# State-space modeling of the dynamics of temporal plant cover using visually determined class data (#46532)



First submission

## Guidance from your Editor

Please submit by **23 Mar 2020** for the benefit of the authors (and your $200 publishing discount) .

**Structure and Criteria**
Please read the 'Structure and Criteria' page for general guidance.

**Raw data check**
Review the raw data.

**Image check**
Check that figures and images have not been inappropriately manipulated.

Privacy reminder: If uploading an annotated PDF, remove identifiable information to remain anonymous.

**Files**
Download and review all files from the materials page.

1 Latex file(s)

## Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this PDF and upload it as part of your review

When ready submit online.

## Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your guidance page.

### BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to PeerJ standards, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (see PeerJ policy).

### EXPERIMENTAL DESIGN

- Original primary research within Scope of the journal.
- Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

### VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. Negative/inconclusive results accepted. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- All underlying data have been provided; they are robust, statistically sound, & controlled.
- Speculation is welcome, but should be identified as such.
- Conclusions are well stated, linked to original research question & limited to supporting results.

# Standout reviewing tips

The best reviewers use these techniques

| **Tip** | ***Example*** |
| --- | --- |
| **Support criticisms with evidence from the text or from other sources** | *Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.* |
| **Give specific suggestions on how to improve the manuscript** | *Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).* |
| **Comment on language and grammar issues** | *The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult.* |
| **Organize by importance of the issues, and number your points** | *1. Your most important issue*<br>*2. The next most important item*<br>*3. …*<br>*4. The least important points* |
| **Please provide constructive criticism, and avoid personal opinions** | *I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC* |
| **Comment on strengths (as well as weaknesses) of the manuscript** | *I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.* |

# State-space modeling of the dynamics of temporal plant cover using visually determined class data

**Hiroki Itô** Corresp. [1]

[1] Hokkaido Research Center, Forestry and Forest Products Research Institute, Sapporo, Japan

Corresponding Author: Hiroki Itô
Email address: abies.firma@gmail.com

A lot of vegetation-related data have been collected as an ordered plant cover class that can be determined visually. However, they are difficult to analyze numerically as they are in an ordinal scale and have uncertainty in their classification. The author constructed a state-space model to estimate unobserved plant cover proportions (ranging from zero to one) from such cover class data. The model assumed that the data were measured longitudinally, so that the autocorrelations in the time-series could be utilized to estimate the unobserved cover proportion. The model also assumed that the quadrats where the data were collected were arranged sequentially, so that the spatial autocorrelations also could be utilized to estimate the proportion. Assuming a beta distribution as the probability distribution of the cover proportion, the model was implemented with a regularized incomplete beta function, which is the cumulative density function of the beta distribution. A simulated dataset and a real dataset, with one-dimensional spatial structure and longitudinal survey, were fit to the model, and the parameters were estimated using the Markov chain Monte Carlo method. Then, the validity was examined using posterior predictive checks. As a result of the fitting, the Markov chain successfully converged to the stationary distribution, and the posterior predictive checks did not show large discrepancies. For the simulated data, the estimated values were close to the values used for the data generation. The estimated values for the real data also seemed to be reasonable. These results suggest that the proposed state-space model was able to successfully estimate the unobserved cover proportion. The present model is applicable to similar types of plant cover class data, and has the possibility to be expanded, for example, to incorporate a two-dimensional spatial structure and/or zero-inflation.

# State-space modeling of the dynamics of temporal plant cover using visually determined class data

**Hiroki Itô**[1]

[1]**Hokkaido Research Center, Forestry and Forest Products Research Institute, Sapporo 062-8516, Japan**

Corresponding author:
Hiroki Itô[1]

Email address: abies.firma@gmail.com

## ABSTRACT

A lot of vegetation-related data have been collected as an ordered plant cover class that can be determined visually. However, they are difficult to analyze numerically as they are in an ordinal scale and have uncertainty in their classification. The author constructed a state-space model to estimate unobserved plant cover proportions (ranging from zero to one) from such cover class data. The model assumed that the data were measured longitudinally, so that the autocorrelations in the time-series could be utilized to estimate the unobserved cover proportion. The model also assumed that the quadrats where the data were collected were arranged sequentially, so that the spatial autocorrelations also could be utilized to estimate the proportion. Assuming a beta distribution as the probability distribution of the cover proportion, the model was implemented with a regularized incomplete beta function, which is the cumulative density function of the beta distribution. A simulated dataset and a real dataset, with one-dimensional spatial structure and longitudinal survey, were fit to the model, and the parameters were estimated using the Markov chain Monte Carlo method. Then, the validity was examined using posterior predictive checks. As a result of the fitting, the Markov chain successfully converged to the stationary distribution, and the posterior predictive checks did not show large discrepancies. For the simulated data, the estimated values were close to the values used for the data generation. The estimated values for the real data also seemed to be reasonable. These results suggest that the proposed state-space model was able to successfully estimate the unobserved cover proportion. The present model is applicable to similar types of plant cover class data, and has the possibility to be expanded, for example, to incorporate a two-dimensional spatial structure and/or zero-inflation.

