Bayesian Spatiotemporal Modelling: An Application to Glacier Satellite Imagery

by

Michael James McCurdy

B.Sc. Hons, University of British Columbia, 2015

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
MATHEMATICS

UNIVERSITY OF NORTHERN BRITISH COLUMBIA

December 2020

Abstract

Glaciers hold 1.7% of the Earth's total water supply, but they contain 68.7% of its freshwater. Given the global warming trend, accurate and recent inventory is necessary to assess glacial changes over time. However, frequent cloud and debris cover often make it difficult to determine the glacier's exact edge. Multispectral Landsat 8 imagery along with data from the Global Land Ice Velocity Extraction (GoLIVE) project are combined to to create a Bayesian multivariate general additive model of the glaciers surrounding Mount Rainier, with Autoregressive Moving Average (ARMA) and Gaussian processes used to model the temporal and spatial autocorrelations. Using root mean square error and Watanabe-Akaike information criterion, all 42 combinations of ARMA models up to 4 total parameters and exponential, Matérn $\frac{1}{2}$, and spherical covariance kernels were compared. The ARMA(3,1) processes with the exponential Gaussian process kernel was determined to be the best fit model. Gaussian mixture models, hierarchical clustering, hard and soft K-means clustering, and support vector machines are used to classify the posterior distribution. The hard K-means algorithm was the best classifier, and it accurately predicted 85.1% of the glaciers, compared to 68.8% from a univariate classification on the Red/SWIR band ratio.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

AR	Autoregressive
ARD	Analysis Ready Data
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
BLUE	Best Linear Unbiased Estimator
DEM	Digital Elevation Model
EM	Expectation-Maximisation
GAM	General Additive Model
GP	Gaussian Process
GLIMS	Global Land Ice Measurements from Space
GoLIVE	Global Land Ice Velocity Extraction from Landsat 8
HMC	Hamiltonian Monte Carlo
MA	Moving Average
MAR	Missing at Random
MCAR	Missing Completely at Random
MCMC	Markov Chain Monte Carlo
MNAR	Missing Not at Random
NIR	Near Infrared
NDSI	Normalised Difference Snow Index
PCA	Principal Components Analysis
RMSE	Root Mean Squared Error

SAR	Synthetic Aperture Radar
ST	Surface Temperature
SUR	Seemingly Unrelated Regression
SWIR	Short Wave Infrared
TIRS	Thermal Infrared
USGS	United States Geological Survey
WAIC	Watanabe-Akaike Information Criterion

Acknowledgements

To my parents, thank you for your love and support. I would not be here without you both. Jeff and Laura, I always look forward to speaking with you, I'm looking forward to many more trips into the mountains. Sarah, I wish you the best in SF, I'm very proud of you. Kevin, thank you for your mentorship; your patient guidance was crucial to finishing this work. Neil and Roger, thank you for your treasured counsel and wisdom. Gillian, Ryan, Andy, and Tiegan thank you for the many many miles, they were sorely needed.

And finally to Julia. Thank you for being a ray of sunshine. I'm looking forward to our life in Calgary.

Chapter 1

Introduction

Glaciers hold less than 2% of the world's total water volume; however they contain more than 68% of its freshwater supply. [1] Given the large and often remote nature of glaciers, it can be difficult to obtain reliable outlines of glacier extents. Manual delineation of the glacial outline can be time consuming and relies on the competence and experience of the technician [2]. Furthermore, clouds and debris cover may obscure the glacier, making the underlying ice hard to identify and model.

This modelling can be performed in a Bayesian framework. Bayesian models are based upon developing a model structure for the data and the underlying parameters. These models can be used to make new predictions or understand the underlying process, but they are often expensive computationally, so tact with model complexity and approximation methods is essential.

The work outlined in this thesis focuses on investigating methods to perform

Figure 1.1: Debris-covered toe of Emmons Glacier on Mount Rainier, August 1984 (Photo by Topinka Lyn).

efficient Bayesian multivariate spatiotemporal analysis with application to the glaciers surrounding Mount Rainier (46.8523° N, 121.7603° W), an isolated stratovolcano in Washington State, USA. This work can be summarised by two parts. The first part creates a multivariate analysis that models the spatiotemporal date to fill in gaps in the missing data. The analysis is done under a Bayesian paradigm, with Gaussian process and autoregressive moving averages modelling the spatial and temporal data. Seemingly unrelated regression is used to model the multivariate response variables. The second section is a classification of the previous section's model output. The K-means, support vector machine, hierarchical clustering algorithms, and Gaussian mixture models are used for clustering the multivariate structure into glacier. The second chapter of this work gives a summary

on current data and methods used within the realm of geostatistics. The third chapter presents the data used in this analysis. The fourth chapter formalises the spatiotemporal model used in this work. The fifth chapter presents the results, and finally the sixth chapter provides a discussion and outlines future work.

Chapter 2

Literature Review

2.1 Methods for Glacier Mapping

Glacier mapping, or determining the outline of a glacier, is done through multiple ways with a variety of data. The available data can be grouped as passive and active data. Active data includes lidar and radar data. While some success has been found mapping glaciers using lidar, initially a portmonteau of "light" and "radar" [3, 4], the data acquisition can be expensive. Synthetic aperature radar involves repeated pulses of radio waves, of which the echo is recorded and used to create a reconstruction of the landscape. While the radar data is able to "see" through clouds, there can often be quality issues in mountainous terrain. [5].

Passive data includes optical and thermal satellites. Optical sensors collect solar radiation reflected by the earth's surface in the visible and near infrared bands of the electromagnetic spectrum. Satellites short revist times and large swaths

make them useful for glacier mapping [6]. The Landsat 8 satellite, launched in 2013, has 11 multi spectral bands with resolutions of 15m x 15m and 100m x 100m, and a revisit period of 16 days, to minimise the seasonal snow cover, acquisition dates are typically limited to between August and mid September. This limits the viable images per year to 3-4 images. [7]. Table 2.1 shows the bands, wavelengths, and spatial resolutions of the Landsat 8 satellite.

Band ratios or standardised difference ratios, typically calculated from the green, red, Near Infrared (NIR), and Short Wave Infrared (SWIR) bands, are commonly used as indices to differentiate between glaciers and the surrounding bedrock [8–11]. With this multispectral data, thresholding is a common technique, in which either a predetermined threshold is declared [9, 10], or the sample is inspected and a threshold is determined to bisect the sample [5].

As glaciers continue to melt, rock and debris cover increasing amounts of a glacier's surface. Depending on the thickness of the debris cover, it can have

Bands	Wavelength(nm)	Resolution(m)
Band 1 - Coastal aerosol	0.43-0.45	30
Band 2 - Blue	0.45-0.51	30
Band 3 - Green	0.53-0.59	30
Band 4 - Red	0.64-0.67	30
Band 5 - Near Infrared (NIR)	0.85-0.88	30
Band 6 - Short Wave Infrared (SWIR) 1	1.57-1.65	30
Band 7 - SWIR 2	2.11-2.29	30
Band 8 - Panchromatic	0.50-0.68	15
Band 9 - Cirrus	1.36-1.38	30
Band 10 - Thermal Infrared (TIRS) 1	10.6-11.19	100
Band 11 - Thermal Infrared (TIRS) 2	11.50-12.51	100

Table 2.1: Multispectral Landsat 8 data

both an albedo effect and an insulation effect. Where debris cover is thin, it can decrease the albedo, or proportion of incoming radiation that is absorbed by a surface, causing the glacier to melt more quickly. As the thickness of the debris cover increases, the debris forms barrier between the underlying ice and the atmosphere, causing the glacier to melt more slowly. Withe the current multispectral thresholding techniques, debris cover is commonly a confounding variable, making surrounding bedrock and glacier debris indistinguishable.

To help with the debris cover, various techniques and new data sources have been developed. Digital Elevation Models (DEM) have been used to create a slope or elevation variable to help model debris cover [10,12]. Synthetic Aperture Radar (SAR) data has also been used with offset tracking, in which motion between two images is calculated using a cross-correlation algorithm [13], or interferometry. In interferometry, the difference in two images is used to model the spatial change [14, 15].

The National Snow and Ice Data Center released a dataset as part of the glacial velocity titled Global Land Ice Velocity Extraction from Landsat 8 (GoLIVE) project. This covers all glaciers with an area over 5 km^2. Pairs of Landsat 8 images are superimposed upon one another. Using an image correlation algorithm, pixels corresponding to ice velocity are generated. [16]. Some recent success has been found using this data source [17, 18]; however, the data is noisy, and given it is a product of optical satellite data, the data can be frequently obstructed by cloud cover.

Given a set of spatial images aquired over multiple years, mapping glaciers

can be done through a spatiotemporal paradigm. The simplest method uses an average at each spatial site to flatten the spatiotemporal data into spatial data [17, 19]. In a similar nature, some researchers use the year as a coefficient in linear regression [20, 21]. These simple methods help efficiently model the glacier data, although there is a trade off in complexity. More complex models such as Autoregressive Moving Average (ARMA) [22] and (local) polynomial regression [23, 24] are also in use for deliniation of glacier outlines.

Despite multiple variables sampled at each pixel, few true multivariate models are used to model the glacier data. Frequently, some supervised classification techniques are employed, such as multivariable logistic regression [25]; however, caution must be taken due to frequent high correlation among the covariables. In some research, the multivariate data is reduced through principal components analysis (PCA) [26], similar to the previously mentioned band ratios and differences. Co-Kriging is also used, where the undersampled singular variable of note is modelled by highly sampled multivariate covariates.

Classification methods are varied and include the previously mentioned thresholding and multiple logistic regression. In addition, object-based clustering [12, 27], in which objects are based upon similar sample values and proximity, is often used and is akin to hierarchical clustering. Multivariate clustering techniques such as K-means and K-medoids, which minimise the error around cluster mean or median respectively, [28, 29] are also employed. Machine learning techniques, such as support vector machines and neural nets, have also been applied to glacier classification with some success [28, 30]. Instead of directly classifying data, the

Expectation-Maximisation (EM) algorithm is often used when there is missing or unlabelled data. The theory is that it is easier to estimate and model using a missing latent parameter, than to model the distribution directly. In this case, instead of classifying the glacier directly, the EM algorithm attempts to identify the unknown parameters of a mixture distribution, which then can be used to classify the glacier [15, 31]. Manual delineation is also used to define a glacier's extent, in which an expert defines the glacier edge based on a visual interpretation of the glacier edge. This method is typically performed using colour composites from optical satellites. Manual deliniation may get high results; however, it can be time consuming and not applicable for the delineation of multiple glaciers [32].

2.2 Contribution of this Work

Overall, spatial methods are well utilised in current glacier remote sensing research, typically through a Kriging variant. The spatiotemporal framework is typically used to impute missing data with a temporally adjacent sample. Additionally, the majority of research models are developed with a univariate response.

The contribution of this work is to broaden the current methods in use in three ways. Firstly, this work is done within a Bayesian paradigm that allows the model parameters to vary and captures the uncertainties involved in the modelling and final classification of the glacier rather than relying on the user's fixed knowledge of the spatiotemporal processes, as implicit in more commonly used methods. Secondly, this work avoids flattening the data in order to maintain its spatiotem-

poral nature. Finally, the model developed in this work has a multivariate response, to account for correlations within the errors of the response variables and to assist with the final glacier classification.

Chapter 3

Data

The glaciers surrounding Mount Rainier were chosen for a variety of reasons. Firstly, it is a fairly isolated group of glaciers, so there is a clear split between glacier and surrounding area. Secondly, there was a recent glacier inventory of Mount Rainier in 2015 shown in Figure 3.1. Finally, as it is located in the United States, Landsat 8 Analysis Ready Data (ARD) is available. The United States Geological Survey (USGS) releases the ARD as preprocessed data that has been tiled, and top of atmosphere and atmospherically corrected for direct analysis and immediate use. The data is also georeferenced so that the internal coordinate reference system matches between scenes. [33]

The initial plan for this research involved incorporating radar data and interferometry; however, there were issues with coregistering the SAR scans due to poor quality DEM data in mountainous regions. As a result only optical data from the Landsat 8 satellite and the GoLIVE project, also derived from Landsat 8

satellite, were used for this work. The data was clipped to a bounding box 26 x 20 km immediately surrounding Mount Rainier.

