*Journal of the Royal Statistical Society Series C: Applied Statistics*, 2024, **00**, 1–23
https://doi.org/10.1093/jrsssc/qlae060

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society
**Series C**
Applied Statistics

**Original Article**

# Object oriented data analysis of surface motion time series in peatland landscapes

**Emily G. Mitchell[1], Ian L. Dryden[1,2], Christopher J. Fallaize[1], Roxane Andersen[3], Andrew V. Bradley[4], David J. Large[5] and Andrew Sowter[6]**

[1]School of Mathematical Sciences, University of Nottingham, Nottingham NG7 2RD, UK
[2]Department of Statistics, University of South Carolina, Columbia, SC 29208, USA
[3]Environmental Research Institute, University of Highlands and Islands, Castle Street, Thurso, Scotland KW14 7JD, UK
[4]Department of Chemical and Environmental Engineering, Faculty of Engineering, Nottingham Geospatial Institute, Innovation Park, Jubilee Campus, Nottingham NG7 2TU, UK
[5]Department of Chemical and Environmental Engineering, Faculty of Engineering, University of Nottingham, Nottingham NG7 2RG, UK
[6]Terra Motion Limited, Ingenuity Centre, Innovation Park, Jubilee Campus, University of Nottingham, Nottingham NG7 2TU, UK

*Address for correspondence*: Ian L. Dryden, School of Mathematical Sciences, University of Nottingham, Nottingham NG7 2RD, UK; Department of Statistics, University of South Carolina, Columbia, SC 29208, USA. Email: dryden@mailbox.sc.edu

## Abstract

Peatlands account for 10% of UK land area, 80% of which are degraded to some degree, emitting carbon at a similar magnitude to oil refineries or landfill sites. A lack of tools for rapid and reliable assessment of peatland condition has limited monitoring of vast areas of peatland and prevented targeting areas urgently needing action to halt further degradation. Measured using interferometric synthetic aperture radar (InSAR), peatland surface motion is highly indicative of peatland condition, largely driven by the eco-hydrological change in the peatland causing swelling and shrinking of the peat substrate. The computational intensity of recent methods using InSAR time series to capture the annual functional structure of peatland surface motion becomes increasingly challenging as the sample size increases. Instead, we utilize the behaviour of the entire peatland surface motion time series using object oriented data analysis to assess peatland condition. Bayesian cluster analysis based on the functional structure of the surface motion time series finds areas indicative of soft/wet peatlands, drier/shrubby peatlands, and thin/modified peatlands. The posterior distribution of the assigned peatland types enables the scale of peatland degradation to be assessed, which will guide future cost-effective decisions for peatland restoration.

**Keywords:** InSAR, peatland condition mapping, spatial, square root velocity function, time series, warping

## 1 Introduction

For over 25 years, global peatland restoration has been actively promoted in the hope that the degradation of peatlands can be reversed (Verhoeven, 2014). Accounting for one third of Earth's soil carbon despite only covering 3% of the land area, peatlands contain up to 95% water and 5% organic matter and provide a full spectrum of ecosystem services such as flood regulation (Grayson et al., 2010), carbon sequestration (Pawson et al., 2012), and water quality (Evans & Lindsay, 2010). Erosion and organic matter loss have a detrimental impact on provision of these services as the peatland system state, or condition, is highly vulnerable to societal pressure,

**Received:** September 18, 2023. **Accepted:** September 28, 2024

(e.g. grazing, burning, agriculture, forestry, recreation, development, and extraction), and susceptible to climate change (Andersen et al., 2017; Ise et al., 2008; Rochefort & Andersen, 2017).

The UK has 2 Mha of peatlands (10% land area), mostly as blanket bog, a globally rare type of peatlands. However, up to 80% of these peatlands are degraded to some degree (Bain et al., 2011). It is estimated that degraded UK peatlands emit 10 million metric tonnes of $CO_2$ per year, a similar magnitude to oil refineries or landfill sites (DBEIS, 2019), placing the UK among the top 20 countries for emissions of carbon from degrading peat (Joosten, 2012).

To date, condition assessment of all UK peatlands has been impeded by the lack of cost-effective ways to monitor large and often remote areas of peatland. However, recently the use of interferometry synthetic aperture radar (InSAR) signals (Sowter et al., 2013) to measure peatland surface motion has proved fruitful as a relatively low cost alternative to taking ground measurements (Marshall et al., 2022). Recent literature has found peatland surface motion, a direct consequence of eco-hydrological change in the peatland, to be highly indicative of peatland condition (Alshammari et al., 2020, 2018; Bradley et al., 2022; Marshall et al., 2022). Surface motion of peatland is mostly driven by the seasonal fluctuation in the water table causing swelling and shrinking of the peat substrate manifesting as an up-down motion (e.g. Howie & Hebda, 2017). In particular, the characteristics (e.g. timing and amplitude of seasonal peaks, overall trend) of these time series are indicative of the peatland condition (Bradley et al., 2022). This annual cycle of peatland displacement has been termed 'bog breathing' (Ingram, 1983), and we expect a single peak and a single trough each year in the annual cycle for peatland in good condition (Bradley et al., 2022). The large volume of data means that current hands-on approaches to data analysis become computationally challenging as study sites increase in area, and so it is of interest to develop a more objective statistical learning method that would bring more consistency and efficiency to mapping peatland condition.

Our raw data are surface motion time series at a set of spatial locations, and we can view each time series as observations from a smooth underlying function. More detail on the data and processing is given in Section 2.

Our primary objective is to cluster the locations into a small number of classes, which are informative regarding the condition of the peat. Therefore, at heart our goal is clustering of functional data. We use the Object Oriented Data Analysis (OODA) framework (Marron & Alonso, 2014; Marron & Dryden, 2022; Wang & Marron, 2007), which provides a general framework for the analysis of complex-structured data. Rather than perform clustering directly on the functions, we extract a small number of key features from the functions (which are estimated from the raw time series via smoothing splines), guided by the scientific knowledge of the 'bog-breathing' behaviour of peat and its relation to peat condition. This provides a low-dimensional, parsimonious representation of the data which we show to be effective for the task at hand. More discussion of OODA in general, as well as how it informs our analysis, is given in Section 2.2.

As mentioned above, a time series is observed at each of $N$ spatial locations. We extract the features independently for each site, giving a $D$-dimensional vector at each of the $N$ sites, and we model these data as a mixture of multivariate normal distributions. We then perform clustering via Bayesian inference, jointly modelling the data, parameters in the likelihood and the cluster labels (which identify the mixture components). Spatial smoothing is achieved via a Potts model (Green & Richardson, 2002) as a prior distribution over the cluster labels. This is presented formally in Section 3.1, but intuitively, this model encourages nearby spatial locations to belong to the same cluster. Conditional on cluster labels, we model each location independently of the rest.

We now briefly summarize the key steps and ideas in our process.

1. The raw data are InSAR time series, giving surface motion recorded every 6 or 12 days. There is one time series observed at each of $N$ spatial locations. We view the underlying object giving rise to each time series as a smooth, continuous process, observed at discrete time points (the observed time series).
2. Two levels of smoothing are applied to each time series: a higher degree of smoothing to estimate an underlying **longer term trend**, and a lower amount of smoothing to estimate a **seasonal oscillations** component of the underlying process.
3. From these two components, we derive 3 measures which are designed to capture important information regarding trend and oscillations of the peat surface motion. Crucially, these measures are defined with direct reference to expert scientific knowledge about the behaviour

of peat surface motion and its relationship with peat condition. Thus, each location is then described by a vector $x_i \in \mathbb{R}^3$, $i = 1, \ldots, N$.

4. Further modelling and inference is then performed using the $\{x_i\}_{i=1}^N$. Specifically, a Bayesian formulation is used, with a Gaussian likelihood on the $\{x_i\}$, conditional on unknown cluster labels. Spatial smoothing of the cluster labels is achieved via a Potts model on the cluster labels, and joint posterior inference for the model parameters and the cluster labels is carried out via MCMC. Crucially, the clusters have meaningful physical interpretations corresponding to peat condition due to the direct reference our measures $\{x_i\}$ make to the available scientific knowledge. Posterior probabilities of cluster membership can be estimated from the MCMC output, thereby quantifying uncertainty in condition type.

As mentioned above, our task is in essence one of clustering functional data. For a comprehensive review of clustering of functional data, see Jacques and Preda (2014). A key contribution of this article is to show how OODA principles can be used to extract key features from functional data for the required task—in our context, features related to peatland condition, grounded in expert knowledge and scientific evidence, which are meaningful and interpretable to peatland experts and which are also highly effective for classifying peatland condition.

The rest of the article is organized as follows. In Section 2, we describe the data and the processing methods used to extract the key features following OODA principles. In Section 3, we describe the Bayesian spatial model used to perform clustering on the features extracted. The results obtained from applying our analysis to data from the Flow Country are presented in Section 4, where we also compare with a recent functional clustering method. We conclude with a discussion.

## 2 The Flow Country and OODA

### 2.1 The Flow Country data

Covering approximately 4,000 km$^2$ of land in the Caithness and Sutherland counties of northern Scotland, the Flow Country is known for being the largest expanse of blanket peatland in Europe (Andersen et al., 2018; Lindsay et al., 1988) and the largest carbon store naturally occurring in the UK (Chapman et al., 2009). The remote nature of the Flow Country also serves as a refuge for many bird species (Lindsay et al., 1988). The current consideration to make the Flow Country a World Heritage site highlights the conservation importance of this area. Compared to other UK peatlands, the Flow Country peatlands remain in good condition overall. However, there remains evidence of past interference from land use changes over the last centuries and more recently, including peat cutting, drainage, burning, and afforestation. The preservation of the current peatland condition and the reversal of peatland degradation in this region is key for the UK to reduce carbon emissions from peatland.

Classifying peatland condition for the Flow Country will indicate which areas are to be targeted for conservation, preservation, and restoration. Based on local expert field knowledge, a diverse range of near-natural peatland conditions are captured by five sub-sites in the Flow Country (Figure 1), each roughly 10–15 km$^2$ with unique environmental and management properties (Bradley et al., 2022). We assess peatland condition for these five sub-sites using InSAR time series between 12th March 2015 and 1st July 2019. Cross Lochs (58.39°N, −3.94°E) at 180 metres above sea level (m.a.s.l), Loch Calium (58.44°N, −3.68°E) at 120 m.a.s.l, and Balavreed (58.38°N, −3.50°E) at 180 m.a.s.l have evidence of low levels of grazing. Cross Lochs and Balavreed consist of flat pool systems whilst Loch Calium has gentle slopes leading down into a central loch. Munsary (58.39°N, −3.35°E) at 100 m.a.s.l has been more intensely drained and grazed in the past. Finally, Knockfin (58.32°N, −3.80°E) at a much higher altitude (360 m.a.s.l) is an upland plateau with pool systems amongst wind-eroded peat islands, with past grazing and drainage. These peatland areas spanning a wide range of naturally occurring peatlands make them ideal for our analysis based on peatland surface motion.

