spaced dates of the contrasts.

The time-series models of Method 2 capture the time-varying variance of the contrast, and while the contrast estimates themselves result from a model which accounts for spatial autocorrelation, the final models fail to reflect the spatial error. Hence, we experimented with precisions for the random walk model, which might reflect the full error which includes the within date variability of the contrast, in an attempt to satisfy objective 3 of the Introduction.

### Subsection 3: Method 3

Finally, using the full dataset, we fitted two additive structured models using the full 90,720 observations. The first model was defined as

$$y_{tid} = \sum_j f_j(d) + \sum_j f_j(t) + \psi_{id} + \epsilon_{tid}, \quad \epsilon_{tid} \sim N(0, \sigma^2),$$

$$\psi_{id} | \psi_{i'd}, i \neq i', \quad \psi_{id} \sim N\left(\sum_{i' \in \partial i} \frac{\psi_{i'd}}{n_i}, \frac{\tau_d^2}{n_i}\right), \tag{7}$$

The second model is analogous to that of Method 1 and was defined by

$$y_{tid} = \sum_t \sum_j f_{j(t)}(d) + \psi_{tid} + \epsilon_{tid}, \quad \epsilon_{tid} \sim N(0, \sigma_t^2),$$

$$\psi_{tid} | \psi_{i'd}, i \neq i', \quad \psi_{tid} \sim N\left(\sum_{i' \in \partial i} \frac{\psi_{i'd}}{n_i}, \frac{\tau_{id}^2}{n_i}\right), \tag{8}$$

Thus, for each timepoint, this model fitted a penalised curve for each treatment at each depth. It was, again, a layered model, which modeled site correlations using CAR models within each layer, and hence, allowed different variances at each depth and day, together with a final unstructured residual whose variance differed by day. Thus, the model from Eq 8 was again a considerable simplification of the model of Method 1. The CAR residual structure was the same, but was coupled with an unstructured variance common to each day, whereas the unstructured variance components of Method 1 (Eq 1) differed by date and by depth. As in Model 1, it fitted a series of penalised smooths across the depth dimension (by date and by treatment).

### Subsection 4: Priors

Priors for the Method 1 precisions for both structured and unstructured residual component precisions were Gamma(5, .005), with priors for the fixed coefficients being normal with mean zero and variance 10.

For the Method 2 models, which fitted the contrast across the time dimension, the priors for the coefficients were generally specified as a diffuse normal prior  $N(0, 10^6)$ . Priors for the precision terms of these models were initially set as Gamma( $10^{-6}, 10^{-6}$ ). However, almost all the models of Method 2 were rerun with priors for the precisions of Gamma( $10^{-4}, 10^{-4}$ ), and final model choice was made using models with this prior.

However, given that we wanted a meaningful temporal description of the contrast together with appropriate credible intervals, we experimented with various ways of apportioning estimates of the total error observed in the model from Method 1. Tables 1 and 2 give the settings

**Table 1. Various priors used for the precisions of the timeseries models of Method 2.**

	<i>Precision for observational error</i>	<i>Precision for random walk error*</i>
Prior 1	~ Gamma(.000001, .000001)	~ Gamma(.000001, .000001)
Prior 2	~ Gamma(.0001, .0001)	~ Gamma(.0001, .0001)
Prior 3	mean $\tau$	~ Gamma(.000001, .000001)
Prior 4	<i>total * r</i>	<i>total * (1-r)</i>
Prior 5	~ Gamma(.000001, .000001)	mean $\tau$

*total* ~ Gamma(*a*, *b*), *r* ~ Beta(1, 1)

\*Priors 1 & 2 were also used for other timeseries models. See the Priors Section.

doi:10.1371/journal.pone.0141120.t001

**Table 2. Various priors used for the precisions of the timeseries models of Method 2 (cont).**

<i>Depth (cm)</i>	<i>Mean <math>\tau</math></i>	<i>Gamma(a,b)</i>	
100	1395	6.934	.004971
120	1759	6.024	.003425
140	2241	12.413	.005538
160	3019	52.316	.017327
180	3226	87.249	.027045
200	3201	180.410	.056354
220	2175	82.412	.037894

*a,b* calculated via method of moments from mean  $\tau$  & 95%CI for posterior in Method 1

doi:10.1371/journal.pone.0141120.t002

for the 5 different priors used for the various models of Method 2, and Priors 3–5 show three schemes for apportioning the error.

Priors for the Method 3 models were set as the default BayesX software priors, with all precisions having a Gamma(.001, .001) prior.

### Subsection 5: Model Comparisons

We adopted the Deviance Information Criterion (DIC) of [29] as the method for model comparison, despite the deficiencies outlined in [30]. Thus, we planned to compare the the full models of Method 1 and Method 3 via the DIC, in addition to choosing a model from the many models of Method 2. Within the Method 2 models, only the models fitted using WinBUGS were compared.

When comparing random walk models of order one (See Results), we considered the root mean square of predictive error as a basis for model choice. This is defined as

$$\sqrt{\sum (y_{t+1} - E(y_{t+1}|y_1, y_2, \dots, y_t, \theta))^2}.$$

### Subsection 6: Computational details

The model of Method 1 which produced the contrast estimates (also used by Method 2) was fitted using custom built software, pyMCMC [31], which used the Gibbs sampler of [7] with its block updating and linear splines. Its daily models had a 6,000 iterate burnin and 16,000 iterates in all. Fewer burnin iterates were needed because of the block updating.

The models of Method 2 were fitted using BayesX [32, 33] or WinBUGS [25]. The BayesX software was used because it offered penalised smooths over time, and because actual dates could be used in the fit. It was thought that such models would offer insight into the seasonality and/or trends in the data. WinBUGS was used because of its transparency and its robustness as a well established software.

The univariate time-series models of Method 2 were run with 2 chains, 120,000 iterates with a 100,000 burnin when using WinBUGS and Gelman-Rubin statistics were checked. This burnin was unnecessarily large. Models fitted using BayesX, Eq 6, were run with a 10,000 iterate burnin and 60,000 iterates in all (again probably unnecessarily large, given that this software uses block updating [28]).

Geweke diagnostics [34] for convergence and Raftery-Lewis estimates for accuracy [35] were checked and found to be satisfactory for all models, except where otherwise noted.

## Results

### Method 1

Not unexpectedly, the variances at the shallower depths show greater variability across the sampling dates, than those at greater depths. [Fig 1](#) illustrates this, showing the square roots of the spatial and the unstructured variances for each date at depths 100 cm and 220 cm based on the model of Method 1. The comparable graphs across all depths show decreasing variability with increasing depth of these parameters across the sampling dates (Figs Y-AE in [S1 File](#)). This decreasing variability with depth is also seen in the contour smooth for the variance components across days and depth shown in the two lower panels of [Fig 2](#). The variability in these parameters justifies the choice to fit the same model across all sampling dates thereby allowing the parameters of the original model to vary by date, since a description of their evolution across time was not obvious a priori.

The estimates for the contrasts at all depths and their 95% credible intervals from Method 1 satisfy the first objective of the analysis, and are given in the online supplement, as are graphs of the fits for the time-series from all seven depths (Figs A-X in [S1 File](#)).

The left panel of [Fig 3](#) shows the point estimates from Method 1 for the contrasts at depths 100 cm to 220 cm. This graph exhibits an apparent continuity of the contrast estimates across time and depth. The same estimates are graphed again as a contour graph of moisture across day and depth (top panel [Fig 2](#)) in order to show the continuity across time and depth.

Two fits for the contrasts at depth 100 cm are shown (Figs [4](#) and [5](#)). [Fig 4](#) shows the fit for Method 1. The corresponding estimates can be seen to have much wider credible intervals compared to the time-series fits shown in [S1 File](#) and the penalised spline fit of [Fig 5](#).