3.1 Preprocessing

Some further processing for the ARD and GoLIVE data was needed. The USGS uses the Function of Mask (Fmask) algorithm to determine cloud cover on pixels. Theoretically, the SWIR band is able to distinguish between clouds and underlying ice; however, clouds made of ice crystals may cause the two to be indistinguishable [34]. Along with the multispectral data outlined in Table 2.1, the USGS releases a quality assessment band, which show the cloud and cloud shadow cover results of the Fmask algorithm. This was used to remove cloud and cloud shadows. Figures 3.2 and 3.3 show the pattern of missing data, and Table 3.1 shows the proportion of missing data per year. The data is relatively complete in years 2013, 2016, and 2017. In 2015, a pair of clouds covers up most of the glacier, causing most of the image to be removed. In 2019, the most cloud-free image is one that bisects the glacier, leaving roughly a third of the image complete. In 2014, the western half of the image is covered in clouds causing half of the image to be removed.

The resolutions between the ARD and GoLIVE data also differed. the ARD data is 30m x 30m, while the GoLIVE data is 100m x 100m. Given that three quarters of the data is at a higher resolution, the GoLIVE data was resampled to 30m x 30m using a nearest neighbour algorithm. The data was clipped to a approximate

20km x 20km area surrounding Mount Rainier, resulting in approximately 800,000 observations.

The additional processing of the ARD and GoLIVE data is as follows. Firstly, all ARD and GoLIVE images were downloaded from 2013-2019 for the period between August 01 to September 21. Data access data was March 2019 for years 2013-2018, and October 2019 for 2019. Secondly, data was clipped to the area immediately surrounding Mount Rainier Thirdly, cloud and cloud shadows were removed using the quality assessment band. Fourthly, for each year, the most complete image was selected. Fifthly, ARD and GoLIVE images from each year were combined to create a multivariate spatial matrix. Finally, images were stacked together to create a multivariate spatiotemporal matrix.

Note that the release date for the ARD program was November 2017. Given its relative infancy, by March 2019 not all Landsat 8 imagery for 2013-2018 was updated into the ARD program. As a result, more cloud-free images for 2013-2018 have been added after the data download and processing was completed.

Year	Aquisition Date	Proportion of Cloud Cover over Area of Interest
2013	August 20	0.28
2014	August 07	0.56
2015	September 20	0.53
2016	September 13	0.20
2017	August 15	0.17
2018	August 18	0.40
2019	August 14	0.68

Table 3.1: Aquisition Dates and proportion of incomplete observations due to cloud cover for 2013-2019.

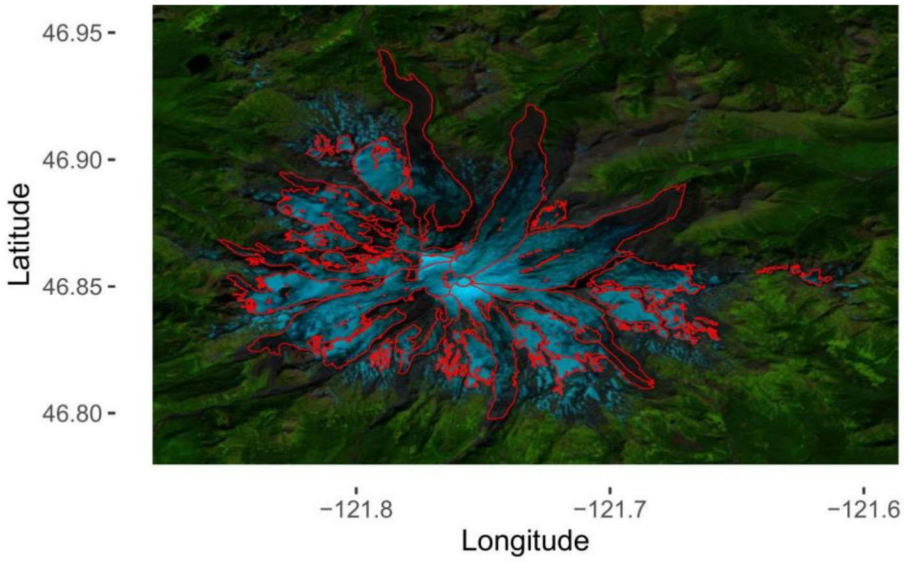


Figure 3.1: SWIR1-NIR-red composite of Mount Rainier on August 23, 2020. The red outline represents the extent of Mount Rainier's glaciers in 2015. This extent was obtained from Andrew Fountain, Portland State University.

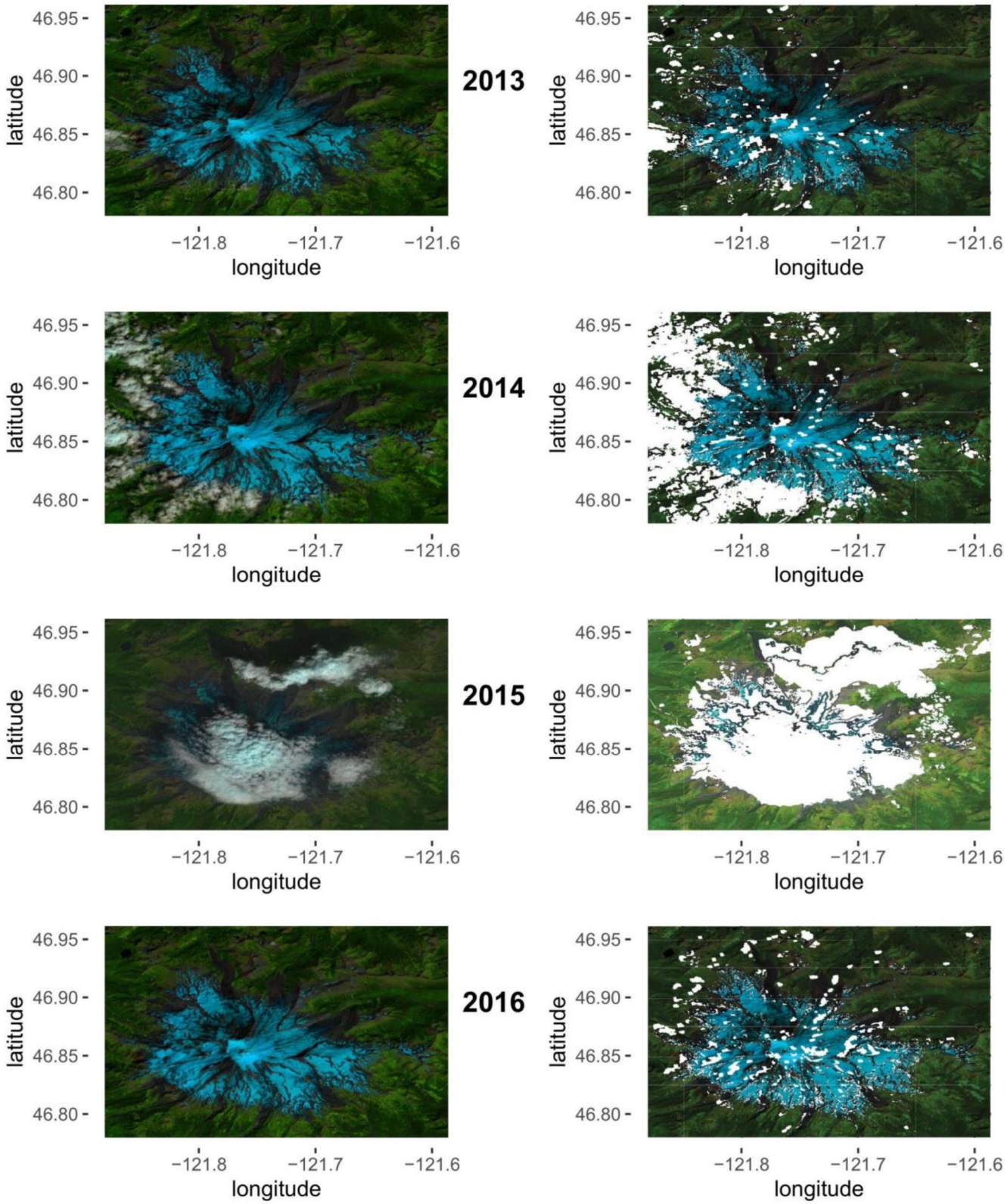


Figure 3.2: SWIR1-NIR-red composite image pairs of Mount Rainier for 2013-2016. Left image is complete image, right image has clouds and cloud shadows removed.

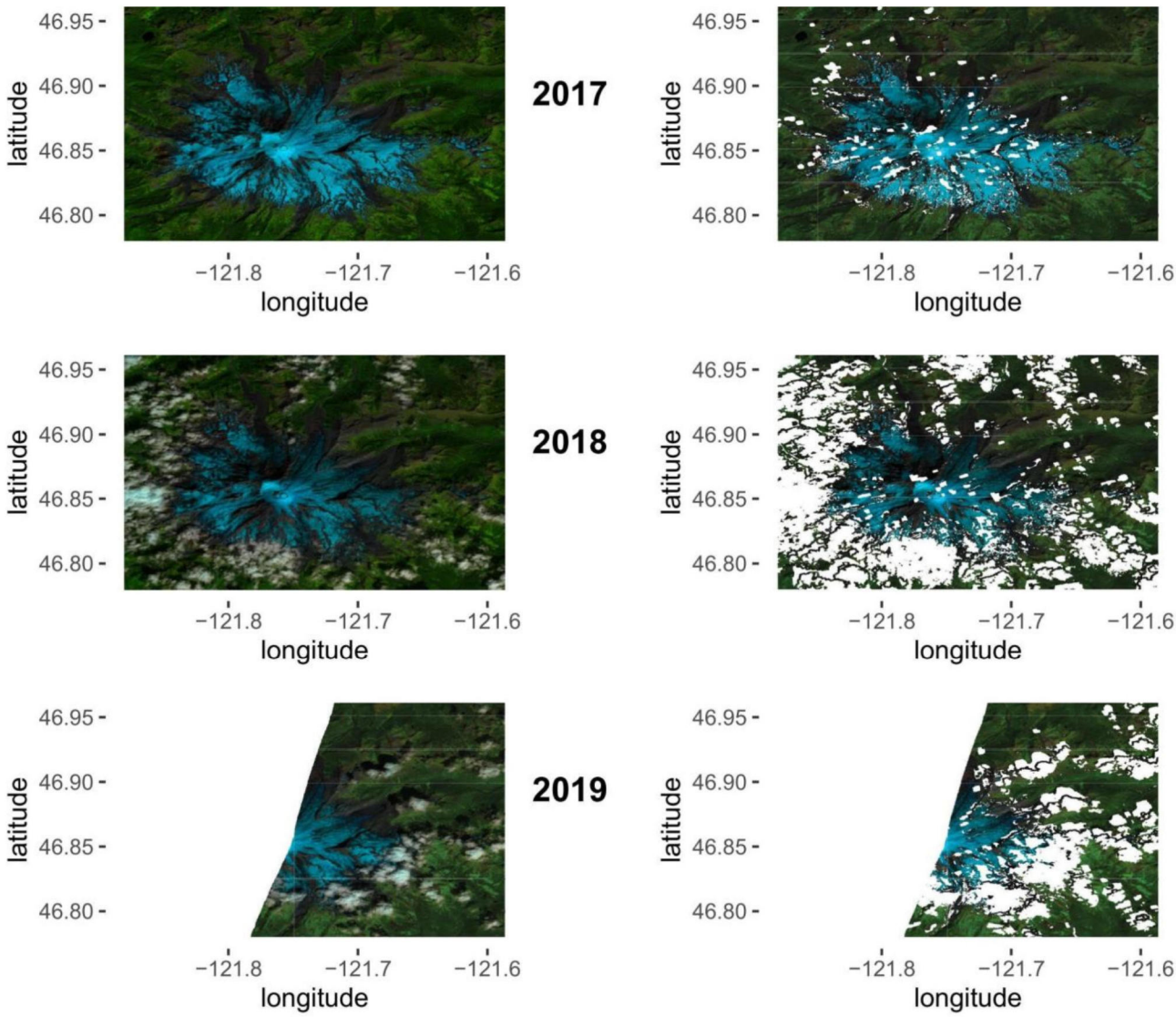


Figure 3.3: SWIR1-NIR-red composite image pairs of Mount Rainier for 2017-2019. Left image is complete image, right image has clouds and cloud shadows removed.

