



Improving Detection of Changepoints in Short and Noisy Time Series with Local Correlations: Connecting the Events in Pixel Neighbourhoods

Tuomas RAJALA , Petteri PACKALEN, Mari MYLLYMÄKI, and Annika KANGAS

Detecting changepoints in time series becomes difficult when the series are short and the observation variance is high. In the context of time series of environmental resource maps, it is often safe to assume that the abrupt events are spatially continuous, and so are the changepoints. We propose to utilise this assumption by means of hierarchical models where the changepoints are modelled using a spatial model. We demonstrate utility of the approach by constructing a Bayesian model based on the Potts model, with additional assumptions relevant to changepoint detection in national multi-source forest inventory maps. We discuss implementation issues and demonstrate the idea's performance using a simulation study. We then apply the model to forest resource maps in order to detect felling events.

Key Words: Changepoint detection; National Forest Inventory; Spatial correlation; Bayesian model; Potts model.

1. INTRODUCTION

Maps are often utilised as tools in science–policy interfaces such as REDD+, IPCC or IPBES to communicate state of environments to policy makers (Ojanen et al. 2021). Forest resources maps are also needed for local decision making, for instance regarding forest management, timber trade and planning of harvest operations (e.g. Kangas et al. 2018). Increased availability of remote sensing data has enabled large area maps of forest resources. In Finland, the first national-level thematic maps describing the status of forests were published already in the 1990's (Tomppo 1996). These maps were based on satellite

T. Rajala (✉) · P. Packalen · M. Myllymäki · A. Kangas, Natural Resources Institute Finland (Luke), Latokartanonkaari 9, 00791 Helsinki, Finland (E-mail: tuomas.rajala@luke.fi). P. Packalen (E-mail: petteri.packalen@luke.fi). M. Myllymäki (E-mail: mari.myllymaki@luke.fi). A. Kangas (E-mail: annika.kangas@luke.fi).

© 2023 The Author(s)

Journal of Agricultural, Biological, and Environmental Statistics, Volume 28, Number 3, Pages 564–590
<https://doi.org/10.1007/s13253-023-00546-1>

images and National Forest Inventory (NFI) plots. Currently similar maps are produced also based on laser scanning data (e.g. Nord-Larsen et al. 2017; Nilsson et al. 2017; Hauglin et al. 2021).

Besides the current forest resources, assessing the changes in them is of great importance (e.g. Hansen et al. 2013; Healey et al. 2018). Global Forest Watch, for instance, monitors changes in the forest canopy cover globally using satellite images. Changes observed in forests can be due to both natural disturbances (forest fire, wind damage, etc.) and human-induced disturbances (harvests), and their impacts can be either persistent or resilient (Kennedy et al. 2014). Besides the above-mentioned abrupt changes, also more subtle and continuous disturbances exist, such as those related to drought stress or insect infestation (Coops et al. 2020).

The assessment of changes has been made possible through the vast archives of satellite images freely accessible for researchers (Zhu 2017; Kennedy et al. 2018). In addition, time series of more detailed airborne data have been used to detect changes (Zhao et al. 2018). The change analyses vary in both temporal and spatial scale. Changes are detected in either tree level (Yu et al. 2004), pixel level of satellite images (Hansen et al. 2013) or forest stand level (Pitkänen et al. 2020). The extent of analysis varies from individual plots to global level (Duncanson and Dubayah 2018; Hansen et al. 2013). In temporal scale, the analyses vary from detection of changes between two time points to the analyses of trends over decades (e.g. Pitkänen et al. 2020; Katila et al. 2020; Wulder et al. 2020).

One important application of the change analysis is to analyse a time series of satellite images in order to detect the time point of a disturbance, and assess the resilience of the landscape after the disturbance. Kennedy et al. (2014) define five different shapes of change that can be observed: decreasing, increasing, jump, jump with resilience and cyclic. Moisen et al. (2016) also mentions flat, vee or down-up, inverted vee or up-down and double jump. The time series can also be analysed to detect the persistence of the effect of disturbance (e.g. Coops et al. 2020). Figure 1 shows four basic shapes alongside examples of these shapes from the Finnish growing stock volume series that we will be analysing. It is evident that the shorter the time series data available, the more challenging it is to detect more complex behaviour, such as double jump or vee and inverted vee, compared to simpler behaviour, i.e. increasing, decreasing or just a single jump.

A number of algorithms exist for analysing such time series, for instance the LandTrendr developed by Kennedy et al. (2010) to analyse Google Earth Engine time series (Kennedy et al. 2018). Usually, the time series analysis is carried out based on a time series of satellite images or satellite image composites. In the case of Finnish multi-source National Forest Inventory (NFI), a time series of interpreted forest resources maps from over two decades is available, and they provide an opportunity for detecting trends and disturbances directly in terms of various forest characteristics such as growing stock, biomass or tree species proportions (Katila et al. 2020); Fig. 1 illustrates the data series and Fig. 7 the spatial raster. In contrast to a typical satellite image analysis with monthly or weekly images, the number of these maps is quite low for the requirements of most time series analysis algorithms. In addition, the year-to-year variability in the resource maps can be high, posing additional challenges to the analysis. Many of the recent techniques are developed for series with

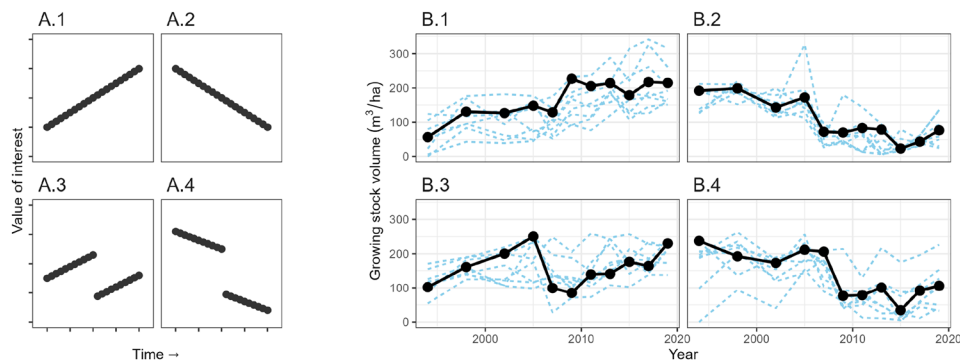


Figure 1. **A** Idealised series of data with high sampling frequency, low error variance, increasing (A.1) or decreasing trend (A.2), and series with changepoints (A.3, A.4). **B** Examples of Finnish National Forest Inventory growing stock volume series data at different locations, with low sampling frequency, high error variance, and with increasing (B.1) or decreasing (B.2) trends and with probable changepoints (B.3, B.4). Solid lines with dots show the values for each focal pixel, and dashed lines show the values in adjacent pixels.

hundreds of time points (e.g. Messer et al. 2018; Jewell et al. 2022) or moderate length series with high signal-to-noise ratio (Moisen et al. 2016).

Vast majority of the methods in the literature analyse the time series in each pixel separately. Our hypothesis is that when checking for a changepoint due to an event that is spatially continuous, the low number of highly variable observations in a time series can be augmented with information from the surrounding pixels. That is, with reference to Fig. 1, we suggest exploiting potential changepoints of the dashed lines to better understand the focal line. Previously, for instance Hughes et al. (2017) have addressed change for a map that is spatially segmented to patches for utilising the neighbourhood in the analysis. Others have used neighbouring pixels to stabilise focal time series (Hamunyela et al. 2016) or to construct predictors (Sebald et al. 2021) before pixelwise changepoint analysis. But to the best of our knowledge, no studies exist that explicitly account for the potential changepoints in the spatial neighbourhood of focal location. The goal of this study is to demonstrate how the neighbourhood information can be modelled and to assess the usefulness of such information.

The work is organised as follows. In Sect. 2, we introduce the NFI data. In Sect. 3, we formalise the changepoint detection problem in a probabilistic setting, introduce our proposal for incorporating spatial correlation and describe simulation trials for benchmarking the proposed method. In Sect. 4, we discuss the simulation trial results and apply the method to the NFI data. Section 5 summarises the work and discusses its limitations, wider applicability and future development.

2. MATERIAL

The Finnish multi-source National Forest Inventory thematic maps used in this study are from the MS-NFI-8 (Tomppo et al. 1998), MS-NFI-9 (Tomppo 2008), MS-NFI-2002 Mäkisara et al. (2001), MS-NFI-2005 (Tomppo et al. 2009), MS-NFI-2007 (Tomppo et al. 2012), MS-NFI-2009 (Tomppo et al. 2013), MS-NFI-2011 (Tomppo et al. 2014), MS-

NFI-2013 (Mäkisara et al. 2016), MS-NFI-2015 (Mäkisara et al. 2019), MS-NFI-2017 and MS-NFI-2019 (Mäkisara et al. 2022). NFI field plot data from 4- to 5-year time period were typically used in the multi-source NFI. During the MS-NFI-8 and MS-NFI-9, the NFI progressed by region and field plots was mostly from the same year as the satellite images. Since the tenth NFI (2004–2008), one-fifth of the clusters have been measured annually in major parts of Finland. Therefore, the NFI plots from the last five years have been employed in the multi-source NFI as training data since MS-NFI-2005.

In addition to NFI field plots, the data sources behind the multi-source NFI estimates are medium-resolution multi-spectral satellite images with pixel sizes of about 10–30 m (Landsat TM, ETM+ and OLI, Spot XS HRV, IRS-1 LISS-III, Sentinel-2A MSI) and digital map data of land use from the National Land Survey of Finland. The survey’s topographic database has been the digital map data source since MS-NFI-2005. The digital map data are used to separate forestry land (FRYL) from other land classes (an output is a FRYL mask) as well as to stratify FRYL and corresponding field plots into open bogs and fens, mineral soil and peatland strata for k-NN estimation purposes (Katila and Tomppo 2002; Tomppo 2008).

The resulting 11 maps have spatial resolution of 16 m and cover all of Finland. For this study, we consider the 11 map rasters in an example sub-region with size approximately 15×15 km and process them by mean-aggregating the images to a 32 m resolution. The resulting raster stack with resolution $480 \times 480 \times 11$ will be referred to as the (multi-source) NFI data. The region and the pixelwise mean over the 1994–2019 period of the NFI growing stock volume maps are illustrated in Fig. 7.

3. METHODS

3.1. PRELIMINARIES

Let $\mathcal{I} = \{i : i = 1, \dots, N\}$ denote a set of units with representative spatial locations x_i where the variable of interest, y_{il} , is available at $T > 0$ time points t_l , $l = 1, \dots, T$. For our NFI example, \mathcal{I} represent $32 \text{ m} \times 32 \text{ m}$ square cells in a regular grid covering a square sub-region of Finland in projection ETRS-TM35FIN (Fig. 7), $N = 230400$, y_{il} is the predicted growing stock volume (m^3/ha) in each cell i , and the $T = 11$ time points are non-equidistant $t_l = 1994, 1998, 2002, 2005, 2007, 2009, 2011, 2013, 2015, 2017, 2019$.

In the multi-source NFI data, the pixel values are predicted forest characteristics (e.g. growing stock volume, biomass, height, diameter), not satellite image channel values or their modifications (e.g. NDVI). This approach can be partly seen as a process to remove differences between the used images due to, e.g. light and cloud conditions, and, more importantly, to facilitate direct interpretation of changes in terms of variables of interest. Motivated by the properties of the NFI time series, we make the following assumptions on the data:

1. Series are short, the number of time steps $T \ll 100$. In the examples of this paper, $T = 11$.

2. The unstructured variability due to, for example, observation noise is relatively high, coefficient of variation $CV \gg 5\%$.
3. The changepoint events are temporally correlated in spatial neighbourhoods.