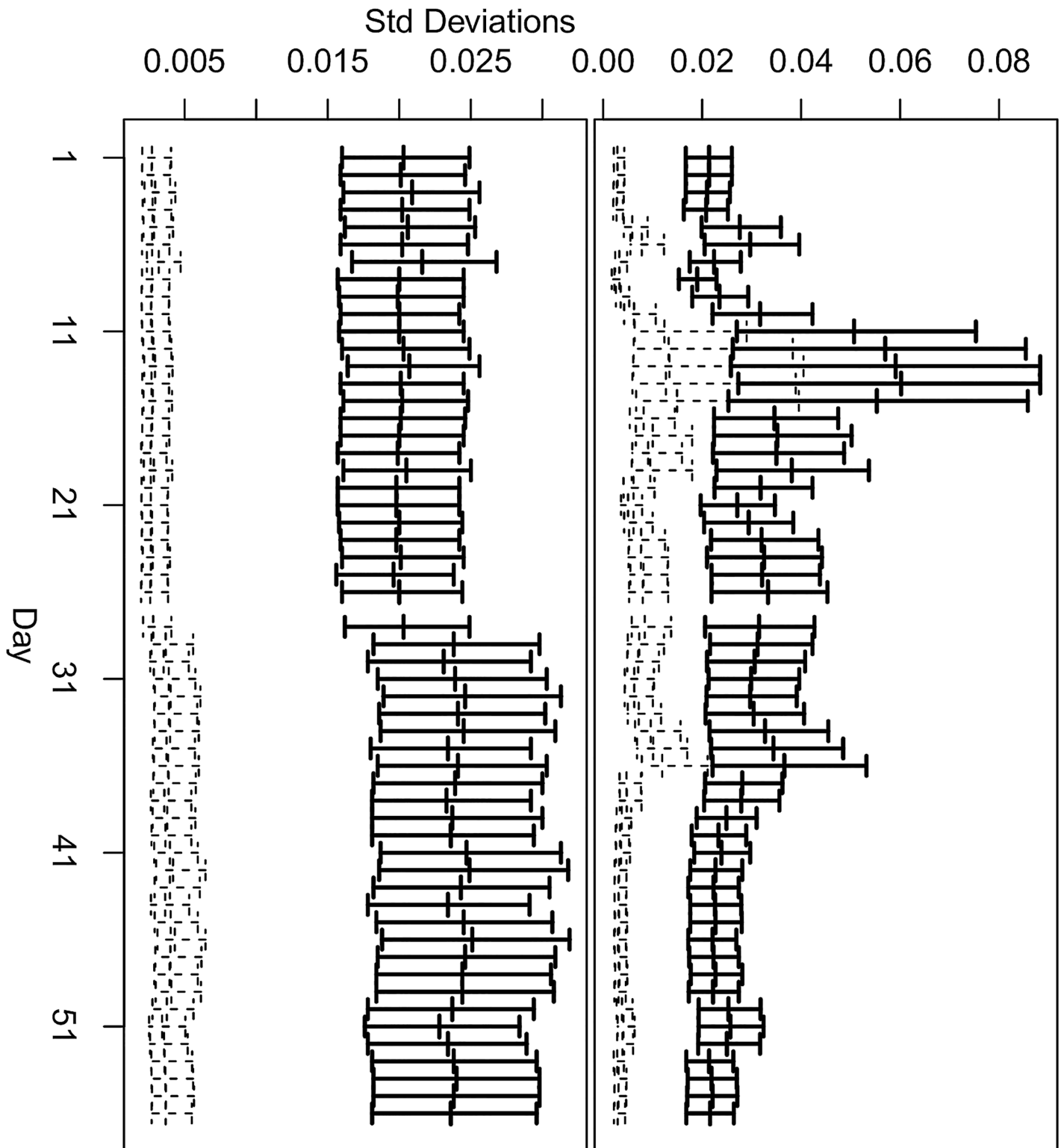
### Method 2

[Fig 5](#) shows the penalised smooth from Method 3 [Eq 8](#), and illustrates the seasonality observed in all models at the shallower depths. The 28 term regression fit from [Eq 2](#) of Method 2 (shown in [S1 File](#)) echoes the penalised spline fit, but with the discontinuities expected in a model with interactions of year by periodics. The random walk of order two model of Method 2 (shown as Fig B in [S1 File](#)) is very similar. These three fits over time show comparable 95% credible intervals.

Comparisons of the time-series models of Method 2 are given in [Table 3](#) for the contrast at depth 140 cm and were used to consider various models and ways of dealing with the unequal time spacing. [Table 3](#) indicates that the AR(1) and AR(2) models are essentially equivalent, and that the AR (1)(12) model is a poor model (with its negative estimate for the number of parameters). [Table 3](#) shows the DIC and pD varying for differing priors for the random walk models, but not varying for the AR models. This issue is discussed further in the Model Comparisons section.

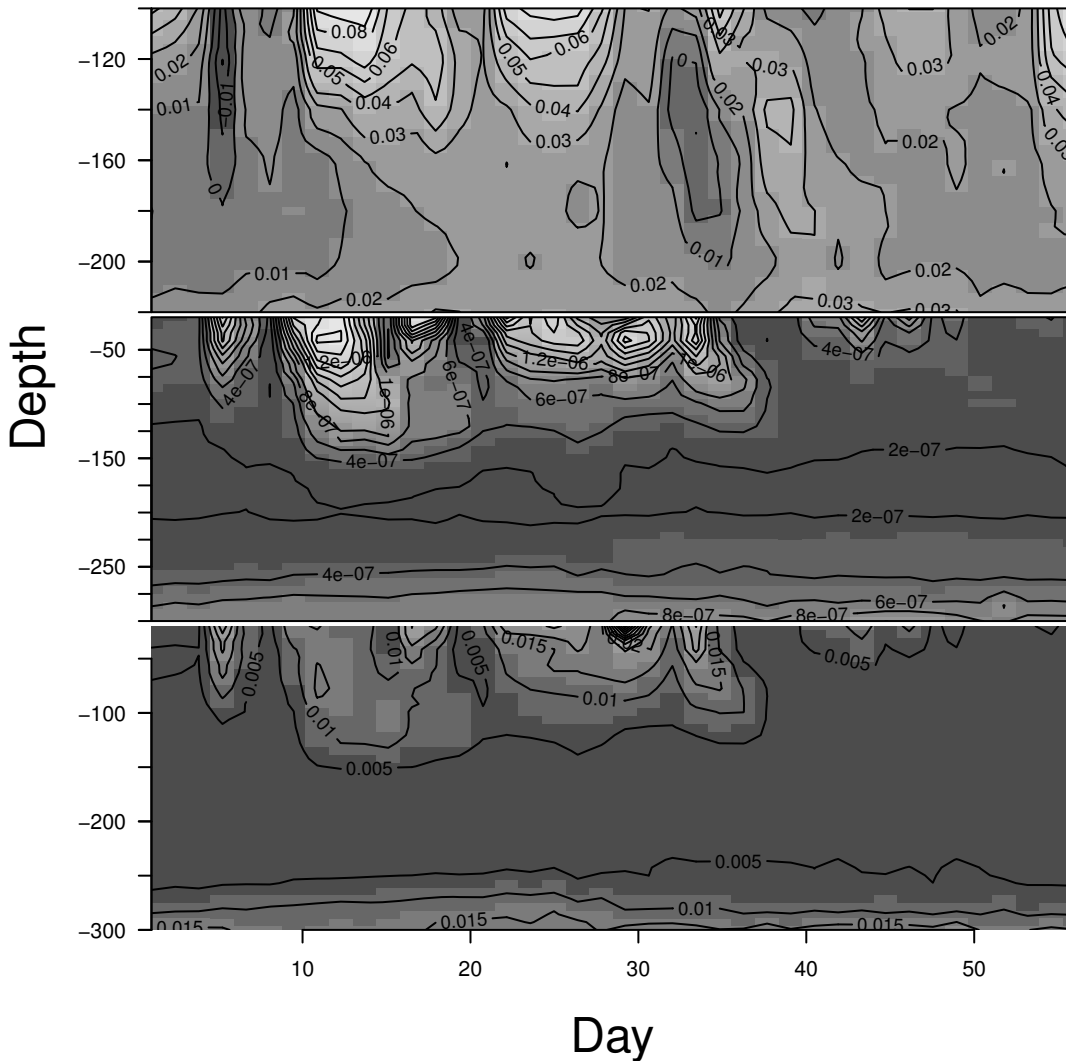
[Table 4](#) compares the models fitted using [Eqs 2](#) and [5](#). This table shows that additional periodic covariates improve the fit of the AR1 model at depths of 100 and 120 cm, but for all other depths the simple AR1 model accounts adequately for the data without the need for rainfall or periodics. This is not surprising, since the time-series models take out the random shocks that might be explained by such terms. Note that the model 'AR1+5', a covariate model in combination with an AR1 model, posits (somewhat unrealistically) the same amplitudes across the years for the cyclical behaviour, but (realistically) posits a common time of year for the yearly maxima and minima. This table also gives the DICs for the regression model with interaction terms of year by periodics which permits differing amplitudes for different years, together with a cubic over time, giving 27 time covariates. Not surprisingly, with 28 parameters to fit 56 observations, these models from [Table 4](#) show the regression model doing better than the AR1 models for all depths (except 200 and 220 cm, where they do equally well).