3.2 Variables of Interest

As previously mentioned, with multispectral data, there are many band ratios and differences to choose from. NDSI $(\frac{\text{Green}-\text{SWIR}}{\text{Green}+\text{SWIR}})$ [10, 35] and the Red/SWIR band ratio [8,11] were frequently suggested. Surface temperature (ST) was added as a frequent accompaniment to multispectral indices [30, 36]. The pixel speed component of the glacial velocity was added from the GoLIVE project. Due to memory and computation time constraints, only 4 variables were chosen. Figures 3.4, 3.5, 3.6, and 3.7 show the spatial distribution of the variables of interest.

This work makes a few assumptions about the spatiotemporal data. Firstly, due to georeferencing of the ARD imagery, the measurements are taken at the same set of locations over different instances in time. Secondly, despite being captured anytime between the dates of Aug 01 and Sept 21, the time measurements are assumed to be equally spaced, i.e. one year apart. If the time measurements were not equally spaced apart, continuous, rather than discrete, time modelling would have been more appropriate.

Missing data is said to be missing at random (MAR) if the distribution of the missing data does not depend on the missing values, $p(\theta|y_{\text{observed}}, y_{\text{missing}}) = p(\theta|y_{\text{observed}})$. Missing data is missing completely at random (MCAR) if $p(\theta|y_{\text{observed}}, y_{\text{missing}}) = p(\theta)$ and not at random (MNAR) if the distribution of θ is conditional upon y_{missing}. The MAR (but not the MCAR) assumption is further supported by including as many observations as possible, as the increased number of observations decreases the degree on which the missing data is modelled by an unobserved parameter [37]. With application to this work, the weaker

assumption of MAR is sufficient for Bayesian inference without further requiring modelling of a missing data parameter [37], so the largely missing years of 2016 and 2019 do not explicitly change the modelling process.

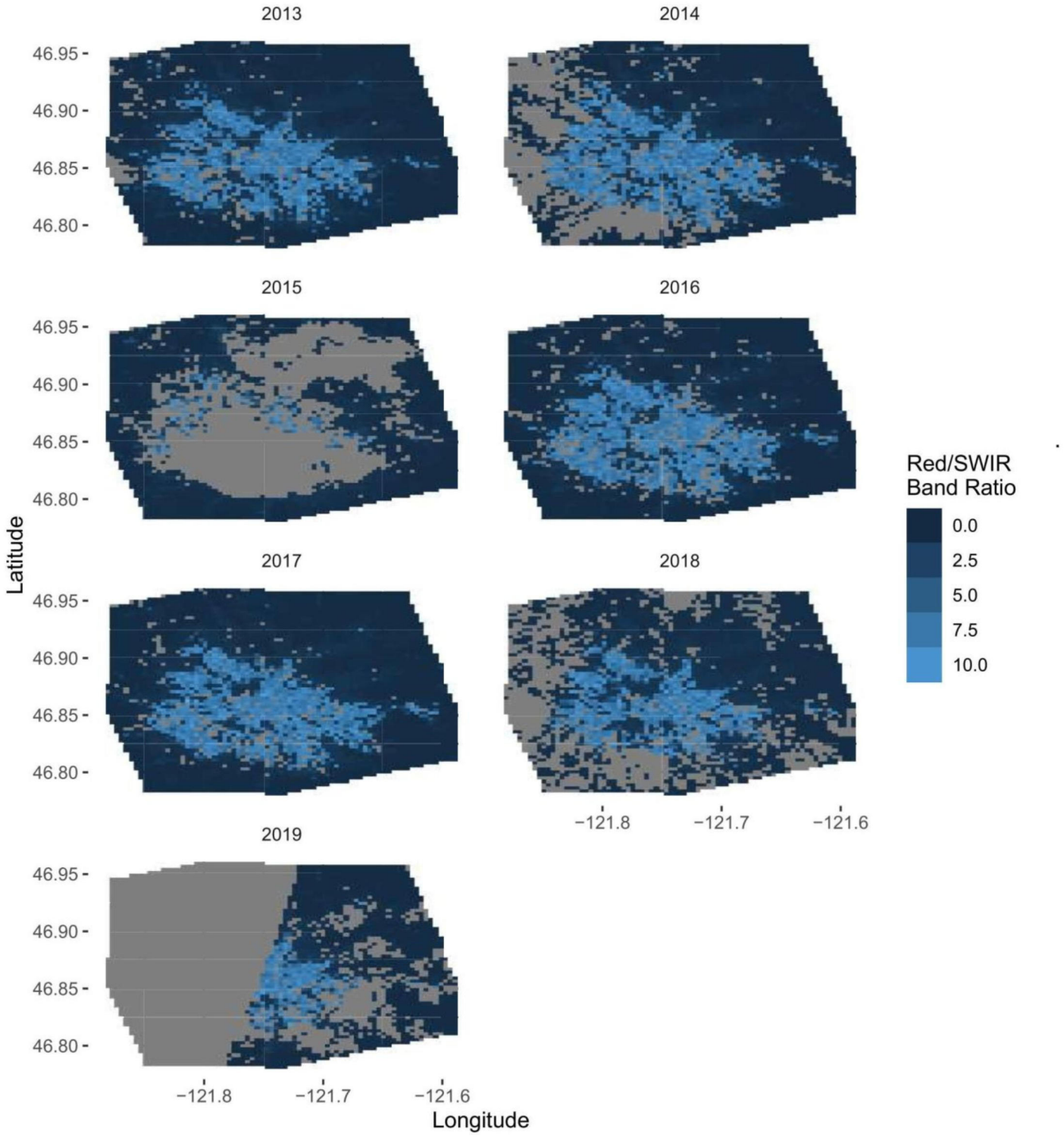


Figure 3.4: Grid of Red/SWIR Band ratio of Mount Rainier's glaciers 2013-2019. The grey indicates missing data due to clouds or cloud shadows.

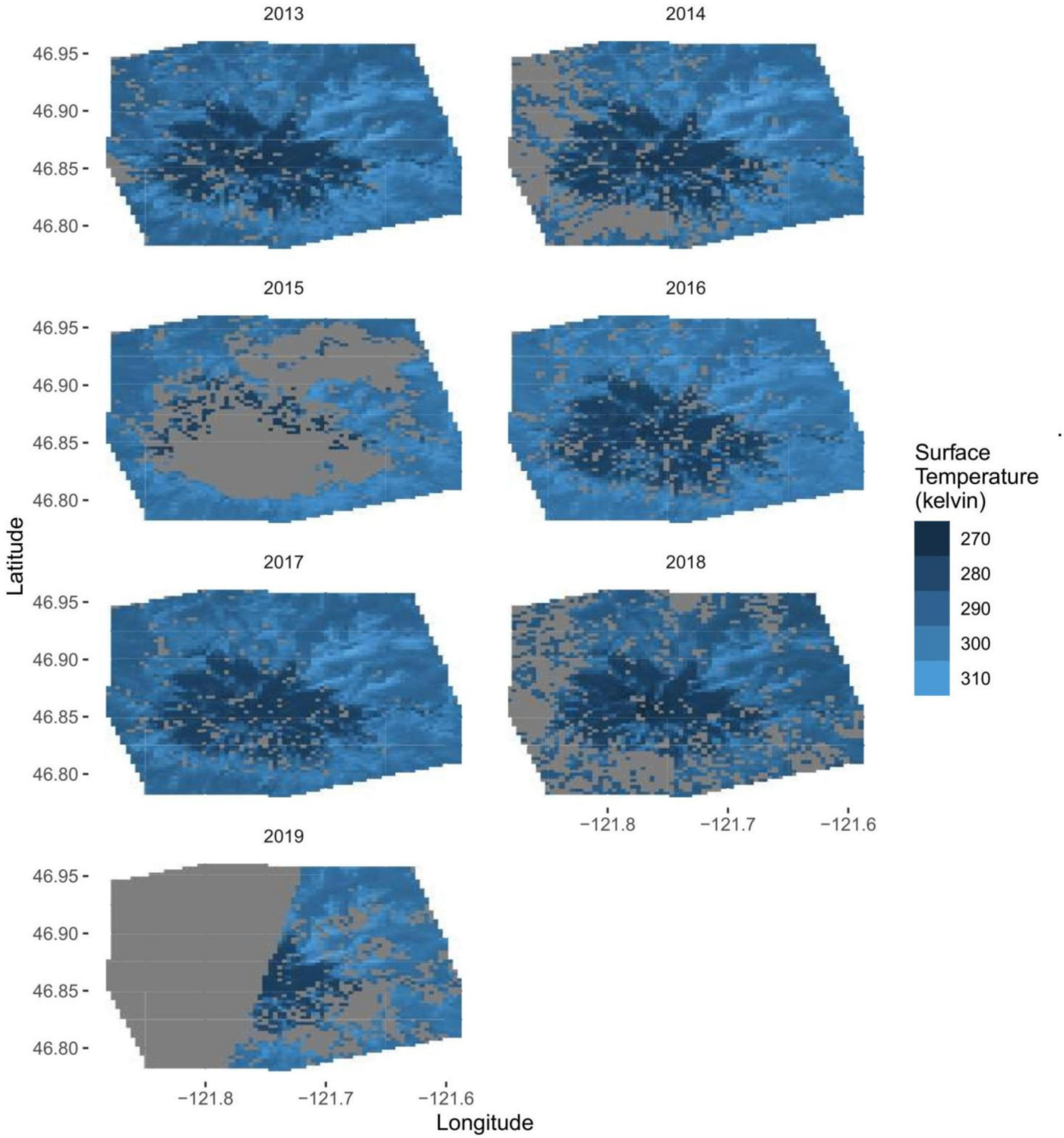


Figure 3.5: Grid of surface temperature of Mount Rainier's glaciers 2013-2019. The grey indicates missing data due to clouds or cloud shadows.