The coupling between ecohydrological condition and motion of the peat surface (Mahdiyasa et al., 2022; Marshall et al., 2022) make the characteristic time series of surface motion measured by the InSAR technique highly suited to quantifying peatland condition (Alshammari et al., 2018; Bradley et al., 2022). The InSAR signals that measure surface deformation are from the European

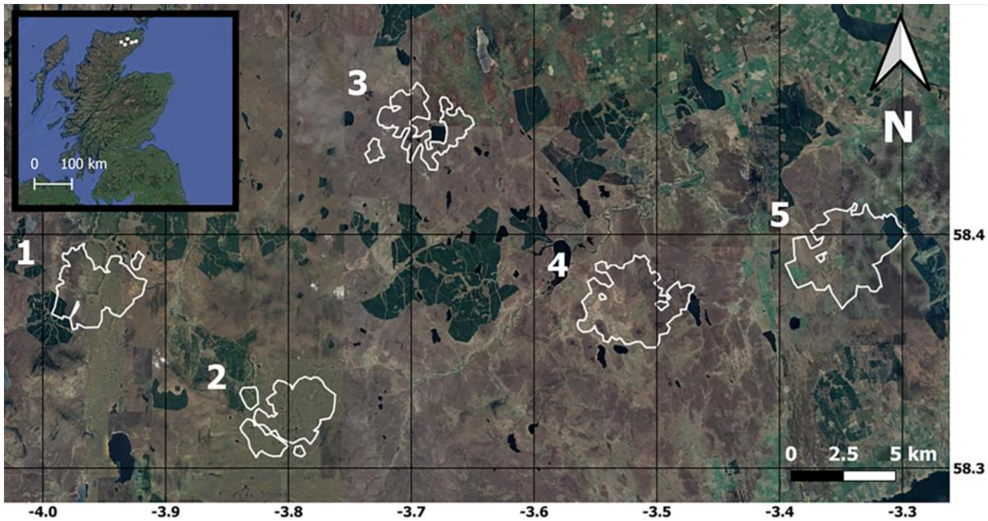


**Figure 1.** Location of five peatland sub-sites in the Flow Country, Scotland (inset): Cross Lochs (1), Knockfin (2), Loch Calium (3), Balavreed (4), and Munsary (5). Map grid units in decimal degrees. Base map data: ©2022 Google. Base map imagery: ©2022 TerraMetrics.

Space Agency Sentinel-1A and Sentinel-1B satellites and are processed by Terra Motion Limited (Sowter et al., 2013) to generate a peatland surface motion time series. The surface motion time series are measured at high spatial ($80 \times 90\ m^2$ units) and temporal (every 6–12 days) resolution across the UK. From 12th March 2015, the surface motion time series are every 12 days and an increase in resources with two satellites in operation enabled measurements to be taken every 6 days from 26th September 2016 until 1st July 2019. The characteristics of these surface motion time series are indicative of the peatland condition. In total, there are $N = 9{,}662$ geographical locations in the five regions.

## 2.2 Object oriented data analysis

From Marron and Dryden (2022), OODA involves the analysis of complex data that often lie in non-Euclidean spaces, such as functions, shapes, images, graphs, and trees. The OODA framework involves asking a series of questions to domain expert collaborators to guide the most appropriate analysis. In particular, first it is important to decide what actually are the **data objects** under study, as there are frequently many choices available leading to different analyses. Second the choice of **object space** in which the objects lie will determine which metrics and mathematical tools are available. Third the space itself may be complex, and possibly infinite dimensional, and so we need to decide what are the important **features** for statistical analysis. Finally, we must decide what types of **methods** are appropriate for statistical analysis.

In extensive discussions we have formulated some initial answers to the following four OODA questions:

1. 'What are the data objects?'
   The recorded data are noisy ground displacement time series observed each 6–12 days at a set of spatial locations (where each location is a pixel of area $80 \times 90\ m^2$). However, these are just partial observations of the underlying continuously moving surface. We choose to consider the idealized data objects as functions of continuous time located at discrete locations in space. It is natural to consider the peatland bogs as analogous to 'breathing' continuously in time, at discrete locations in space.
2. 'In what space do the objects lie?'
   The idealized object space is a product of function spaces, where each function measures the ground displacement versus time at a particular location. We also require the functions to be differentiable among other properties, and details are given in Section 2.5.

3. 'What are the important features for practical data analysis?'

   From past studies (Bradley et al., 2022), it is evident that the condition of the peat is reflected by changes of the peat levels over two different time scales. Over several years there is a smooth trend in the level of the peat, and on a shorter time scale there is often an oscillating behaviour on an annual basis with a single peak and single trough in the peat levels (the 'bog breathing'). The features of importance are the peak amplitude, peak timing, and trend gradient. Higher amplitude and earlier (Winter) peaks are indicative of 'soft/wet' peat; lower amplitude and later (Summer) peaks are indicative of 'drier/shrubby' peat; and inconsistent oscillations are indicative of 'thin/modified' peat. In addition, a positive trend gradient can indicate improving peat and a negative trend gradient can indicate deteriorating peat.
4. 'What methods will be used?'

   After the features have been obtained, we develop a Bayesian spatial model for clustering each spatial location into similar types of peat. A joint prior smoothing model for the cluster labels will be specified for all locations to encourage neighbours to have similar labels. The posterior distribution is obtained by Markov chain Monte Carlo simulation. The aim is to produce a map of the area, with estimated class labels indicating similar types of peatland condition, and in addition a measure of the uncertainty in the estimates.

The initial answers to the OODA questions have been developed over a sustained period by numerous interactions and collaborations with experts on peatland monitoring. The responses to the questions are developed further in the rest of the article.

## 2.3 Smoothing the peat surface motion time series

To obtain relevant features from each time series we fit two smoothing splines (Green & Silverman, 1993; Hastie et al., 2008) with two different levels of smoothing: larger smoothing to obtain the **trend**, and smaller smoothing to obtain the annual cycles. The difference between the two splines is the de-trended peat level, which we call the **oscillations**.

An example of a peat surface motion time series with estimated smooth functions for the oscillations and trends is given in Figure 2. We use the `smooth.spline` command in R (R Core Team, 2023) with parameter `spar` (a smoothing parameter, with higher values giving more smoothing). When `spar` = 0.7 the estimated function, $f_j$ say, (Figure 2a) captures the annual oscillations

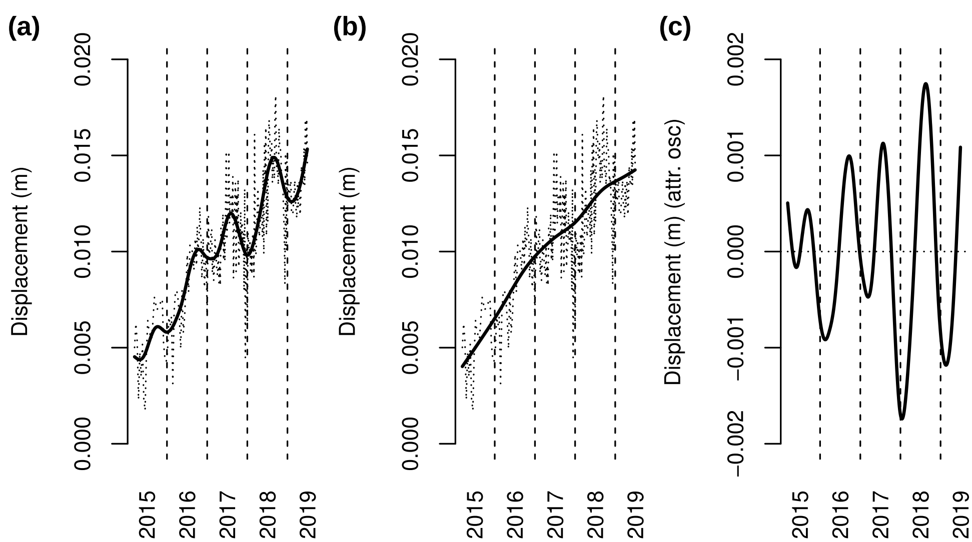


**Figure 2.** Example of a smoothed peatland surface motion time series. (a) small level of smoothing with `spar` = 0.7 to capture trends and oscillations combined, (b) large level of smoothing with `spar` = 1 to capture the overall trend, and (c) oscillations extracted once the trend has been removed.

combined with an overall trend at location $j$. When $\texttt{spar} = 1$, the estimated function $f_j$ estimates the overall trend at location $j$ (Figure 2b) without the annual oscillations. These choices of smoothing were informed using simulation experiments over a range of parameter values and calibrating with the scientific knowledge about expected bog-breathing behaviour. Full details of these experiments are given in the online supplementary material (Mitchell et al., 2024). The difference between the two smoothed functions gives the de-trended oscillations attributed to bog breathing (Figure 2c). The smoothed functions of the oscillations and trends will be used hereafter. Note that we estimate the oscillation plus trend, and trend separately. An alternative approach to decomposing the data is to estimate the seasonal and trend components simultaneously allowing for warping, which is the approach considered by Tai et al. (2017).

Since interest lies in the rise and fall of a surface within the time series itself, the gradient function of the trend is also considered for the following analysis. The gradient accentuates features interrupting the constant rate accumulation/degradation of peatlands, such as their response to extreme weather events and restoration.

For the surface motion time series in the Flow Country sub-sites, we smooth the function between 12th March 2015 ($t_1$) and 1st July 2019 ($t_{202}$), where for all locations $L_j = 202$ is the number of time points we have measurements for with either 6 or 12 day spacings. Fitting cubic splines to the data collected every 6–12 days allows for interpolated daily data to be estimated and outliers to be smoothed over.

## 2.4 Examples of different types of peat surface motion time series

In Figure 3, we see some example surface motion time series in the left-hand column; in the middle column are the oscillations for these example locations; and in the right-hand column, we see the trends of peatland motion for these examples. The rows of Figure 3 indicate example locations of (first row) soft and wet peatland, (second row) drier, shrubby peatland, and (third row) thin,

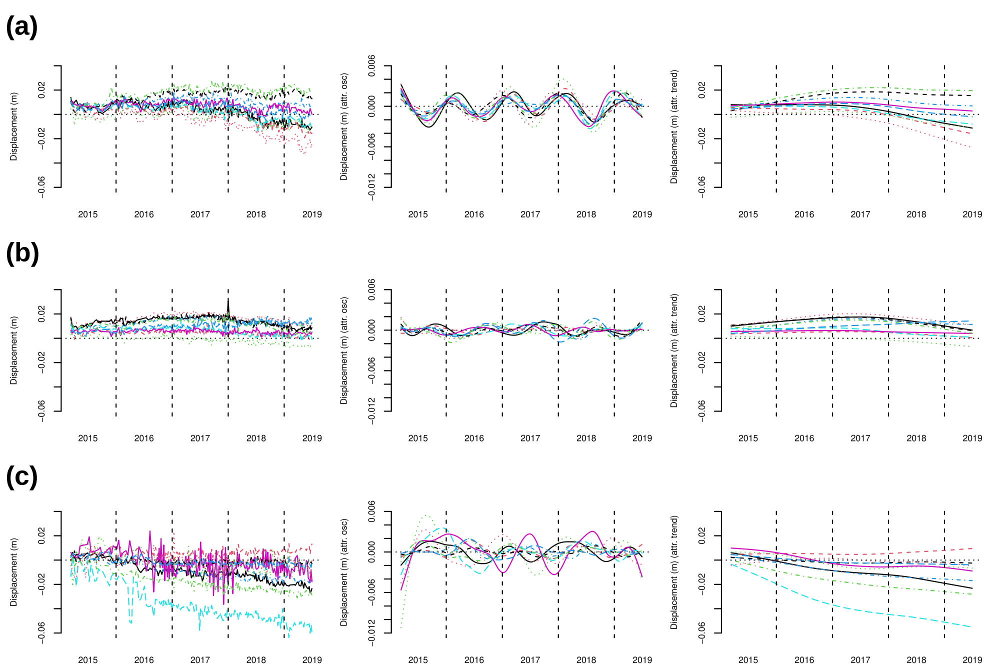


**Figure 3.** Left: InSAR time series of peatland surface motion in the Flow Country from 12th March 2015 to 1st July 2019, centre: examples of oscillations, right: examples of trends of peatland surface motion. Time series in each row share oscillatory or trend features, found to be important from past studies. The rows indicate example locations of (first row) soft/wet peatland, (second row) drier/shrubby peatland, and (third row) thin/modified peatland. (a) Example locations of soft/wet peatland. (b) Example locations of drier/shrubby peatland, and (c) example locations of thin/modified peatland.

modified peatland. (These locations are randomly sampled from each of the identified peat types (see Section 4) covering all of the sub-sites shown in Figure 1, and we observe no discernible difference between the peat types between the sub-sites.) For brevity, we denote these types of peatland as ‘soft/wet’, ‘drier/shrubby’ and ‘thin/modified’ throughout the article. It can be seen that the soft/wet peatland has larger amplitude oscillations with earlier timed peaks (in Winter, due to swelling), the drier/shrubby peatland has smaller amplitude peaks with later timed peaks (in Summer, due to a build-up in vegetation from the growing season). The thin/modified peaks do not have a consistent structure of oscillations. The trend is more downward for the thin/modified peatlands whereas it is more flat for the soft/wet and drier/thinner peat here. It is not surprising to see a downward trend towards the end of the series, even in the good ‘soft/wet’ peatlands, due to the drought of 2018 affecting the water table depth over the whole region. The ability to respond to such climactic conditions in subsequent wetter years is a sign of peat in good condition: if we followed the series further into the future, we would expect to see the trend rise again in response to wetter years for peat in good condition. Hence, the features of amplitude, timing and trend in these examples highlight our chosen features of interest.