## INTRODUCTION

There is a vast amount of historical data regarding plant abundance that were recorded as plant abundances in an ordered cover class, e.g., the Braun-Blanquet classification (Podani, 2006; Irvine and Rodhouse, 2010; Damgaard, 2014), much of which was determined visually. In many cases, such data are difficult to treat numerically; they are typically recorded in an “ordinal scale” so that standard arithmetic operations, such as addition or subtraction, are not applicable (Dale, 1989; Podani, 2006). In addition, the uncertainty derived from the visual classification of such data tends to be ignored in analyses.

However, attempts to estimate unobserved “true” plant cover (the proportion in a unit area) from the ordered class data have been developed along with progress in statistical methods in the field of ecology (Irvine and Rodhouse, 2010; Damgaard, 2014; Herpigny and Gosselin, 2015). Ordered class data are typically modeled using ordered logit (cumulative logit) models, but the interpretation of the models has been known to be rather complicated (Herpigny and Gosselin, 2015).

However, some attempts have been made to model the plant cover proportion, assuming that this proportion follows the beta distribution (Chen et al., 2008; Damgaard, 2014). For example, Damgaard (2014) modeled the plant cover class as determined visually using the incomplete beta function based on the beta distributions of the plant cover, and then Herpigny and Gosselin (2015) incorporated zero-inflation, accounting for the excess zeros in the class data, into the model.

In recent decades, state-space models have been applied to many subjects in ecology, such as population dynamics (Clark and Bjørnstad, 2004; Iijima et al., 2013), metapopulation dynamics (Harrison et al., 2011), and tree growth (Shimatani and Kubota, 2011; Hiura et al., 2019). The state-space model consists of two types of sub-model, the observation model and the system model; the former describes the relationships that exist between the observed data and unobserved systems, and the latter describes the processes in the unobserved latent system, such as the temporal changes of the system. Notably, this class of models has a hierarchical structure and can explicitly describe the observation processes and the latent system processes separately.

The state-space model has also been used for dealing with time-series pin-point cover data (Damgaard, 2012). However, cover class data treated in the present study typically have less information than that of the pin-point cover data. Few studies have applied state-space modeling to cover class data, but if the class data were collected longitudinally, we would be able to utilize the information; i.e., the value of the latent state at a survey occasion should be similar to those at temporally adjacent occasions. In addition, if the class data were surveyed in quadrats that are arranged sequentially, we could also utilize information from the spatial autocorrelation.

In this study, a state-space model was constructed to estimate the unobserved proportion of plant cover from ordered class data using the incomplete beta function, combining information from temporal and spatial autocorrelations. This type of model would help to utilize visually determined plant cover data with temporal and spatial autocorrelation.

# METHODS

## Statistical model

### *Model basis*

Beta distribution has been used to describe statistical variations in plant cover, because the distribution has a boundary from zero to one, and because it can describe various shapes (Chen et al., 2008; Eskelson et al., 2011; Damgaard, 2012, 2013, 2014; Herpigny and Gosselin, 2015; Wright et al., 2017; Takarabe and Iijima, 2019). In this approach, the proportion of cover $p$ $(0 \leq p \leq 1)$ is assumed to follow the beta distribution:

$$p \sim \mathrm{Beta}(\alpha, \beta),$$

where $\alpha$ $(> 0)$ and $\beta$ $(> 0)$ are the parameters. Another parameterization using the mean of the proportion $\mu$ $(0 < \mu < 1)$ as a parameter is available,

$$p \sim \mathrm{Beta}\left(\frac{\mu}{\delta} - \mu, \frac{(1-\mu)(1-\delta)}{\delta}\right),$$

where $\delta$ $(0 < \delta < 1)$ denotes the intra-quadrat correlation of the plant distribution Damgaard (2012, 2013, 2014). However, the parameter $\delta$ can be regarded as that to control the variance or uncertainty of the observation of the cover proportion when $\delta$ is rather smaller (Fig. 1). In the case of $\mu = 0.5$, the distribution stays unimodal when $\delta$ is smaller than 1/3. In contrast, when $\delta$ becomes larger, the distribution tends to become bimodal (zero and one), or unimodal at zero or one (depending on $\mu$). In this parameterization set, the variance was given as $\delta\mu(1-\mu)$.

The probability that $p$ falls between $x_0$ and $x_1$ $(0 \leq x_0, x_1 \leq 1$, and $x_0 < x_1)$ can be described as follows:

$$\Pr(x_0 < p < x_1 | \alpha, \beta) = B(x_1, \alpha, \beta) - B(x_0, \alpha, \beta),$$

or

$$\Pr(x_0 < p < x_1 | \mu, \delta) = B\left(x_1, \frac{\mu}{\delta} - \mu, \frac{(1-\mu)(1-\delta)}{\delta}\right) - B\left(x_0, \frac{\mu}{\delta} - \mu, \frac{(1-\mu)(1-\delta)}{\delta}\right),$$

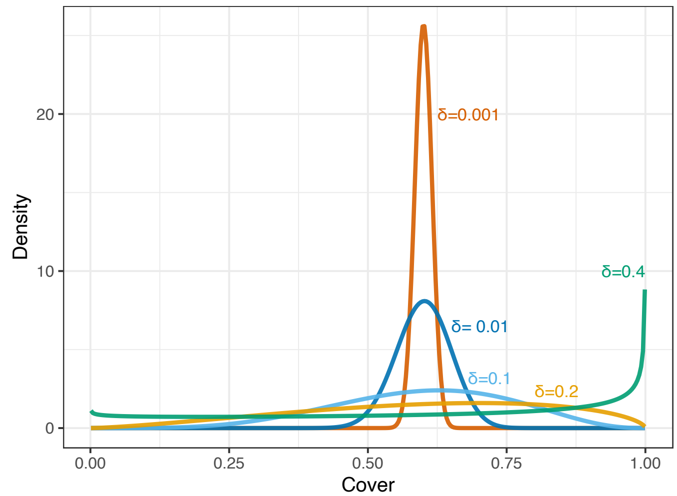


**Figure 1.** Probability densities of beta distributions corresponding to the cover proportion with a fixed mean (=0.6) and varying the value of $\delta$ (0.01, 0.1, 0.2, and 0.4).