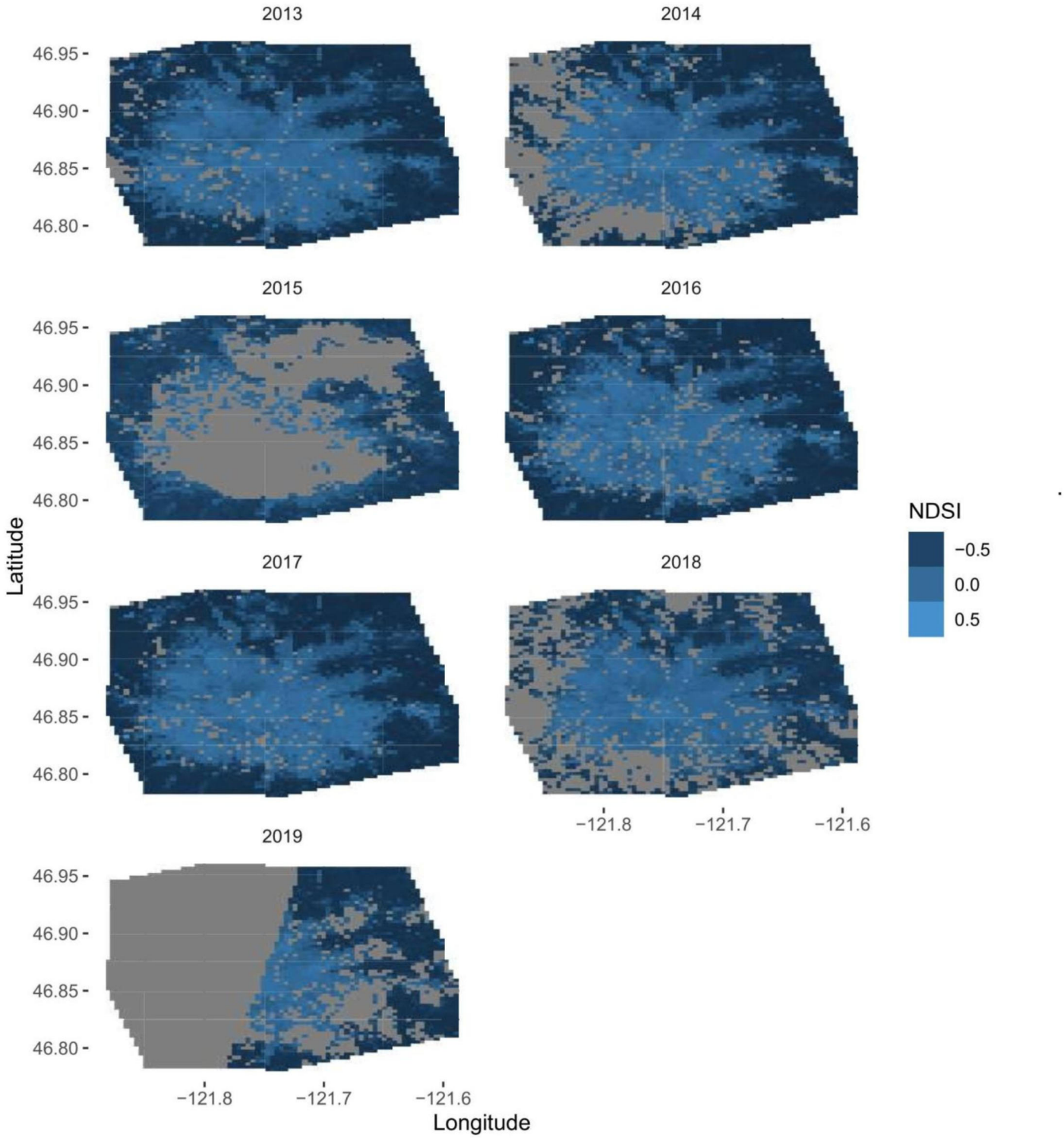


Figure 3.6: Grid of NDSI of Mount Rainier's glaciers 2013-2019. The grey indicates missing data due to clouds or cloud shadows.

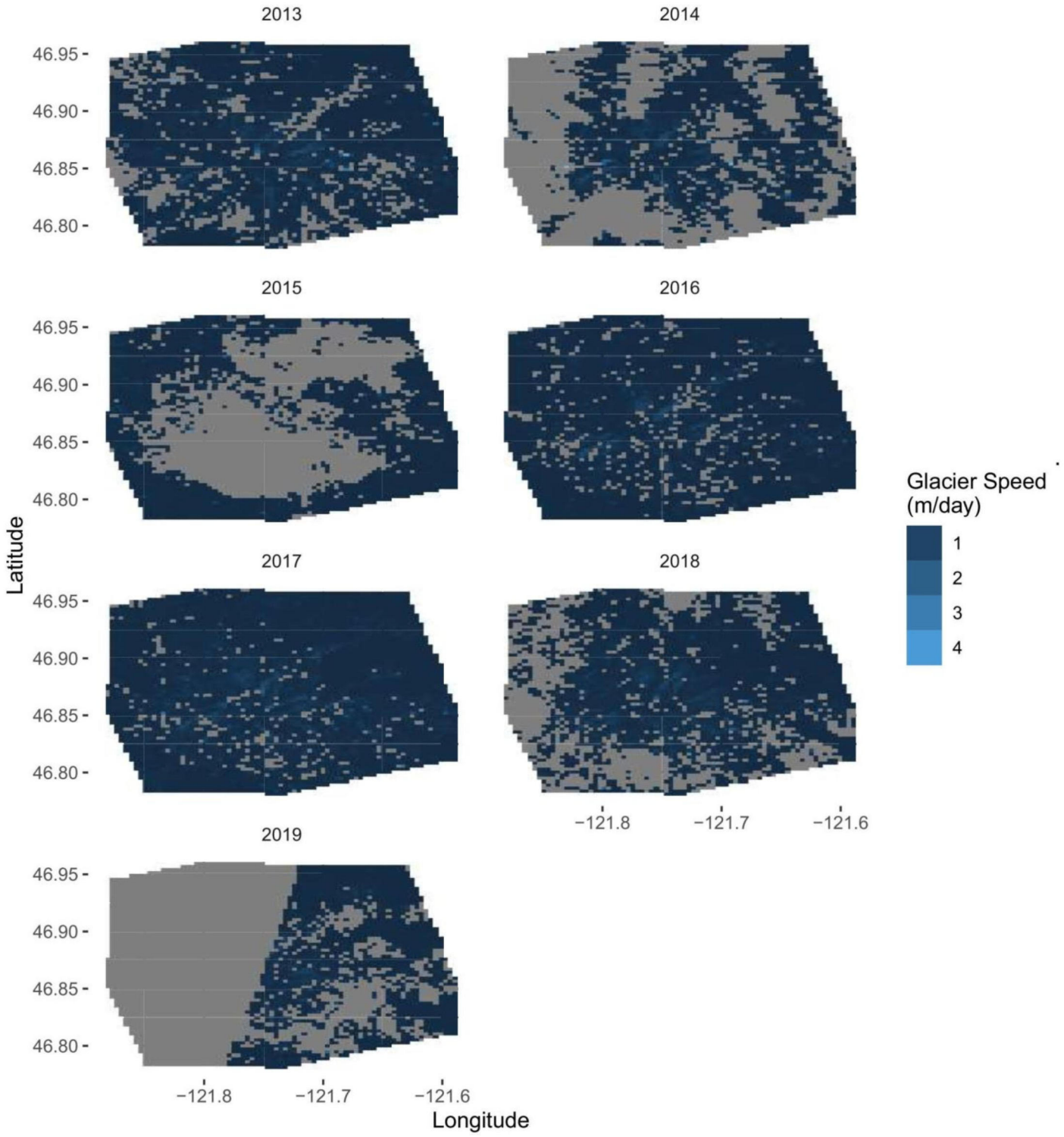


Figure 3.7: Grid of glacier speed of Mount Rainier's glaciers 2013-2019. The grey indicates missing data due to clouds or cloud shadows.

Chapter 4

Methods

> Everything is related to everything else. But near things are more related than distant things.
>
> -Tobler's First Law of Geography

This quote largely summarises the problems with spatial, and by extension spatiotemporal modelling. The independence and identically distributed assumption of observations is no longer applicable, and thus the development and study of new techniques is necessary. There are a few different ways of modelling spatiotemporal data. One way is to approximate the spatiotemporal process as a collection of spatial processes taken at different time indices. This is essentially a temporal extension of the spatial process. This way represents most of the research done within remote sensing. Historically, obtaining satellite images was expensive [38], so many applied methods were developed with a minimal amount of different time indices. Currently this is seen with three-dimensional

data being flattened, or techniques such as co-Kriging using adjacent temporal images to model a variable of interest. Alternatively, spatiotemporal modelling can be perceived as a spatial collection of time series. This is the paradigm in which this this thesis research is approached. The underlying assumption is that the temporal dependence is as natural to model as the spatial dependence [39].

4.1 Modelling Overview

The modelling is done under a Bayesian paradigm, described in Section 4.2. This will allow the uncertainties to be captured through the distribution of the posterior and then predict the missing data. As the Bayesian posterior has no analytical solution, the Bayesian inference is performed using Markov chain Monte Carlo methods, described in Section 4.3. Given the highly correlated variables of interest, a multivariate modelling approach is necessary. The four response variables are modelled through a generalised additive model, defined in Section 4.4. The link between the four response variables is done through seemingly unrelated regression, which is described in Section 4.5. Autoregressive moving averages and Gaussian processes are respectively used to model the temporal and spatial autocorrelations within the generalised additive model. These methods are described in Sections 4.6 and 4.7. Each of these methods have varying levels of complexity, and the methods for choosing the best model is described in Section 4.8. The model chosen via Section 4.8 is then used to predict the cloud covered, missing data. This complete dataset is then classified using techniques described

in Section 4.9. Finally, a principal components analysis, defined in Section 4.10, is used to support the choice in multivariate response variables for this work.

4.2 Bayesian Methodology

The Bayesian approach to modelling takes its name from the 18th Century English non-conformist minister Thomas Bayes, well before the advent of modern statistics. Despite the name, the present form of Bayes's theorem comes from French mathematician Pierre-Simon Laplace, who generalised Bayes's theorem to any probability distribution:

$$p(\theta \mid x) \propto p(x \mid \theta) \times p(\theta) \tag{4.1}$$

$$\text{Posterior} \propto \text{likelihood} \times \text{prior}. \tag{4.2}$$

θ is the unknown parameter of interest and x corresponds to the data. Bayesian inference is performed upon the posterior, in which prior information is updated via the likelihood. However, for most situations, this posterior integral is intractable, meaning it has no closed-form solution. Techniques have been developed to estimate these posterior distributions through approximations such as integrated nested Laplace approximations or variational Bayes, or through Markov Chain Monte Carlo (MCMC) algorithms [40, 41].

The strength of Bayesian inference, with regards to this thesis research, is how it handles missing data. Bayesian inference supports a general approach through which missing data is represented as one or more parameters estimated via the

posterior [37].

4.3 Markov Chain Monte Carlo

Becoming popular in recent decades, MCMC methods date back to World War II, when they were used to solve nuclear diffusion problems. The MCMC name combines two concepts: Markov chain and Monte Carlo. The Markov chain represents that the samples are generated by a sequential process. Each sample is used in an iterative process to generate the next random sample. The Markov property states that each sample depends on only the previous sample, not any samples before then. A Monte Carlo process is the practice of estimating the properties of a distribution through random sampling of the distribution, rather than symbolic analysis.

There are several MCMC algorithms available, the one used in this work is a Hamiltonian Monte Carlo (HMC) algorithm coded in Stan, a programming language named after Polish scientist Stanislaw Ulam, a creator of Monte Carlo methods. The Hamiltonian, often used in physics, is an operator that relates the changes of potential and kinetic energies of a system over time. In this case the potential energy is the log-likelihood of the distribution and the kinetic energy is a random parameter introduced to determine the relative location of the next MCMC sample [42].

The HMC algorithm starts at a random set of parameters, then for a previously determined number of iterations a new momentum vector is sampled and θ is

updated according to Hamiltonian dynamics. To avoid being stuck in local minima, a Metropolis acceptance step is introduced, where the probability of keeping updated θ' and ρ' is min $\left(1, e^{(H(\rho,\theta)-H(\rho',\theta'))}\right)$. If θ' and ρ' are not accepted, then the previous parameters are returned, and another random draw is taken [37,42].

4.4 Generalised Additive Model

GAMs are a class of regression models in which the predicted variable depends linearly on smooth functions of continuous covariates [43].

$$Y \sim N(\mu, \Sigma), \tag{4.3}$$

$$\mu_m \sim XB + \sum_{j=1}^{P} f_{jm}(x_j). \tag{4.4}$$

In this model, we assume Y has a multivariate Gaussian distribution with two parameters, mean vector μ and variance-covariance matrix Σ. Each response, μ_m is distributed additively. The first parameter. B is a least-squares regression parameter based upon covariates X. In this work, The B parameter corresponds to an intercept. Each f_{jm} is a smoothing parameter used to model autocorrelations within the data. The errors of the $\{y_m\}$ response variables are correlated as per Section 4.5. ARMA processes are used to model the temporal autocorrelations, and Gaussian processes are used to model spatial autocorrelations. This combination of ARMA and Gaussian process framework was developed by Murray-Smith in 2001 [44] and has been used in application with weather forecasting and other

spatiotemporal processes [45, 46].