Assumption 1. rules out many of the modern change detection algorithms. Assumption 2. reduces the efficiency of statistical changepoint detection, but, as will be demonstrated, assumption 3. can compensate and improve the efficiency to useful levels.

For the purposes of demonstrating the approach in the context of NFI data ,we will set two further simplifying assumptions:

4. There is at most 1 changepoint event.
5. The events represent harvest events and are therefore negative shifts.

Assumption 4. can be relaxed if longer series are available, see e.g. [Killick et al. \(2012\)](#). Assumption 5. is not strictly necessary for the developed model, but it improves identifiability.

3.2. PROBABILISTIC MODEL FOR TIME SERIES CHANGEPOINTS

We formalise a changepoint in the context of probabilistic modelling. We will consider time series observed on a fixed set of time points $t_1 < \dots < t_T$. Without great loss of generality, let us consider a parametric model connecting the variable of interest at location (pixel) i at time t_l , $y_{il} \in \mathbb{R}$, to its mean, μ_{il} , $l = 1, \dots, T$, by assuming for the vector of values $\mathbf{y}_i = (y_{i1}, \dots, y_{iT}) \in \mathbb{R}^T$ the relationship

$$\mathbf{y}_i = \boldsymbol{\mu}_i + \boldsymbol{\epsilon}_i \quad \boldsymbol{\epsilon}_i \sim F_{\boldsymbol{\epsilon}}(\boldsymbol{\Sigma}_i) \tag{1}$$

with a mean vector $\boldsymbol{\mu}_i = \boldsymbol{\mu}_i(\boldsymbol{\theta}_i) = (\mu_{i1}(\boldsymbol{\theta}_i), \dots, \mu_{iT}(\boldsymbol{\theta}_i))$ depending on the vector of parameters $\boldsymbol{\theta}_i$, and some distribution $F_{\boldsymbol{\epsilon}}$ with parameters $\boldsymbol{\Sigma}_i$ for the error terms $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{iT})$.

Without change events, the parameters $\boldsymbol{\theta}$, $\boldsymbol{\Sigma}$ are constant in time. A time point t_k , with index $k \in \{1, \dots, T - 1\}$, is defined to be a *changepoint* for \mathbf{y}_i if the distribution before (inclusive) and after the changepoint is different, that is, $\boldsymbol{\theta}_{i,t_l \leq t_k} \neq \boldsymbol{\theta}_{i,t_l > t_k}$ or $\boldsymbol{\Sigma}_{i,t_l \leq t_k} \neq \boldsymbol{\Sigma}_{i,t_l > t_k}$. If we consider only changes in the mean,

$$y_{il} = \begin{cases} \mu_{il}(\boldsymbol{\theta}_{i,t_l \leq t_k}) + \epsilon_{il}, & \text{when } l = 1, \dots, k \\ \mu_{il}(\boldsymbol{\theta}_{i,t_l > t_k}) + \epsilon_{il}, & \text{when } l = k + 1, \dots, T. \end{cases} \tag{2}$$

We reserve the special case $k = T$ to indicate “no changepoint”. Note that auto-correlation can still exist between the errors over the changepoint. A changepoint could affect also the variance parameters; for example, switching satellite data to higher quality could reduce the variance considerably for following observations.

For applications such as the NFI harvest detection, we are content to consider the linear model, with the mean function

$$\mu_{il} = \mathbf{X}_{il} \boldsymbol{\theta}_{it_l} \tag{3}$$

where relevant trend components such as environmental factors, nonparametric spline basis coefficients, or, particularly, (polynomials of) time variable t_l , are included in the row vector X_{il} . Then we can formulate the parameters and their change as

$$\theta_{it} = \theta_i \quad t \leq t_k, \quad \theta_{it} := \theta_i + \delta_i \quad t > t_k \tag{4}$$

where δ_i quantifies the change in the parameters of the mean function at location i .

In this paper, we consider the simple independent and identically distributed Gaussian error model with a constant variance σ^2 , together with a linear trend in time with $X_{il} = [1 \ t_l]$. We stack X_{il} rowwise to create the regressor matrix X_i . The vector θ_i contains the parameters corresponding to the constant level and the slope in time. We will further restrict ourselves to studying changes only in the intercept of the mean function; thus, $\delta_i = [\delta_i \ 0]^\top$. The observation model for each location $i = 1, \dots, N$ then reduces to the T -dimensional normal distribution,

$$y_i \mid \theta_i, \delta_i, k_i, \sigma \sim \text{Normal}_T \left(X_i \theta_i + \delta_i \psi^k, \sigma^2 \mathbf{I}_{T \times T} \right) \tag{5}$$

where ψ^k is the vector with k 0's followed by $(T - k)$ 1's and $\mathbf{I}_{T \times T}$ is the identity matrix. Additional properties such as auto-regressive errors and higher-order trends can be modelled, when data permit, by modifying Σ and X_i , respectively. Extensions to multiple changepoints are more involved, but roughly follow the pattern in Eq (2) with additional δ_i 's in Eq (4).

We will estimate the parameters within Bayesian framework (Gelman et al. 2014) in order to quantify uncertainty of the estimates, not always possible in special algorithms based on optimisation formulations of the problem (Jewell et al. 2022). The optimisation formulation methods are usually implicitly based on a probabilistic models but trade quantifiable uncertainty for computational efficiency. In this paper, we aim to show utility of the neighbourhood correlation, and sacrifice computational efficiency for methodological transparency.

Using a hierarchical modelling approach, we can also incorporate external information via the prior distributions, usefulness of which is evident from being able to set a truncated Gaussian prior for the δ_i s according to our desire to discover only negative changes (Assumption 5.). At this stage of the model specification, we set the priors

$$\theta_i \sim \text{Normal}_2(\mathbf{m}_\theta, \mathbf{S}_\theta) \tag{6}$$

$$\delta_i \sim \text{trunc-Normal}(m_\delta, s_\delta^2, l_\delta, u_\delta) \tag{7}$$

$$\sigma^2 \sim \text{Inv-Gamma}(a, b) \tag{8}$$

where, if not otherwise stated, we set the tuning parameters $\mathbf{m}_\theta = \mathbf{0}$, $\mathbf{S}_\theta = 10^5 \mathbf{I}_{2 \times 2}$, $m_\delta = 0$, $s_\delta^2 = 10^5$, $l_\delta = -\infty$, $u_\delta = 0$, $a = 5$, $b = 5000$.

3.3. INCLUDING NEIGHBOURHOOD INFORMATION

We introduce correlation between the series of neighbouring locations via hierarchical modelling of the changes. Smoothing the image, as, for example, done by Hamunyela et al.

(2016), is akin to connecting θ_i to neighbouring θ_j , and we could alternatively achieve this in our model by setting a spatially correlated prior on the θ_i s. But instead of (or in addition to) connecting the trends, we suggest connecting the actual *changepoints* variables $\{k_i\}$.

Our approach is similar to that of panel data analysed by [Bardwell et al. \(2019\)](#). Consider that the area represented by each pixel $i \in \mathcal{I}$ is small enough so that each event of interest, say final felling between the years t_k and t_{k+1} , affects a group of pixels (Assumption 3.). With the restriction of at most one event per pixel, the object of interest is then the latent image, say $\mathcal{I}^k = \{k_i : i \in \mathcal{I}\}$, where each pixel describes the presence or absence of a changepoint event with values $k_i = 1, \dots, T$.

The goal of changepoint detection is to recover the image \mathcal{I}^k . Consider first assuming that there is no spatial dependency. Then we could add to the hierarchical model (5) independently across the pixels a categorical prior model $k_i \sim \text{Cat}(T; \boldsymbol{\pi})$. The unit simplex vector $\boldsymbol{\pi} = (\pi_1, \dots, \pi_T) \in [0, 1]^T$ defines prior probabilities for the change to occur at each time point, with the “no jump” probability given by π_T . Then in order to estimate \mathcal{I}^k with the prior assumption of spatial correlation, we can extend the categorical model to a multi-variate model with the Markovian form

$$p(k_i = k \mid \{k_j : j \sim i\}, \boldsymbol{\Gamma}, \boldsymbol{\pi}) \propto \pi_k \exp \left[\sum_{j \sim i} \gamma_{kk_j} \right], \quad (9)$$

where $j \sim i$ are neighbours of i and $\boldsymbol{\Gamma} = [\gamma_{kk'}]$ is a $T \times T$ -dimensional matrix of *interaction* parameters. Each component $\gamma_{kk'}$ describes the interaction of the changepoint of focal pixel i at the time point t_k with the changepoints of neighbouring pixels at time point $t_{k'}$. If $\gamma_{kk'} > 0$, then the probability of $k_i = k$ given $k_j = k'$ is adjusted upwards from the prior π_k . If $\boldsymbol{\Gamma} = 0$, neighbours do not add to the likelihood of a changepoint, but in any other case the model introduces correlation across the image \mathcal{I}^k and thus in the changepoint estimate of the pixel i and the changepoint estimates of its neighbours.

In this work, we consider the simple diagonal $\boldsymbol{\Gamma} \equiv \gamma \mathbf{I}_{T \times T}$ where $\gamma \in \mathbb{R}$, and, if not otherwise stated, neighbourhood of focal pixel consists of the 8 adjacent pixels. Then the conditional probability model (9) describes the well studied Potts model ([Wu 1982](#)). In general, the model (9) can be viewed as the auto-logistic model discussed by [Besag \(1974\)](#), and, for example, the $\boldsymbol{\Gamma}$ could be non-diagonal. An example of non-diagonal $\boldsymbol{\Gamma}$ is to consider $\gamma_{k,k+1} \neq 0$ to model a temporal lag component which might help identify events on a forest patch edges next to another patch with an event of the previous or following year. An obvious example for such a situation is a storm damage at the edge of a final felling carried out previous year, or a bark beetle damage following a storm damage.

Another extension is to consider different neighbourhood “kernel” shapes and size, say by adding a scalar weight w_{ij} into the sum in Eq. (9). The 8-adjacent-pixel kernel is the uniform kernel using max norm, and we will briefly discuss some alternatives to it. The shapes and sizes are illustrated in [Fig. 8](#). In order to have comparable results with a fixed γ , we scale the kernels so that the sum over all pixels that have nonzero weight in Eq. (9) is at maximum 8γ . Note that the neighbourhood together with $\boldsymbol{\Gamma}$ can be understood as geo-statistical co-variance model. In our example model, however, the “smoothing” occurs on the multi-nomial timing variables k instead of the response variables (growing stock

volume). As a result, the mean, the trend, and the event magnitude can vary freely over the pixels.

3.4. COMPUTATION

In the presented work, we consider the variables $\Gamma = \gamma \mathbf{I}_{T \times T}$ and $\boldsymbol{\pi}$ fixed and given as tuning parameters. We will discuss the parameter $\pi^1 := 1 - \pi_T$ and assume otherwise flat prior $\pi_k = \pi^1 / (T - 1)$, $k < T$. Inference on the Γ is difficult as the normalising constant of the likelihood is numerically intractable for any $\Gamma \neq 0$, and needs to be approximated with significant computation cost, even though recent development seems promising (Moore et al. 2020).

With γ , π^1 and the data \mathbf{y} fixed, the posterior distribution of the parameters $\{\boldsymbol{\theta}_i, \delta_i, k_i\}_1^N$ and σ can be sampled using a Gibbs sampler (Gelman et al. 2014). The details of the algorithm are provided in Appendix Sect. B. A more efficient algorithm would optimise the model with, for example, gradient descent or an EM approach, with required simplifications, but care must be taken where to cut corners: we tested a mean field variational approximation with deteriorated quality of the results. The purpose for this paper is not optimal computational efficiency; thus, the Gibbs sampler suffices.