where $B(x, \alpha, \beta)$ is the cumulative density function of beta distribution, identical to the regularized incomplete beta function $I_x(\alpha, \beta)$. Note that $B(0, \alpha, \beta) = 0$ and $B(1, \alpha, \beta) = 1$.

Fig. 2 shows changes in probabilities that each cover class is chosen according to the true proportion. When the value of $\delta$ is small, the chosen cover class would be rather precise. In contrast, the larger $\delta$ becomes, classes other than the correct one tend to be chosen more frequently.

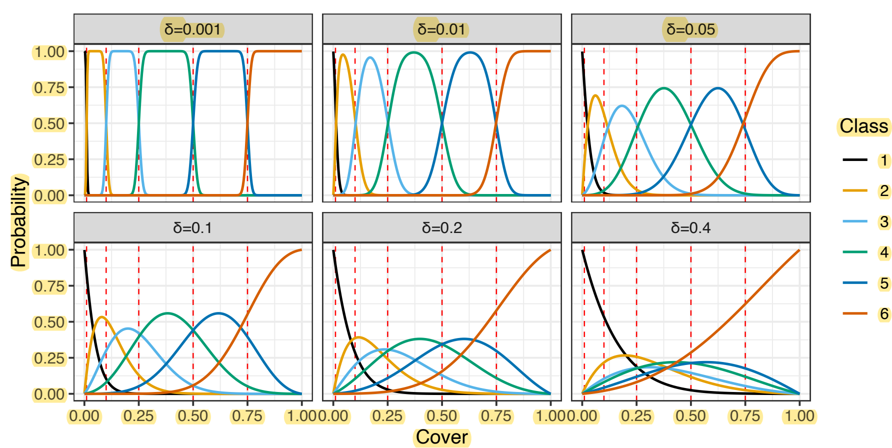


**Figure 2.** Relationships between the plant cover proportion and the probability that the proportion is classified to each class, with varying values of $\delta$ (0.01, 0.1, 0.2, and 0.4). Red dashed lines show cut points (inner boundaries of the classes).

### *State-space model*

**Observation model** Assume that surveys on plant cover were conducted $N_T$ times in $N_Q$ quadrats. In the present study, quadrats were assumed to be arranged on a line.

Cover class, $Y$, was defined as six classes corresponding with the proportion of plant cover as follows:

$$Y = \begin{cases} 1 & \text{if } 0 \le \text{cover} \le 0.01, \\ 2 & \text{if } 0.01 < \text{cover} \le 0.1, \\ 3 & \text{if } 0.1 < \text{cover} \le 0.25, \\ 4 & \text{if } 0.25 < \text{cover} \le 0.5, \\ 5 & \text{if } 0.5 < \text{cover} \le 0.75, \\ 6 & \text{if } 0.75 < \text{cover} \le 1. \end{cases}$$

In reality, $Y$ would be typically determined with visual measurements.

In this study, estimating the cover proportion of a particular species was the primary purpose rather than the presence/absence of the species. Thus, for simplicity, the model did not distinguish the absence of the species (or more precisely, the absence of the detection of the species) from the smallest plant cover class. When the plant species richness of the area is the study purpose, both states should be modeled separately. In those cases, incorporating zero-inflation (Herpigny and Gosselin, 2015) and the correction of false-negative errors (Chen et al., 2009, 2013) into the model is required.

The relationship between the observation $Y_{t,q}$, cover class at time $t \in \{1,2,\ldots,N_\mathrm{T}\}$ and in quadrat $q \in \{1,2,\ldots,N_\mathrm{Q}\}$, and $p_{t,q}$, the proportion of plant cover at time $t$ in quadrat $q$, was defined after Damgaard (2014) and Herpigny and Gosselin (2015) as follows:

$$\begin{aligned} \Pr(Y_{t,q} = Y \mid p_{t,q}, \delta) & \\ &= B\left(d_Y, \frac{p_{t,q}}{\delta} - p_{t,q}, \frac{(1-p_{t,q})(1-\delta)}{\delta}\right) \\ &\quad - B\left(d_{Y-1}, \frac{p_{t,q}}{\delta} - p_{t,q}, \frac{(1-p_{t,q})(1-\delta)}{\delta}\right). \end{aligned}$$

In this study, cut points $d_Y$ were defined as $\{0.01, 0.1, 0.25, 0.5, 0.75\}$ for $Y \in \{1,2,\ldots,5\}$, corresponding to the definition of $Y$. In addition, $d_0$ and $d_6$ were defined to be 0 and 1, respectively, so that $B(d_0, \alpha, \beta) = 0$ and $B(d_6, \alpha, \beta) = 1$.