4.5 Seemingly Unrelated Regression

Seemingly unrelated regression (SUR) is used to model the multivariate response variables of the model. SUR involves a set of regression equations with correlated error terms having different variances. Algebraically, the SUR model is written as

$$y_{m,s,t} \sim N(\eta_{m,s,t}, \sigma^2_{m,s,t}), \tag{4.5}$$

$$\eta_{m,s,t} = \underbrace{\alpha_{m,s,t}}_{\text{Intercept}} + \underbrace{f_{GP}(x)_{m,s,t}}_{\text{Gaussian Process}} + \underbrace{\sum_{j=1}^{p} \phi_j y_{m,s,t-j}}_{\text{Autoregressive}} + \underbrace{\sum_{j=1}^{q} \theta_j \epsilon_{m,s,t-j}}_{\text{Moving Average}} + \underbrace{\epsilon_m}_{\text{Error}}. \tag{4.6}$$

Each of the m responses has a univariate regression. The system of equations is linked through their correlated error structure

$$Var(\epsilon_{mi}) = \sigma^2_{mm}, \tag{4.7}$$

$$Cov(\epsilon_{mi}, \epsilon_{m'i}) = \sigma^2_{mm'}, \tag{4.8}$$

$$Cov(\epsilon_{mi}, \epsilon_{m'i'}) = 0. \tag{4.9}$$

With $i \neq i'$ and $m \neq m'$. The SUR model assumes that the errors within each of the m equations are uncorrelated and homoscedastic. However, the errors of an observation are allowed to be correlated across regression equations, that is, if two

variables are positively correlated and one is has a negative residual, the other is likely to be have a negative residual too. Bayesian SUR models are well used as a relatively simple model in multivariate regression [47–50].

4.6 Autoregressive Moving Average

Time series data typically comes from sequential data collection. The data may take discrete or continuous values. The structure of a time series is usually identified through patterns of trends, cyclic effects, and irregular fluctuations.

Early models assumed that time series were deterministic [51], the idea of a stochastic parameter with a more complex structure was first developed by Yule [52], with the formation of the autoregressive and moving average models. Until the 1970s, most of the time series model structure was decided by expert opinion, when a maximum likelihood estimator was developed for parameters in time series models [53].

The link between autoregressive and moving average models was first developed by Wold in 1938 [54]. From his decomposition theorem it follows that AR processes can be written as MA processes. However these two can be combined to create ARMA processes, the advantage being that usually ARMA process can describe a time series with fewer parameters than entirely AR or MA models. The

ARMA process is written as

$$y_{s,t} = \underbrace{\alpha_{s,t}}_{\text{Intercept}} + \underbrace{\sum_{j=1}^{p} \phi_j y_{s,t-j}}_{\text{Autoregressive}} + \underbrace{\sum_{j=1}^{q} \theta_j \epsilon_{m,s,t-j}}_{\text{Moving Average}} . \tag{4.10}$$

Here, ϕ is restricted to be in $(-1, 1)$ as ARMA models are not identifiable if the roots of the moving average characteristic polynomial lie outside the unit circle.

Most time series models require stationary with $E(y) = c$ for $c \in \mathbb{R}$. This seems like an overly severe assumption, especially as Autoregressive Integrated Moving Average (ARIMA(p, d, q)) models have been developed for non-stationary processes. However, ARIMA(p, d, q) can be approximated as ARMA$(p + d, q)$ models.

In an ARIMA model the autoregressive characteristic polynomial can be written as the following, with B the backshift operator:

$$\phi(B) = (1 - B)^d (1 - \phi_1 B - \cdots - \phi_p B^p) \tag{4.11}$$

,

$$\phi^*(B) = \left(1 - \frac{B}{r}\right)^d (1 - \phi_1 B - \cdots - \phi_p B^p) \approx \left(1 - \phi_1^* B - \cdots - \phi_{p+d}^* B^{p+d}\right). \tag{4.12}$$

Here, the root of $1 + \epsilon$ for $\epsilon > 0$, is taken. Therefore an ARIMA(p, d, q) process can be approximated as an ARMA$(p + d, q)$ process, and we can approximate a non-stationary process with our ARMA model.

There are a few other ways of determining ARMA model order. The first is a fully Bayesian model, where a prior is put on the order of parameters, and the

MCMC samples are allowed to transition between the various orders. [55] The second is an empirical Bayesian model, where models of different orders, p, q, are used to fit to the data and the best model is chosen. This is the method that is used for this work. Given that additive ARMA Gaussian prior models have been used with up to ARMA(2,2) parameters [44, 56, 57], all 14 combinations of ARMA processes up to 4 total parameters $p + q$ are fit.

4.7 Gaussian Process

Usually in simple linear regression analysis, the form of a regression is $y = \beta_0 + \beta_1 x + \epsilon$, and β_0, β_1 are the intercept and slope parameters. The core of a Gaussian process is that the regression function is not fixed, but rather the process has a prior distribution. We can look at this as $y = f(x) + \epsilon$, with $f(x) \sim GP(m(x), K(X)))$ where $GP()$ is the Gaussian prior distribution for the functional f. The GP has parameters of mean function m and covariance kernel K.

Non-Bayesian Gaussian processes in the geostatistics world have been referred to as Kriging [58], named after the South African mining engineer Danie Krige, who developed this method to look for iron ore deposits [59]. Traditionally, Kriging consists of multiple steps, including exploratory statistical analysis of the data, variogram modelling, creating the surface, and optionally exploring a variance surface and yielding a predicted mean. If the output is Gaussian and covariance function is exactly known, it produces the best linear unbiased estimator (BLUE). Bayesian methods, as previously noted, do not require a completely known vari-

ance and do not yield a mean estimate, but rather a posterior distribution from which a mean can be extracted.

The mean function of a Gaussian process $m(x)$ defines the base function around which all of the realizations of the Gaussian process will be distributed. In this case, we will assume it be 0, as there is a non-zero intercept $A_{s,t}$ for each sample location. If we were to have a more complex Gaussian process with a non-zero mean function it could simply be added to the intercept. The Gaussian process is used to model the deviation due to surrounding points in this regression model.

The covariance kernel $K(X) = k(x, x')$, for $x \neq x'$ controls how f varies around the mean function. The larger the covariance between x and x' is the more their outputs $f(x)$ and $f(x')$ vary. [60, 61]

In this work, three covariance kernels are compared. Firstly, the squared exponential: $\sigma^2 exp\left(-\frac{(x-x')^2}{2l^2}\right)$, secondly the spherical: $\sigma^2\left(1 - 1.5\frac{|x-x'|}{l} + 0.5\left(\frac{|x-x'|}{l}\right)^3\right)$, and finally the Matérn $\frac{1}{2}$: $\sigma^2 exp\left(-\frac{|x-x'|}{l}\right)$. Here, l is length parameter, controlling the smoothness of the response. In terms of the application in this work, we can view x and x' as spatial coordinates and $l = |x - x'|$ [37]. Figure 4.1 shows an application of Gaussian processes on a toy dataset, each blue line is a random sample of a Gaussian process.

The strength of Gaussian processes comes from the prediction. If there is a Gaussian observation model, $y \sim N(\mu, \sigma^2)$, we can estimate $\tilde{\mu}$ at new locations

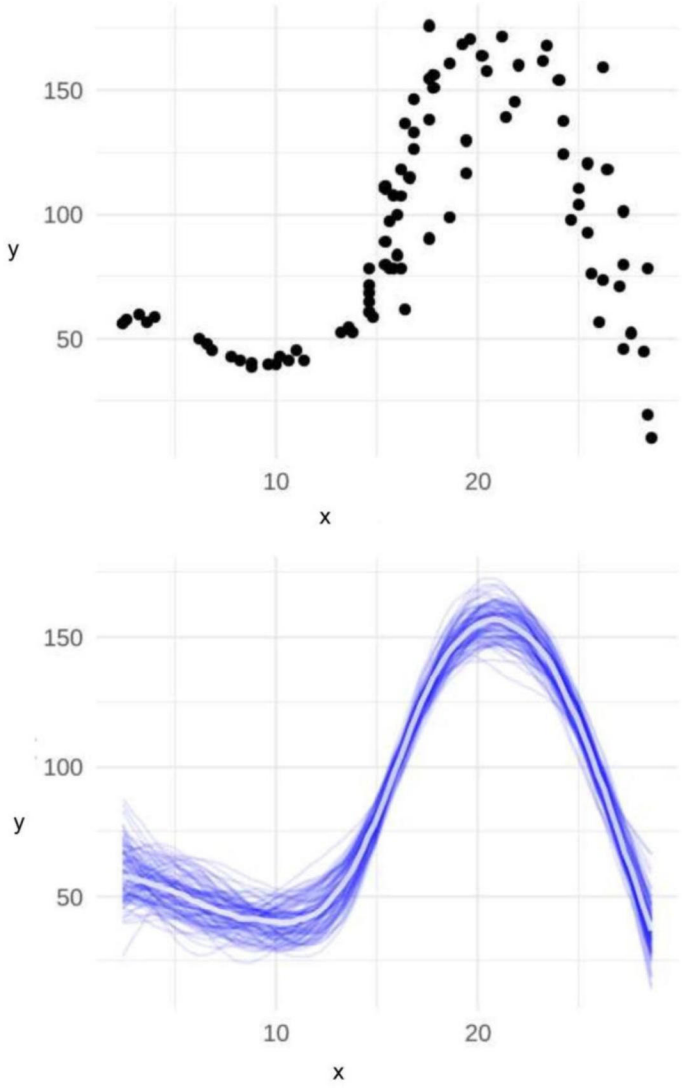


Figure 4.1: Demonstration of Gaussian process regression on a toy dataset. Top figure is the data, made of sums of sine functions with random noise. Bottom figure is 100 samples from the posterior of a Gaussian process regression function.

through a joint density and then find the marginal density of μ_{new} [37]

$$\begin{pmatrix} y \\ \tilde{\mu} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} K(x,x) + \sigma^2 I & K(\tilde{x},x) \\ K(x,\tilde{x}) & K(\tilde{x},\tilde{x}), \end{pmatrix}\right) \tag{4.13}$$

with,

$$\mu' \mid x, y, \tau, l, \sigma \sim N(E(\mu'), \mathrm{cov}(\mu')), \tag{4.14}$$

$$E(\mu') = K(x', x)\left(K(x, x) + \sigma^2 I\right)^{-1} y, \tag{4.15}$$

$$\mathrm{cov}(\mu') = K(x', x') - K(x', x)\left(K(x, x) + \sigma^2 I\right)^{-1} K(x, x'). \tag{4.16}$$

On first glance, estimating the posterior distribution of μ and the new μ' may seem simple; however each matrix inversion requires $O\left(n^3\right)$ computations for each MCMC iteration. And thus the computation becomes very expensive with n being more than a few thousand, which is the case here. In a fully Bayesian analysis, the length parameter, l, would be a distribution and be allowed to vary over MCMC samples. In this work, l was fixed to be 12, this brought the memory and time required from three weeks and 64 gigabytes of RAM to one week and 28 gigabytes of RAM. The Gaussian process was coded in Stan, modified from code in the Stan user's guide [62].

4.8 Model Choice

As discussed in the previous sections, 14 orders of ARMA(p, q) models and 3 different Gaussian process kernels are being compared. A choice needs to be made that balances the goodness of fit with model complexity.

Given that this model will be used to make predictions of the missing data, evaluation of the predicted mean should be measured [63]. Root Mean Squared

Error (RMSE) has been employed for measuring this predictive efficacy in Bayesian models [64, 65]. While the output of a model is a posterior distribution, any statistic may be calculated from this distribution. In this case, the predicted mean can be calculated from the distribution and RMSE calculated henceforth.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(x - E(x))^2}{n}} \tag{4.17}$$

Note, due to the lack of independence between each of the predicted variables, RMSE is calculated for each of these variables.

When assessing model performance, complex models with more parameters typically out perform simpler models. [66]. The Akaike Information Criterion (AIC), developed by Japanese statistician Hirotugu Akaike in 1974, is a statistic frequently used to balance goodness of fit of a model with model complexity [66, 67]. In 2010, the Bayesian extension, Watanabe-Akaike Information Criterion or Widely Available Information Criterion (WAIC) was developed as an extension for Bayesian models [68]. Note that the WAIC asymptotically approximates the AIC [68]:

$$WAIC = T_n + \frac{V_n}{n}. \tag{4.18}$$

Where T_n corresponds to how the model fits the data, and V_n corresponds to the functional variance.

$$T_n = -\frac{1}{n}\sum_{i=1}^{n} \log p^*(X_i|w) \tag{4.19}$$

$$V_n = \sum_{i=1}^{n} \left\{ \mathbb{E}_w \left[(\log p(X_i|w))^2 \right] - \mathbb{E}_w [\log p(X_i|w)]^2 \right\}. \tag{4.20}$$

Similarly with the RMSE, in this work the WAIC is calculated for each response variables. The WAIC calculation was coded in Stan, modified from code in the Stan user's guide [62]. The RMSE was calculated using R calculated from the mean model response.