Since the Potts prior theoretically connects each pixel with all other pixels (weakly, but still), simultaneous estimation of large rasters becomes difficult both in computational time and in memory requirements: The traces of the sampler should be kept at least for the N probability vectors, each of size T , to access the posteriors of k_i 's. Mixing of the chain is also typically slow with discrete latent variables. For convenience of running our prototype estimation algorithm, we opted for a mutually independent partitioned estimation. In this approach, the original raster is divided into $n_x \times n_y$ sub-rasters ("tiling") and the model is fitted independently in each of them. The pixels at the each sub-raster edges have lower neighbour counts than in the full raster, and to alleviate this, we overlap spatially adjacent sub-rasters by including a buffer ("buffering"). The buffer size was set to the number of pixels used in the neighbourhood definition. The model was then estimated independently on these buffered sub-rasters. When collating the posterior estimates on the full raster from the sub-raster estimates, the estimates for pixels in each sub-raster's buffer, those that belong to neighbouring sub-rasters, are excluded. This ensures that each pixel will be considered in only one sub-raster and will have a unique posterior estimate. Figure 9 illustrates the tiling and the buffering.

In our trivially parallelisable implementation, each sub-raster will have its own variance parameter σ , so as a side effect the tiling relaxes the homogeneity assumption (cf. Equation (5)) that is unlikely valid over large landscapes anyway. Figure 15 depicts Monte Carlo traces for key variables in three example cells of the NFI data and is provided as evidence for the convergence of our proof-of-concept algorithm implementation. All processing and computations were done using the R software (R Core Team 2022), and the code for reproducing the experiments is distributed through the first author's Github profile.

3.5. BENCHMARKING THE MODEL

We constructed a simulated dataset for which the change events were known and for which we could then assess the detection quality. The idea was to emulate the NFI application, with $T = 11$. An *in silico* map of change events, the k_i 's in \mathcal{I}^k , was created by simulating the Potts model using an extended number of times $T' = 29$ (using Gibbs sampling on a 625^2 grid with $\gamma = 2.0$, “external field constant” 2, 5 M iterations) and then identifying times 12–29 as 11. This was done to get the amount of pixels with a jump event, i.e. $k_i < 11$, to around 35%, a reasonably realistic estimate over a 25-year period. To emulate linear features (road, power line), a 1-pixel-wide line was added diagonally through the region. Illustration of the resulting \mathcal{I}^k is given in the top panel of Fig. 3.

To generate many series similar to the NFI data, a selection of 27 series with no harvest events was selected visually from the NFI dataset, and the linear regression coefficients (“pool”) and the median error standard deviation were recorded. Then for each pixel over the change map, a time series was generated by sampling the regression coefficients from the pool and adding Gaussian noise using the median standard deviation. For each pixel with a harvest event, $k_i < 11$, a negative shift was added to the mean after t_{k_i} . The series was rejected and re-sampled if the trend took the series mean below 0. The values of the noisy series were also truncated to be non-negative to imitate the growing stock volume data. We simulated three sets with $f = 30\%$, 50% and 70% harvest at the time of the jump, $\delta_i = -f \cdot (\theta_{i1} + \theta_{i2}t_{k_i})$, using 11 time steps $t = 0, 4, 8, 11, 13, 15, 17, 19, 21, 23, 25$. The three levels were chosen to provide a decent coverage of realistic detection tasks. Since the spatial effects are fairly localised, we expect datasets with mixed jump levels to have correspondingly mixed scores.

After the model was fitted to one of the synthetic dataset with known k_i 's, we predicted the harvest events using the maximal estimated jump timing $\hat{k}_i = \operatorname{argmax}_k P(k_i = k | \text{data})$ and for overall having a jump event $\hat{p}_i := \hat{P}(i \text{ has a jump}) = 1 - P(k_i = 11 | \text{data})$, thresholding at 0.5. We assessed the harvest detection quality by means of true positive rate (TP: fraction of pixels with $\hat{p}_i > .5$ out of pixels with $k_i < 11$) and the false positive rate (FP: fraction of pixels with $\hat{p}_i > .5$ out of pixels with $k_i = 11$).

For comparison, we ran pixel-wise changepoint detection using the CUSUM test (Brown et al. 1975, R function `structchange: :efp`) and an alternative Bayesian changepoint analysis (Barry and Hartigan 1993, “BCP”, R function `bcp: :bcp`). The latter was computed with time-wise prior-jump probabilities 0.1 and 0.5.

4. RESULTS

With the jump event timings known in the three synthetic datasets, we can explore the detection quality and their response to changing parameters. In the following, we note the 30, 50 and 70% reduction of target variable by “hard”, “medium” and “easy” case.

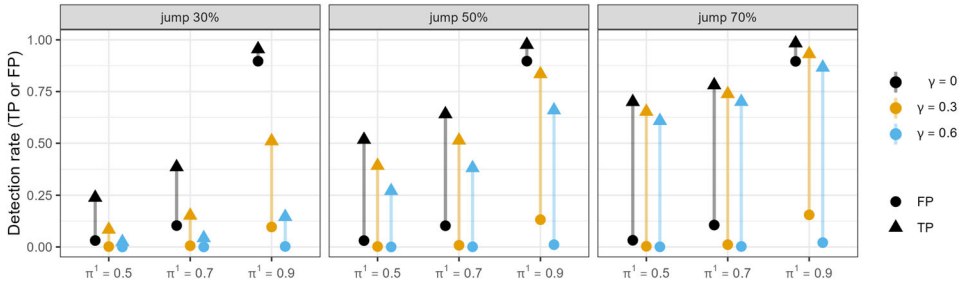


Figure 2. True positive (TP) and false positive (FP) rates in the simulation trial. Three change event sizes (30, 50, and 70% drop in the mean level of y_{it}), 9 combinations of tuning parameters. The model with $\gamma = 0$ refers to a model where jump timers k_i are not correlated in neighbourhoods. Estimated using a 8×8 tiling with 2 pixel buffers. Larger separation between TP (triangle) and FP (circle) is better .

4.1. SYNTHETIC BENCHMARKS

The TP and FP rates for all three datasets were estimated using a grid of tuning parameters, and the results for a chosen 9 combinations are illustrated in Fig. 2. The hard case with mere 30% reduction events turned out to be quite challenging: At best, we reached 50% TP rates with an elevated 10% FP rate ($\gamma = 0.3, \pi^1 = 0.9$). With no neighbourhood correlations, $\gamma = 0$, there is no good balance between TP and FP: Either the TP rate is low or the FP rate is high. With the easier medium and easy cases of size 50% and 70% reductions, the quality predictably improves. Most importantly, adding information from neighbourhoods by setting $\gamma > 0$ reduces the FP rate considerably. For example, in the medium case it is possible to have TP rate of around 80% with FP rate around 13%, and if we accept a TP rate reduction down to around 65%, the FP rate can drop to 1% (Fig. 2, middle panel).

Increasing γ produces conservative predictions and lowers the TP rate, and one might imagine this being emphasised, for example, at the border pixels of some “harvested” patch. Too large a γ should therefore be avoided. As expected, increasing π^1 increases TP with the usual trade-off of spurious false detection occurring more frequently. The model seems to be conservative in terms of the “jump-anywhere” prior probability π^1 , as the FP rates start increasing only after $\pi^1 > 0.5$.

Classical CUSUM testing had no power in any of the cases (0% TP and FP; excluding trend the rates were 14% TP and 9% FP for easy case, but the model is then clearly misspecified). The approach in BCP is similar to ours and reached a TP 46% with FP 4% when jump-prior was 0.1. Increasing the prior to 0.5 increased TP 79% but, as with $\gamma = 0$ case, also increased FP to 29%. Results for more difficult cases are illustrated in 10; the pattern is the same, with lower TP rates.

To understand the lacklustre classification performance in the hard case a bit more, we grouped the pixels with a jump event by “signal-to-noise ratio” $|\delta_i|/\sigma$ to below and above 2, which roughly group the pixels to those with low or high just-before-jump levels with respect to the variability (true $\sigma \approx 22$). With $\gamma = 0.3, \pi^1 = 0.9$, the TP rate in the below/low class is at best 36% but in the above/high class reaches 72%. This indicates that the model can be adjusted to be reasonably sensitive to 30% jump events while at the same time maintaining a low FP overall rate as long as the coefficient of variation at the time of the event is below

$30\%/2 = 15\%$. Setting the same coefficient of variation threshold for the medium and easy cases, the above/high class TP rates reach 98% and 99%, although more FPs occur at this high level $\pi^1 = 0.9$.

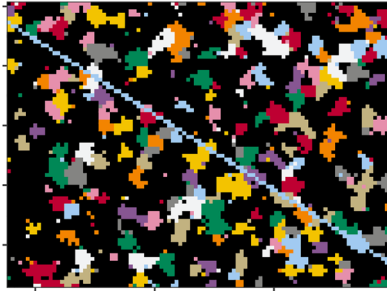
To check the biases from parallelising the calculation into sub-rasters, we computed the quality for a small sub-raster of size 62×63 with varying divisions and buffers. Overall, the bias seems to be small, few percentage point differences in the true and false positive rates (Fig. 12). If partitioning is required, then buffering should be included, or else the pixels on the edges can be expected to have a higher FP rate. The amount of edge pixels increases with more divisions, but buffering mere 2 pixels seems to remove the biases well for both 4×4 and 8×8 mosaics in our tests.

To visually compare the detection rates, in Fig. 3 we zoom in on the central 100×75 pixel sub-image and inspect the per-pixel posterior modes, jump probabilities and the misclassifications of the medium case. Neither $\gamma = 0$ nor $\gamma = 0.3$ estimates are able to fully detect each changepoint cluster, but most clusters are partially detected. Comparing $\gamma = 0$ and $\gamma = 0.3$ (left v right columns), the neighbourhood correlation seems to “fill in” holes, for both jump clusters and non-jump regions. The estimated jump probabilities are more polarised, and pixels within clusters are pulled towards majority. Spurious false positives are mostly removed, but the correlation also increases false negative count and their clustering. The linear “road” feature passing through the area is mere 1pxl wide, so pixels on it face 3 against 6 votes in the isotropic neighbourhoods, and as a result, the road is poorly identified.

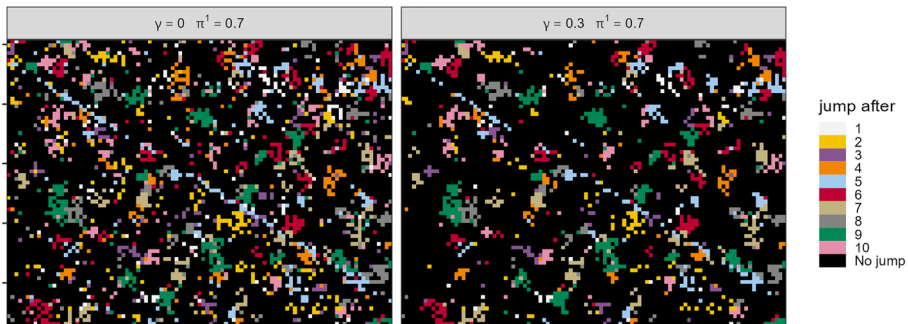
To explore what effects changing the neighbourhood definition has, we compared the medium difficulty case rates using a square neighbourhood, a circular neighbourhood and a radially weighted (squared exponential of cell centre distance) neighbourhood. We also altered the radius of the neighbourhood from 1 to 3 pixels away; see Fig. 8 for illustration. Recall that we scale the neighbourhoods to be cross-comparable for a given γ . The results across combinations of $\gamma = 0.3, 0.6$ and $\pi^1 = 0.7, 0.9$ (Fig. 11) indicate that kernel size is more important than the shape. Increasing radius from 1 to 3 pixels decreases TP rates (average change -10.8%). Circular neighbourhoods, with less diagonal neighbours, lead to slightly less conservative predictions (on average $+3.5\%$ to FP rate). Distance-weighting did not provide any benefits over the square. Altering the kernel might have more benefits when events are expected to be occur in larger patches, or when we want to target anisotropic shapes such as linear features.