## 2.5 Square root velocity functions

Recall that the OODA approach has led to the data being treated as functions at discrete spatial locations. Functional data analysis (Ramsay & Silverman, 2005; Srivastava & Klassen, 2016) provides a powerful statistical framework to analyse data as functions of time, where possible warping of the time axis may be needed to register similar features of curves. In our case, variability in timing and amplitude of the oscillation functions is of interest, and functional data analysis allows us to capture this. We analyse the trend functions separately, where variability in amplitude but not timing is of interest—our treatment of the trend component is discussed briefly at the end of this section, and in the next.

Warping involves matching each peatland surface motion oscillation component to another template function by time warping to minimize a distance metric such as the $L_2$ distance. A popular method for warping functions involves first converting them to their respective square root velocity functions (SRVFs) (Srivastava, Klassen, et al., 2011; Srivastava, Wu, et al., 2011). As well as performing well in a wide variety of applications and being straightforward to calculate, the method has appealing theoretical properties. In particular, the $L_2$ distance between the SRVFs is equivalent to an elastic metric which is right invariant under simultaneous warping of both functions.

Without loss of generality, time points $t$ are assumed to lie in [0, 1] (Srivastava, Klassen, et al., 2011). For each $j \in \{1, \ldots, N\}$, suppose the continuous function $f_j$ is defined by $f_j : [0, 1] \to \mathbb{R}$ and the square-root velocity function $q_j : [0, 1] \to \mathbb{R}$ of $f_j$ is given by

$$q_j(t) = \frac{\dot{f}_j(t)}{\sqrt{|\dot{f}_j(t)|}} \tag{1}$$

when $|\dot{f}_j(t)| \neq 0$ and 0 otherwise, where $\dot{f}_j$ is the derivative of $f_j$ (Srivastava, Klassen, et al., 2011). Note that the functions $f_j$, $j = 1, \ldots, N$ are in a product of first-order Sobolov spaces (Bauer et al., 2016).

Then, the warping function $\gamma_j \in \Gamma$ is chosen by minimizing the squared Euclidean distance between $q_j$ and a template square-root velocity function $q^*$, with a penalty placed on how far the warping function $\gamma_j$ is from the identity:

$$\hat{\gamma}_j = \underset{\gamma_j \in \Gamma}{\operatorname{arginf}} \left|\left| q^* - (q_j \circ \gamma_j)\sqrt{\dot{\gamma}_j} \right|\right|_2^2 + \lambda \left|\left| 1 - \sqrt{\dot{\gamma}_j} \right|\right|_2^2, \tag{2}$$

where $\lambda \geq 0$ and the $L_p$ norm is $\|f\|_p = \{ \int_0^1 f(t)^p dt\}^{1/p}$, $p \geq 1$. Being a popular choice for penalization (Srivastava & Klassen, 2016), the penalty placed on the warping function is based on the squared $L_2$ distance between the first-order derivative of the warping function $\gamma_j$ and 1, and so this captures the distance to the identity warp (where there is no warping). We denote the $j$th registered function to the template $q^*$ as $\tilde{q}_j$, where

$$\tilde{q}_j = (q_j \circ \hat{\gamma}_j)\sqrt{\dot{\hat{\gamma}}_j}, \quad j = 1, \ldots, N. \tag{3}$$

In our context, the functions to be registered will be oscillation components of the peatland surface motion time series transformed to their SRVF representations. This will extract the variability in peak timing of the oscillations, captured by $\gamma_j$, from the variability in peak amplitude, captured by $\tilde{q}_j$ from (3).

In practice, the function $f_j$, $j \in \{1, \ldots, N\}$, is obtained by smoothing the raw data to give estimated values at discrete daily time points $\{1, \ldots, 1, 573\}$, with $t = 1$ being the 12th March 2015 and $t = 1, 573$ being the 1st July 2019. To improve computational efficiency, each SRVF is sub-sampled at every 20th day when computing the warping and distances.

The warping penalty, i.e. the second term of Equation (2), restricts the level of warping permitted when matching each component of peatland surface motion to their respective reference functions. This allows for some movement in the timings but avoids over-warping to depict the reference functions exactly, which may no longer represent the features in the original function (Wu & Srivastava, 2011). Recall that we expect one cycle each year, and over-warping can manifest itself by squeezing together peaks should more than one peak occur annually in the oscillation function. We choose $\lambda = 0.1$ to fulfil the trade-off between over-warping to the reference function and negligible warping.

In order to find the warping function associated to each registered function, a choice needs to be made for the reference function $q^*$. We use a sinusoidal wave with peaks occurring in the middle of astronomical spring every year (approximately May 5th) and troughs appearing in the autumn (Figure 4a). According to the individual warping functions, registering to peaks in mid-spring will split those peaking in winter, indicative of soft/wet peatlands, and those peaking in summer, indicative of drier/shrubby peatlands. Warping functions used to register to the reference function can be found using pair_align_functions in the R package fdasrvf (Tucker, 2021; Tucker et al., 2013). The corresponding registered SRVFs can be found by first converting each function to their SRVF using `f_to_srvf`, then warping the SRVF according to the warping function already found using `warp_q_gamma`. These functions are also included in the R package `fdasrvf`. In summary, to assess the peak timing we carry out the warping using the square-root velocity function of the oscillation curve to an annual sine wave template with peak at May 5 approximately.

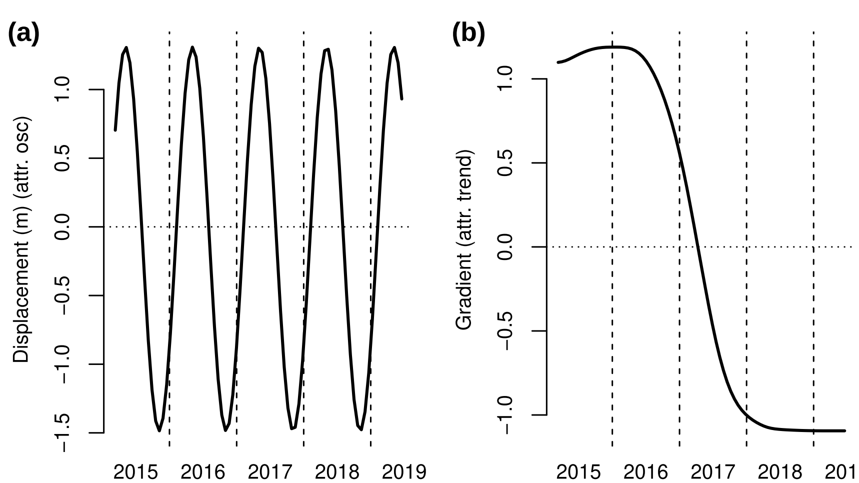


**Figure 4.** (a) Standardized sinusoidal template with peaks at the middle of astronomical spring (roughly 5th May) and troughs at the middle of autumn (roughly 5th November). (b) Standardized trend gradient mean.

In terms of the trend, we work with the gradient of the trend component, i.e. $f_j$ is the first derivative of the trend in this case. As with the oscillations, we then convert each $f_j$ to its SRVF $q_j$, and measure distances of each to a template. Little is known about the overall trend for peatlands except there should be very little difference in the timing of the trends within the region, so the arithmetic mean of the observed SRVFs, $q^* = \sum_{j=1}^{N} q_j/N$ (Figure 4b), is a natural choice of reference template often used in the alignment of functional data (e.g. Tucker et al., 2013). Since very little difference in the timing of the trends is expected, warping is not considered within this procedure ($\gamma_j$ is fixed as the identity function). The procedure is further discussed in the next section, where we describe how our key features are computed from the SRVFs and (in the case of oscillations) warping functions.

## 2.6 Distance measures

Registering each peatland surface motion time series oscillation function to the sinusoidal wave results in two further functions of interest: the warping function reflecting the difference in timing between the peatland oscillations and the sinusoidal wave, and the registered oscillations once the difference in timing has been removed.

For analysis of the trend, we compare the SRVF of each trend function's first derivative to the overall arithmetic mean of all such curves, without warping. Based on these functions, we compute three key measures to be used to classify peatland condition, which we now describe.

Table 1 gives the formal definition of each measure. The first measure (oscillation amplitude distance) is the $L_1$ distance between the registered oscillation function and the sinusoidal wave once standardized in $q$-space, which captures how peat reacts to water storage change (Roulet, 1991; Waddington et al., 2010). In order to provide comparable distances over the region, each SRVF is standardized to have mean zero and standard deviation 1. An example of a curve, its warp to the template, and the respective standardized SRVF functions is given in Figure 5.

The second measure (integrated warp difference, Table 1 (2)) captures the differences in timing between the oscillation functions and the sinusoidal wave with peaks in mid-Spring, which is related to peatland ecohydrology (Alshammari et al., 2020; Tampuu et al., 2020). Large positive/negative values of this measure correspond to warping functions which lie above/below the identity warping function, and thus the measure discriminates between regular peaks in winter and summer—see Figure 6, top row. Note that this measure does not discriminate between no warping and warping which switches between above/below the identity—see Figure 6, bottom row. A measure such as $L_1$ of $(\hat{\gamma} - \gamma_{\text{id}})$ could be used to separate such cases. We found this was not necessary in our context, as signals had either strong, regular peaks in summer or winter, else low amplitude oscillations, and these three situations are characteristic of the three clusters we identify—see Section 4. Over the timescales in our data, the condition of the peat, and hence peak timing, is not expected to change appreciably. Over longer timescales, where it is of interest to also detect changes in peat condition, a measure such as $L_1$ of $(\hat{\gamma} - \gamma_{\text{id}})$ could be used to discriminate between no warping and warping which switches between ahead/behind the identity.

The third measure (trend gradient amplitude distance) measures how far each trend gradient SRVF is from the mean trend gradient SRVF, and we use $L_2$ distance so that large departures have more weight. The oscillation measures are calculated using the $L_1$ norm. This will not penalize a peatland surface motion function as a result of our choice of timing of the peaks of the sinusoidal wave as much as if the $L_2$ norm was used.