4.9 Classification

As a part of the Global Land Ice Measurements from Space (GLIMS) initiative, a census of Mount Rainier's glaciers was done by Andrew Fountain from Portland State University in 2015 [69]. In this thesis a variety of different classification techniques are used to classify the multivariate matrix of predicted variables as glacier. Hard algorithms, in which the classification is expressed as a binary outcomes and soft algorithms, in which the classification is expressed as a probability, are described in the following subsections.

4.9.1 Hard K-means

Classical (Hard) K-means is a iterative clustering algorithm in which n observations are partitioned into K clusters. The clusters are created to minimise the intra-

cluster sum of squares,

$$W = \sum_{k=1}^{K} \sum_{i=1}^{N} \frac{I_{i,k}(y_i - \mu_k)^2}{N} \tag{4.21}$$

where μ_k is the mean of each cluster, and I is an indicator function for a particular observation belonging to a cluster. The hard K-means algorithm was performed on the two clusters of glacier and non-glacier.

4.9.2 Soft K-means

Soft K-means, is similar to Hard K-means, being an iterative classification technique with K clusters; however, each y_n may be assigned to different cluster, with probability $p(I_k)$:

$$W = \sum_{k=1}^{K} \sum_{i=1}^{N} p(k)(y_i - \mu_k)^2. \tag{4.22}$$

In classical K-means, the iterative process assigns each observation to a cluster with an index function. This is akin to assigning each observation with a probability of 1. Soft K-means differs in that the probability of an observation belonging to any cluster is no longer 1, but can range between 0 and 1. The Soft K-means algorithm was performed on the two clusters of glacier and non-glacier.

4.9.3 Support Vector Machines

Support vector machines (SVM) are a classification technique that has gained popularity over last few decades through machine-learning development. In SVM,

the data is mapped to a higher dimension via a kernel and an optimal hyperplane segments the data into groups.

The radial basis function kernel used in this analysis is written as,

$$K\left(y_i, y_j\right) = \exp\left(-\gamma\left(y_i - y_j\right)^2\right). \tag{4.23}$$

The gamma parameter defines the influence of a singular observation. Higher gamma functions are more influenced by a single outlier, and thus more prone to overfitting. The gamma function of 0.25 was chosen via an iterative search from 0 to 1. The threshold balances correct classification with maximisation of the decision function. As the threshold increases, the classification becomes less sensitive to errors. The threshold value of 0.1 was chosen through an iterative search from 0 to 1 to minimise the sum of type I and II errors.

4.9.4 Hierarchical Clustering

Hierarchical clustering builds a hierarchy of nested clusters. The agglomerative technique here is a bottom-up algorithm pairs of observations are combined together until one final cluster remains. In order to choose the pairs of clusters being merged, a dissimilarity matrix is created, and a linkage criterion is applied to create clusters.

While it is not necessary to predetermine the number of clusters in hierarchical clustering, given it is a series of nested observations, once observations are assigned to a cluster, they will never separate. If the clusters are not truly nested,

hierarchical clustering is not appropriate. Complete, single and average linkages were investigated in the analysis, with average linkage performing the best and used for the final analysis.

4.9.5 Gaussian Mixture Models

A mixture model is a probabilistic model for representing groups within a population. If we assume a Gaussian mixture, we have the following

$$f(x) = \sum_{i=1}^{N} p(x_i) g_i(x), \tag{4.24}$$

$$g_i \sim N(\mu_i, \sigma_i^2). \tag{4.25}$$

The EM algorithm is an iterative process often used to estimate missing data. In this case, we will assume that the data is a mixture of two distributions, one for the glacier and another for the surrounding area. This Gaussian mixture model uses the EM algorithm to estimate the mean μ_k and variance σ_k^2 for each of the mixtures. The EM algorithm alternates between an expectation step, which creates a function to for the expectation of the log-likelihood given the parameters, and the maximisation step, which computes new parameters to maximise the expected log-likelihood found in the expectation step. These new parameters are then used in subsequent expectation steps and the algorithm repeats until the values converge. The Gaussian mixture model was coded in Stan, using modified code from the Stan users guide [62] and executed in R using the RStan package.

4.10 Principal Components Analysis

Principal components analysis (PCA) is a statistical method that aims to find the principal directions in which the data varies. It takes linear combinations of the data in order to capture the maximal variance.

The principal components themselves are found though calculating the eigenvalue and eigenvectors of either the variance-covariance matrix or the correlation matrix. The eigenvector with the largest eigenvalue denotes the direction with the highest variation, and subsequently the second highest eigenvalue's eigenvector is the second highest variation. If the set of all eigenvectors are normalised, they form an orthonormal basis for the multivariate data structure. If only the eigenvectors that capture the most variance are selected. PCA provides a tool for dimension reduction that limits the loss in variance. Due to the differences in units and scale of the variables of interest in this work, PCA is performed on the correlation matrix. The principal components analysis was performed on the correlation matrix in R.

Chapter 5

Results

5.1 Glacier Model Choice

All 42 combinations of ARMA(p, q) models up to 4 parameters and 3 Gaussian process kernels were modelled. The RMSE and WAIC for each of the four variables of interest are shown in Tables 5.1 and 5.2. Note that lower RMSE and lower WAIC correspond to better model fit.

No single model was lowest in each of the response variables, so the RMSE were combined via an Euclidean distance, and the WAIC were combined via a sum. The ARMA(3,1) model with exponential Gaussian process kernel was lowest in both of these metrics. This model is highlighted in red in Tables 5.1 and 5.2 and is used for prediction of the missing data and subsequent classification.

ARMA Process	Gaussian Process Kernel	Surface Temperature RMSE	Band Ratio RMSE	NDSI RMSE	Glacier Speed RMSE	Pooled RMSE
ARMA(0,1)	Spherical	8.73	7.56	0.153	0.175	11.556
ARMA(0,2)	Spherical	7.89	7.20	0.149	0.170	10.692
ARMA(0,3)	Spherical	7.11	6.66	0.138	0.166	9.755
ARMA(0,4)	Spherical	6.84	6.51	0.131	0.163	9.456
ARMA(1,0)	Spherical	6.68	6.58	0.130	0.169	9.387
ARMA(1,1)	Spherical	6.69	6.46	0.126	0.158	9.310
ARMA(1,2)	Spherical	6.88	6.47	0.125	0.157	9.449
ARMA(1,3)	Spherical	6.92	6.42	0.123	0.158	9.451
ARMA(2,0)	Spherical	6.57	6.49	0.128	0.164	9.243
ARMA(2,1)	Spherical	6.70	6.57	0.125	0.158	9.397
ARMA(2,2)	Spherical	5.29	6.47	0.125	0.156	8.363
ARMA(3,0)	Spherical	6.84	6.44	0.123	0.160	9.405
ARMA(3,1)	Spherical	6.92	6.42	0.123	0.158	9.450
ARMA(4,0)	Spherical	7.00	6.42	0.123	0.159	9.504
ARMA(0,1)	Matern	5.92	8.52	0.209	0.162	10.386
ARMA(0,2)	Matern	5.55	8.00	0.195	0.160	9.747
ARMA(0,3)	Matern	5.26	7.30	0.177	0.158	9.008
ARMA(0,4)	Matern	5.11	7.13	0.163	0.156	8.784
ARMA(1,0)	Matern	4.93	7.19	0.160	0.159	8.731
ARMA(1,1)	Matern	4.83	7.08	0.154	0.153	8.578
ARMA(1,2)	Matern	4.83	7.08	0.154	0.153	8.583
ARMA(1,3)	Matern	4.80	7.03	0.151	0.153	8.524
ARMA(2,0)	Matern	4.87	7.09	0.156	0.157	8.614
ARMA(2,1)	Matern	4.82	7.18	0.156	0.153	8.662
ARMA(2,2)	Matern	4.78	7.06	0.153	0.153	8.531
ARMA(3,0)	Matern	4.80	7.06	0.152	0.154	8.544
ARMA(3,1)	Matern	4.72	6.99	0.152	0.153	8.445
ARMA(4,0)	Matern	4.82	7.03	0.152	0.154	8.531
ARMA(0,1)	Exponential	5.22	7.74	0.150	0.156	9.345
ARMA(0,2)	Exponential	4.86	7.35	0.146	0.155	8.822
ARMA(0,3)	Exponential	4.71	6.78	0.134	0.153	8.264
ARMA(0,4)	Exponential	4.62	6.65	0.126	0.152	8.110
ARMA(1,0)	Exponential	4.44	6.70	0.128	0.154	8.046
ARMA(1,1)	Exponential	4.40	6.60	0.123	0.150	7.936
ARMA(1,2)	Exponential	4.38	6.60	0.122	0.150	7.925
ARMA(1,3)	Exponential	4.30	6.56	0.119	0.150	7.910
ARMA(2,0)	Exponential	4.41	6.61	0.125	0.152	7.958
ARMA(2,1)	Exponential	4.33	6.66	0.124	0.150	7.955
ARMA(2,2)	Exponential	4.39	6.56	0.122	0.150	7.905
ARMA(3,0)	Exponential	4.39	6.57	0.120	0.150	7.912
ARMA(3,1)	Exponential	4.37	6.55	0.120	0.150	7.881
ARMA(4,0)	Exponential	4.39	6.55	0.120	0.150	7.894

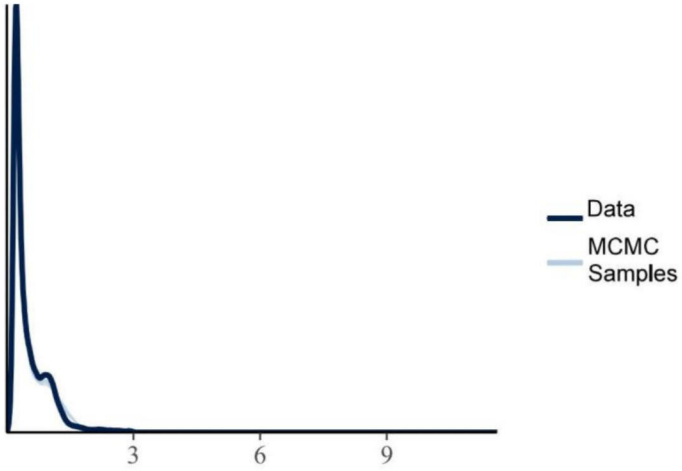
Table 5.1: RMSE for all potential models and variables of interest. The chosen model is highlighted in red.