The proportion of plant cover $p_{t,q}$ was defined by incorporating the latent state $\theta_t$ at time $t \in \{1,2,\ldots,N_\mathrm{T}\}$,

$$p_{t,q} = \mathrm{logit}^{-1}(\theta_t + r_{t,q}),$$

where $r_{t,q}$ denotes the spatial random effect incorporating spatial autocorrelation.

**System model** The latent state $\theta_t$ at time $t$ denotes the states related to the proportion of plant cover, and the expected proportion of plant cover at time $t$ for the overall plots, $\phi_t$, is given as

$$\phi_t = \mathrm{logit}^{-1}(\theta_t).$$

The transition of the latent state $\theta_t$ was defined using second-order differences with normal error as follows:

$$\theta_t \mid \theta_{t-1}, \theta_{t-2}, \sigma_\mathrm{T} \sim \mathrm{Normal}(2\theta_{t-1} - \theta_{t-2}, \sigma_\mathrm{T}^2) \quad \text{for } t \in \{3,4,\ldots,N_\mathrm{T}\},$$

where $\sigma_\mathrm{T}$ denotes the standard deviations.

Priors of the latent states at time $t \in \{1,2\}$ were defined as weakly informative (Gelman et al., 2017) but wide enough for the logit-scaled parameters (in case $\theta = 5$, $\phi$ is 0.99),

$$\theta_t \sim \mathrm{Normal}(0, 10) \quad \text{for} \quad t \in \{1,2\}.$$

The spatial random effect $r_{t,q}$ at time $t$ of quadrat $q$ was defined as follows:

$$r_{t,q} - r_{t,q-1} \mid \sigma_R \sim \text{Normal}(0, \sigma_R^2) \quad \text{for}\, q \in \{2,3,\ldots,N_Q\}$$
$$r_{t,1} \sim \text{Normal}(0, 10),$$

where $\sigma_R$ denotes the standard deviation among the spatial random effects. The value of the random effect $r_{t,q}$ was assumed to be affected by those of the adjacent quadrats. This formulation was equivalent to a process model of a state-space model with a first-order difference in the state changes. Then, the values were updated so that their sum should be zero for each survey time to avoid affecting the overall intercept and the identifiability of the model.

$$r_{t,q} \leftarrow r_{t,q} - \frac{1}{N_Q}\sum_{j=1}^{N_Q} r_{t,j}.$$

Priors for standard deviation parameters $\sigma_R$ and $\sigma_T$ were defined as weakly informative but wide enough for the changes in the logit-scaled parameters as follows:

$$\sigma \sim \text{HalfNormal}(0, 10).$$

## Application to simulated data

### *Generation of simulated data*

Assume that there were $N_T = 10$ quadrats that settled sequentially, and plant cover classes were surveyed for $N_Q = 15$ times in each quadrat. A simulated dataset was generated according to this assumption. In the simulated data, the parameter $\theta_t$, which denotes the latent state at time $t$, was generated following the relationship below:

$$\theta_1 = -6$$
$$\theta_t \sim \text{Normal}(\theta_{t-1} + 0.3, 0.5^2) \quad \text{for}\, t \in \{2,3,\ldots,N_T\}.$$

The latent state $\theta_t (t \in \{2,3,\ldots,N_T\})$ was randomly generated following the above normal distribution. Note that the first-order difference was used in this data generation, for simplicity, while the second-order difference was adopted in the model defined above. The spatial random effects $r_q(1 \in \{2,3,\ldots,N_Q\})$ were also generated randomly, with the assumption of following the above normal distribution. In this simulation, the spatial random effects were assumed to be invariant through time.

$$r_1 = 0$$
$$r_q \sim \text{Normal}(r_{q-1}, 0.5^2).$$

Proportions of plant cover $p$ were generated according to the model defined in the previous subsection:

$$p_{t,q} = \text{logit}^{-1}(\theta_t + r_q).$$

Then, the plant cover classes were generated with an uncertainty $\delta$. In this simulation, the value of $\delta$ was set to 0.05. The cover classes adopted in this simulated data were as follows: 1 (for proportion 0–0.01, including 0), 2 (0.01–0.1), 3 (0.1–0.25), 4 (0.25–0.5), 5 (0.5–0.75), and 6 (0.75–1). The generated data are shown in Fig. 3, and the cover proportion averaged for each time is shown in Fig. 5 (black curve) as well as the class data (black dots). The data generation code is available at the GitHub repository (`https://github.com/ito4303/ssmcover`).

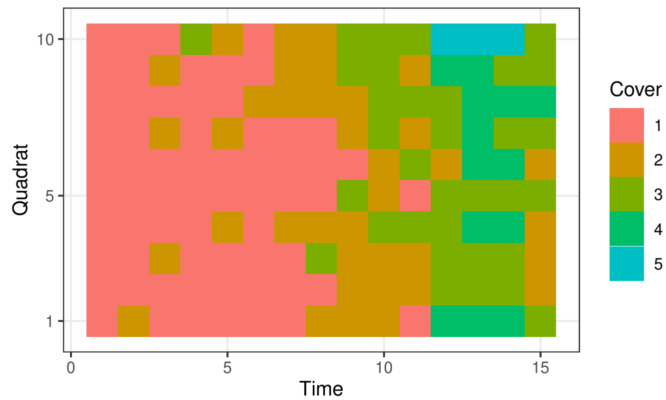


**Figure 3.** Simulated data that were generated for 10 sequential quadrats and 15 survey times. Classes denoted as follows, 1: 0–0.01 (including 0), 2: 0.01–0.1, 3: 0.1–0.25, 4: 0.25–0.5, 5: 0.5–0.75. Class 6 (0.75–1) was absent from these data.