**Table 1.** Distance measures based on the registered and warping functions for the oscillations and the registered functions for the trend gradients

| | | | |
|---|---|---|---|
| **Oscillations** | | | |
| 1 | Amplitude distance | $\left\lVert \frac{(\tilde{q}_{\text{osc}} - \text{mean}(\tilde{q}_{\text{osc}}))}{\text{sd}(\tilde{q}_{\text{osc}})} - \frac{(q^*_{\text{osc}} - \text{mean}(q^*_{\text{osc}}))}{\text{sd}(q^*_{\text{osc}})} \right\rVert_1$ | Figure 5 |
| 2 | Integrated warp difference | $\lVert(\hat{\gamma} - \gamma_{\text{id}})^+\rVert_1$ - $\lVert(\hat{\gamma} - \gamma_{\text{id}})^-\rVert_1$ | Figure 6 |
| **Trend gradients** | | | |
| 3 | Amplitude distance | $\left\lVert \frac{(q_{\text{tre}} - \text{mean}(q_{\text{tre}}))}{\text{sd}(q_{\text{tre}})} - \frac{(q^*_{\text{tre}} - \text{mean}(q^*_{\text{tre}}))}{\text{sd}(q^*_{\text{tre}})} \right\rVert_2^2$ | |

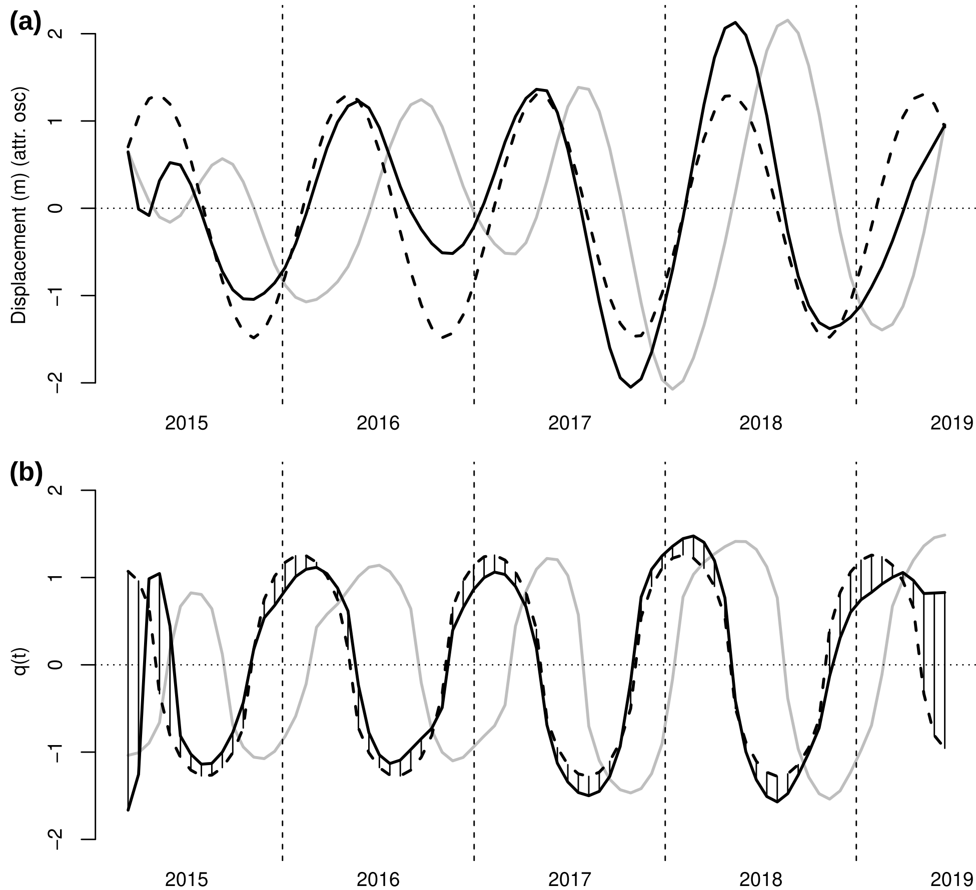


**Figure 5.** (a) the warped oscillations (darker solid line) from the original oscillations (lighter solid line) to the sinusoidal wave (dashed line) in the original function space. (b) the warped SRVF (darker solid line) from the original SRVF (lighter solid line) to the SRVF representation of the sinusoidal wave (dashed line). The distance measure is calculated between the SRVFs of the sinusoidal wave and the warped oscillations (vertical solid lines). The $L_1$ norm is taken between the two functions. All functions have been standardized.

Clustering can now be directly applied to these three measures for the peatland surface motion time series using a common clustering algorithm such as k-means clustering or hierarchical clustering using Ward's method (Ward, 1963). However, these clusters will not account for the spatial dependency between neighbouring peatland locations, whereas an observed site is more likely to be in the same cluster as another one within its vicinity. In our Bayesian framework, we include spatial dependency using a Potts prior on the cluster labels—see Section 3.1.

## 3 Bayesian cluster and uncertainty analysis

### 3.1 Likelihood, prior, and posterior distribution

We consider a multivariate normal model for each peat condition measurement (key feature). We write $\mathbf{x}^{(m)}$ for the $N$-vector of measurement type $m$ taken at each of the $N$ sites, and $m = 1, \ldots, D$. We use the $N \times K$ indicator matrix $\mathbf{Z}$ to denote the cluster membership for each site, which has a 1 in the $v$th column of the $j$th row if site $j$ is in cluster $v$, $j = 1, \ldots, N$, $v = 1, \ldots, K$, and 0 otherwise. The cluster means for measurement type $m$ are written as $\boldsymbol{\mu}^{(m)}$, which is a $K$ vector. Each measurement vector $\mathbf{x}^{(m)}$ given the cluster labels $\mathbf{Z}$ has a multivariate normal distribution:

$$\mathbf{x}^{(m)} \sim N_N(\mathbf{Z}\boldsymbol{\mu}^{(m)}, (\sigma^2)^{(m)}\mathbf{I}_N),$$

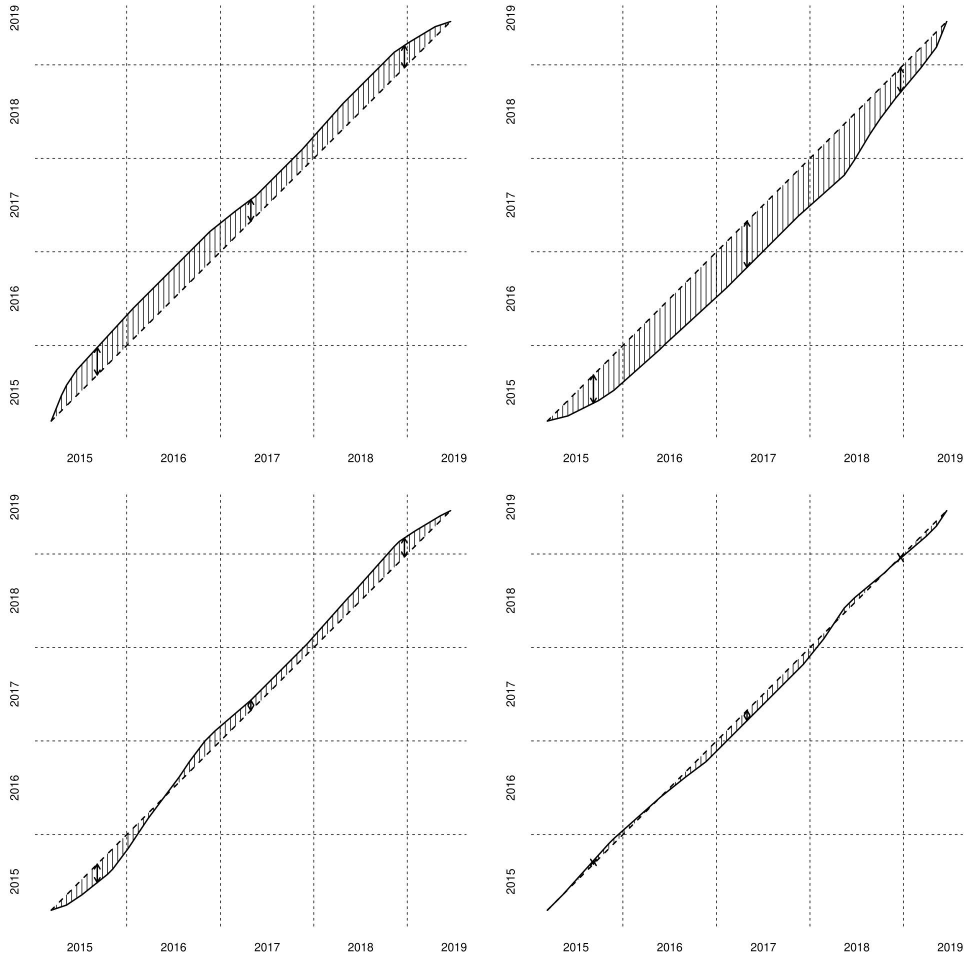


**Figure 6.** Warping function (black solid line) to register the time series to the sinusoidal wave, where the identity warping function (black dashed line) represents no warping. The integrated warp difference (Table 1) measures the difference between the shaded areas above and below the identity.

independently for each $m = 1, \ldots, D$, where $\mathbf{I}_N$ is the $N \times N$ identity matrix. So, the log-likelihood for the data is:

$$\log L(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(D)} | \mathbf{Z}, \boldsymbol{\mu}^{(1)}, \ldots, \boldsymbol{\mu}^{(D)}, (\sigma^2)^{(1)}, \ldots, (\sigma^2)^{(D)}) = -\frac{ND}{2}\log(2\pi) - \frac{N}{2}\sum_{m=1}^{D}\log((\sigma^2)^{(m)}) - \sum_{m=1}^{D}\frac{1}{2(\sigma^2)^{(m)}}(\mathbf{x}^{(m)} - \mathbf{Z}\boldsymbol{\mu}^{(m)})^T(\mathbf{x}^{(m)} - \mathbf{Z}\boldsymbol{\mu}^{(m)}).$$

The prior distribution for the cluster labels is a Potts model (Green & Richardson, 2002), which encourages neighbouring locations to have similar labels. We also use the alternative notation $Z_j = v_j$ to denote that site $j$ has cluster label $v_j$, which is equivalent to the $j$th row of $\mathbf{Z}$ having a 1 in column $v_j$ and 0 otherwise, $v_j \in \{1, \ldots, K\}$. The Potts model prior has log density of $\{Z_j = v_j : j = 1, \ldots, N\}$ equal to

$$\log \pi(\mathbf{Z}) = \text{constant} + \sum_{j=1}^{N}\frac{\lambda}{|\Delta_j|}(\#\text{neighbours of site } j \text{ in cluster } v_j),$$

where $\Delta_j$ is the set of neighbours of site $j$, $|\Delta_j|$ is the number of neighbours of site $j$, and the cluster label at site $j$ is $v_j \in \{1, \ldots, K\}$, for $j = 1, \ldots, N$. Two sites are neighbours when the distance between the sites is less than some chosen distance $d_{\max}$; we take $d_{\max} = 0.00157$ in our results of Section 4, giving an average of 7 neighbours per location. The parameter $\lambda$ is non-negative, and the greater the value of $\lambda$, the more labels are encouraged to be in the same cluster as their neighbours. Note that the marginal distribution of each $Z_j$ is uniform across the labels (Green & Richardson, 2002).

The prior distribution for each $(\sigma^2)^{(m)}$ is an inverse Gamma$(\alpha, \beta)$ distribution, and $\boldsymbol{\mu}^{(m)}$ are uniformly distributed on a large bounded region, and so have a uniform prior density. Finally, note that $\mathbf{Z}, \boldsymbol{\mu}^{(1)}, \ldots, \boldsymbol{\mu}^{(D)}, (\sigma^2)^{(1)}, \ldots, (\sigma^2)^{(D)}$ are *a priori* mutually independent.