The final observation we make is independent of the neighbourhood definition: the timing of the changepoint affected in our examples the detection power. Generally, changepoints at the extremes of a series are harder to detect because the increased before–after sample size imbalance. But because in our examples the magnitude of the jump was relative to the mean at the time of the jump, it follows that the detection rate depends on the trend of a series (Fig. 13). The trends in our examples were mostly positive (i.e. growth), so the absolute jump sizes of changepoints later in time became higher and were therefore easier to detect. For example, in the medium case (50% jumps) using model $\gamma = 0.0, \pi^1 = 0.5$, the true positive rates for changepoints at time point $k = 2$ were quite similar, 34% and 35%, for pixels with negative and positive trends, respectively. But for changepoints at $k = 8$, the

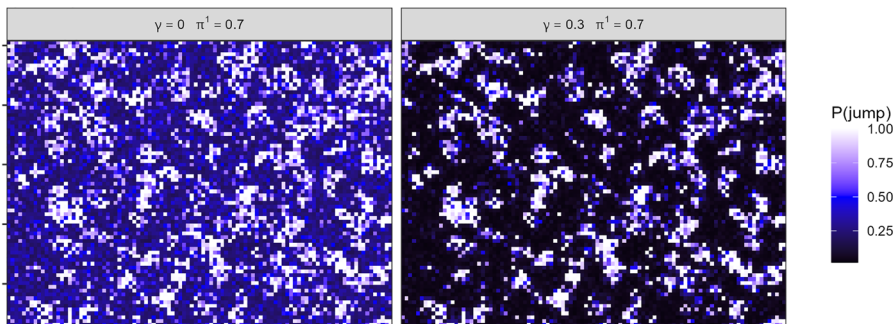
True changepoint pixels



Posterior mode



Posterior probability of a jump



FN and FP pixels

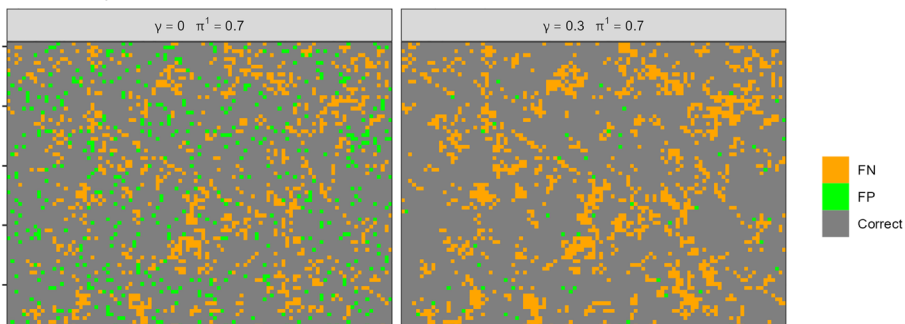


Figure 3. *Top row:* 100×75 pxl central region of the true changepoint image used in the simulation trial. Colours encode time steps, same encoding as on the 2nd row. *2nd row:* Posterior mode, i.e. predicted changepoints. *3rd row:* Posterior probability of a jump. *4th row:* False negatives (FN) and false positives (FP) of detecting a jump if using posterior probability threshold $P(\text{jump}) > 0.5$. Estimates are presented for the medium case of 50% jumps, and for two models with prior $\pi^1 = 0.7$ and either $\gamma = 0$ (left) or $\gamma = 0.3$ (right).

rates diverge and changepoints of series with positive trends were detected with rate 72% while for negative trend series the detection rate stayed closer to 30%.

4.2. NFI RESULTS

We then applied the modelling approach to part of the multi-source NFI data (cf. Figure 7) using 8×8 tiling with 2-pixel-wide buffers, and interaction parameters $\gamma = 0, 0.3, 0.6$ and jump priors $\pi^1 = 0.5, 0.7, 0.9$. Figure 15 illustrates the algorithm's convergence. The estimated fraction of pixels with a changepoint ranges with $\pi^1 = 0.5$ from 17% ($\gamma = 0.6$) to 27% ($\gamma = 0$), and with $\pi^1 = 0.9$ from 34% ($\gamma = 0.6$) to 97% ($\gamma = 0$). The last estimate is most likely a gross overestimate due to false positives. We will report in more detail the results using $\gamma = 0.6$ and $\pi^1 = 0.7$, for which the estimated rate was 20%, and based on the simulations, should have a low false positive rate.

Figure 4 depicts the estimated changepoints by pixel for tuning parameters $\gamma = 0.6$, $\pi^1 = 0.7$. The Finnish Forest Act regulates that a forest owner has to file the forest use declaration before a certain stand or stands are harvested. We have overlaid these declarations from the years 1997–2018 with the estimated changepoints in Fig. 4. It is important to note that some of the declared harvests are never implemented, so we do not expect them to represent “true positives”. But we do observe several detected clusters outside the planned regions. With changepoint years 1994 and 1998, the actual NFI input data predate the first declaration register year 1997, and in any year, the area delineated in the declaration is not always precise, but more study is needed to understand why large regions appear to have been harvested outside the register. In Finland, around 3 per cent of harvests have been done without a proper declaration, which is considered illegal harvest (even if the owners harvest their own stands); in Sweden, the level reached 10 per cent before monitoring started (Pitkänen et al. 2020; Sawyer et al. 2015).

For a visual check of the detection, we provide a close up of the multi-source NFI raster data together with aerial photographs in a small sub-region at two time points. (The sub-region is highlighted in Fig. 7.) Figure 5 shows the predicted change overlaid on the NFI growing stock data rasters from years 2011 and 2019, and Fig. 6 shows aerial photographs from closest available years 2012 and 2019. Visual inspection of the NFI data confirms a general trend of growth, with large final felling regions visible (e.g. lower right corner) and also well detected by the algorithm. The algorithm is also capable of detecting intermediate thinnings, such as the 2015–2018 event near the centre of the region. (We checked also $\gamma = 0.6$, $\pi^1 = 0.9$ with which the corresponding registry event is discovered nearly completely.) Some small events are harder to justify based on the aerial photographs (e.g. individual pixel 2011–2014 events in the top right corner), possibly due to misalignment of the aerial photo dates and the NFI dates, or a road construction appearing in some special way with the 32 m raster resolution.

Figure 14 provides scales of the estimated harvesting events by mapping the relative amount of growing stock removed at each changepoint event pixel. Most estimated events are large in scale, with reduction of growing stock estimated to be between -100% and -80% , but some smaller events of scale -40% to -20% are also predicted.

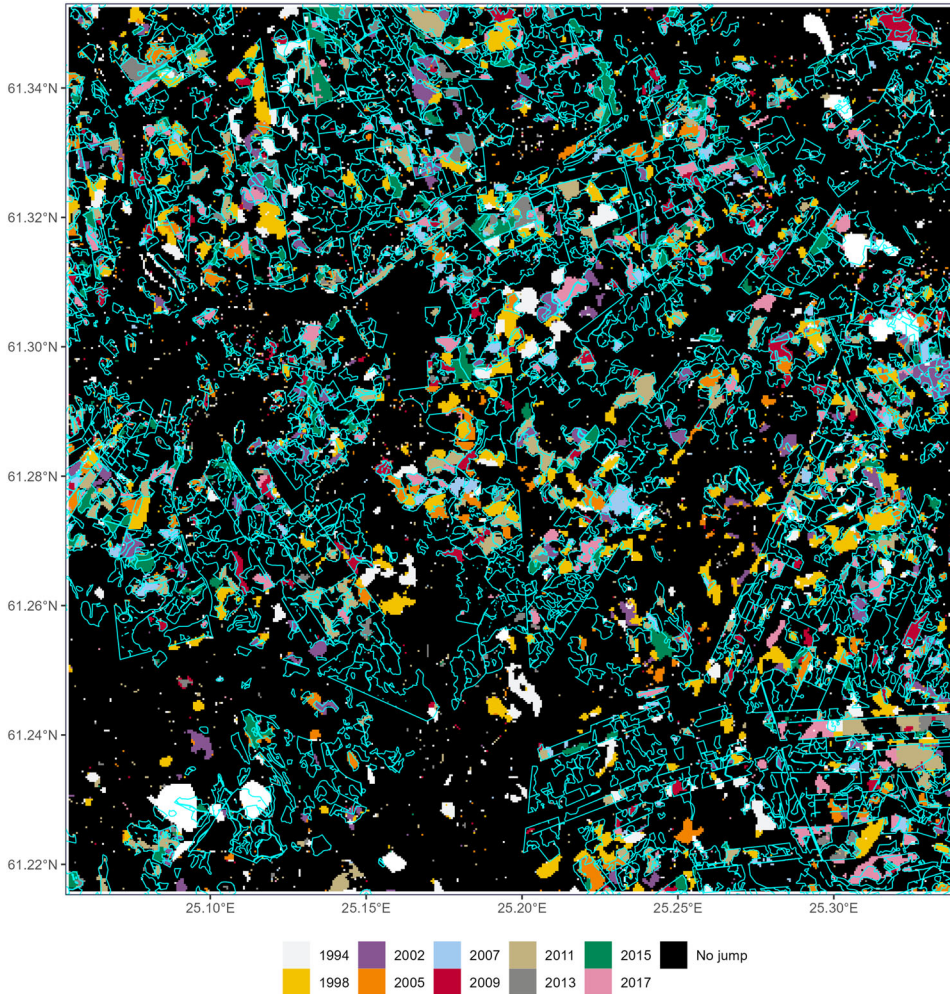


Figure 4. Changepoint estimates for NFI example data using tuning parameters $\gamma = 0.6$, $\pi^1 = 0.7$. Colours provide estimated years after which harvesting event occurred. Overlaid with cyan lines are all forest use declarations over the period 1997–2018.

With tuning parameters $\gamma = 0.6$, $\pi^1 = 0.7$, the standard deviation estimates $\hat{\sigma}$ ranged from 27 to 40, median 37, over the $8 \times 8 = 64$ sub-rasters in the computational tiling. The spatial pattern of the noise variability estimates was similar across all tuning parameters, with a group of lower values in the western and north-western tiles.