### *Fitting to the model*

The generated data were fit to the Bayesian state-space model defined in the above subsection, and the posterior distributions of each parameter were estimated using the Markov chain Monte Carlo (MCMC) method. The model was implemented using Stan version 2.21.0 (Carpenter et al., 2017) with the re-parameterization of the model for the stability and efficiency of the Hamiltonian Monte Carlo algorithm, which was adopted in the Stan software. The Stan model code is also available at the GitHub repository. Posterior samples were drawn from 1,000 iterations after 1,000 warm-up (burn-in) iterations from each of 4 chains, and the posterior distributions of the parameters were estimated. Then, posterior predictive checks were conducted to evaluate the fitting to the model using the 'bayesplot' package (Gabry et al., 2019) in the statistical software R version 3.6.2 (R Core Team, 2019). In the posterior predictive check, the data drawn from the posterior predictive distribution that was calculated with the model were compared to the observed data (Gabry et al., 2019). If there are considerable discrepancies between them, it indicates that the model poorly explains the observed data.

## Application to real data

Real data to be fitted to the model were taken from the long-term vegetation monitoring following a catastrophic windthrow (Itô et al., 2018). The data were collected from the period of 1957 to 2017 in the headwater region of the Ishikari River, Hokkaido, northern Japan. Six plots were settled in the region in 1955. Quadrats sized two meters $\times$ two meters were settled sequentially, and the number was 15–25 for each plot. The visually determined cover classes were recorded for species that occurred in each quadrat.

The classes used in the surveys were as follows: + (proportion: 0–0.01, excluding zero), 1 (0.01–0.1), 2 (0.1–0.25), 3 (0.25–0.5), 4 (0.5–0.75), and 5 (0.75–1). Species that were not detected (i.e., the cover was 0) did not appear in the dataset. However, in the analysis, the notation was changed to be identical to the simulated data shown above for the sake of simplicity in numerical treatments so that the absence (more precisely, non-detection) was combined to the class 1 (0–0.01, including zero). The dataset is also available at the GitHub repository since it was published under the license CC BY 4.0.

From this dataset, cover classes of a species of dwarf bamboo, *Sasa senanensis*, in the shrub layer of a plot (No. 27) was used as the real data to be fit to the Bayesian state-space model. The data had a wide range in the cover class measurements and were suitable for model evaluation. The plot (No. 27) had 19 quadrats, and the survey was conducted 20 times (in 1957–1968, 1972, 1976, 1980, 1984, 1988, 2002, 2009, and 2017). Though the measurements were not conducted in all years during the period, the latent state could be estimated using the state-space model. Fig. 4 shows the changes in cover classes.

The posterior distributions were estimated using the MCMC method. Stan was also used for the estimation, and the posterior samples were drawn from 2,000 iterations after 2,000 warm-up (burn-in) iterations from each of 4 chains. Then, the posterior predictive checks were conducted using the 'bayesplot' package.

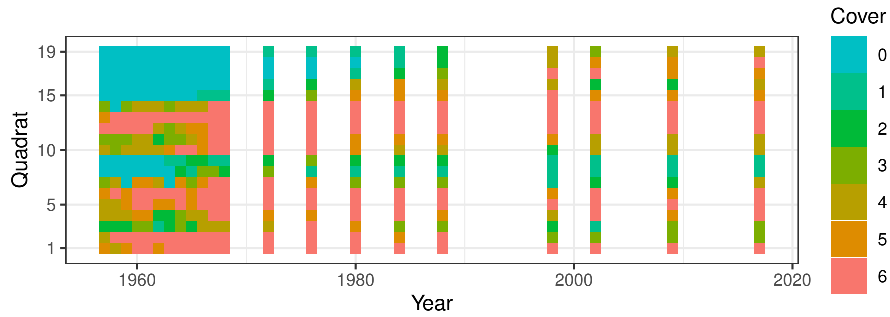


**Figure 4.** Changes in plant cover classes of *Sasa senanensis* following a catastrophic windthrow. Classes are denoted as follows, 0: 0, 1: 0–0.01 (excluding 0), 2: 0.01–0.1, 3: 0.1–0.25, 4: 0.25–0.5, 5: 0.5–0.75, 6 0.75–1. Class 0 and 1 were combined in the model fitting.

# RESULTS

## Simulated data

Gelman-Rubin statistics, $\hat{R}$, (Gelman and Rubin, 1992; Brooks and Gelman, 1998) were smaller than 1.1 for all the parameters, suggesting that the Markov chain successfully converged to the stationary distribution.