The log-posterior density is therefore given by:

$$\begin{aligned}
&\log \pi(\mathbf{Z}, \boldsymbol{\mu}^{(1)}, \ldots, \boldsymbol{\mu}^{(D)}, (\sigma^2)^{(1)}, \ldots, (\sigma^2)^{(D)} | \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(D)}) \\
&\quad = \log L(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(D)} | \mathbf{Z}, \boldsymbol{\mu}^{(1)}, \ldots, \boldsymbol{\mu}^{(D)}, (\sigma^2)^{(1)}, \ldots, (\sigma^2)^{(D)}) \\
&\qquad + \log \pi(\mathbf{Z}) + \log \pi(\boldsymbol{\mu}^{(1)}, \ldots, \boldsymbol{\mu}^{(D)}) + \log \pi((\sigma^2)^{(1)}, \ldots, (\sigma^2)^{(D)}) + \text{constant}
\end{aligned}$$

## 3.2 MCMC sampling

Once initial clusters have been formed using Ward's clustering algorithm (Ward, 1963), a Markov chain Monte Carlo (MCMC) sampler (Geman & Geman, 1984) is constructed based on the three measures used for clustering, enabling samples to be generated approximately from the marginal posterior distributions for the cluster centres and cluster labels.

There are actually three Gibbs steps in each iteration.

1. *Cluster assignments*: Let us write $\mathbf{Z}^{(jv)}$ for the matrix which is equal to $\mathbf{Z}$ except in the $j$th row, which has a 1 in the $v$th column. This matrix represents the assignment of the cluster label at site $j$ to cluster $v$.

$$\begin{aligned}
\log\big(p(Z_j = v | \text{rest})\big) = &-\sum_{m=1}^{D} \frac{1}{2(\sigma^2)^{(m)}} (\mathbf{x}^{(m)} - \mathbf{Z}^{(jv)} \boldsymbol{\mu}^{(m)})^T (\mathbf{x}^{(m)} - \mathbf{Z}^{(jv)} \boldsymbol{\mu}^{(m)}) \\
&+ \frac{\lambda}{|\Delta_j|} \sum_{h \in \Delta_j} I(Z_h = v), + \text{constant}.
\end{aligned}$$

   and therefore,

$$Z_j | \text{rest} \sim \text{Multinomial}\{p(Z_j = 1 | \text{rest}), \ldots, p(Z_j = i | \text{rest}), \ldots, p(Z_j = K | \text{rest})\}. \tag{4}$$

   This Gibbs step is repeated for each observation $j$ in a randomized order resampled each time a cycle of the MCMC sampler has been run.

2. *Cluster means*: Once all cluster labels have been updated, next the cluster means of these are updated with a Gibbs step for each cluster $i$, $i = 1, \ldots, K$. The conditional distribution of $\{\boldsymbol{\mu}^{(m)} | \text{rest}\}$ is

$$\boldsymbol{\mu}^{(m)} | \text{rest} \sim N_K(\mathbf{B}\mathbf{Z}^T \mathbf{x}^{(m)}, (\sigma^2)^{(m)} \mathbf{B}) ,$$

   where $\mathbf{B} = (\mathbf{Z}^T \mathbf{Z})^{-1}$.

3. *Cluster variance*: The final Gibbs step of the MCMC sampler is to update each $(\sigma^2)^{(m)}$ with log conditional distribution

$$\begin{aligned}
\log\Big(p((\sigma^2)^{(m)} | \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(D)}, \mathbf{Z}, \boldsymbol{\mu}^{(1)}, \ldots, \boldsymbol{\mu}^{(D)})\Big) = -\left(\frac{N}{2} + \alpha + 1\right) \log\Big((\sigma^2)^{(m)}\Big) - \frac{\beta}{(\sigma^2)^{(m)}} \\
- \frac{1}{2(\sigma^2)^{(m)}} (\mathbf{x}^{(m)} - \mathbf{Z}\boldsymbol{\mu}^{(m)})^T (\mathbf{x}^{(m)} - \mathbf{Z}\boldsymbol{\mu}^{(m)}) + \text{constant}
\end{aligned}$$

i.e. an inverse Gamma distribution,

$$(\sigma^2)^{(m)} \sim \mathrm{IG}\left(\frac{N}{2}+\alpha, \beta+\frac{1}{2}(\mathbf{x}^{(m)}-\mathbf{Z}\boldsymbol{\mu}^{(m)})^T(\mathbf{x}^{(m)}-\mathbf{Z}\boldsymbol{\mu}^{(m)})\right).$$

The MCMC sampler is repeated $T$ times to get samples for $Z_j$ for $j = 1, \ldots, N$ and $\boldsymbol{\mu}^{(m)}$, $(\sigma^2)^{(m)}$ for $m = 1, \ldots, D$, which approximately come from the joint posterior distribution and their respective marginal distributions. For the cluster allocations $Z_j$, $j = 1, \ldots, N$, the approximation to the posterior distribution for $Z_j$ allows for probabilities of belonging to each cluster to be estimated. We shall also look at the means of the approximate posterior distributions for the cluster centres to assess what each cluster may represent.

### 3.3 Posterior inference

We carry out Markov chain Monte Carlo simulation, and after many iterations the values of the chain will be (approximate) dependent samples from the posterior distribution. Trace plots are used to identify an appropriate burn-in period, with the remainder of the simulated values used for inference.

The estimated probability of each location $j$ being in cluster $k$ is simply the proportion of time after burn-in that its cluster label spends in each cluster. Where a single allocation of each location to a particular cluster is desired, the clustering is estimated by using the marginal maximum *a posteriori* (MMAP) estimate (Doucet et al., 2002), which is the labelling that gives the largest posterior probability from the marginal distribution of $Z_j$, $j = 1, \ldots, N$. The results will be displayed graphically using a map of the MMAP cluster labels, and individual maps for the estimated probability of the location being in each cluster.

Note that there is a potential for label switching between the clusters, and hence in general it may be necessary to post-process the MCMC output using an appropriate loss function for the labels. We did not observe label switching: the clusters are identified by clearly separated mean vectors, and switching would be evident in the observed chains, which we did not see any evidence of.

## 4 Results

We run the MCMC chain for 160,000 iterations and inspect trace plots of the parameters. For these data, the burn-in period only needs to be short and the chains mix well quickly, but being cautious we take 10,000 as the burn-in period leaving 150,000 observations for the sample. We take every 5th sample to reduce dependency between the samples, which leaves 30,000 samples to base our analysis on. Plots of the sample paths can be found in the online supplementary material document (Mitchell et al., 2024). We take $\lambda = 0.1$ for the parameter controlling spatial dependency in the Potts prior, and we set $\alpha = \beta = 0.001$ for the priors on the $(\sigma^2)^{(m)}$.

The posterior means for each cluster mean $\boldsymbol{\mu}^{(m)}$ are given in Figure 7. These posterior means enable us to interpret the type of peatland which each cluster represents. The posterior mean indicated by the blue density and red density have similar amplitude distances on average for oscillations and trend gradient, but the timing measure as per the integrated warp difference is clearly very different with the blue density being earlier in the year than the sine template peak at May 5th, and the red density being later in the year after May 5th. Hence, the blue density cluster is indicative of soft/wet peatlands, whose peaks occur in winter due to swelling from the hydrological recharge following the growing season (Alshammari et al., 2020). The red density cluster, with later peaks, is indicative of drier/shrubby peatlands due to a build-up in vegetation from the growing season. The gradient trends for the red density cluster are similar to those in the blue density cluster, both being close to the arithmetic mean and similar long-term change in peatland depth.

The posterior mean of the cluster represented by the white density has a large oscillation amplitude distance suggesting that regardless of how much warping is applied to the oscillations, the

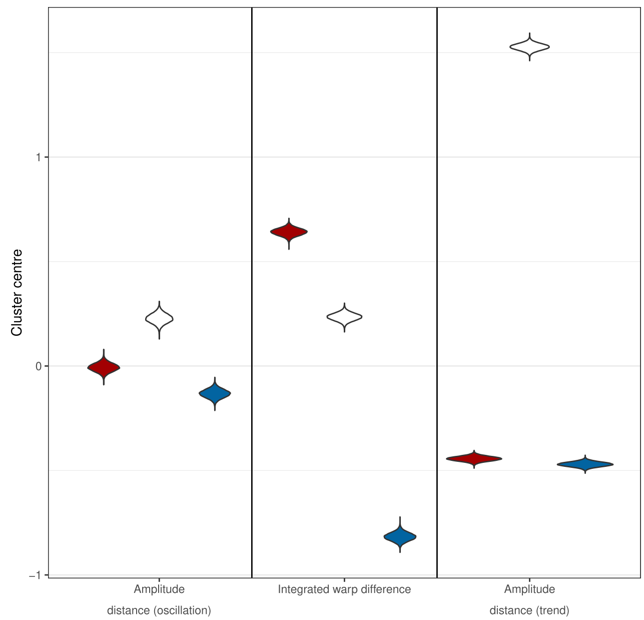


**Figure 7.** Violin plots of the posterior means for each of the cluster centres. From the positioning of the posterior means, in each panel the right-hand density represents soft/wet peatland, the left-hand density represents drier/shrubby peatland, and the middle density represents thin/modified peatland.

oscillations will not closely resemble the sinusoidal wave. The integrated warp difference is quite close to zero suggesting that warping will rarely achieve oscillations close to the sinusoidal wave and so there is little warping to avoid further penalization. There is a very large trend gradient amplitude distance, and so the trend gradient is far from the mean for the region. This cluster is indicative of thin/modified peatlands. Examples for each class can be found in the Figure 3.

Samples approximately drawn from the posterior distribution for the cluster labels are also provided by the MCMC sampling scheme. For each of the $N = 9662$ peatland locations, the proportion of samples assigned to each cluster is the estimated probability of being assigned to that peatland type. We also find the MMAP estimate, the peatland type with the highest probability, for each peatland location. For Balavreed, one of the five sub-sites in the Flow Country, the MMAP estimates and the corresponding probabilities for each peatland type are plotted in Figure 8. The plots for the remaining four sub-sites can be found in the online supplementary material (Mitchell et al., 2024).

Though there is no formal classification with which to quantify the accuracy of our classifications (according to MMAP) in the traditional way, some of the authors (RA,AB,DL) and colleagues have extensive experience with these areas of the Flow Country. Extensive field-based assessment of site condition has been undertaken by one of the authors on the same locations at a suitable spatial scale, as part of previous work specifically seeking to relate InSAR with vegetation, hydrology, and land-use (Alshammari et al., 2020; Bradley et al., 2022). As part of this work, it was clearly demonstrated that InSAR-based surface motion represents real, spatially organized dynamic behaviours relatable to eco-hydrological condition on the ground (Bradley et al., 2022). Notably, Bradley et al. (2022) demonstrated that the InSAR-based surface motion was extremely effective at predicting areas where pool systems—the most dynamic part of the system corresponding to a 'soft/wet' class—were likely to occur. The authors are confident that the current classification carries meaning in a similar way, displaying the same correspondence to expert knowledge gathered from field surveys but grounded in a more robust and principled statistical methodology.