**Fig 1. Method 1: Spatially structured and unstructured standard deviations.** Top panel: and 95% credible intervals at depths 100 cm. Bottom panel: 95% credible intervals at depths 220 cm. The spatial standard deviations are shown using heavy black lines and bars, the unstructured standard deviations using dotted lines and bars.

doi:10.1371/journal.pone.0141120.g001



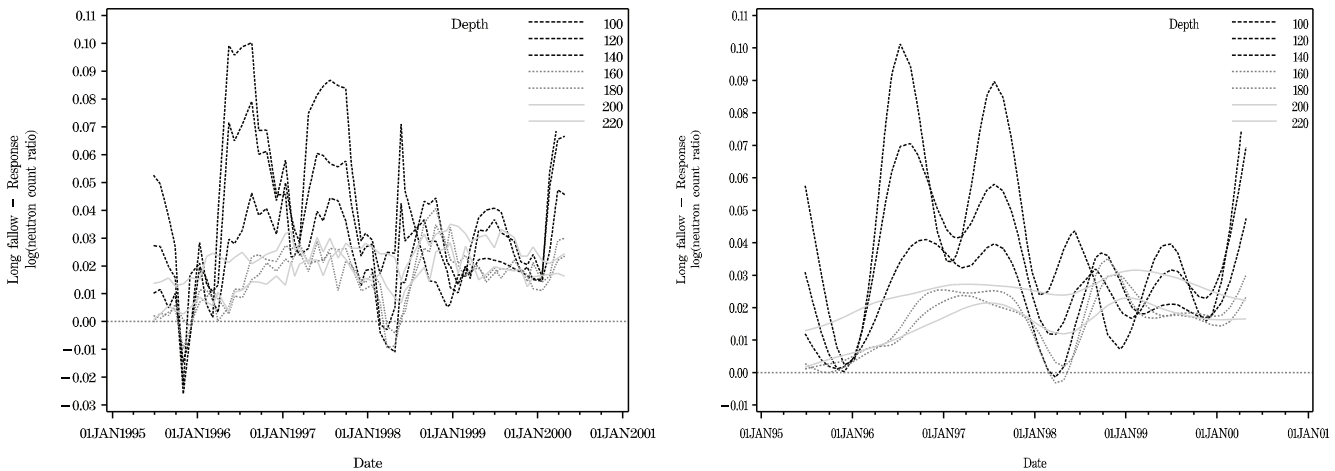
**Fig 2. Method 1: Top panel—Long following vs Response cropping.** Contour graphs from the point estimates from the MCMC iterates. Middle panel—the square root of the unstructured variance components. Bottom panel—the square root of the spatial variance components. Method 1 model.

doi:10.1371/journal.pone.0141120.g002

Table 5 shows the DICs and the estimated number of parameters (pD) for the random walk models of order one and two, both weighted and unweighted. However, these models were not used in our final model comparisons, as we did not get stable solutions for the untransformed contrast variable, for reasons explained in the Model Comparisons section.

The random walk models of order two (RW(2)) are identifiable models, and when the contrast variable was standardised, gave estimates which were not dependent on the choice of priors. (However, the transformation of the variable meant that the DICs could not be compared with the models having the unstandardised response variable.) The RW(2) models allowed the calculation of the ratio (W/V) between the two types of variance in the model (Eqs 3 and 4 which is the signal to noise ratio [20]. The signal to noise ratio for the different depths is tabulated in Table 6 and shows that at the lower depths the observational error outweighs the signal.

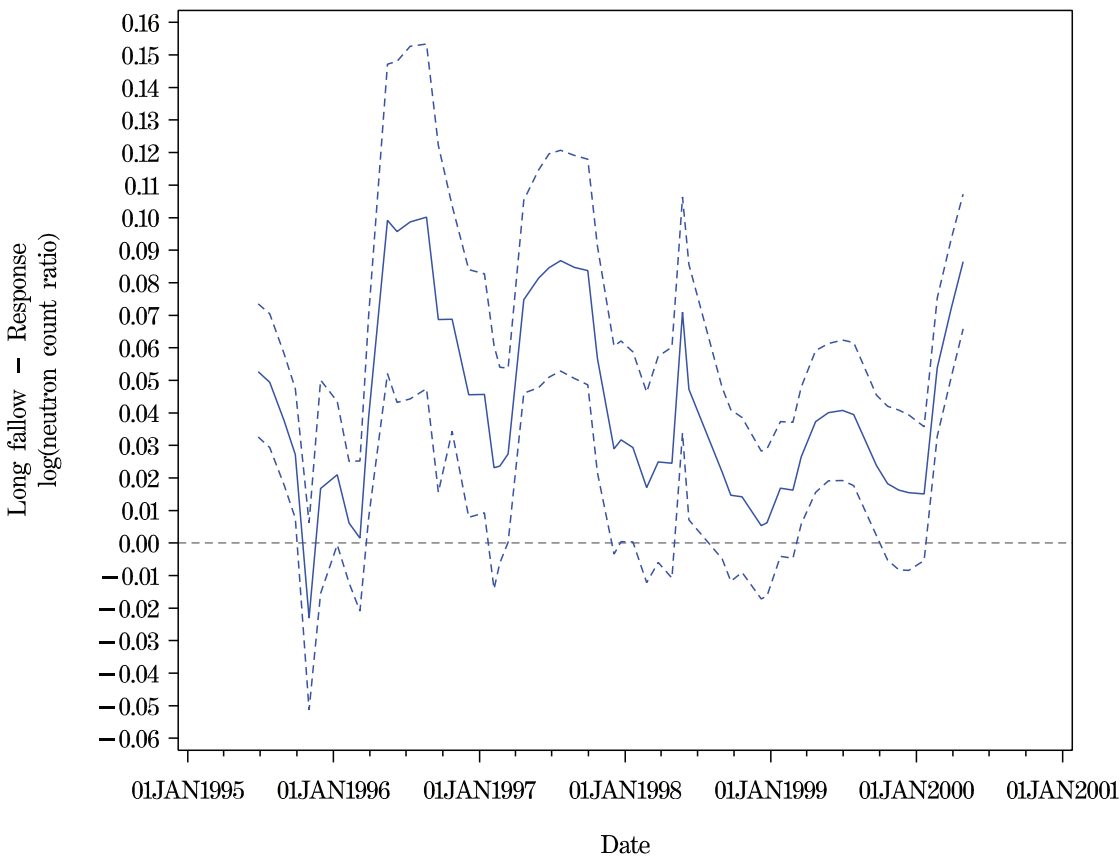
The penalised spline fits from Eq 6 (right panel of Fig 3) show seasonal peaks and troughs which vary in amplitude across the years, thereby suggesting interactions of year by sine and



**Fig 3. Long fallowing vs Response cropping at all depths.** Left panel: Point estimates from the MCMC iterates of Method 1. Right panel: Spline curves from BayesX pspline estimation (Method 2, Eq 6).