ARMA Process	Gaussian Process Kernel	Surface Temperature WAIC	Band Ratio WAIC	NDSI WAIC	Glacier Speed WAIC	Pooled WAIC
ARMA(0,1)	Spherical	217,735	140,614	-20,621	-18,234	319,494
ARMA(0,2)	Spherical	216,023	138,278	-22,745	-18,775	312,781
ARMA(0,3)	Spherical	215,083	136,887	-26,053	-19,185	306,731
ARMA(0,4)	Spherical	213,836	136,585	-27,270	-19,394	303,756
ARMA(1,0)	Spherical	214,078	136,565	-26,422	-18,686	305,535
ARMA(1,1)	Spherical	213,881	136,399	-27,771	-19,746	302,763
ARMA(1,2)	Spherical	213,263	136,387	-27,774	-19,724	302,152
ARMA(1,3)	Spherical	213,191	136,373	-27,657	-19,224	302,683
ARMA(2,0)	Spherical	213,992	136,413	-27,434	-19,243	303,728
ARMA(2,1)	Spherical	213,562	136,803	-27,768	-19,735	302,861
ARMA(2,2)	Spherical	213,346	136,350	-28,009	-19,816	301,872
ARMA(3,0)	Spherical	213,152	136,382	-28,056	-19,554	301,924
ARMA(3,1)	Spherical	213,137	136,338	-27,671	-19,227	302,577
ARMA(4,0)	Spherical	213,136	136,372	-27,687	-19,151	302,669
ARMA(0,1)	Matern	217,914	140,080	-19,399	-18,199	320,396
ARMA(0,2)	Matern	216,181	137,784	-21,658	-18,743	313,565
ARMA(0,3)	Matern	215,227	136,419	-24,980	-19,166	307,500
ARMA(0,4)	Matern	214,009	136,148	-26,268	-19,379	304,509
ARMA(1,0)	Matern	214,214	136,158	-25,593	-18,646	306,133
ARMA(1,1)	Matern	214,017	135,987	-26,859	-19,727	303,419
ARMA(1,2)	Matern	213,408	135,979	-26,864	-19,729	302,793
ARMA(1,3)	Matern	213,339	135,965	-27,246	-19,731	302,327
ARMA(2,0)	Matern	214,127	136,000	-26,553	-19,224	304,351
ARMA(2,1)	Matern	213,702	136,351	-25,732	-19,752	304,569
ARMA(2,2)	Matern	213,367	135,917	-26,885	-19,802	302,596
ARMA(3,0)	Matern	213,308	135,974	-27,135	-19,533	302,614
ARMA(3,1)	Matern	216,597	135,422	-27,421	-19,709	304,889
ARMA(4,0)	Matern	213,288	135,959	-27,279	-19,650	302,318
ARMA(0,1)	Exponential	217,840	140,716	-20,595	-18,184	319,777
ARMA(0,2)	Exponential	216,102	138,363	-22,727	-18,738	313,000
ARMA(0,3)	Exponential	215,155	136,957	-26,036	-19,148	306,929
ARMA(0,4)	Exponential	213,909	136,655	-27,248	-19,362	303,953
ARMA(1,0)	Exponential	214,142	136,631	-26,397	-18,647	305,729
ARMA(1,1)	Exponential	213,948	136,468	-27,748	-19,711	302,956
ARMA(1,2)	Exponential	213,335	136,454	-27,751	-19,699	302,338
ARMA(1,3)	Exponential	213,266	136,442	-28,129	-19,686	301,892
ARMA(2,0)	Exponential	214,056	136,480	-27,413	-19,200	303,923
ARMA(2,1)	Exponential	213,669	136,884	-26,508	-16,491	307,554
ARMA(2,2)	Exponential	213,331	136,396	-27,767	-19,771	302,189
ARMA(3,0)	Exponential	213,228	136,449	-28,029	-19,521	302,126
ARMA(3,1)	Exponential	213,219	136,403	-28,142	-19,686	301,795
ARMA(4,0)	Exponential	213,211	136,438	-28,161	-19,625	301,862

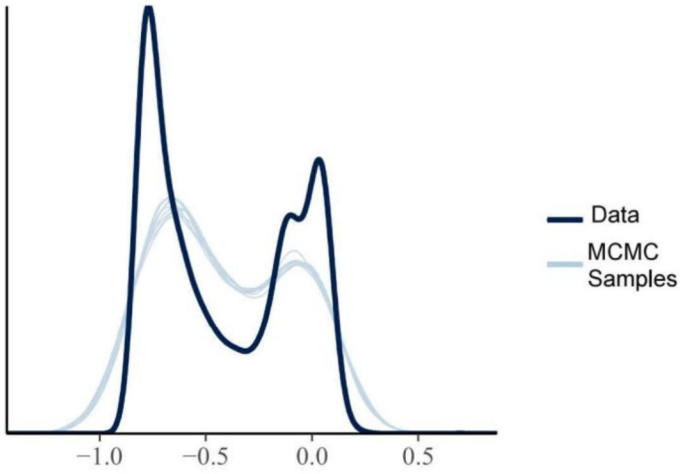
Table 5.2: WAIC for all potential models and variables of interest. The chosen models is highlighted in red.

5.2 Model Prediction

Posterior predictive checks are a standard model check procedure for Bayesian models. Data is repeatedly simulated from the fitted model and compared to the observed data. This repeated sampling allows the check to include the uncertainty associated with the estimate of the parameter of the model, along with the estimate itself. Figures 5.1 and 5.2 shows the posterior predictive checks for the model. Overall the posterior distribution approximates the observed data distribution well except for the NDSI variable. The NDSI distribution is bimodal, with the modes of the posterior distribution far apart. As the MCMC samples are taken in sequence, with the next sample being dependent upon the single previous sample. If the modes are far apart, the MCMC samples may have difficulty transitioning between the two modes. However, as the NDSI distribution modes are far apart, this may make for easier classification, as the two mixtures can be easily identified. In contract with the Red/SWIR band ratio or the surface temperature, the small deviations from the Gaussian distribution may indicate a mixture with the modes much closer together. Apart from the NDSI variable, the posterior predictive distribution estimates the observed distribution well.

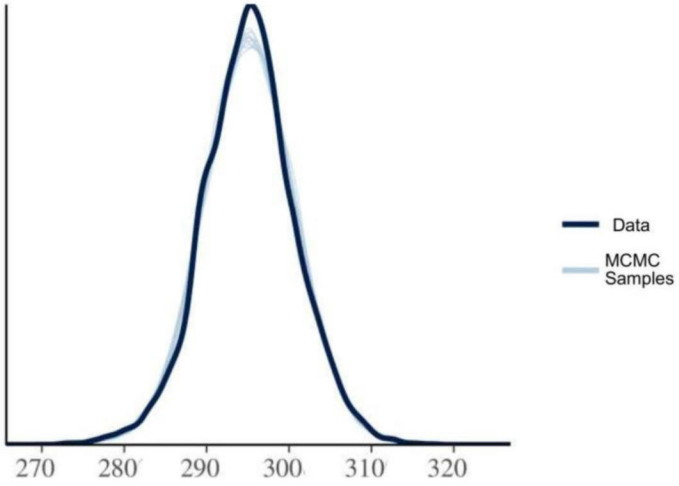


(a) Posterior predictive distribution of the Red/SWIR band ratio

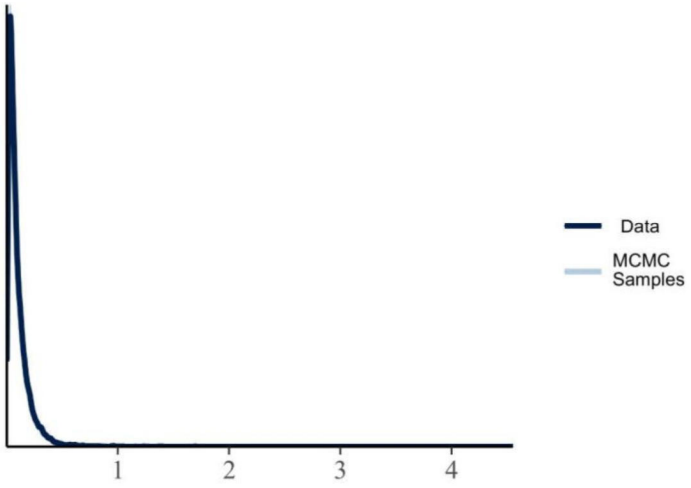


(b) Posterior predictive distribution of the NDSI

Figure 5.1: Density plots of the posterior predictive distributions of the Red/SWIR band ratio and NDSI. The samples from the posterior simulate the observed distribution fairly well with the exception of the bottom NDSI figure. The NDSI observations have a bimodal distribution, which MCMC sampling has trouble converging upon.



(a) Posterior predictive distribution of the surface temperature



(b) Posterior predictive distribution of the glacier speed

Figure 5.2: Density plots of the posterior predictive distributions of the surface temperature and glacier speed.

The mean predictions of this model are shown in Figures 5.3, 5.4, 5.5, and 5.6. Overall the model shows clear images of the glaciers without large artifacts with some exceptions. In 2013, the forward predicting ARMA time series has no support, so the 2013 prediction is solely made with the approximate Gaussian process. 2015 has a large amount of missing data, leading to poor prediction, especially in the NDSI variable. Despite missing half the image, 2019 was modelled well compared to 2015. This may be due to the increased number of preceding years. Due to there being no 2012 observations, in 2015 the third and fourth lags in an AR(4) model add nothing, whereas in 2019 more previous years are available.

Figures 5.7, 5.8, 5.9, and 5.10 show the residual plots calculated from the posterior means. Due to the prevalent missing data, they are somewhat difficult to inspect. However, apart from 2013 and 2016, the residual plots appear to be randomly distributed. As previously discussed, the forward predicting time series had no support in 2013, so the residuals are not randomly distributed and large artifacts are found throughout the image. The residuals in 2016 appear to suffer from a similar problem, the large amount of missing data in 2015 again affects the time series predictions and there is some non-randomness seen in the western half of the Red/SWIR band ratio image. While the predictions in 2015 appear to be imperfect, due to the missing data, there are no residuals to calculate and inspect.

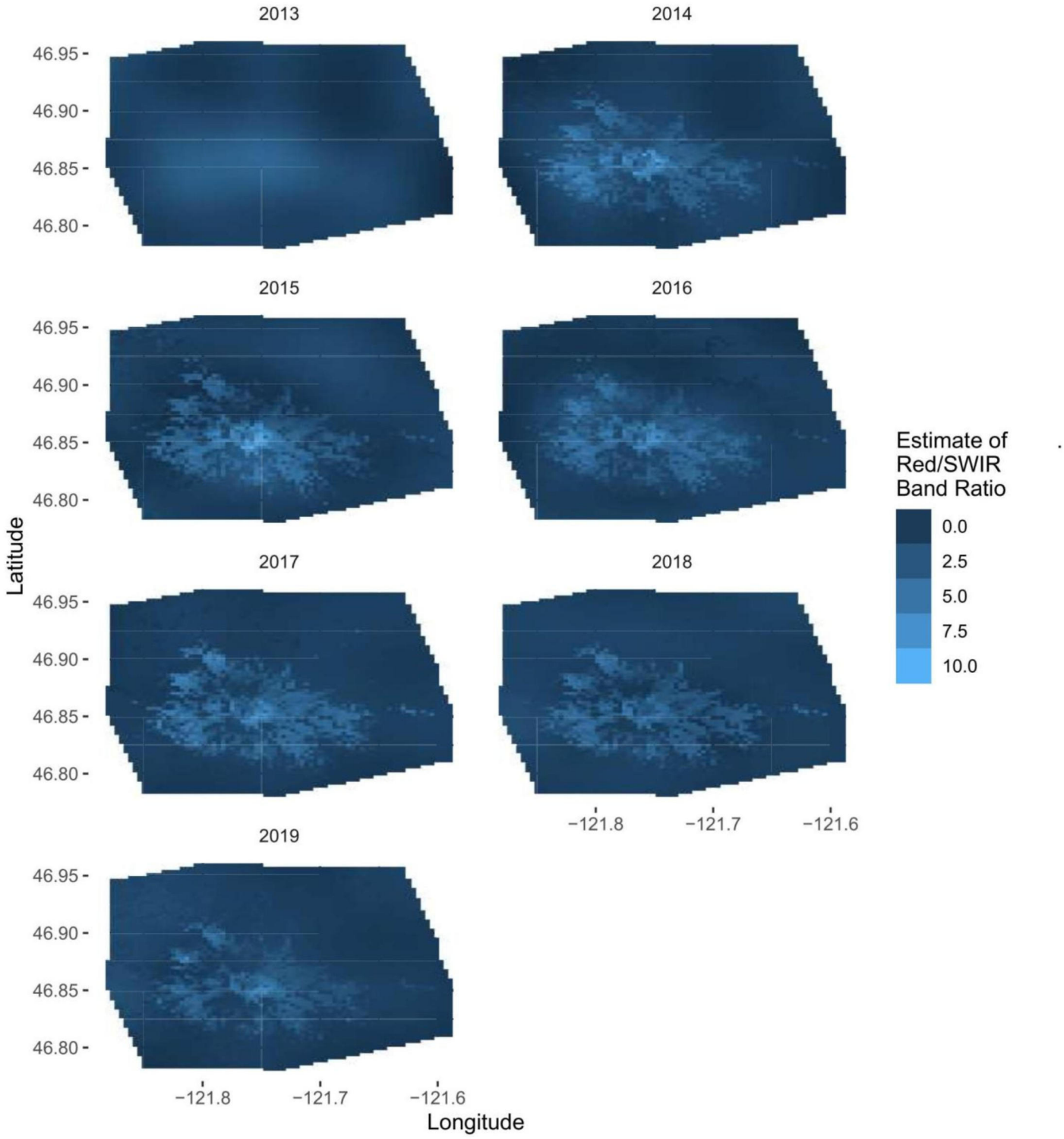


Figure 5.3: Grid of mean estimate of Red/SWIR Band ratio of Mount Rainier's glaciers 2013-2019.