**Table 1.** Summary of the posteriors of the parameters $\delta$, $\sigma_{\mathrm{R}}$, and $\sigma_{\mathrm{T}}$ for the simulated data.

| Parameter | Mean | Percentile | | | $\hat{R}$ |
|---|---|---|---|---|---|
| | | 2.5% | 50% | 97.5% | |
| $\delta$ | 0.06 | 0.03 | 0.06 | 0.09 | 1.00 |
| $\sigma_{\mathrm{R}}$ | 0.34 | 0.08 | 0.34 | 0.62 | 1.01 |
| $\sigma_{\mathrm{T}}$ | 0.75 | 0.32 | 0.71 | 1.43 | 1.00 |

Table 1 shows the summary of the posteriors for the parameters $\delta$, $\sigma_{\mathrm{T}}$, and $\sigma_{\mathrm{R}}$. The posterior means (and 95% credible intervals) of these parameters were estimated as 0.06 (0.03–0.09) for $\delta$, 0.34 (0.08–0.62) for $\sigma_{\mathrm{R}}$, and 0.75 (0.32–1.43) for $\sigma_{\mathrm{T}}$ (Table 1). The values used for the data generation were 0.05, 0.5, and 0.5, respectively.

Fig. 5 shows the overall cover proportion ($\phi = \mathrm{logit}^{-1}(\theta)$) calculated from the posterior median (the red line) and the 95% credible intervals (the red region). The posterior predictive check showed no conflicts between the observed value and the predicted distribution for each time. Fig. 6 shows the result at time 15.

## Real data

$\hat{R}$ values were smaller than 1.1 for all parameters, and the Markov chain seemed to converge to the stationary distribution.

**Table 2.** Summary of the posteriors of the parameters $\delta$, $\sigma_{\mathrm{R}}$, and $\sigma_{\mathrm{T}}$ for the real data.

| Parameter | Mean | Percentile | | | $\hat{R}$ |
|---|---|---|---|---|---|
| | | 2.5% | 50% | 97.5% | |
| $\delta$ | 0.14 | 0.05 | 0.13 | 0.25 | 1.00 |
| $\sigma_{\mathrm{R}}$ | 3.21 | 2.51 | 3.20 | 3.91 | 1.00 |
| $\sigma_{\mathrm{T}}$ | 0.05 | 0.01 | 0.04 | 0.12 | 1.00 |

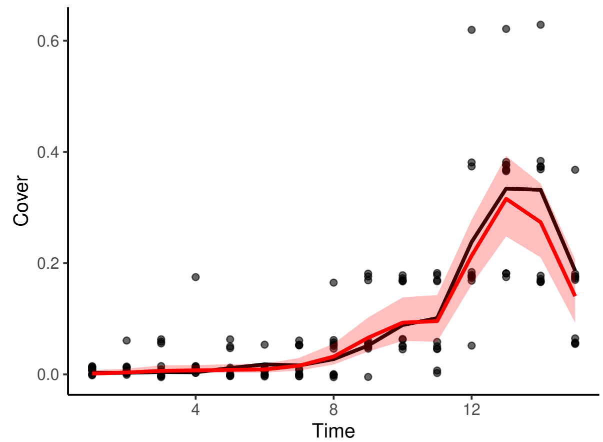


**Figure 5.** Estimated values of the cover proportion with simulated cover class data. Black curve: mean cover proportions that were used to generate cover class simulated data (averaged within each time). Black dots: cover classes in the simulated data (dots are on the medians of the classes and are jittered vertically). Red curve: estimated overall cover proportion. Red region: 95% credible intervals of the estimated cover proportion.

Table 2 shows a summary of the posteriors of the parameters $\delta$, $\sigma_{\mathrm{T}}$, and $\sigma_{\mathrm{R}}$. The posterior means (and 95% credible intervals) of these parameters were estimated at 0.14 (0.05–0.25) for $\delta$, 3.21 (2.51–3.91) for $\sigma_{\mathrm{R}}$, and 0.05 (0.01–0.12) for $\sigma_{\mathrm{T}}$ (Table 2).

Fig. 7 shows the estimated overall cover proportion for each year. The posterior predictive check showed no conflicts between the observed values and the predicted distribution for each year. Fig. 8 shows the result in 2017.

## DISCUSSION

For the simulated data, the posterior means (and medians) for the three major parameters did not differ much from the values that were used in the data generation (Table 1). The estimated curve of the plant cover proportion was similar to that which generated the simulated data (Fig. 5). The result of the posterior predictive check (Fig. 6) also suggests little discrepancy between the fitted model and the simulated data. The differences between the posterior means and the original values in parameters $\sigma_{\mathrm{R}}$ and $\sigma_{\mathrm{T}}$ may be at least partially due to variations in the randomly generated data. However, the slightly smaller value of $\sigma_{\mathrm{R}}$ may be attributable to small variations of cover classes among quadrats in the first several surveys (Figure 3). Over the period, the small value of $\theta_t$ overwhelmed the value of $r_t$. In addition, the small variations in the period would affect the narrow credible intervals of the posteriors (Fig. 5). Also, the assumption of the second-order differences in the system model rather than the first order used in the data generation may have affected the difference in $\sigma_{\mathrm{T}}$.

For the real data, the value of $\sigma_{\mathrm{R}}$ was large (Table 2), although the range was smaller than the prior ($\mathrm{HalfNormal}(0,10)$). This is likely due to the somewhat more considerable variation in the measurements among the adjacent quadrats (Fig. 4). The estimated curve of the plant cover proportion seems to be reasonable when comparing the measured cover classes (Fig. 7). The result of the posterior predictive check (Fig. 8) also suggests few discrepancies between the fitted model and the real data. The credible intervals were wider in the later period than those in the earlier period. This may be because of less information due to sparse survey intervals in the later period rather than variations in the measurements. The estimated posterior mean of $\delta$, or the uncertainty of the observation, seems understandable since the class data were visually determined (Fig. 2). However, the large variations in cover classes within a year may have increased the value.