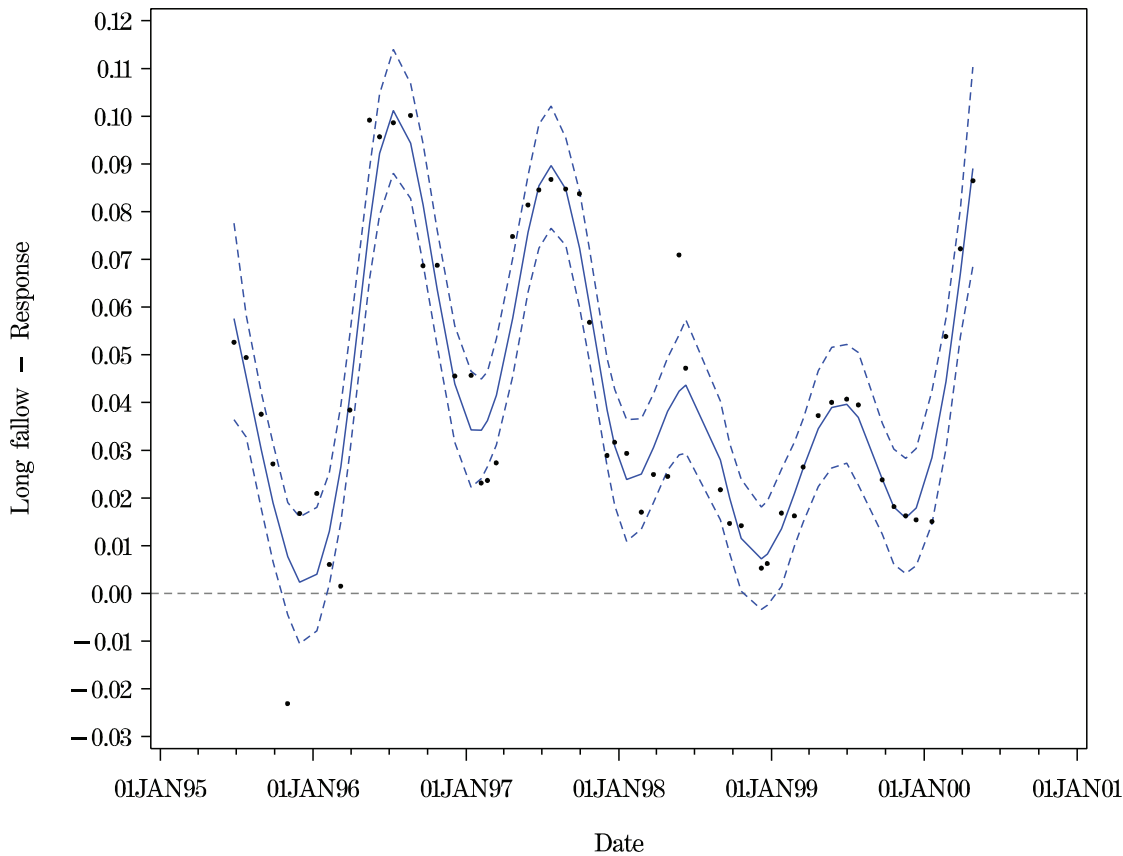
doi:10.1371/journal.pone.0141120.g003

cosine terms with periods of a year and half year. These curves also show the periodic behaviour dampening with increasing depth. We found significant terms in the rainfall and interactions of year by half-yearly periodic simple regression model, but such models showed the expected problems of sine curves being too smooth at their peaks and troughs with disjuncts at



**Fig 4. Long fallowing vs Response cropping at depth 100 for all trial dates.** Point estimates and 95% CIs from MCMC iterates from Method 1.

doi:10.1371/journal.pone.0141120.g004



**Fig 5. Long fallowing vs Response cropping at depth 100 for all trial dates.** Penalised spline smooth across dates. Point estimates and 95% CIs (Method 2, Eq 6).

doi:10.1371/journal.pone.0141120.g005

**Table 3. Summary of DICs for Contrast 1 (Long fallowing vs Response cropping) at Depth 140.**

Model	Prior 1		Prior 2	
	pD	DIC	pD	DIC
Regression	30	-377		
AR(1)	4	-343	4	-343
AR(1)(12)	-2	-356	-2	-355
AR(2)	4	-343	5	-342
RW(1)	69	-435	36	-379
RW(1) (weighted)	73	-468	40	-392
RW(1) (t <sub>10</sub> distribution)	73	-450	39	-378
RW(1) (t <sub>4</sub> distribution)	74	-451	41	-375
RW(2)	20	-370	23	-373
RW(2) (weighted)	26	-390	43	-395
RW(1) (1768 time points)	49	-304	(Prior 5)	

Prior 1: both precision priors Gamma(0.000001,0.000001)

Prior 2: both precision priors Gamma(0.0001,0.0001)

doi:10.1371/journal.pone.0141120.t003

**Table 4. DICs for Long fallowing vs Response cropping: 1st order autoregressive models vs simple regression model.**

Depth	AR1 With rainfall **		AR1		AR1+5		Regression(28)	
	pD	DIC	pD	DIC	pD	DIC	pD	DIC
100	5	-279	4	-278	9	-289	30	-315*
120	5	-301	4	-303	9	-306	30	-344*
140	5	-341	4	-343	9	-342	30	-377*
160	5	-386	4	-386	9	-386	30	-425*
180	5	-414	4	-414*	9	-410	30	-433
200	5	-449	4	-450*	9	-444	30	-449
220	5	-457	4	-458*	9	-450	30	-455

\* Better models

\*\* Covariate:  $\log(\text{rainfall} + 1)$

t AR1+5: Covariates:  $\log(\text{rainfall} + 1)$ ,  $\sin(x)$ ,  $\cos(x)$ ,  $\sin(2x)$ ,  $\cos(2x)$ ,  $x = \text{date} * 2\pi / 365$

Regression(28): Covariates:  $x, x^2, x^3, x^4, \text{year} * (\sin(x), \cos(x), \sin(2x), \cos(2x))$

doi:10.1371/journal.pone.0141120.t004

the year breaks, thus giving rise to serially correlated errors. We included these models because they gave a basis for comparison with the WinBUGS models where equality of time intervals was assumed, and allowed a possibility of a correction when used in combination with the AR models.

The missing data model, motivated by objective 3 in the Introduction, was based on Eq 3 since this was found to be close to the best model of those fitted. This was fitted because the approximation of equally spaced time intervals in the WinBUGS models seemed a considerable oversimplification. The graph for one missing data model, which attempted to construct credible intervals to reflect the spatial variability by adjusting the priors for the two variances of an RW1 model, is shown in Fig L in S1 File. This fit used prior 5 and the credible intervals for the missing values reflect the prior. Table 7 shows the apparently arbitrary nature of this undertaking, with the goodness of fits (as demonstrated by  $R^2$ ) showing highly differing values depending on the prior. However, prior 3 allocates the total variance from Model 1 to the observational error (a very informative prior) and thus allows no variation to be absorbed by the random walk. This leads to the poor fit ( $R^2$ ) and the low number of estimated parameters

**Table 5. DICs for Long fallowing vs Response cropping: random walk model comparisons, using Prior 2.**

Depth	RW1		RW1 (W)		RW2		RW2 (W)	
	pD	DIC	pD	DIC	pD	DIC	pD	DIC
100	47	-342	47	-346	21	-332	38	-340
120	39	-348	42	-360	26	-323	40	-360
140	36	-379	40	-392	23	-373	43	-395
160	34	-413	38	-424	25	-417	43	-419
180	32	-433	36	-434	25	-439	43	-419
220	28	-457	34	-448	24	-458	42	-424
220	28	-461	35	-452	24	-463	43	-427

(W): inverse time interval weights.

doi:10.1371/journal.pone.0141120.t005

**Table 6. Square root of the Signal to Noise ratio for the RW models.**

Depth (cm)	SN ratio	95% CI
100	1.06	(.58, 2.30)
120	.59	(.23, 1.09)
140	.63	(.19, 1.46)
160	.92	(.37, 2.10)
180	.71	(.29, 1.43)
200	.23	(.08, .48)
220	.13	(.05, .38)

doi:10.1371/journal.pone.0141120.t006

of Table 7 for this prior. Prior 4 partitions the total variance (estimated as a gamma prior) from Model 1 between the observational error and the random walk error. This leads to models with meaningful, but high, estimated numbers of parameters (pD). Based on the DIC, these models are poorer than the best models from Table 4. Prior 5 effectively allocates all of the error from Model 1 to the random walk component, forcing the observational error to be very small. This gives a an unsatisfactory ‘join the dots’ representation, signalled by the excessively large pD (greater than the number of degrees of freedom of the data). (The negative DICs also show that there is a problem with this model.) The issue of the error in RW(1) models is discussed further in the Model Comparisons section.

As described in the Methods, we also compared models with respect to the root mean square predictive error. In particular, we compared the the RW(1) models under Prior 1 and Prior 2 (Table 8). This criterion shows no sensitivity to the choice of prior.

Figs A–X in S1 File show the fits from Method 1, and for the time-series from Method 2, including the penalised smooths of Method 2, Eq 6. These show periodicity at the shallower depths decreasing with increasing depth. The RW(2) model fits (Method 2) are also shown.

*Convergence*

Tests for convergence (using Geweke’s test [34]) for the first model of Method 3 Eq (7) showed failure to converge. This is not surprising since this model is a major simplification of the model, Eq 1, from Method 1. See Methods. The more complex model from Method 3 Eq (8) also failed to converge. Hence, the modelling strategy of Method 3 was not pursued. However, the BayesX penalised spline models along the time dimension which modelled the 56 contrast estimates from Method 1 (Method 2, Eq 6) converged satisfactorily.

Models from Methods 1 and 2, with the exception of the random walk models, showed successful convergence using Geweke statistics [34] and Raftery-Lewis statistics [35] for the quantities of interest.