5. DISCUSSION

We have introduced a novel approach to incorporate spatial correlation into (probabilistic) time series changepoint detection in the many series, few-time-point setting often present in such applications as national forest resource mapping. We benchmark a proof-of-concept instance of the proposed model on simulated datasets, and the results clearly show the benefits of incorporating spatial information, especially when compared to non-spatial

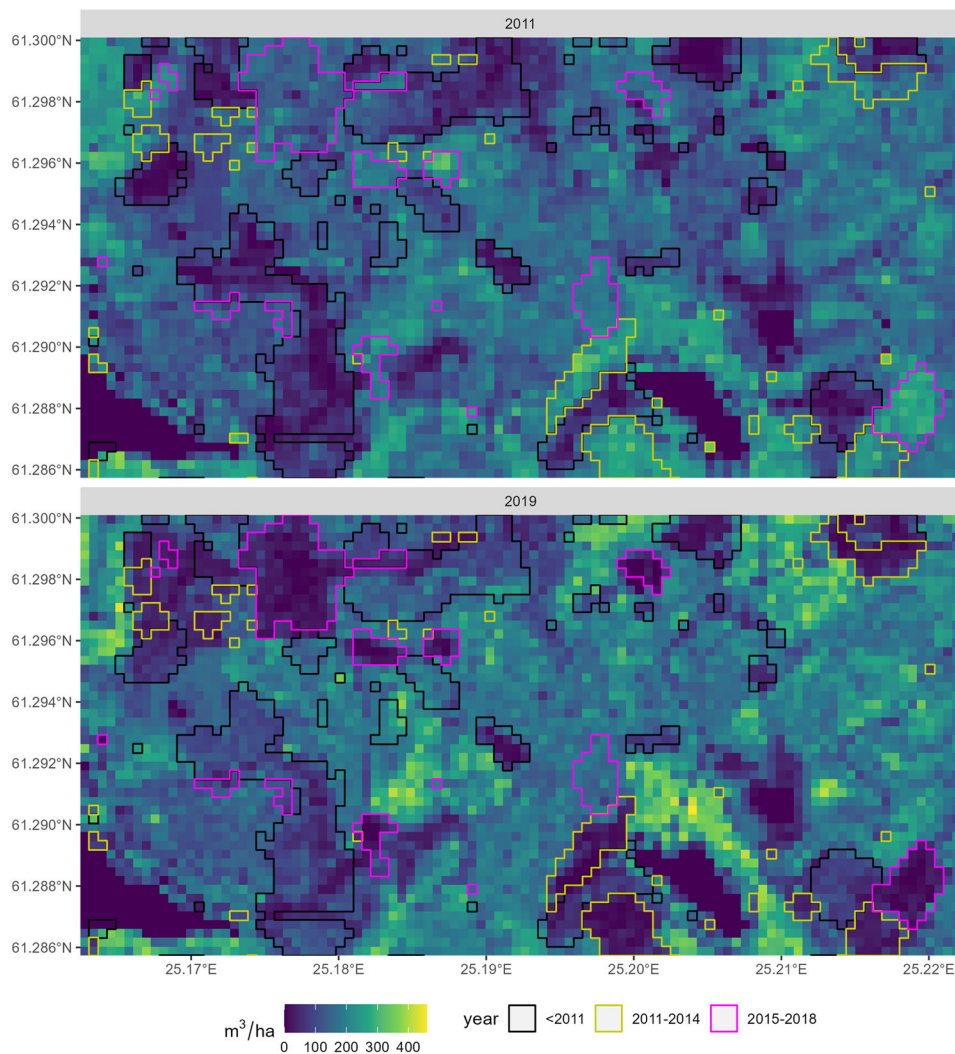


Figure 5. Illustration of the changepoint detection in a subset of the main Häme region. The multi-source NFI growing stock volume data, years 2011 and 2019, pixel clusters with detected changepoints using $\gamma = 0.6$, $\pi^1 = 0.7$ overlaid. Changes before the first image are overlaid with black, and in these regions the natural positive trend of the growing stock is clearly visible. Compare to the aerial photographs in Fig. 6.

counterparts which, in the worst case, have no detection power whatsoever. We also applied the approach to a real-world forest inventory data to detect sudden changes in growing stock.

The probability of observing a change in the forests from satellite images is the harder, the smaller the change. For instance, [Pitkänen et al. \(2020\)](#) analysed both thinnings and final fellings. The accuracy of detecting the final fellings was good, but that of thinnings considerably lower. Thus, it can be assumed that the final felling areas or large severe natural disturbances are found without utilising information from neighbouring pixels, but in case of thinnings or less severe natural disturbances the neighbouring information is especially useful. The driving tracks within a stand are a clear sign of a carried out thinning, but these tracks only affect a small part of the thinned area and the effect on each pixel is not very

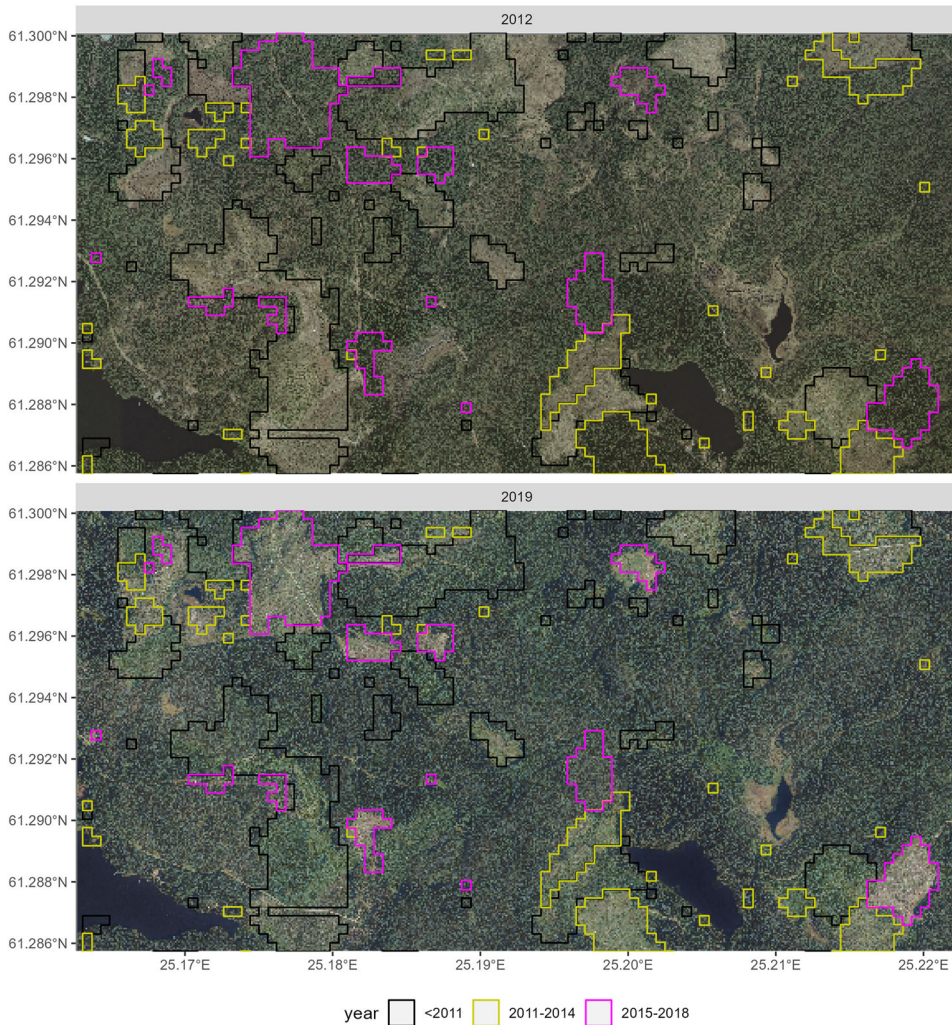


Figure 6. Illustration of the changepoint detection in a subset of the main Häme region. Aerial photographs courtesy of National Land Survey of Finland, years 2012 and 2019, pixel clusters with detected changepoints using $\gamma = 0.6$, $\pi^1 = 0.7$ overlaid. Changes before the first image are overlaid with black, and in such regions the natural positive trend of the growing stock is clearly visible.

large nor very long-lasting. Then, information of occurrence of such a track or other small opening in the neighbourhood adds to the evidence that harvest has probably also happened in the target pixel.

In this study, we assessed the changepoints of time series at the 32×32 m pixel level. When we wish to detect the changes at the stand level, it is enough for a large part ($>50\%$) of the pixels to show change within a same year for us to conclude that the whole stand has been thinned or that it has experienced some natural hazards. So, even if the thinned stands are only partially detected, it is possible to conclude that a harvest has happened. Delineating such thinned stands, which show only partial change, would possibly improve the analysis. This remains to be examined in the future.

The quantity of surfaces used in change detection is typically the pixel values, which have some physically meaningful unit such as reflectance (Pasquarella et al. 2022). In this study, we used predicted values of growing stock volumes (m^3/ha) instead of pixel values of satellite images or their transformations. The stock volume prediction model uses field plot data in addition to satellite images. Similar approach of using modelled surface instead of pixel values of satellite images has been used, for example, in Yin et al. (2018) and Main-Knorn et al. (2013). In our case, the use of a modelled surface provided certain advantages over the direct use of pixel values of satellite images. Growing stock was estimated image by image using field plots from the region that the image covers. This way conversion from the raw pixel values to bottom-of-atmosphere reflectance and seasonality issues due to time differences between training plots and target population are avoided. It also makes sense to compare the quantity of interests, here growing stock volume. The downside of using predicted growing stock volume is that predictions have error and it is not guaranteed that errors in successive predictions are correlated.

We showed by analysis of simulated and multi-source NFI time series data that modelling interaction between changepoints of neighbouring pixels can improve the detection of changes in challenging cases where the time series are relatively short and observation noise is relatively high. We constructed a statistical model which was inspired by the environmental application, and it has clear limitations. Straightforward extensions include geo-statistical generalised linear models for the observed variables (such as spatially correlated Poisson regression), temporally auto-regressive errors and pixel-wise over-dispersion. Including more than one potential changepoint is more difficult and requires care, as the discrete latent space of multi-jump variables grows exponentially, the Gibbs sampler needs to be well designed (or swapped for a more efficient algorithm) and even more data are required; see, for example, Kennedy et al. (2010) and Zhao et al. (2019) for how to approach the issue. Using the current model, the most likely missed behaviour in the short series NFI example is the double thinning, which the model probably classifies as negative overall trend instead of two abrupt changes. For improved modelling of the NFI application, one could tie the changepoint probability to the growing stock volume as forest management is tied to the life stages of the forest stand. Because of modular construction, the model can be tuned for specific purposes, for example by using elongated neighbourhoods to detect linear features, or by designing non-diagonal Γ to add temporal correlation to discover events such as storm damages at edges of final fellings.

The thinning event can appear in multiple variables of interest simultaneously within each location, e.g. growing stock by species, and combining them in a joined model can improve detection (Chen et al. 2022). In our approach, we intentionally did not correlate the observed variables directly, which means that any multi-variate series with mixtures of counts, binary values or continuous values can be modelled jointly with the least amount of added computational complexity. In its simplest, we need only formulate additional observation models in Eq. (5) for each additional observed variable, and the latent jump variables can then be estimated based on them all.

Declarations

Conflict of interest The authors declare that there is no conflict of interest.

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

[Received January 2023. Revised April 2023. Accepted April 2023. Published Online May 2023.]

A. ADDITIONAL FIGURES

See Figs. 7, 8, 9, 10, 11, 12, 13, 14, 15.

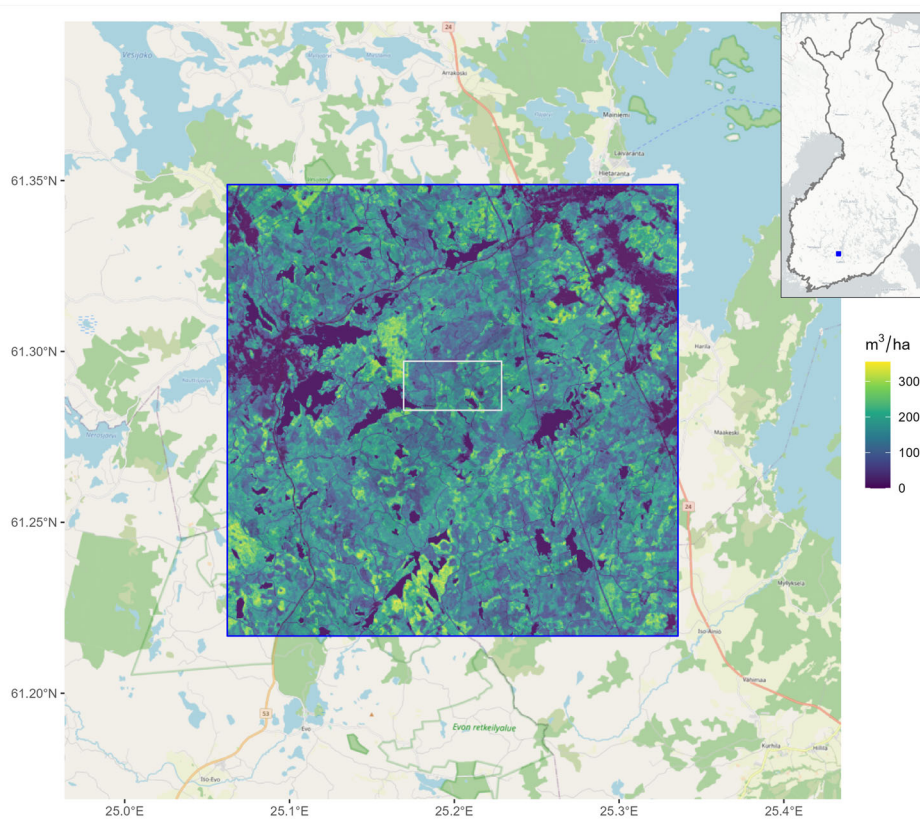


Figure 7. NFI example application region in Finland. The central square depicts the NFI pixelwise averaged growing stock values over the series runtime 1994–2019. The rectangle inside the square is described in more detail in Results. Background maps from OpenStreetMap .