The state-space modeling seems to have successfully estimated the changes in the latent states in the years that the surveys were not conducted. These results suggest that the present model is applicable to this type of plant cover class data.

Though the model proposed in this study is rather simple, more elaborate models can be constructed.

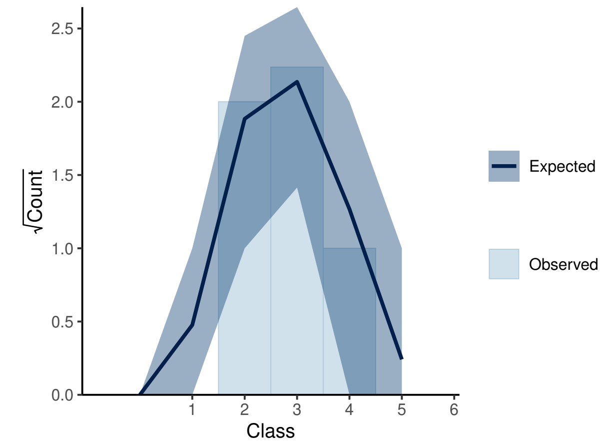


**Figure 6.** Rootogram showing the posterior predictive check for the Bayesian estimation of the simulated data. The bars show the measurements for each cover class observed at time 15; the curve shows the expected value of the posterior predictive distribution, and the dark region shows the 90% credible intervals at time 15.

For example, the one-dimensional structure of the present model can be expanded to two dimensions. To incorporate a two-dimensional spatial autocorrelation, conditional autoregressive (CAR) models can be utilized, and they are available in Stan (Joseph, 2016; Morris et al., 2019)

Another possible expansion is to incorporate zero-inflation. Herpigny and Gosselin (2015) has already provided modeling of plant cover classes with zero-inflation. When incorporating this, false-negative errors should be considered (Chen et al., 2009, 2013).

## CONCLUSION

State-space modeling for plant cover class data can successfully estimate the unobserved cover proportion by utilizing spatial and temporal autocorrelations that are contained within the data. The present model can be applicable to similar types of plant cover class data, and then can be expanded to incorporate two-dimensional field data and/or zero-inflation.

## ACKNOWLEDGEMENTS

I thank Dr. H. Iijima (Forestry and Forest Products Research Institute, Japan) for useful comments on a previous version of this paper.

## REFERENCES

Brooks, S. P. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76:1–32, ISSN: 1548-7660, DOI: 10.18637/jss.v076.i01.

Chen, G., Kéry, M., Plattner, M., Ma, K., and Gardner, B. (2013). Imperfect detection is the rule rather than the exception in plant distribution studies. *Journal of Ecology*, 101:183–191, DOI: 10.1111/1365-2745.12021.

Chen, G., Kéry, M., Zhang, J., and Ma, K. (2009). Factors affecting detection probability in plant distribution studies. *Journal of Ecology*, 97:1383–1389, DOI: 10.1111/j.1365-2745.2009.01560.x.

Chen, J., Shiyomi, M., Bonham, C. D., Yasuda, T., Hori, Y., and Yamamura, Y. (2008). Plant cover estimation based on the beta distribution in grassland vegetation. *Ecological Research*, 23:813–819, ISSN: 1440-1703, DOI: 10.1007/s11284-007-0443-3.

Clark, J. S. and Bjørnstad, O. N. (2004). Population time series: process variability, observation errors, missing values, lags, and hidden states. *Ecology*, 85:3140–3150, DOI: 10.1890/03-0520.

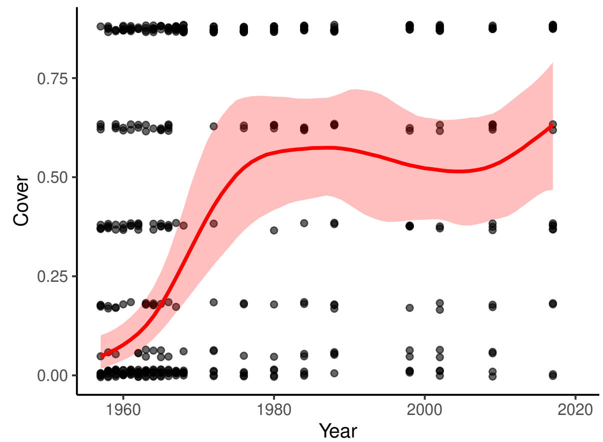


**Figure 7.** The real data and the estimated values of the plant cover proportion. Black curve: cover proportion to generate cover classes, averaged among quadrats. Black dots: generated cover classes in the data (dots are on the medians of the classes and are jittered vertically). Red curve: estimated overall cover proportion. Red region: 95% credible intervals of the estimated cover proportion.


Dale, M. B. (1989). Dissimilarity for partially ranked data and its application to cover-abundance data. *Vegetatio*, 82:1–12, ISSN: 1573-5052, DOI: 10.1007/BF00217977.

Damgaard, C. (2012). Trend analyses of hierarchical pin-point cover data. *Ecology*, 93:1269–1274, ISSN: 0012-9658, DOI: 10.1890/11-1499.1.