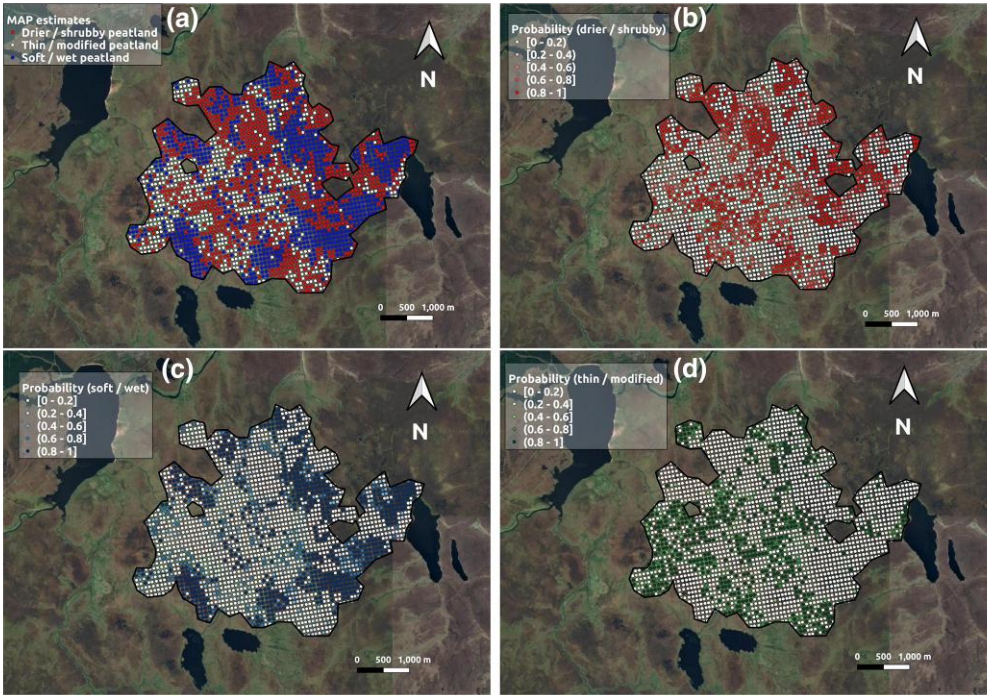


**Figure 8.** (a) MMAP estimate for peatland condition in Balavreed, one of the five sub-sites in the Flow Country. Different shadings are indicative of soft/wet peatland, drier/shrubby peatland, and thin/modified peatland. (b) probability of drier/shrubby peatland, (c) probability of soft/wet peatland, and (d) probability of thin/modified peatland (here darker represents higher probability and white low probability). Base map data: ©2022 Google. Base map imagery: ©2022 CNES/Airbus, Getmapping plc, Landsat/Copernicus, Maxar Technologies.

## 4.1 Model checking

To check our model is a good fit to our data, we simulated replicate data sets from the model using parameter configurations from our posterior samples, and inspected plots to check for consistency with our observed data. Figure 9 shows pairwise plots of draws from our model for one randomly chosen parameter configuration from the MCMC run, along with the observed data. (Five hundred randomly chosen points are plotted for clarity.) The plot shows good agreement, evidencing that our model gives a faithful representation of the data.

## 4.2 Comparison with FPCA with warping

A standard method to perform dimension reduction with functional data is functional principal component analysis (FPCA) (Ramsay & Silverman, 2005). This assumes the functions are *aligned* in time (or phase), and extracts information about *amplitude* variability. This is not the case in our application, where peak timings differ and contain valuable information about peat condition, as discussed earlier. In other words, functions should be warped before comparing amplitude across functions via standard FPCA, and the warping contains information about peak timing—recall how we have used information from the warping functions to define two of the key features in our approach. Tucker et al. (2013) proposed a framework for exploring the amplitude variability (via *vertical* FPCA) and warping variability (via *horizontal* FPCA), and this is a natural method to compare with our approach.

To investigate this, we first smoothed our time series, similar to the first step in our approach. However, we did not specify the `spar` parameter, instead using the R default value for each time series. We chose the endpoint of each series as the '$f(0)$' parameter of Tucker et al. (2013), which enables reconstruction of $f$ from $q$ since $q$ is invariant to vertical shifts.

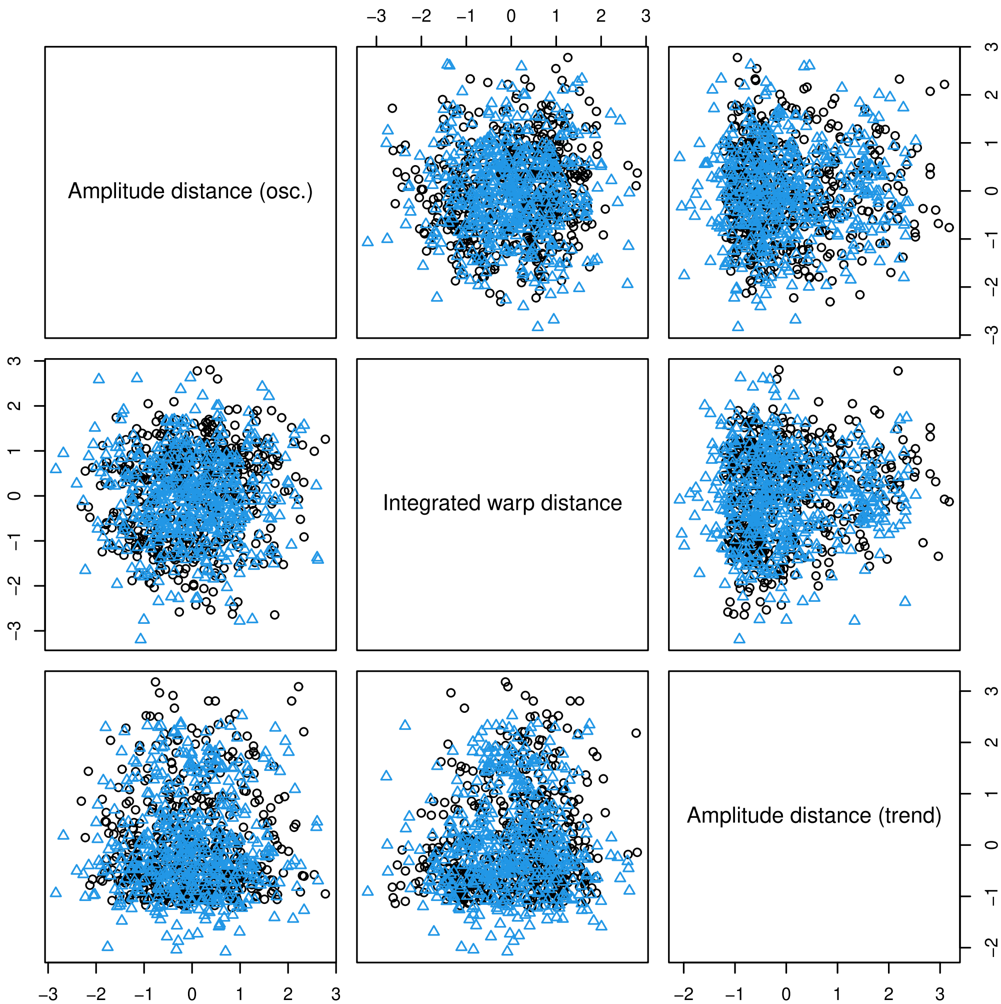


**Figure 9.** Simulated samples (triangles) and observed data (circles).

Figure 10 shows the first 4 principal components from the vertical FPCA, in the original function space. The FPCA itself is carried out in $q$-space (on the warped $q$ functions), where principal-geodesic paths corresponding to each principal component are also defined. Functions in the original function space can then be obtained via integration. The middle function in each plot of Figure 10 corresponds to the mean, and the upper and lower functions correspond to $\pm$ one step along the geodesic path for each component. These first four components account for 94.2%, 1.2%, 0.71%, and 0.59% of the amplitude variability, respectively, so the first mode of variation overwhelmingly dominates. This mode essentially reflects the variability in the endpoint of the series, which is analogous to the variability our trend measure captures.

Figure 11 shows the first 4 principal components from the horizontal FPCA. These components account for 24.0%, 15.1%, 11.8%, and 8.9% of the warping variability, respectively. Here, the first 9 components are needed in order to explain 90% of the variability, and they are not particularly easy to interpret in a simple and meaningful way for peatland scientists. It is not surprising that a relatively large number of components are needed to explain all the observed variation in warpings over a number of years, but in fact what is important is the broad classification of peak timings into winter and summer. Our simple measures effectively achieve this in a low-dimensional, parsimonious manner.

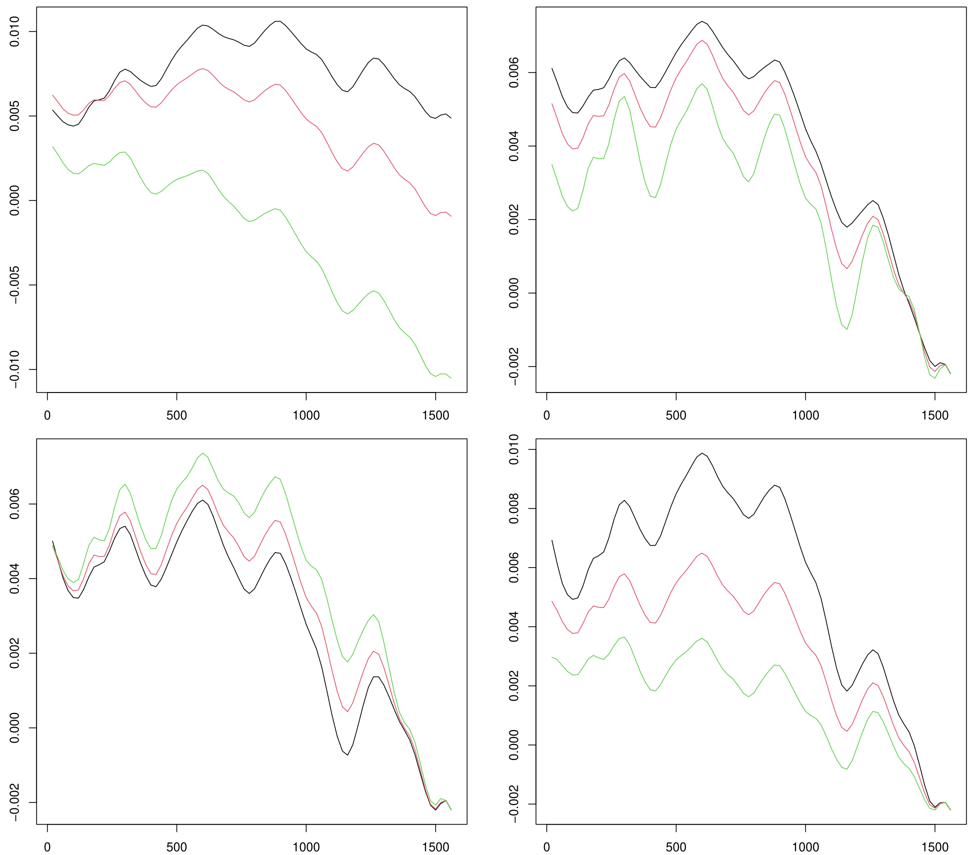

**Figure 10.** Vertical FPCA principal components and geodesic paths for the first four components, clockwise from top left. The middle line in each plot corresponds to the mean, and the upper and lower functions correspond to ± one step along the geodesic path for each component.

To produce a classification, we used the first 3 vertical FPCA coefficients and the first 9 horizontal FPCA coefficients to represent each location as a vector in $\mathbb{R}^{12}$, and obtained a classification via Ward's method on the Euclidean distances between these points. To compare the clusters with ours, we relabelled in such a way as to maximize the number of locations classified to the corresponding clusters from each method.

Overall, we find that 54% of locations are classified the same, but interestingly we find quite a strong agreement between our soft/wet cluster and the corresponding cluster from FPCA (see Table 2). If one delves deeper, we find that the locations which 'disagree' are generally the ones which are more uncertain according to our MMAP probabilities, whereas the ones which 'agree' tend to have very high MMAP probabilities (>0.9) of being soft/wet. The classification map from FPCA for Balavreed can be seen in Figure 12, which can be compared to those from our method. There is less obvious agreement between the other two classes, with thin/modified having the poorest agreement.