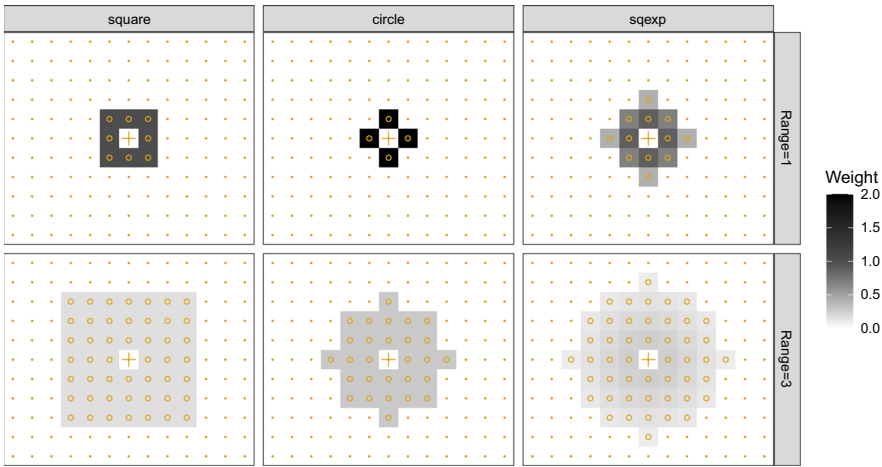


Figure 8. Illustration of the considered neighbourhood definitions (kernels), giving weights as functions of the cell centre locations. The larger circles (o) are the neighbours of the central cell (+), having nonzero weight. The shading shows the weight given to the neighbour. Top row has range 1 and bottom row range 3. If not otherwise stated, square with range 1 is used in this work, and to allow direct comparisons, all other kernels are scaled to also sum to 8. The *square*: Uniform kernel with max norm. The *circle*: Uniform kernel with L^2 -norm. The *sqexp*: Squared exponential kernel with L^2 -norm, truncated at (range + 1).

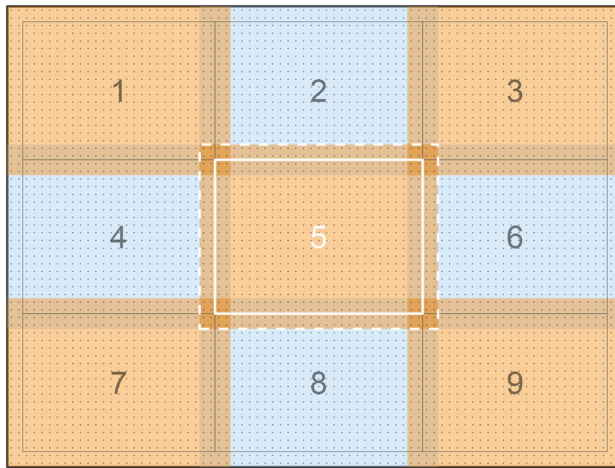


Figure 9. Illustration of the tiling used in the partitioned estimation. The illustration is for a spatial raster with size 60×80 square cells. The outer line is the bounding box of the raster. The dots represent cell (pixel) centroids. The division by lines and the numbers describe a 3×3 tiling, and the shading describes the overlapping of the buffered tiles when using a 2 pixel buffer. The white dashed rectangle highlights the cells used in estimation corresponding to tile No. 5, whereas the white solid rectangle highlights the cells whose posterior estimates are kept only from the estimation corresponding to tile No. 5. Cells in buffer overlaps provide information to multiple analysis, but only one set of posteriors is kept for each cell. Cells at the edges of the raster, e.g. left-hand side buffer of tile No. 4, are suspect to edge effects and will be discarded .

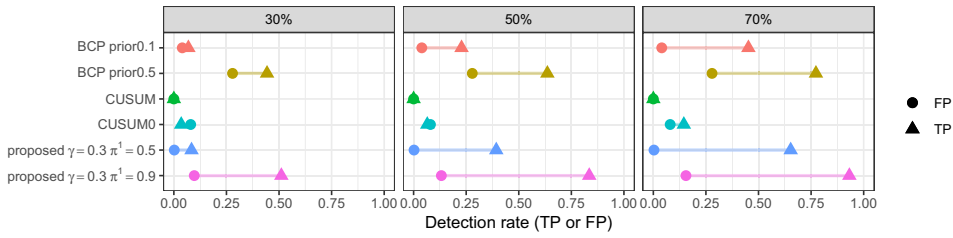


Figure 10. Simulation trial changepoint detection quality comparison. Comparing the proposed method to off-the-shelf methods CUSUM and BCP. “CUSUM0” is the CUSUM test based on the intercept-only model, otherwise the models have also linear trend. Three different difficulty levels, ranging from 30 (left) to 70% (right) reductions in response variable mean after change.

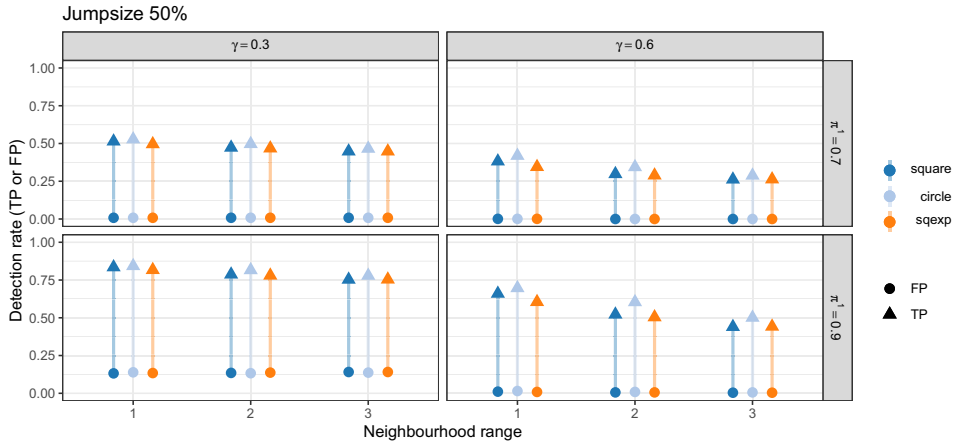


Figure 11. Comparing different neighbourhood kernel definitions. Dataset is simulation trial case “medium” with 50% reductions in response variable after change. For illustration of the different kernels, see 8.

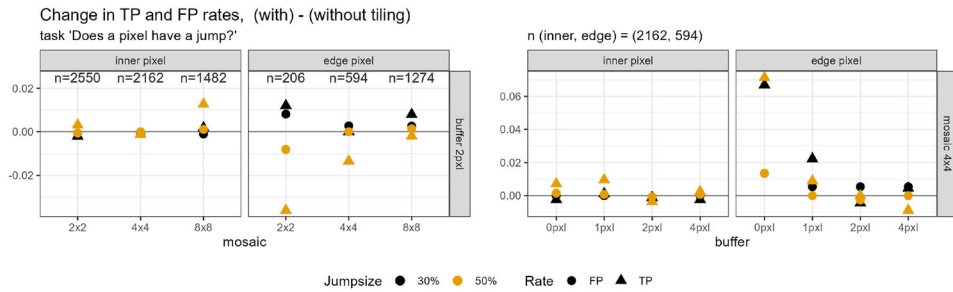


Figure 12. Change in the true positive (TP) and false negative (FP) rates due to tiling. Left: With 2pixel buffer and increasing number of divisions. Right: With a 4 × 4 division and increasing buffer size. Value above (below) zero indicate higher (lower) rate than without tiling. Model estimate using $\gamma = 0.5, \pi^1 = 0.8$, tested on the 62 × 63 sub-raster at the centre of the simulation region, with rates computed after removing 4-pixel-wide slice from each edge in order to compare the same set of pixels.

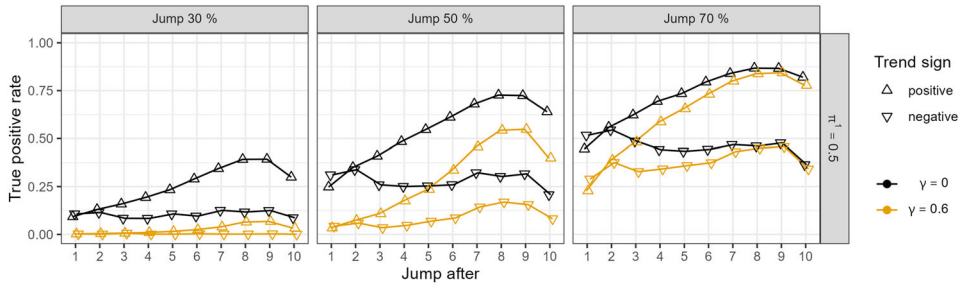


Figure 13. True positive rates in changepoint occurrence classes, separated by time series trend. Time series lengths $T = 11$.

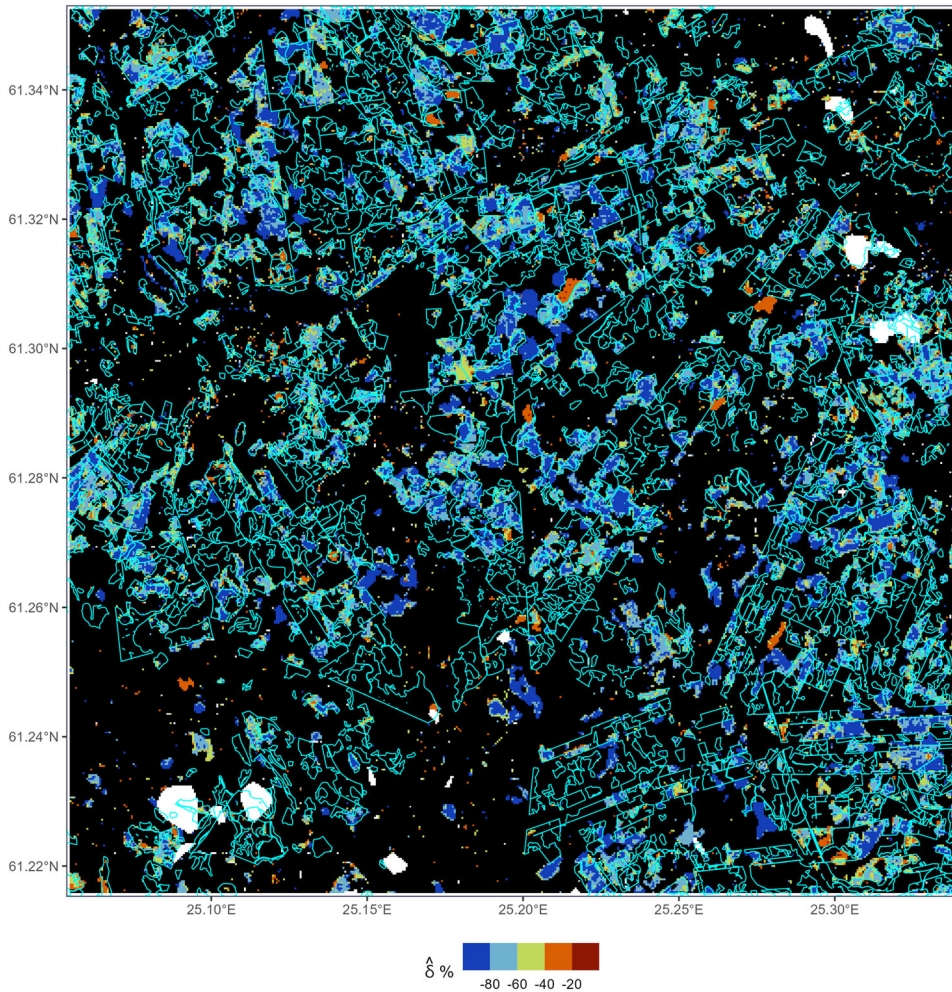


Figure 14. Estimates of reduction in growing stock volume after changepoint using tuning parameters $\gamma = 0.6$, $\pi^1 = 0.7$. Overlaid with cyan lines are all forest use declarations over the period 1997–2018.