**Table 7. R<sup>2</sup>, pD and DIC for the RW(1) weighted models using priors 3–5.**

Depth (cm)	Prior 3			Prior 4			Prior 5		
	R <sup>2</sup>	pD	DIC	R <sup>2</sup>	pD	DIC	R <sup>2</sup>	pD	DIC
100	33%	13	–255	80%	36	–258	99%	100	–411
120	23%	9	–277	79%	35	–279	99%	97	–421
140	12%	6	–299	80%	35	–271	99%	94	–433
160	16%	5	–323	83%	35	–251	100%	90	–446
180	19%	5	–332	85%	34	–246	99%	89	–448
200	27%	4	–337	86%	34	–239	99%	89	–448
220	18%	3	–318	82%	34	–225	99%	94	–434

doi:10.1371/journal.pone.0141120.t007

**Table 8. Root mean square predicted error for RW1 models under Priors 1 & 2.**

<i>Depth</i>	<i>Prior</i>	<i>Median</i>	<i>25%ile</i>	<i>75%ile</i>
100	Prior 1	0.020	0.019	0.020
100	Prior 2	0.019	0.019	0.020
120	Prior 1	0.016	0.015	0.016
120	Prior 2	0.015	0.014	0.016
140	Prior 1	0.011	0.011	0.011
140	Prior 2	0.010	0.010	0.011
160	Prior 1	0.0075	0.0073	0.0077
160	Prior 2	0.0073	0.0070	0.0077
180	Prior 1	0.0057	0.0054	0.0059
180	Prior 2	0.0056	0.0054	0.0059
200	Prior 1	0.0037	0.0035	0.0039
200	Prior 2	0.0041	0.0039	0.0043
220	Prior 1	0.0035	0.0033	0.0037
220	Prior 2	0.0039	0.0037	0.0041

doi:10.1371/journal.pone.0141120.t008

Table 4 indicates that the regression models with periodic terms are the better descriptions for the behaviour of moisture contrast over time at the shallower depths, while simple AR(1) models are sufficient to capture the damped periodicity at the deeper depths. Table 5 shows that the random walk of order 2 models give an almost equivalently good fit as the best models from Table 4. For all models and methods, it is the case that point estimates and 95% credible intervals for this contrast are generally positive, leading to the conclusion that long fallow cropping generally led to moister soils than response cropping. Thus, if we wish to reduce the possibility of deep drainage and increased groundwater salinity, the recommended cropping system is response cropping.

## Discussion

This paper has provided a careful evaluation of a range of spatio-temporal models for a real case study. Some remarks are now made about the choice of models in the context of the experimental design and about the inferences drawn about the problem itself.

The experimental design was a balanced lattice design. Two issues arise here. First, the models considered in this paper treated the data as observational since the observations reflect the randomisation inherent in the design. An alternative approach is to incorporate the design structure explicitly in the models, for example by fitting the replicates and obtaining an error term with respect to them. The two approaches have merit and are arguably equally appropriate. Second, even though the observations embody the design characteristics, the spatial character of the data is not necessarily adequately described by the randomisation and a subsequent design-based analysis. This provides further motivation for our modelling strategy, in which the replicates are dealt with as neighbours, so that in the CAR-layered model the replicate error forms part of the spatial error. The comparative characteristics and performance of the modelling methods considered in this paper are discussed in further detail below.

As a reviewer comments, an alternative modelling strategy would have been to fit the replicates and obtain an error term with respect to them. However, we feel that our modelling strategy, in which the replicates were treated as neighbours, is equally appropriate. Thus, in the CAR-layered model used, the replicate error formed part of the spatial error.

It should be noted that the comparisons are between systems and not crops: systems of long fallowing v. response cropping and so on, which have the potential to use soil water, or lose it

as deep drainage depending on the crop planting and sequencing rules inherent in the design of each system. In this context, the crops are merely tools with which to suck soil water given the crop planting decision rules and rotational constraints (For example, disease carry over) which characterise each system. We take the crop rotations and differences in crops into account in that we have 9 treatments (although an opportunity crop in one year may differ from the opportunity crop on another year within a treatment, and this is what response cropping is all about—the farmer plants (or fails to plant) the most appropriate crop given the soil water status, time of year, and the potential for disease carry over from previous crops. The aim in managing the experiment was to grow a healthy crop appropriate for the system. A comparison between crops would be a separate exercise to this system comparison and would be fraught with confounding issues as the experiment was not designed for such a comparison.

The layered CAR analysis [6, 7] of these data showed that response cropping delivers lower moisture levels for most times of the year, in contrast to long fallow cropping. At the shallower depths, not surprisingly, this contrast exhibited considerable periodicity which attenuated with depth. It appears that the temporal component adds little additional uncertainty to the contrasts which has not been captured by model 1.

As discussed below, there are very few papers which deal with four-dimensional data, and frequently, the work is simply descriptive in one of the dimensions. The methods we used allowed us to fully explore the complexity of these four dimensional agricultural data.

## Modelling spatio-temporal data

Most spatio-temporal modelling deals with two spatial dimensions (surface) and time. Within the context of two spatial dimensions  $\times$  time much work is largely descriptive. For example, Teschke et al [36] and Bell et al [37] use maps at several timepoints, with the maps essentially being descriptive devices. The more complex models are generally Bayesian and use either geostatistical methods or the convolution CAR prior of [8]. Models with CAR priors typically partition the error term in the model,  $\epsilon_{it}$  as  $e_{is}$ ,  $e_t$  and  $e_{it}$ , where the first two error terms capture the structured spatial random effects and the structured temporal random effects, and  $e_{it}$  is a simple unstructured random effect with  $e_{it} \sim N(0, \sigma^2)$ . See, e.g., [38–42], where the last three papers use BayesX software [32, 33] to conduct the analyses.

Assuncao et al [43] fit quadratics over time which differ for each spatial location, and for which the coefficients are smoothed using CAR priors, but both space and time are separable. This elegant solution to modelling very short time sequence data allows the possibility of seeing increasing and decreasing rates, while accounting for spatial closeness. Assuncao and others [44, 45] use space-varying regression coefficients on their quadratic models in time. Yan and Clayton [46] use the space-time interaction to define a set of space-time separable clusters carrying a specific risk, and fit a final unstructured random effect.

Abellan et al [47] decompose the error term into a structured temporal effect, a structured spatial random effect, and a time by space interaction random effect which is a mixture of two Gaussians and thus equivalent to an outlier or contaminant model. This allows the identification of sites (areas) and times which fail to fit the common temporal and common spatial patterns.

Within an agricultural context, (and again considering 2D  $\times$  time analyses), the analysis of Trought and Bramley [48] considers the quality of grape juice by site across time. Their strategy is to fit different curves across time for each site, and then to look at spatial outcomes of their model by mapping. In considering longitudinal agricultural experiments, [49–52] and [53] use mixed models within a REML framework to analyse their spatio-temporal data, and fit state-space models via standard software and REML. The fixed part of their models is generally straightforward and the data are measured on two spatial dimensions.



Moving to three spatial dimensions and time, Higdon's method of convolution with Gaussian kernels [13] allows for non-stationary spatial smoothing. The analyses of [54, 55] only give snapshots over time. Lemos et al [56], having failed to find much influence in terms of spatial proximity, model time sequences for each site using time-series methods. The convolutional approach is very attractive but possibly three dimensional spatial observations of ocean waters are smoother than soil observations over three spatial dimensions.

In an agricultural context, most studies involving three spatial dimensions and time do not attempt to model the data across all three spatial dimensions or across time. Thus, the soil profile study of [57] does not use spatial information in the analysis. Other studies composite the soils from different depths across soil types or treatment [58] thereby returning to a 2-dimensional  $\times$  time analysis. However, Piepho et al [49] present a four dimensional REML mixed modelling approach which is used by [59].

A major difference between agricultural data and epidemiological data which is so often modelled using the CAR prior of [8], together with an additive common spatially structured error, an additive common structured temporal random term (and an unstructured error with a variance common over both space and time), is that the spatial units of epidemiological data tend to vary slowly over time scales of a few years. Additionally, administrative time shocks may often be constant across a map, and hence this simple modelling structure works well. In contrast, the agricultural data modelled here vary markedly from sampling date to sampling date, and it is clear that the simple separable variance decomposition used by so many epidemiological models does not describe the data well.