The functional PCA approach can be viewed as a pure exploration of the variability in the functional data. A potential disadvantage is that the modes of variation may not have an obvious simple interpretation, and clustering on the coefficients may not produce useful clusters for end users. In our approach, we tackle this head-on, by defining measures that are designed to extract features that will be physically interpretable and relate to the scientific knowledge of peat motion behaviour. A pure exploration of the variability is undoubtedly an extremely valuable exercise, which we view very much as a complementary tool to our proposed method.

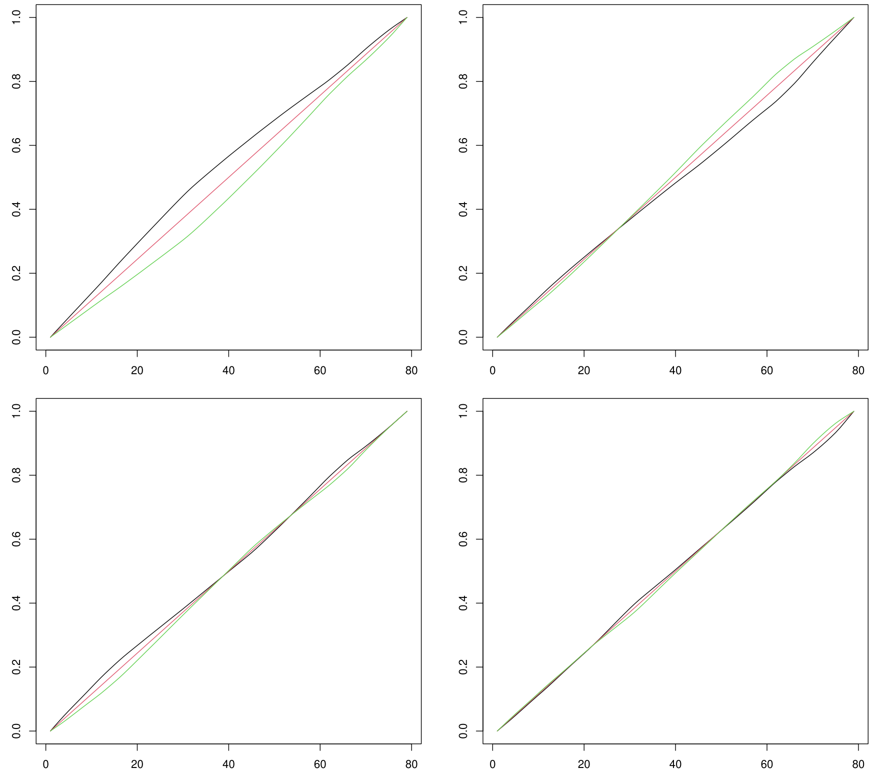

**Figure 11.** Horizontal FPCA principal directions.

**Table 2.** Contingency table of classifications from OODA and FPCA

| | | OODA | | | |
|---|---|---|---|---|---|
| | | Drier/shrubby | Thin/modified | Soft/wet | |
| | Drier/shrubby | 1,786 | 713 | 895 | 52.6% |
| FPCA | Thin/modified | 1,525 | 1,006 | 223 | 36.5% |
| | Soft/wet | 585 | 500 | 2,429 | 69.1% |
| | | 45.8% | 45.3% | 68.5% | 54.0% |

*Note*. The final column gives the percentage of each FPCA class that agrees with OODA, and the final row gives the percentage of each OODA class that agrees with FPCA. The overall percentage agreement is 54.0%.

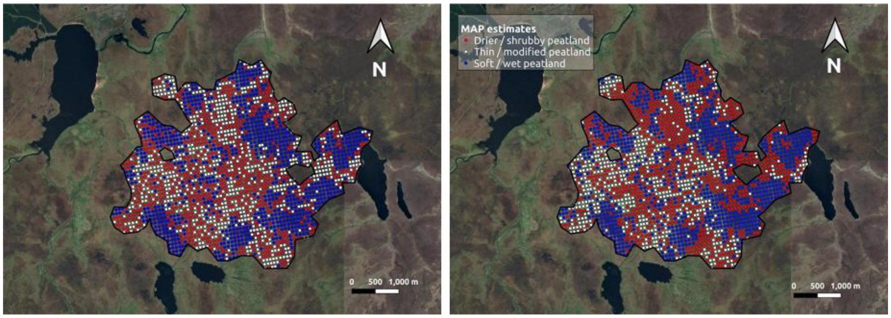


**Figure 12.** FPCA classifications (left) and OODA classifications (right) for Balavreed. Base map data: ©2022 Google. Base map imagery: ©2022 CNES/Airbus, Getmapping plc, Landsat/Copernicus, Maxar Technologies.

## 5 Discussion and conclusions

Our analysis of peatland surface motion between 12th March 2015 and 1st July 2019 from InSAR data of five sub-sites in the Flow Country identified areas of soft/wet, drier/shrubby, and thin/modified peatland. Areas identified as thin/modified with a high level of certainty can be earmarked for future restoration and monitored in the meantime. Identified areas of soft/wet and drier/shrubby peatlands must be protected to continue to act as a carbon sink. The areas identified are sensible from ground observations whilst time and expense of conducting field studies are saved. Additionally, this method is far less computationally demanding for larger areas compared to previous methods to assess peatland condition (Bradley et al., 2022).

The first part of the method involved constructing measures based on distances between the registered oscillations and an oscillatory template and the trend gradients and a trend template. To register the oscillations, a reference function was required to base our analysis on. Guided by expert knowledge of peatlands, we used a sinusoidal template with peaks in the middle of astronomical spring and troughs in the middle of autumn which is known to be midway between distinct types of peat: shrubby marginal peats peak in summer and soft wet sphagnum peaks in winter. It would be unusual for the peatland surface motion to be exactly represented by this wave once warped, nevertheless we would expect something close to this for the warped versions of the soft/wet and drier/shrubby peatlands due to the hydrological recharge of the shrubs in the dry peatlands during the summer growing season and the soft/wet peatlands during the winter.

To find the peatland oscillation motion attributed to the peat substrate and not the motion from the underlying hydrology, it would be useful to gather precipitation data for the area and incorporate this, alongside topographic information, into our preconception of the local oscillations for specific years.

An alternative approach, which would avoid selecting the reference function, is to take the reference function as the Karcher mean (Srivastava, Wu, et al., 2011) of the set of oscillation SRVF functions. However, some concerns include the influence of non-oscillatory regions such as the thin/modified peat, and also the effect of outliers.

We restricted the level of warping for the oscillations using smoothness penalty $\lambda = 0.1$ to allow peak timing to be shifted, at most, approximately half a year. It would be unreasonable to suggest the timing between two peatland surface motion time series differed by over half a year since we expect the oscillatory cycle to repeat annually. Further work could entail switching to a Bayesian approach for functional data analysis (Cheng et al., 2016) and putting a prior such as half-Cauchy on $1/\lambda$ to favour larger levels of $\lambda$. A fully Bayesian approach could propagate other sources of uncertainty due to variations in the registration of the functional data (Kurtek, 2017; Lu et al., 2017; Tucker et al., 2021).

In addition to the oscillations, our analysis is also based on the trend gradients, where we chose to not include any warping. Little is known about the trends in the region except that there should be very little difference in the timing of the trends. Hence, there is very little difference between taking the arithmetic mean or the Karcher mean as the reference function, but finding the Karcher mean is far more computationally demanding.

An extension of the model is to consider spatially correlated noise rather than i.i.d noise. For example, we have experimented with a conditional autoregression (CAR) model (Clayton & Kaldor, 1987) where there is correlation between errors if they are from neighbouring sites. However, this error model does not perform well, as there will often be correlations across cluster boundaries. The two types of smoothness (cluster label Potts model and CAR error model across the whole region) are in competition, and from experiments this results in less smooth cluster labels than when using the i.i.d. noise model. A CAR model that is dependent on neighbours being in the same cluster is computationally challenging, as the normalizing constant for the likelihood would need recomputing at each iteration of the MCMC scheme. We will investigate this model in future work, but note that the correlations captured through the Potts model and i.i.d errors work well in this application.

One important question is whether the OODA part of our method, based on the five sub-sites in the Flow Country, can be extended to other regions. These sites were selected because they contained varied blanket peatland conditions in a near-natural state. For other regions, the underlying analysis of the oscillations based on a fixed sinusoidal template can be transferred. The overall trend was based on the arithmetic mean of the trend gradients of the surface motion time series in the five sub-sites. The speed at which the overall trend declines within these regions increases until late 2017, where it continues to rapidly decrease at the same pace. This could be a result of extreme weather events during 2017 and 2018 from which the peatland has yet to recover (Fenner & Freeman, 2011; Stirling et al., 2020; Undorf et al., 2020). If the trend gradients were included in the clustering, a different region, which may not have exhibited such weather events, with different trend gradients would be classed as degraded when compared to the Flow Country. Topography, climatology, and geography of a landscape can all influence peatland surface motion. A potentially fruitful area of research work would be to study different sites with varied peatland conditions, climate and topography and compare their means.

Related to the above point, there is the question of how to validate results in new areas where such excellent local knowledge is not available. In other work, we have demonstrated that the same classification system can effectively be applied to areas where there was no prior knowledge, using a simple field-validation tool and management information. In terms of further validation in completely new areas, the simplicity of the wide applicability of our physically and ecologically grounded conceptual framework gives us confidence that the algorithm can be reliably deployed for blanket bogs and raised bogs within a similar climate space. Should we wish to deploy a similar approach to a different peatland system (e.g. fen peat in which the hydrology is controlled by groundwater, tropical peat with a different pattern of wet and dry season, or peat experiencing severe freezing in winter), then we would start by first developing a new conceptual framework reflecting expected peatland behaviour, and adapt our choice of measures and condition classes to reflect scientific knowledge of key behaviours.

We fixed the number of clusters in this article to be $K = 3$, to reflect our knowledge of the peatlands largely consisting of soft/wet peatlands, drier/shrubby peatlands, and thin/modified peatlands. If we were to move to a different region, those which have significantly different trend gradients may be classed as thin/modified, although they may still contain the oscillatory behaviour found in soft/wet or drier/shrubby peatlands. A potentially fruitful area of work would be to investigate the possibility of further sub-clusters by varying the number of clusters, either by penalizing for the number of clusters included or creating a tree structure, for example by first classifying according to trend gradients followed by classifying according to oscillations. Penalization can be incorporated as a Poisson or Gamma prior placed on the number of clusters in the prior model when smoothing to account for neighbours.

Another interesting question, which is beyond the scope of the present paper, is to assess the peatlands' response to restoration by examining the behaviour of the function before restoration and the behaviour afterwards. Testing the success of restoration would instead require a sliding window approach to find if and when the functional behaviour changes. In addition to restoration success, a sliding window would allow peatland condition to be assessed through time. If we continue to use a sinusoidal wave as the template for the oscillations and cluster according to the measures inside the sliding window, the changepoint could be identified as the point where the peatland type probabilities significantly change. This would not only enable reflection on the success of past restoration, it would also indicate best restoration practices for varied types of peatland environments, to support future action.

We have provided a new method for the analysis of environmental time series which would also be appropriate in many more general applications, such as monitoring ocean temperatures, El Niño and the Southern Oscillation (ENSO), phenology, and gas emissions, where seasonal cycles and trends are present. Using the same Object Oriented Data Analysis approach to building appropriate methodology is anticipated to be a valuable approach in such applications.

*Conflict of interests:* The authors have no conflicts of interest to declare.

## Funding

This work has been funded by the Natural Environment Research Council (NERC) [grant number NE/T010118/1], as part of the project 'Developing a statistical methodology for the assessment and management of peatland (StAMP)' in the Landscape Decisions Programme. Roxane

Andersen is also supported by a Leverhulme Research Leadership Award from the Leverhulme Trust [grant number RL-2019-002].