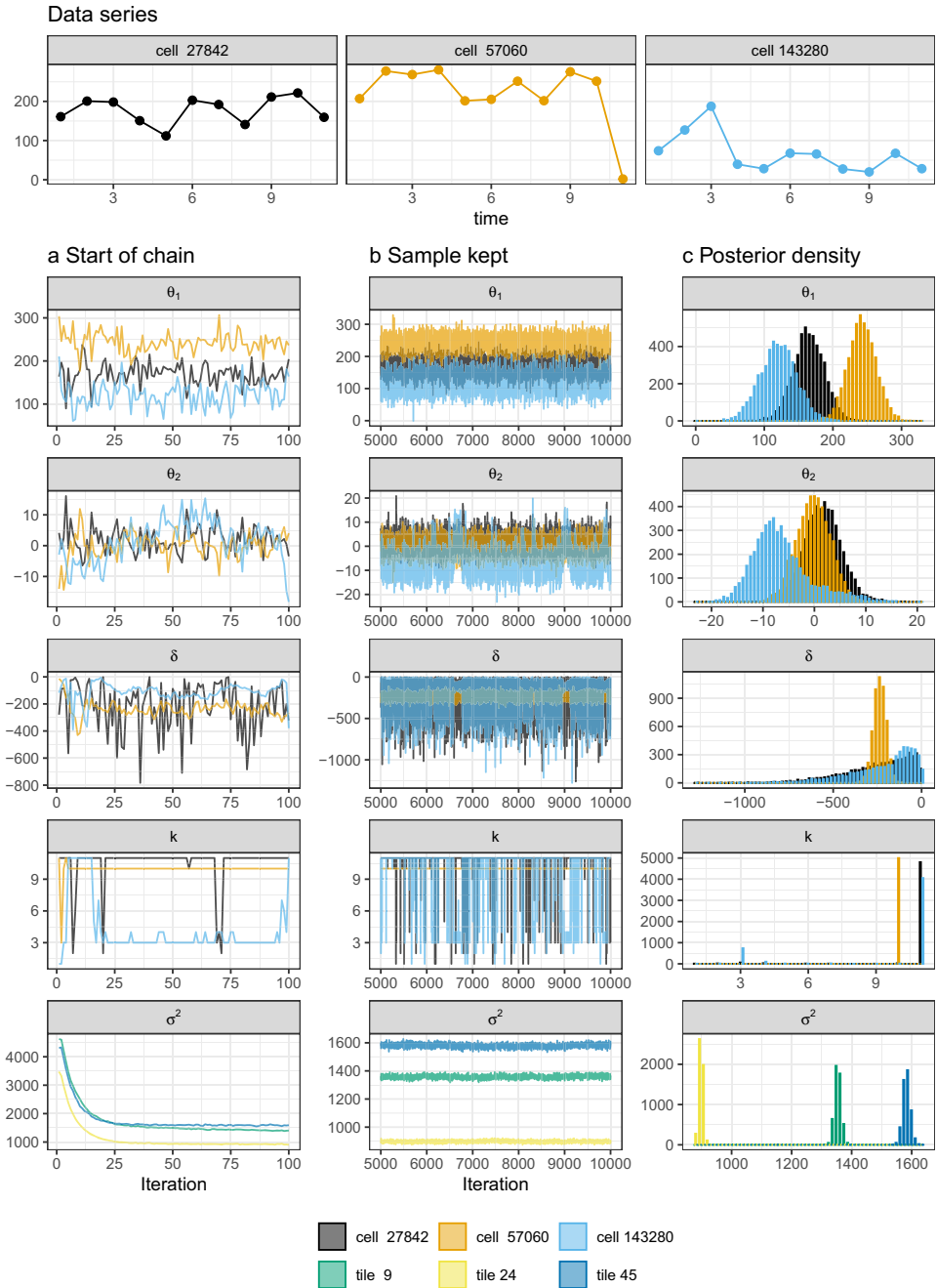


Figure 15. Top row depicts the data series of three cells (pixels) of the NFI dataset. Below that are the MCMC traces at the start of the estimation Monte Carlo chain (a), the kept traces after burn-in period (b), and the corresponding approximate posterior densities (c). Estimated using $\gamma = 0.6$, $\pi^1 = 0.7$, square neighbourhood of size 1 and 8×8 tiling with 2 pixel buffer. The σ^2 estimates are the per-tile estimates in tiles which contain the example cells .

B. DESCRIPTION OF THE GIBBS SAMPLER

For the numerical demonstrations in this work, we sampled the model parameters conditional on data using a Gibbs sampler, that is, we approximate the posterior distributions of model parameters given data using a Markov chain Monte Carlo algorithm. Let $y_{i\ell}$ denote the observation (e.g. wood stock volume) at cell i at time $t_\ell \in \{t_1, \dots, t_T\}$. We will write (column) vectors in bold, for example $\mathbf{y}_i := [y_{i1} \dots y_{iT}]^T$ and $\boldsymbol{\theta}_i = [\theta_{i1} \theta_{i2}]^T$. Define $\mathbf{X}_i := [1 \ \mathbf{t}]$ for the $T \times 2$ matrix with first column 1s and second column the times of observation. Define also $\boldsymbol{\psi}_k := [0 \ 0 \dots 1 \ 1]^T$ with k 0s followed by $T - k$ 1s, and $\boldsymbol{\pi}$ a T -dimensional simplex vector.

The probabilistic model we considered in this work is given by the following definitions:

$$\begin{aligned} \mathbf{y}_i | k_i, \mathbf{X}_i, \boldsymbol{\theta}_i, \delta_i, \sigma_i^2 &\sim \text{Normal}(\mathbf{X}_i \boldsymbol{\theta}_i + \delta_i \boldsymbol{\psi}_{k_i}, \sigma_i^2 \mathbf{I}_{T \times T}) \\ \boldsymbol{\theta}_i &\sim \text{Normal}(\mathbf{m}_x, \mathbf{S}_x) \\ \delta_i &\sim \text{truncated-Normal}(m_d, s_d^2, l_d, u_d) \\ \{k_i : i = 1 \dots N\} &\sim \text{Potts}(\boldsymbol{\pi}, \Gamma) \\ \sigma^2 &\sim \text{Inverse-Gamma}(a, b) \end{aligned}$$

where $\mathbf{m}_x, \mathbf{S}_x, m_d, s_d > 0, l_d < u_d, a > 0, b > 0, \boldsymbol{\pi}$ and Γ are fixed and known. We used $\Gamma = \gamma \mathbf{I}_{T \times T}$, with fixed γ .

The full probability density, omitting prior constants from notation, is

$$p(\{\mathbf{y}_i, \boldsymbol{\theta}_i, k_i, \delta_i : i = 1, \dots, N\}, \sigma^2) = \left[\prod_{i=1}^N p(\mathbf{y}_i | \boldsymbol{\theta}_i, k_i, \sigma^2) p(\boldsymbol{\theta}_i) p(\delta_i) \right] \cdot p(\{k_i : i = 1, \dots, N\}) p(\sigma^2)$$

Next we provide the full conditional densities used in the Gibbs sampler.

B.1. UPDATING THE LATENT TIMERS $\{k_i\}$

The timer variables $k_i \in \{1, \dots, T\}$ are modelled using a Potts model. We will use shorthand $i \sim j$ for neighbours j of cell i . The joint density for all k_i 's is

$$p(\{k_i : i = 1, \dots, N\}) = Z_N(\Gamma, \boldsymbol{\pi}) \exp \left[\sum_{i=1}^N \log \boldsymbol{\pi}_{k_i} + \sum_{i=1}^N \sum_{j \sim i} \Gamma_{k_i k_j} \right]$$

The normalising constant $Z_N(\Gamma, \boldsymbol{\pi})$ is a sum over all possible configurations $\{k_i : i = 1 \dots N\}$ and cannot be directly evaluated.

The full conditional of interest for the Gibbs sampling, for each k_i , writing $\boldsymbol{\eta}_i := \mathbf{X}_i \boldsymbol{\theta}_i$, is

$$\begin{aligned} \log q_i(k) &:= \log p(k_i = k | \{k_l : l \neq i\}, \boldsymbol{\theta}_i, \delta_i, \sigma^2, \mathbf{y}_i) \\ &= \text{const} + \end{aligned}$$

$$-\frac{1}{2\sigma^2} \left[\sum_{l=1}^k (y_{il} - \eta_{il})^2 + \sum_{l=k+1}^T (y_{il} - \eta_{il} - \delta_i)^2 \right] + \log \pi_k + \sum_{j \sim i} \Gamma_{k,k_j}$$

Normalising q_i 's to sum to 1 over $k = 1, \dots, T$ leads to the full conditional $Cat(T; \{q_i(1), \dots, q_i(T)\})$ for k_i .

B.2. UPDATING THE REGRESSION AND JUMP SIZE PARAMETERS

The full conditional of θ_i is Normal($\tilde{\mathbf{m}}_x, \tilde{\mathbf{S}}_x$) with

$$\tilde{\mathbf{S}}_x = [\sigma^{-2} \mathbf{X}^T \mathbf{X} + \mathbf{S}_x^{-1}]^{-1}, \quad \tilde{\mathbf{m}}_x = \tilde{\mathbf{S}}_x^{-1} [\sigma^{-2} \mathbf{X} \tilde{\mathbf{y}} + \mathbf{S}_x^{-1} \mathbf{m}_x]$$

The full conditional of the jump size δ_i is truncated-Normal($\tilde{m}_d, \tilde{s}_d^2, l_d, u_d$) with

$$\tilde{s}_d^2 = [(T - k)\sigma^{-2} + s_d^{-2}]^{-1}, \quad \tilde{m} = \tilde{s}_d^{-2} [\sigma^{-2} \sum_{l=k+1}^T (y_{il} - \eta_{il}) + s_d^{-2} m_d]$$

where again we use $\eta_i := \mathbf{X}_i \theta_i$.

B.3. UPDATING THE VARIANCE PARAMETER σ^2

The full conditional for the variance parameter σ^2 is Gamma(\tilde{a}, \tilde{b}) with

$$\tilde{a} = a + N/2, \quad \tilde{b} = b + \sum_{i=1}^N \sum_{l=1}^T e_{il}^2/2$$

where $e_i = y_i - \mathbf{X}_i \theta_i - \psi_{k_i} \delta_i$.

B.4. UPDATE STRATEGY

We ran the chains for a fixed amount of iterations. Initial values were sampled from the priors. We used a random order sweep to go through the N cells per iteration.