## Modelling the case study data

In moving to four dimensions, there are many more possibilities for the decomposition of the fixed and residual parts of the model. Here, given the very different scales at which the spatial measurements were made, we chose to use CAR models for each depth layer, with neighbourhood smoothing occurring across the (roughly) horizontal layer. This choice to treat the third spatial dimension differently is made by others with three dimensional spatial data and for the same reason. Thus, Ridgway et al [60] separate out the depth component in their loess data fit, when modelling ocean temperatures.

For these data, the use of the CAR layered model [6, 7] also gave greater flexibility in modelling error components in the depth dimension, a flexibility which we found was needed. Thus, in the CAR model, depth is excluded from the neighbourhood error structures. (If depth neighbours were to be included as neighbours with equal weights, the horizontal layer information would be downweighted. If weighting uses functions of distance, the horizontal correlations would become effectively irrelevant.) This decision to define neighbours as neighbours only within the same depth layer gave the useful property that the CAR model is then permitted to have differing variances across the depths. For our model (Method 1, Eq 1), both the homogeneous and spatial variance components differ by depth, and while no formal tests were conducted, this flexibility in the model appeared better.

The Method 1 model is a date interaction model with the daily model. Each daily model is independent of each other which allows us to sum the DICs and the pDs [61, 62] over the 56 daily models and thus allows the possibility of a crude comparison of DICs with the models of Method 2, Eqs 6 and 8, where all 90,720 observations are modelled at once. We had planned to use this to compare the Method 3 models fitted to the full dataset, with the fit from the daily models from Method 1. However, the Method 3 models failed to converge, and hence, this was not done.

Method 1 gives appropriate 95% credible intervals for the contrast estimates, but no insight into the way in which these contrasts vary over time. The modelling strategy adopted in

Method 2 attempts to remedy that by fitting time-varying covariates and by using time-series methods. Two-stage models do not account for the treatment effect variation observed in the model 1 fits. However they do allow us to see what level of complexity may be required to account for the time-varying nature of the contrasts.

A missing data model is shown in the supplementary materials (Fig L in [S1 File](#)). However, there is little point in fitting a missing data model when the posterior variance is largely dictated by the choice of a prior for a precision, as it must be with only 3% of the data being observed.

## Model Comparisons

Where competing models are suggested, the preference for model comparison is to use some summary statistic of the analysis fit such as the AIC [63] (used for geostatistical model comparisons by [64]), the BIC [65] or the DIC [29] which remains useful despite various criticisms [30]. When WinBUGS is used for model fitting, an obvious choice for a model comparison criterion is the DIC, which has the added advantage of estimating the number of parameters used by the model. [Table 7](#) shows DICs for random walk models of order 1 where the only modelling difference is in the priors used for the two precisions. The differing priors make differences to the fit (shown by  $R^2$ ) and to the DIC.

In arguing the case for the DIC, a CAR model is explicitly discussed in [29]. Their model has a CAR normal spatial prior, but the unstructured error component is Poisson, and therefore dictated by the estimates for the Poisson rate, which are themselves modified by the spatial CAR prior. For the agricultural data here, the CAR (and RW) models (both of which have two variance components) gave DICs which were highly sensitive to the choice of priors. The DICs of the autoregressive models are unaffected by the prior choice for the error, but the random walk models of order one often have estimates for the number of parameters which are greater than the number of observations used. Additionally, estimates for the number of parameters change with choice of prior, as does the DIC. This is not a problem of the criterion choice. Calculation of the BIC, which is also based on the final fit, gives essentially the same preferred models. The convolution prior of [8] which works so well for spatial epidemiological data, works less well when all the model components are normal and there is a single observation to be partitioned into structured and unstructured error, and a maximum of two neighbours, as is the case for the RW(1) models here. The random walk of order one with error partitioned into a normal random walk component and a normal observational error component is not identifiable. This generally does not lead to difficulties when the priors chosen for both precisions are identical, and sufficiently, but not overly diffuse. Here the partitioning was induced by restrictive priors on the random walk error, which had substantial influence on the model and led to very different model fits ([Tables 3](#) and [5](#)). However, random walks of order two with both observational and random walk error are identifiable, yet we still had problems with them. We believe that the RW(2) models gave unstable results as a consequence of the failure to standardise the outcome variable. When the outcome variable was standardised, the RW(2) models gave stable results for all sets of equal priors used. (These models, however, were not able to be compared with the earlier models via DIC or BIC.) Note that the DIC and its associated estimated number of parameters,  $pD$ , were extremely useful in alerting us to the problem of non-identifiability in the RW(1) models.

Our final view was that the purpose of the modelling across time was to develop insight into the time-varying nature of the contrast estimates. No 'best' model was chosen. Rather, the 'better' models with comparable DICs from [Table 4](#), together with the daily penalised spline models and the Method 1 model were used to inform the description of the contrast of long

following versus response cropping. There is evidence of periodicity at the shallower levels (See [Table 4](#)) and this is also shown by the penalised spline smooths of [Eq 6](#) and illustrated in [Fig 5](#), with the double periodics per year at the shallower levels. At the depths of 200 cm and 220 cm there are fewer obvious peaks and troughs.

## Conclusions

Our purpose was to account for spatial and temporal autocorrelations in the context of four-dimensional data. The model of Method 1 and [\[7\]](#) forms the basis for the final analyses within this paper. It gave the fitted estimates for moisture at every combination of depth, date and treatment, and allowed a complex error structure, with an unstructured error at every depth, date and site, and with variances differing by depth and date. The spatially structured error was fitted across each horizontal layer and ignored depth neighbours. The variance of these structured spatial errors also differed by depth and date. Comparisons with three dimensional CAR neighbourhood models (not shown here) showed that that this separation of the two-dimensional plot arrangement from the depth dimension gave better descriptions of the data. See [\[66\]](#).

The simple expedient of fitting the data as a series of daily models allowed the maximum possible complexity in terms of the experiment and was a useful approach to modelling the full dataset. By fitting an interaction model by date at all levels of the daily model, we were able to explore more fully and more effectively the variability of the data, in particular, the variability of the variance components across the four dimensions. At the greater depths, seasonal variation in both contrasts and variances is less visible. See [Fig 3](#) for the contrast, and the bottom panel of [Fig 1](#) and the two lower panels of [Fig 2](#) for the variances.

These analyses show that response cropping delivers lower moisture levels for most times of the year, in contrast to long fallow cropping. At the shallower depths, not surprisingly, this contrast exhibits considerable cyclicity which attenuates with depth. It appears that the temporal component adds little additional uncertainty to the contrasts which has not been captured by Model 1.

The method of defining neighbours within a horizontal layer has potentially wide applicability in three and four dimensional agricultural datasets, where the plot and treatment are defined by the two-dimensional surface coordinates. It may also be applicable in measurements made over the ocean where variables may also be measured at depth, again a situation where the differences in latitude and longitude between measurements far outweigh the differences in the depth dimension.

## Supporting Information

**S1 File. Supporting Figs A-AE.** Figs A–X show point and 95% credible intervals for various models discussed in the text. Figs Y–AE show the estimated spatial and the unstructured variance components at the different depths over the period of the experiment. (PDF)

## Acknowledgments

We wish to thank Dr. Clair Alston for her useful discussions, and Dr. Chris Strickland for his programming of, and assistance with the Gibbs sampler in PyMCMC.

This work was supported by the Australian Research Council, by QUT and by Anne Clements and Associates. We thank Dr. Alison Bowman of NSW Industry and Investment for her interest in, and support of this work. We also wish to thank the reviewer for very thoughtful comments that have led to a greatly improved manuscript.

## Author Contributions

Conceived and designed the experiments: RRY. Performed the experiments: RRY. Analyzed the data: MRD. Wrote the paper: MRD KLM RRY.