Damgaard, C. (2013). Hierarchical and spatially aggregated plant cover data. *Ecological Informatics*, 18:35–39, DOI: 10.1016/j.ecoinf.2013.06.001.

Damgaard, C. (2014). Estimating mean plant cover from different types of cover data: a coherent statistical framework. *Ecosphere*, 5:1–7, ISSN: 2150-8925, DOI: 10.1890/ES13-00300.1.

Eskelson, B. N. I., Madsen, L., Hagar, J. C., and Temesgen, H. (2011). Estimating riparian understory vegetation cover with beta regression and copula models. *Forest Science*, 57:212–221, DOI: 10.1093/forestscience/57.3.212.

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., and Gelman, A. (2019). Visualization in bayesian workflow. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182:389–402, DOI: 10.1111/rssa.12378.

Gelman, A. and Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7:457–472.

Gelman, A., Simpson, D., and Betancourt, M. (2017). The prior can often only be understood in the context of the likelihood. *Entropy*, 19(10):555, DOI: 10.3390/e19100555.

Harrison, P., Hanski, I., and Ovaskainen, O. (2011). Bayesian state-space modeling of metapopulation dynamics in the Glanville fritillary butterfly. *Ecological Monographs*, DOI: 10.1890/11-0192.1.

Herpigny, B. and Gosselin, F. (2015). Analyzing plant cover class data quantitatively: Customized zero-inflated cumulative beta distributions show promising results. *Ecological Informatics*, 26:18–26, DOI: 10.1016/j.ecoinf.2014.12.002.

Hiura, T., Go, S., and Iijima, H. (2019). Long-term forest dynamics in response to climate change in northern mixed forests in japan: A 38-year individual-based approach. *Forest Ecology and Management*, 449:117469, DOI: 10.1016/j.foreco.2019.117469.

Iijima, H., Nagaike, T., and Honda, T. (2013). Estimation of deer population dynamics using a Bayesian state-space model with multiple abundance indices. *Journal of Wildlife Management*, 77:1038–1047, DOI: 10.1002/jwmg.556.

Irvine, K. M. and Rodhouse, T. J. (2010). Power analysis for trend in ordinal cover classes: Implications for long-term vegetation monitoring. *Journal of Vegetation Science*, 21:1152–1161, DOI: 10.1111/j.1654-1103.2010.01214.x.

Itô, H., Nakanishi, A., Tsuyama, I., Seki, T., Iida, S., and Kawahara, T. (2018). Data on forest regeneration after catastrophic windthrow in the headwater region of the Ishikari River, Hokkaido, Japan. *Bulletin of FFPRI*, 17:265, DOI: 10.20756/ffpri.17.3_265. (in Japanese with English summary).

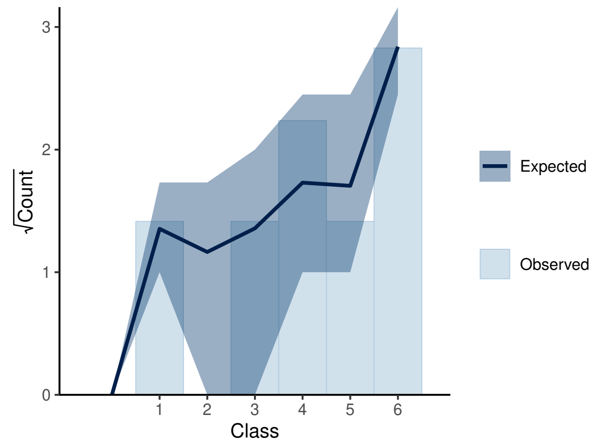


**Figure 8.** Rootogram showing the posterior predictive check for the Bayesian estimation of the real data. The bars show the measurements for each cover class observed in the year 2017; the curve shows the expected value of the posterior predictive distribution, and the dark region shows the 90% credible intervals in 2017.


Joseph, M. (2016). Exact sparse CAR models in Stan. `https://mc-stan.org/users/documentation/case-studies/mbjoseph-CARStan.html`. accessed 31 December 2019.

Morris, M., Wheeler-Martin, K., Simpson, D., Mooney, S. J., Gelman, A., and DiMaggio, C. (2019). Bayesian hierarchical spatial models: Implementing the Besag York Mollié model in stan. *Spatial and Spatio-temporal Epidemiology*, 31:100301, DOI: `10.1016/j.sste.2019.100301`.

Podani, J. (2006). Braun-Blanquet's legacy and data analysis in vegetation science. *Journal of Vegetation Science*, 17:113–117, DOI: `10.1111/j.1654-1103.2006.tb02429.x`.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, `https://www.R-project.org/`.

Shimatani, K. and Kubota, Y. (2011). The spatio-temporal forest patch dynamics inferred from the fine-scale synchronicity in growth chronology. *Journal of Vegetation Science*, 22:334–345, DOI: `10.1111/j.1654-1103.2010.01255.x`.

Takarabe, K. and Iijima, H. (2019). Contrasting effect of artificial grasslands on the intensity of deer browsing and debarking in forests. *Mammal Study*, 44:173 – 181, DOI: `10.3106/ms2018-0082`.

Wright, W. J., Irvine, K. M., Warren, J. M., and Barnett, J. K. (2017). Statistical design and analysis for plant cover studies with multiple sources of observation errors. *Methods in Ecology and Evolution*, 8:1832–1841, ISSN: `2041210X`, DOI: `10.1111/2041-210X.12825`.