Given previous iteration values for σ^2 and $\{\theta_i, \delta_i, k_i\}$ for all pixel, we updated the chain using the full conditional densities in these steps:

1. Draw a random permutation \mathcal{P} of the N raster cell indices
2. For each pixel in order \mathcal{P} , say $i = 1, \dots, N$:
 - (a) Update $k_i \rightarrow k'_i$ given $\mathbf{k}'_{1:(i-1)}, \mathbf{k}_{(i+1):N}$ and θ_i
 - (b) Update $\theta_i \rightarrow \theta'_i$ given k'_i
 - (c) Update $\delta_i \rightarrow \delta'_i$ given k'_i, θ'_i

3. Update σ^2

Here $(a : b) = (a, a + 1, \dots, b)$ for $0 \leq a \leq b \leq N$, \emptyset otherwise.

REFERENCES

- Bardwell L, Fearnhead P, Eckley IA et al (2019) Most recent changepoint detection in panel data. *Technometrics* 61(1):88–98. <https://doi.org/10.1080/00401706.2018.1438926>
- Barry D, Hartigan JA (1993) A Bayesian analysis for change point problems. *J Am Stat Assoc* 88(421):309–319. <https://doi.org/10.1080/01621459.1993.10594323>
- Besag J (1974) Spatial interaction and the statistical analysis of lattice systems. *J R Stat Soc Ser B (Methodol)* 36(2):192–225. <https://doi.org/10.1111/j.2517-6161.1974.tb00999.x>
- Brown RL, Durbin J, Evans JM (1975) Techniques for testing the constancy of regression relationships over time. *J R Stat Soc Ser B (Methodol)* 37(2):149–192 [arXiv:abs/jstor:2984889](https://arxiv.org/abs/jstor:2984889)
- Chen Y, Wang T, Samworth RJ (2022) High-dimensional, multiscale online changepoint detection. *J R Stat Soc Ser B (Stat Methodol)* 84(1):234–266. <https://doi.org/10.1111/rssb.12447>
- Coops NC, Shang C, Wulder MA et al (2020) Change in forest condition: characterizing non-stand replacing disturbances using time series satellite imagery. *For Ecol Manag* 474(118):370. <https://doi.org/10.1016/j.foreco.2020.118370>
- Duncanson L, Dubayah R (2018) Monitoring individual tree-based change with airborne lidar. *Ecol Evol* 8(10):5079–5089. <https://doi.org/10.1002/ece3.4075>
- Gelman A, Carlin JB, Stern HS et al (eds) (2014) 3rd edn. Chapman & Hall, Boca Raton
- Hamunyela E, Verbesselt J, Herold M (2016) Using spatial context to improve early detection of deforestation from Landsat time series. *Remote Sens Environ* 172:126–138. <https://doi.org/10.1016/j.rse.2015.11.006>
- Hansen MC, Potapov PV, Moore R et al (2013) High-resolution global maps of 21st-century forest cover change. *Science* 342(6160):850–853. <https://doi.org/10.1126/science.1244693>
- Hauglin M, Rahlf J, Schumacher J et al (2021) Large scale mapping of forest attributes using heterogeneous sets of airborne laser scanning and National Forest Inventory data. *For Ecosyst* 8(1):65. <https://doi.org/10.1186/s40663-021-00338-4>
- Healey SP, Cohen WB, Yang Z et al (2018) Mapping forest change using stacked generalization: an ensemble approach. *Remote Sens Environ* 204:717–728. <https://doi.org/10.1016/j.rse.2017.09.029>
- Hughes M, Kaylor S, Hayes D (2017) Patch-based forest change detection from Landsat time series. *Forests* 8(5):166. <https://doi.org/10.3390/f8050166>
- Jewell S, Fearnhead P, Witten D (2022) Testing for a change in mean after changepoint detection. *J R Stat Soc Ser B (Stat Methodol)* 84(4):1082–1104. <https://doi.org/10.1111/rssb.12501>
- Kangas A, Astrup R, Breidenbach J et al (2018) Remote sensing and forest inventories in Nordic countries—roadmap for the future. *Scand J For Res* 33(4):397–412. <https://doi.org/10.1080/02827581.2017.1416666>
- Katila M, Tomppo E (2002) Stratification by ancillary data in multisource forest inventories employing k-nearest-neighbour estimation. *Can J For Res* 32(9):1548–1561. <https://doi.org/10.1139/x02-047>
- Katila M, Rajala T, Kangas A (2020) Assessing local trends in indicators of ecosystem services with a time series of forest resource maps. *Silva Fennica* 54(4):19
- Kennedy RE, Yang Z, Cohen WB (2010) Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr—temporal segmentation algorithms. *Remote Sens Environ* 114(12):2897–2910. <https://doi.org/10.1016/j.rse.2010.07.008>
- Kennedy RE, Andréfouët S, Cohen WB et al (2014) Bringing an ecological view of change to Landsat-based remote sensing. *Front Ecol Environ* 12(6):339–346. <https://doi.org/10.1890/130066>
- Kennedy R, Yang Z, Gorelick N et al (2018) Implementation of the LandTrendr algorithm on Google Earth Engine. *Remote Sens* 10(5):691. <https://doi.org/10.3390/rs10050691>

- Killick R, Fearnhead P, Eckley IA (2012) Optimal detection of changepoints with a linear computational cost. *J Am Stat Assoc* 107(500):1590–1598. <https://doi.org/10.1080/01621459.2012.737745>
- Main-Knorn M, Cohen WB, Kennedy RE et al (2013) Monitoring coniferous forest biomass change using a Landsat trajectory-based approach. *Remote Sens Environ* 139:277–290. <https://doi.org/10.1016/j.rse.2013.08.010>
- Mäkisara K, Heikkinen J, Heiler I et al (2001) Updating of the Finnish national forest inventory using remote sensing data (Valtakunnan metsien inventoinnin (VMI) ajantasaistus kaukokartoitusaineiston avulla). Final report, Finnish Forest Research Institute
- Mäkisara K, Katila M, Peräsaari J (2019) The multi-source National Forest Inventory of Finland methods and results 2015. Technical report 8/2019, Natural Resources Institute Finland (Luke)
- Mäkisara K, Katila M, Peräsaari J (2022) The multi-source National Forest Inventory of Finland—methods and results 2017 and 2019. Technical report, Natural Resources Institute Finland (Luke)
- Mäkisara K, Katila M, Peräsaari J, et al (2016) The multi-source National Forest Inventory of Finland—methods and results 2013. Technical report 10/2016, Natural Resources Institute Finland (Luke)
- Messer M, Albert S, Schneider G (2018) The multiple filter test for change point detection in time series. *Metrika* 81(6):589–607. <https://doi.org/10.1007/s00184-018-0672-1>
- Moisen GG, Meyer MC, Schroeder TA et al (2016) Shape selection in Landsat time series: a tool for monitoring forest dynamics. *Glob Change Biol* 22(10):3518–3528. <https://doi.org/10.1111/gcb.13358>
- Moore M, Nicholls G, Pettitt A et al (2020) Scalable Bayesian inference for the inverse temperature of a hidden Potts model. *Bayesian Anal*. <https://doi.org/10.1214/18-BA1130>
- Nilsson M, Nordkvist K, Jonzén J et al (2017) A nationwide forest attribute map of Sweden predicted using airborne laser scanning data and field data from the National Forest Inventory. *Remote Sens Environ* 194:447–454. <https://doi.org/10.1016/j.rse.2016.10.022>
- Nord-Larsen T, Riis-Nielsen T, Ottosen MB (2017) Forest Resource Map of Denmark: mapping of Danish Forest Resource using ALS from 2014–2015. IGN report, Department of Geosciences and Natural Resource Management, University of Copenhagen
- Ojanen M, Brockhaus M, Korhonen-Kurki K et al (2021) Navigating the science-policy interface: forest researcher perspectives. *Environ Sci Policy* 118:10–17. <https://doi.org/10.1016/j.envsci.2021.01.002>
- Pasquarella VJ, Arévalo P, Bratley KH et al (2022) Demystifying LandTrendr and CCDC temporal segmentation. *Int J Appl Earth Obs Geoinf* 110(102):806. <https://doi.org/10.1016/j.jag.2022.102806>
- Pitkänen TP, Sirro L, Häme L et al (2020) Errors related to the automatized satellite-based change detection of boreal forests in Finland. *Int J Appl Earth Obs Geoinf* 86(102):011. <https://doi.org/10.1016/j.jag.2019.102011>
- R Core Team (2022) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
- Sawyer G, Dubost A, de Vries M (2015) Copernicus sentinels' products economic value: a case study of forest management in Sweden. Technical report
- Sebald J, Senf C, Seidl R (2021) Human or natural? Landscape context improves the attribution of forest disturbances mapped from Landsat in Central Europe. *Remote Sens Environ* 262(112):502. <https://doi.org/10.1016/j.rse.2021.112502>
- Tomppo E (1996) Multi-source National Forest Inventory of Finland. New thrusts in forest inventory. In: European Forest Institute EFI proceedings, Tampere, Finland, vol I. IUFRO XX World Congress, 6–12 August 1995, Tampere, Finland
- Tomppo E (2008) Multi-source national forest inventory: methods and applications. No. 18 in managing forest ecosystems. Springer, Dordrecht
- Tomppo E, Katila M, Moilanen J et al (1998) Kunnittaiset metsävaratiedot 1990–94. Metsätieteen aikakauskirja. <https://doi.org/10.14214/ma.6453>
- Tomppo E, Haakana M, Katila M et al (2009) The multi-source National Forest Inventory of Finland—methods and results 2005. Technical report, Finnish Forest Research Institute
- Tomppo E, Katila M, Mäkisara K, et al (2012) The multi-source National Forest Inventory of Finland—methods and results 2007. Technical report 227, Finnish Forest Research Institute

- Tomppo E, Katila M, Mäkisara K, et al (2013) The multi-source National Forest Inventory of Finland—methods and results 2009. Technical report 273, Finnish Forest Research Institute
- Tomppo E, Katila M, Mäkisara K, et al (2014) The multi-source National Forest Inventory of Finland—methods and results 2011. Technical report 319, Finnish Forest Research Institute
- Wu FY (1982) The Potts model. *Rev Mod Phys* 54(1):235–268. <https://doi.org/10.1103/revmodphys.54.235>
- Wulder MA, Hermosilla T, Stinson G et al (2020) Satellite-based time series land cover and change information to map forest area consistent with national and international reporting requirements. *For Int J For Res* 93(3):331–343. <https://doi.org/10.1093/forestry/cpaa006>
- Yin H, Prishchepov AV, Kuemmerle T et al (2018) Mapping agricultural land abandonment from spatial and temporal segmentation of Landsat time series. *Remote Sens Environ* 210:12–24. <https://doi.org/10.1016/j.rse.2018.02.050>
- Yu X, Hyyppä J, Kaartinen H et al (2004) Automatic detection of harvested trees and determination of forest growth using airborne laser scanning. *Remote Sens Environ* 90(4):451–462. <https://doi.org/10.1016/j.rse.2004.02.001>
- Zhao K, Suarez JC, Garcia M et al (2018) Utility of multitemporal lidar for forest and carbon monitoring: tree growth, biomass dynamics, and carbon flux. *Remote Sens Environ* 204:883–897. <https://doi.org/10.1016/j.rse.2017.09.007>
- Zhao K, Wulder MA, Hu T et al (2019) Detecting change-point, trend, and seasonality in satellite time series data to track abrupt changes and nonlinear dynamics: A Bayesian ensemble algorithm. *Remote Sens Environ* 232(111):181. <https://doi.org/10.1016/j.rse.2019.04.034>
- Zhu Z (2017) Change detection using landsat time series: a review of frequencies, preprocessing, algorithms, and applications. *ISPRS J Photogramm Remote Sens* 130:370–384. <https://doi.org/10.1016/j.isprsjprs.2017.06.013>

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.