## References

1. Cressie N, Wikle CK. *Statistics for spatio-temporal data*. Hoboken, New Jersey: John Wiley & Sons; 2011.
2. Berliner LM, Milliff RF, Wikle CK. Bayesian hierarchical modeling of air-sea interaction. *Journal of Geophysical Research: Oceans*. 2003; 108(C4):3104. doi: [10.1029/2002JC001413](https://doi.org/10.1029/2002JC001413)
3. Wikle C, Hooten M. A general science-based framework for dynamical spatio-temporal models. *TEST*. 2010; 19(3):417–451. doi: [10.1007/s11749-010-0214-2](https://doi.org/10.1007/s11749-010-0214-2)
4. Berliner LM. Hierarchical Bayesian time-series models. In: Hanson KM, Silver RN, editors. *Proceedings of the Fifteenth International Workshop on Maximum Entropy and Bayesian Methods*. Dordrecht, N.L.: Kluwer Academic Publishers; 1996. p. 15–22.
5. Wikle CK, Hooten MB. Hierarchical Bayesian Spatio-Temporal Models for Population Spread. In: Clark JS, Gelfand A, editors. *Hierarchical Modelling for the Environmental Sciences*. Oxford University Press; 2006. p. 145–169.
6. Donald M, Alston CL, Young RR, Mengersen KL. A Bayesian analysis of an agricultural field trial with three spatial dimensions. *Computational Statistics and Data Analysis*. 2011; 55(12):3320–3332. Available online 25 June 2011. doi: [10.1016/j.csda.2011.06.022](https://doi.org/10.1016/j.csda.2011.06.022)
7. Donald M, Strickland CM, Alston CL, Young RR, Mengersen KL. Comparison of three dimensional profiles over time. *Journal of Applied Statistics*. 2012; Available online: 31 Jan 2012. doi: [10.1080/02664763.2012.654771](https://doi.org/10.1080/02664763.2012.654771)
8. Besag JE, York J, Mollie A. Bayesian image restoration with applications in spatial statistics (with discussion). *Annals of the Institute of Mathematical Statistics*. 1991; 43:1–59. doi: [10.1007/BF00116466](https://doi.org/10.1007/BF00116466)
9. Besag JE, Green P, Higdon D, Mengersen K. Bayesian Computation and Stochastic Systems. *Statistical Science*. 1995; 10(1):3–41. doi: [10.1214/ss/1177010130](https://doi.org/10.1214/ss/1177010130)
10. Gelfand AE, Vounatsou P. Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics*. 2003; 4(1):11–25. doi: [10.1093/biostatistics/4.1.11](https://doi.org/10.1093/biostatistics/4.1.11) PMID: [12925327](https://pubmed.ncbi.nlm.nih.gov/12925327/)
11. Cressie NAC. *Statistics for spatial data*. Wiley series in probability and mathematical statistics. Applied probability and statistics. New York: John Wiley; 1991.
12. Banerjee S, Carlin BP, Gelfand AE. *Hierarchical modeling and analysis for spatial data*. Monographs on statistics and applied probability. Boca Raton, London, New York, Washington D.C.: Chapman & Hall; 2004.
13. Higdon D. A process-convolution approach to modelling temperatures in the North Atlantic Ocean. *Environmental and Ecological Statistics*. 1998; 5:173–190. doi: [10.1023/A:1009666805688](https://doi.org/10.1023/A:1009666805688)
14. Besag JE, Mondal D. First-order intrinsic autoregressions and the de Wijs process. *Biometrika*. 2005; 92(4):909–920. doi: [10.1093/biomet/92.4.909](https://doi.org/10.1093/biomet/92.4.909)
15. Lindgren F, Rue H, Lindstrom J. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2011; 73(4):423–498. doi: [10.1111/j.1467-9868.2011.00777.x](https://doi.org/10.1111/j.1467-9868.2011.00777.x)
16. Hrafnkelsson B, Cressie N. Hierarchical modeling of count data with application to nuclear fall-out. *Environmental and Ecological Statistics*. 2003; 10:179–200. doi: [10.1023/A:1023674107629](https://doi.org/10.1023/A:1023674107629)
17. Rue H, Tjelmeland H. Fitting Gaussian Markov Random Fields to Gaussian Fields. *Scandinavian Journal of Statistics*. 2002; 29(1):31–49. doi: [10.1111/1467-9469.00058](https://doi.org/10.1111/1467-9469.00058)
18. Harvey AC. *Forecasting, structural time series models and the Kalman filter*. Cambridge England: Cambridge University Press; 1989.
19. Box GEP, Jenkins GM, Reinsel GC. *Time series analysis: Forecasting and control*. 3rd ed. Englewood Cliffs, N.J.: Prentice Hall; 1994.
20. West M, Harrison J. *Bayesian forecasting and dynamic models*. 2nd ed. Springer series in statistics. New York: Springer; 1997.
21. Commandeur JJF, Koopman SJ. *An introduction to state space time series analysis*. Practical econometrics. Oxford New York: Oxford University Press; 2007.
22. Fahrmeir L, Kneib T, Lang S. Penalized structured additive regression for space-time data: A Bayesian perspective. *Statistica Sinica*. 2004; 14:731–761.

23. Kneib T. Geoadditive hazard regression for interval censored survival times. *Computational Statistics and Data Analysis*. 2006; 51:777–792. doi: [10.1016/j.csda.2006.06.019](https://doi.org/10.1016/j.csda.2006.06.019)
24. Ringrose-Voase AJ, Young RR, Payder Z, Huth NI, Bernardi AL, Cresswell HP, et al. Deep Drainage under Different Land Uses in the Liverpool Plains Catchment. *Agricultural Resource Management Report Series, NSW Agriculture Orange*; 2003. 3.
25. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS—A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*. 2000; 10(4):325–337. doi: [10.1023/A:1008929526011](https://doi.org/10.1023/A:1008929526011)
26. Box GEP, Jenkins GM. *Time series analysis: Forecasting and control*. Rev. ed. Holden-Day series in time series analysis and digital processing. San Francisco: Holden-Day; 1976.
27. Gamerman D, Lopes HF. *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. 2nd ed. London; New York: Chapman & Hall; 2006.
28. Brezger A, Lang S. Generalized structured additive regression based on Bayesian P-splines. *Computational Statistics and Data Analysis*. 2006; 50(4):967–991. doi: [10.1016/j.csda.2004.10.011](https://doi.org/10.1016/j.csda.2004.10.011)
29. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*. 2002; 64(4):583–639. doi: [10.1111/1467-9868.00353](https://doi.org/10.1111/1467-9868.00353)
30. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2014; 76(3):485–493. doi: [10.1111/rssb.12062](https://doi.org/10.1111/rssb.12062)
31. Strickland C. pyMCMC: a statistical package for Bayesian MCMC analysis. *Journal of Computational and Graphical Statistics*. 2010;p. 1–46. Submitted August, 2010.
32. Belitz C, Brezger A, Kneib T, Lang S. *BayesX Software for Bayesian Inference in Structured Additive Regression Models Version 2.0.1 Reference Manual*; 2009. Accessed: October 25, 2010. Available from: <http://www.stat.uni-muenchen.de/~bayesx/bayesx.html>
33. Belitz C, Brezger A, Kneib T, Lang S. *BayesX Software for Bayesian Inference in Structured Additive Regression Models Version 2.0.1 Software Methodology Manual*; 2009. Accessed: October 25, 2010. Available from: <http://www.stat.uni-muenchen.de/~bayesx/bayesx.html>
34. Geweke J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Bayesian Statistics*. 1992; 4:169–188.
35. Raftery AL, Lewis S. How many iterations in the Gibbs sampler? In: Bernardo JM, Berger JO, Dawid AP, Smith AFM, editors. *Bayesian Statistics 4*. Oxford: Oxford University Press; 1992. p. 763–773.
36. Teschke K, Chow Y, Bartlett K, Ross A, van Netten C. Spatial and temporal distribution of airborne *Bacillus thuringiensis* var. *kurstaki* during an aerial spray program for gypsy moth eradication. *Environmental Health Perspectives*. 2001; 109(1):47–54. PMID: [11171524](https://pubmed.ncbi.nlm.nih.gov/11171524/)
37. Bell ML, Dominici F, Ebisu K, Zeger SL, Samet JM. Spatial and temporal variation in PM<sub>2.5</sub> chemical composition in the United States for health effects studies. *Environmental Health Perspectives*. 2007; 115(7):989–995. doi: [10.1289/ehp.9621](https://doi.org/10.1289/ehp.9621) PMID: [17637911](https://pubmed.ncbi.nlm.nih.gov/17637911/)
38. Waller LA, Carlin BP, Xia H, Gelfand AE. Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association*. 1997; 92(438):607–617. doi: [10.1080/01621459.1997.10474012](https://doi.org/10.1080/01621459.1997.10474012)
39. Knorr-Held L, Besag J. Modeling risk from a disease in time and space. *Statistics in Medicine*. 1998; 17:2045–2060. doi: [10.1002/\(SICI\)1097-0258\(19980930\)17:18%3C2045::AID-SIM943%3E3.0.CO;2-P](https://doi.org/10.1002/(SICI)1097-0258(19980930)17:18%3C2045::AID-SIM943%3E3.0.CO;2-P) PMID: [9789913](https://pubmed.ncbi.nlm.nih.gov/9789913/)
40. Crook AM, Knorr-Held L, Hemingway H. Measuring spatial effects in time to event data: a case study using months from angiography to coronary artery bypass graft (CABG). *Statistics in Medicine*. 2003; 22(18):2943–2961. doi: [10.1002/sim.1535](https://doi.org/10.1002/sim.1535) PMID: [12953290](https://pubmed.ncbi.nlm.nih.gov/12953290/)
41. Adebayo SB, Fahrmeir L. Analysing child mortality in Nigeria with geoadditive discrete-time survival models. *Statistics in Medicine*. 2005; 24(5):709–728. doi: [10.1002/sim.1842](https://doi.org/10.1002/sim.1842) PMID: [15696506](https://pubmed.ncbi.nlm.nih.gov/15696506/)
42. Poncet C, Lemesle V, Mailleret L, Bout A, Boll R, Vaglio J. Spatio-temporal analysis of plant pests in a greenhouse using a Bayesian approach. *Agricultural and Forest Entomology*. 2010; 12(3):325–332.
43. Assuncao RM, Reis IA, Oliveira CD. Diffusion and prediction of Leishmaniasis in a large metropolitan area in Brazil with a Bayesian space-time model. *Statistics in Medicine*. 2001; 20(15):2319–2335. doi: [10.1002/sim.844](https://doi.org/10.1002/sim.844) PMID: [11468766](https://pubmed.ncbi.nlm.nih.gov/11468766/)
44. Assuncao RM, Potter JE, Cavenaghi SM. A Bayesian space varying parameter model applied to estimating fertility schedules. *Statistics in Medicine*. 2002; 21(14):2057–2075. doi: [10.1002/sim.1153](https://doi.org/10.1002/sim.1153) PMID: [12111887](https://pubmed.ncbi.nlm.nih.gov/12111887/)
45. Assuncao RM. Space varying coefficient models for small area data. *Environmetrics*. 2003; 14(5):453–473. doi: [10.1002/env.599](https://doi.org/10.1002/env.599)