## Data availability

The InSAR surface motion time series for the five sub-sites studied in this article and the code to produce the analyses of the data are available in the online supplementary material (Mitchell et al., 2024).

## Supplementary material

Supplementary material is available online at *Journal of the Royal Statistical Society: Series C*.

## References

Alshammari L., Boyd D. S., Sowter A., Marshall C., Andersen R., Gilbert P., Marsh S., & Large D. J. (2020). Use of surface motion characteristics determined by InSAR to assess peatland condition. *Journal of Geophysical Research: Biogeosciences*, *125*(1), 1–15. https://doi.org/10.1029/2018JG004953

Alshammari L., Large D. J., Boyd D. S., Sowter A., Anderson R., Andersen R., & Marsh S. (2018). Long-term peatland condition assessment via surface motion monitoring using the ISBAS DInSAR technique over the Flow Country, Scotland. *Remote Sensing*, *10*(7), 1–24. https://doi.org/10.3390/rs10071103

Andersen R., Cowie N., Payne R. J., & Subke J. A. (2018). The flow country peatlands of Scotland. *Mires and Peat*, *23*, 1–4. https://doi.org/10.19189/MaP.2018.OMB.381.

Andersen R., Farrell C., Graf M., Muller F., Calvar E., Frankard P., Caporn S., & Anderson P. (2017). An overview of the progress and challenges of peatland restoration in Western Europe. *Restoration Ecology*, *25*(2), 271–282. https://doi.org/10.1111/rec.2017.25.issue-2

Bain C. G., Bonn A., Stoneman R., Chapman S., Coupar A., Evans M., Gearey B., Howat M., Joosten H., Keenleyside C., & Labadz J. (2011). *IUCN UK commission of inquiry of peatlands* (Technical report). IUCN UK Peatland Programme, Edinburgh.

Bauer M., Bruveris M., & Michor P. W. (2016). Why use Sobolev metrics on the space of curves. In P. K. Turaga, & A. Srivastava (Eds.), *Riemannian computing in computer vision* (pp. 233–255). Springer International Publishing.

Bradley A., Andersen R., Marshall C., Sowter A., & Large D. J. (2022). Identifcation of typical ecohydrological behaviours using InSAR allows landscape-scale mapping of peatland condition. *Earth Surface Dynamics*, *10*(2), 261–277. https://doi.org/10.5194/esurf-10-261-2022

Chapman S. J., Bell J., Donnelly D., & Lilly A. (2009). Carbon stocks in Scottish peatlands. *Soil Use and Management*, *25*(2), 105–112. https://doi.org/10.1111/sum.2009.25.issue-2

Cheng W., Dryden I. L., & Huang X. (2016). Bayesian registration of functions and curves. *Bayesian Analysis*, *11*(2), 447–475. https://doi.org/10.1214/15-BA957

Clayton D., & Kaldor J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, *43*(3), 671–681. https://doi.org/10.2307/2532003

DBEIS (2019). *Final UK greenhouse gas emissions national statistics* (Technical report). Department for Business, Energy & Industrial Strategy, available at www.data.gov.uk/dataset/9568363e-57e5-4c33-9e00-31dc528fcc5a/final-uk-greenhouse-gas-emissions-national-statistics.

Doucet A., Godsill S. J., & Robert C. P. (2002). Marginal maximum a posteriori estimation using Markov chain Monte Carlo. *Statistics and Computing*, *12*(1), 77–84. https://doi.org/10.1023/A:1013172322619

Evans M., & Lindsay J. (2010). Impact of gully erosion on carbon sequestration in blanket peatlands. *Climate Research*, *45*(1), 31–41. https://doi.org/10.3354/cr00887

Fenner N., & Freeman C. (2011). Drought-induced carbon loss in peatlands. *Nature Geoscience*, *4*(12), 895–900. https://doi.org/10.1038/ngeo1323

Geman S., & Geman D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-6*(6), 721–741. https://doi.org/10.1109/TPAMI.1984.4767596.

Grayson R., Holden J., & Rose R. (2010). Long-term change in storm hydrographs in response to peatland vegetation change. *Journal of Hydrology*, *389*(3–4), 336–343. https://doi.org/10.1016/j.jhydrol.2010.06.012

Green P. J., & Richardson S. (2002). Hidden Markov models and disease mapping. *Journal of the American Statistical Association*, *97*(460), 1055–1070. https://doi.org/10.1198/016214502388618870

Green P. J., & Silverman B. W. (1993). *Nonparametric regression and generalized linear models: A roughness penalty approach*. Chapman and Hall.

Hastie T., Tibshirani R., & Friedman J. (2008). *The elements of statistical learning: Data mining, inference, and prediction*. Springer.

Howie S. A., & Hebda R. J. (2017). Bog surface oscillation (mire breathing): A useful measure in raised bog restoration. *Hydrological Processes*, *32*(11), 1518–1530. https://doi.org/10.1002/hyp.11622

Ingram H. (1983). Mires: Swamp, bog, fen and moor. In *Ecosystems of the world*. Elsevier Science.

Ise T., Dunn A. L., Wofsy S. C., & Moorcroft P. R. (2008). High sensitivity of peat decomposition to climate change through water-table feedback. *Nature Geoscience*, *1*(11), 763–766. https://doi.org/10.1038/ngeo331

Jacques J., & Preda C. (2014). Functional data clustering: A survey. *Advances in Data Analysis and Classification*, *8*(3), 231–255. https://doi.org/10.1007/s11634-013-0158-y

Joosten H. (2012). *The global peatland $CO_2$ picture* (Technical report). Greifswald University.

Kurtek S. (2017). A geometric approach to pairwise Bayesian alignment of functional data using importance sampling. *Electronic Journal of Statistics*, *11*(1), 502–531. https://doi.org/10.1214/17-EJS1243

Lindsay R., Charman D. D., Everingham F., O'Reilly R. M., Palmer M. A., Rowell T. A., & Stroud D. A. (1988). *The flow country: The peatlands of Caithness and Sutherland* (Technical report). Joint Nature Conservation Committee. https://repository.uel.ac.uk/item/86qqv

Lu Y., Herbei R., & Kurtek S. (2017). Bayesian registration of functions with a Gaussian process prior. *Journal of Computational and Graphical Statistics*, *26*(4), 894–904. https://doi.org/10.1080/10618600.2017.1336444

Mahdiyasa A. W., Large D. J., Muljadi B. P., Icardi M., & Triantafyllou S. (2022). MPeat—A fully coupled mechanical-ecohydrological model of peatland development. *Ecohydrology*, *15*(1), 1–18. https://doi.org/10.1002/eco.2361

Marron J. S., & Alonso A. M. (2014). Overview of object oriented data analysis. *Biometrical Journal*, *56*(5), 732–753. https://doi.org/10.1002/bimj.v56.5

Marron J. S., & Dryden I. L. (2022). *Object oriented data analysis*. Chapman and Hall.

Marshall C., Sterk H. P., Gilbert P. J., Andersen R., Bradley A. V., Sowter A., Marsh S., & Large D. J. (2022). Multiscale variability and the comparison of ground and satellite radar based measures of peatland surface motion for peatland monitoring. *Remote Sensing*, *14*(336), 1–22. https://dx.doi.org/10.3390/rs14020336

Mitchell E. G., Dryden I. L., Fallaize C. J., Andersen R., Bradley A. V., Large D. J., & Sowter A. (2024). Supplementary material for 'Object oriented data analysis of surface motion time series in peatland landscapes'. *Journal of the Royal Statistical Society, Series C*.

Pawson R. R., Evans M. G., & Allott T. E. H. A. (2012). Fluvial carbon flux from headwater peatland streams: Significance of particulate carbon flux. *Earth Surface Processes and Landforms*, *37*(11), 1203–1212. https://doi.org/10.1002/esp.v37.11

Ramsay J., & Silverman B. (2005). *Functional data analysis* (2nd ed.). Springer.

R Core Team (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Rochefort L., & Andersen R. (2017). Global peatland restoration after 30 years: Where are we in this mossy world? *Restoration Ecology*, *25*(2), 269–270. https://doi.org/10.1111/rec.2017.25.issue-2

Roulet N. T. (1991). Surface level and water table fluctuations in a subarctic fen. *Arctic and Alpine Research*, *23*(3), 303–310. https://doi.org/10.1080/00040851.1991.12002849

Sowter A., Bateson L., Strange P., Ambrose K., & Syafiudin M. F. (2013). DInSAR estimation of land motion using intermittent coherence with application to the South Derbyshire and Leicestershire coalfields. *Remote Sensing Letters*, *4*(10), 979–987. https://doi.org/10.1080/2150704X.2013.823673

Srivastava A., & Klassen E. (2016). *Functional and shape data analysis*. Springer.

Srivastava A., Klassen E., Joshi S., & Jermyn I. H. (2011). Shape analysis of elastic curves in Euclidean spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(7), 1415–1428. https://doi.org/10.1109/TPAMI.2010.184

Srivastava A., Wu W., Kurtek S., Klassen E., & Marron J. S. (2011). *Registration of functional data using Fisher–Rao metric* (Technical report). Florida State University.

Stirling E., Fitzpatrick R. W., & Mosley L. M. (2020). Drought effects on wet soils in inland wetlands and peatlands. *Earth-Science Reviews*, *210*, 1–15. https://doi.org/10.1016/j.earscirev.2020.103387

Tai L.-H., Srivastava A., & Gallivan K. A. (2017). *Trend and variable-phase seasonality estimation from functional data* (Technical report). arXiv, https://arxiv.org/abs/1704.07358.

Tampuu T., Praks J., Uiboupin R., & Kull A. (2020). Long term interferometric temporal coherence and DInSAR phase in Northern peatlands. *Remote Sensing*, *12*(10), 1–22. https://doi.org/10.3390/rs12101566

Tucker J. D. (2021). *fdasrvf: Elastic functional data analysis*. R package version 1.9.7.

Tucker J. D., Lyndsay S., & Chowdhary K. (2021). Multimodal Bayesian registration of noisy functions using Hamiltonian Monte Carlo. *Computational Statistics & Data Analysis*, *163*, 1–19. https://doi.org/10.1016/j.csda.2021.107298

Tucker J. D., Wu W., & Srivastava A. (2013). Generative models for functional data using phase and amplitude separation. *Computational Statistics and Data Analysis*, *61*, 50–66. https://doi.org/10.1016/j.csda.2012.12.001

Undorf S., Allen K., Hagg J., Li S., Lott F. C., Metzger M. J., Sparrow S. N., & Tett S. F. B. (2020). Learning from the 2018 heatwave in the context of climate change: Are high-temperature extremes important for adaptation in Scotland? *Environmental Research Letters*, *15*(3), 1–11. https://doi.org/10.1088/1748-9326/ab6999

Verhoeven J. T. A. (2014). Wetlands in Europe: Perspectives for restoration of a lost paradise. *Ecological Engineering*, *66*, 6–9. https://doi.org/10.1016/j.ecoleng.2013.03.006

Waddington J. M., Kellner E., Strack M., & Price J. S. (2010). Differential peat deformation, compressibility, and water storage between peatland microforms: Implications for ecosystem function and development. *Water Resources Research*, *46*(7), 1–12. https://doi.org/10.1029/2009WR008802

Wang H., & Marron J. S. (2007). Object oriented data analysis: Sets of trees. *The Annals of Statistics*, *35*(5), 1849–1873. https://doi.org/10.1214/009053607000000217

Ward J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*(301), 236–244. https://doi.org/10.1080/01621459.1963.10500845

Wu W., & Srivastava A. (2011). An information-geometric framework for statistical inferences in the neural spike train space. *Journal of Computational Neuroscience*, *31*(3), 725–748. https://doi.org/10.1007/s10827-011-0336-x