46. Yan P, Clayton MK. A cluster model for space-time disease counts. *Statistics in Medicine*. 2006; 25(5):867–881. doi: [10.1002/sim.2424](https://doi.org/10.1002/sim.2424) PMID: [16453380](https://pubmed.ncbi.nlm.nih.gov/16453380/)
47. Abellan JJ, Richardson S, Best N. Use of space-time models to investigate the stability of patterns of disease.(Mini-Monograph). *Environmental Health Perspectives*. 2008; 116(8):1111–1119. doi: [10.1289/ehp.10814](https://doi.org/10.1289/ehp.10814) PMID: [18709143](https://pubmed.ncbi.nlm.nih.gov/18709143/)
48. Trought MCT, Bramley RGV. Vineyard variability in Marlborough, New Zealand: characterising spatial and temporal changes in fruit composition and juice quality in the vineyard. *Australian Journal of Grape and Wine Research*. 2011; 17(1):79–89. doi: [10.1111/j.1755-0238.2010.00120.x](https://doi.org/10.1111/j.1755-0238.2010.00120.x)
49. Piepho HP, Buchse A, Richter C. A mixed modelling approach for randomized experiments with repeated measures. *Journal of Agronomy and Crop Science*. 2004; 190(4):230–247. doi: [10.1111/j.1439-037X.2004.00097.x](https://doi.org/10.1111/j.1439-037X.2004.00097.x)
50. Piepho HP, Ogutu JO. Simple state-space models in a mixed model framework. *American Statistician*. 2007; 61(3):224–232. doi: [10.1198/000313007X220426](https://doi.org/10.1198/000313007X220426)
51. Piepho HP, Richter C, Williams E. Nearest neighbour adjustment and linear variance models in plant breeding trials. *Biometrical Journal*. 2008; 50(2):164–189. doi: [10.1002/bimj.200710414](https://doi.org/10.1002/bimj.200710414) PMID: [18383446](https://pubmed.ncbi.nlm.nih.gov/18383446/)
52. Wang LA, Goonewardene Z. The use of mixed models in the analysis of animal experiments with repeated measures data. *Canadian Journal of Animal Science*. 2004; 84(1):1–11. doi: [10.4141/A03-123](https://doi.org/10.4141/A03-123)
53. Brien CJ, Demetrio CGB. Formulating Mixed Models for Experiments, Including Longitudinal Experiments. *Journal of Agricultural, Biological, and Environmental Statistics*. 2009; 14(3):253–280. doi: [10.1198/jabes.2009.08001](https://doi.org/10.1198/jabes.2009.08001)
54. Lemos RT, Sanso B. A Spatio-Temporal Model for Mean, Anomaly, and Trend Fields of North Atlantic Sea Surface Temperature. *Journal of the American Statistical Association*. 2009; 104(485):5–18. doi: [10.1198/jasa.2009.0018](https://doi.org/10.1198/jasa.2009.0018)
55. Sahu SK, Challenor P. A space-time model for joint modeling of ocean temperature and salinity levels as measured by Argo floats. *Environmetrics*. 2008; 19(5):509–528. doi: [10.1002/env.895](https://doi.org/10.1002/env.895)
56. Lemos RT, Sanso B, Huertos ML. Spatially Varying Temperature Trends in a Central California Estuary. *Journal of Agricultural, Biological, and Environmental Statistics*. 2007; 12(3):379–396. doi: [10.1198/108571107X227603](https://doi.org/10.1198/108571107X227603)
57. Macdonald BCT, Reynolds JK, Kinsela AS, Reilly RJ, van Oploo P, Waite TD, et al. Critical coagulation in sulfidic sediments from an east-coast Australian acid sulfate landscape. *Applied Clay Science*. 2009; 46(2):166–175. doi: [10.1016/j.clay.2009.07.020](https://doi.org/10.1016/j.clay.2009.07.020)
58. Sleutel S, Vandenbruwane J, De Schrijver A, Wuyts K, Moeskops B, Verheyen K, et al. Patterns of dissolved organic carbon and nitrogen fluxes in deciduous and coniferous forests under historic high nitrogen deposition. *Biogeosciences*. 2009; 6(12):2743–2758. doi: [10.5194/bg-6-2743-2009](https://doi.org/10.5194/bg-6-2743-2009)
59. Nayyar A, Hamel C, Lafond G, Gossen BD, Hanson K, Germida J. Soil microbial quality associated with yield reduction in continuous-pea. *Applied Soil Ecology*. 2009; 43(1):115–121. doi: [10.1016/j.apsoil.2009.06.008](https://doi.org/10.1016/j.apsoil.2009.06.008)
60. Ridgway K, Dunn J, Wilkin J. Ocean interpolation by four-dimensional weighted least squares-application to the waters around Australasia. *Journal of Atmospheric and Oceanic Technology*. 2002; 19(9):1357–1375. doi: [10.1175/1520-0426\(2002\)019%3C1357:OIBFDW%3E2.0.CO;2](https://doi.org/10.1175/1520-0426(2002)019%3C1357:OIBFDW%3E2.0.CO;2)
61. Gelfand AE, Ecker MD, Christiansen C, McLaughlin TJ, Soumerai SB. Conditional categorical response models with application to treatment of acute myocardial infarction. *Applied Statistics*. 2000; 49(2):171–186.
62. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*. 2002; 64(4):583–639. doi: [10.1111/1467-9868.00353](https://doi.org/10.1111/1467-9868.00353)
63. Akaike H. Information theory and an extension of the maximum likelihood principle. In: Petrox B, Caski F, editors. *Second International Symposium on Information Theory*. Akademia Kiado, Budapest, Hungary; 1973. p. 267–281.
64. Hoeting JA, Davis RA, Merton AA, Thompson SE. Model Selection for Geostatistical Models. *Ecological Applications*. 2006; 16(1):87–98. doi: [10.1890/04-0576](https://doi.org/10.1890/04-0576) PMID: [16705963](https://pubmed.ncbi.nlm.nih.gov/16705963/)
65. Schwarz G. Estimating the dimension of a model. *The Annals of Statistics*. 1978; 6(2):461–464. doi: [10.1214/aos/1176344136](https://doi.org/10.1214/aos/1176344136)
66. Donald MR, Alston CL, Young RR, Mengersen KL. A Bayesian analysis of an agricultural field trial with three spatial dimensions. *Computational Statistics & Data Analysis*. 2011; 12:3320–3332. doi: [10.1016/j.csda.2011.06.022](https://doi.org/10.1016/j.csda.2011.06.022)