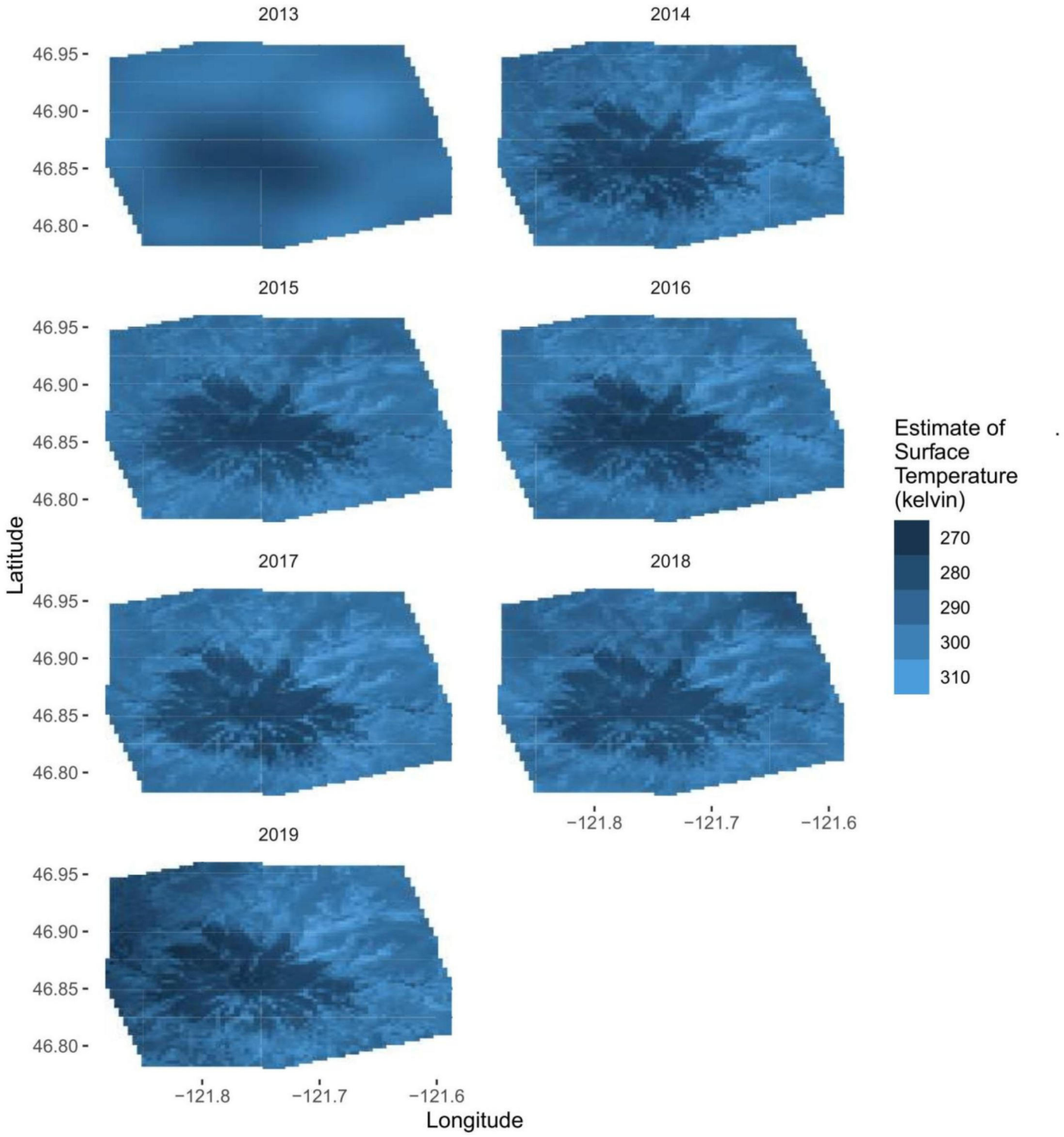


Figure 5.4: Grid of mean estimate surface temperature of Mount Rainier's glaciers 2013-2019.

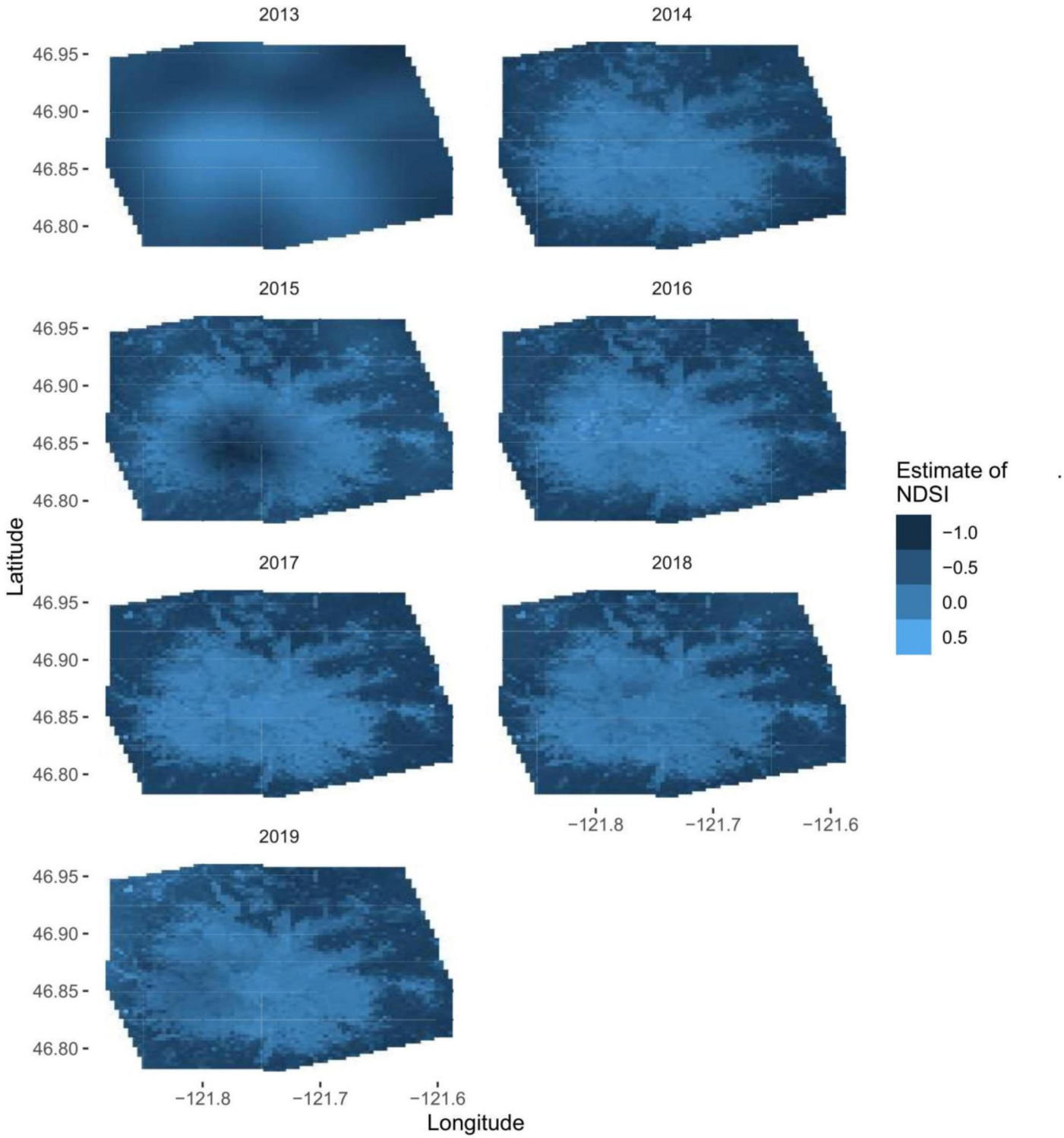


Figure 5.5: Grid of mean estimate NDSI of Mount Rainier's glaciers 2013-2019.

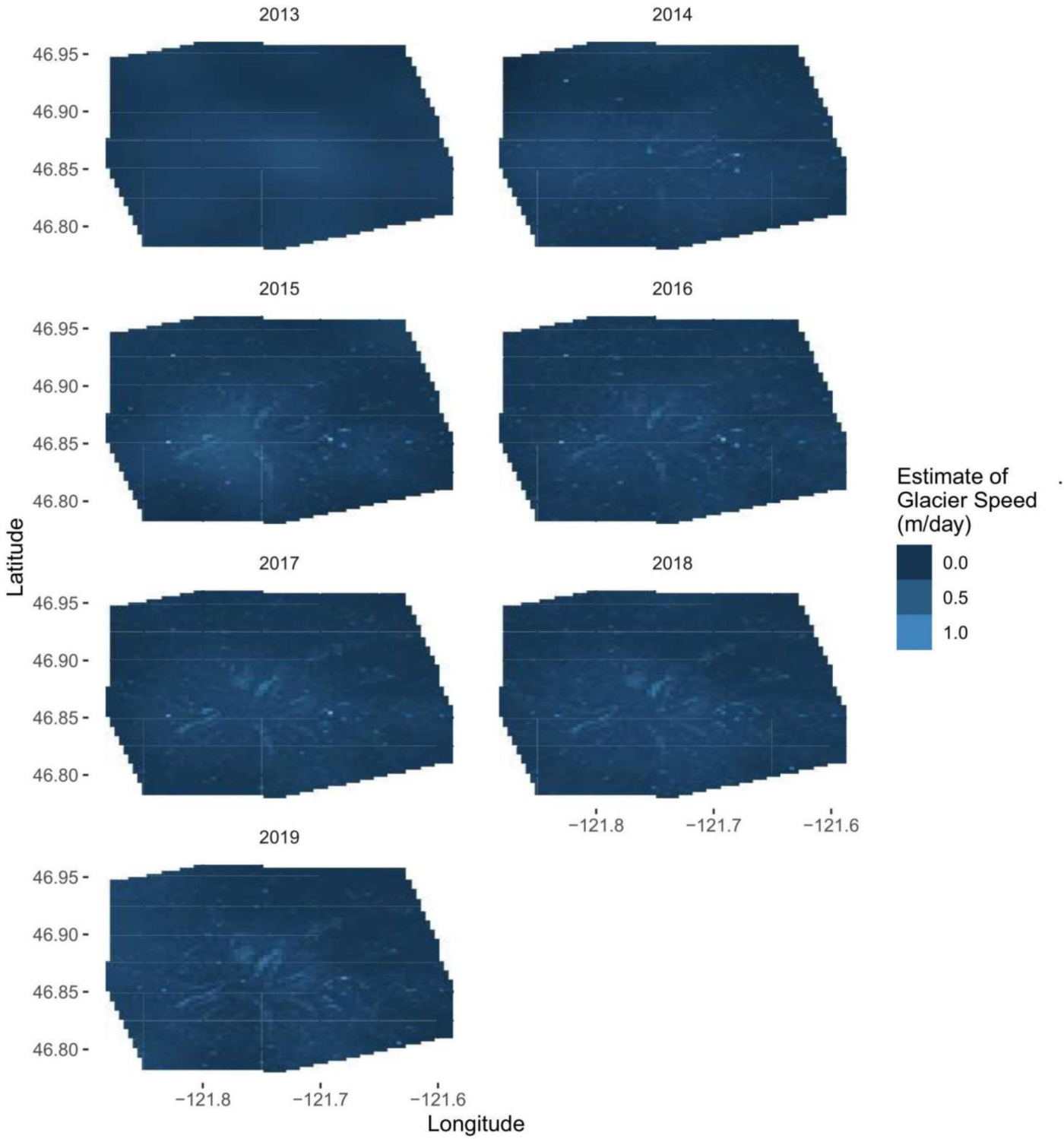


Figure 5.6: Grid of mean estimate of glacier speed of Mount Rainier's glaciers 2013-2019.

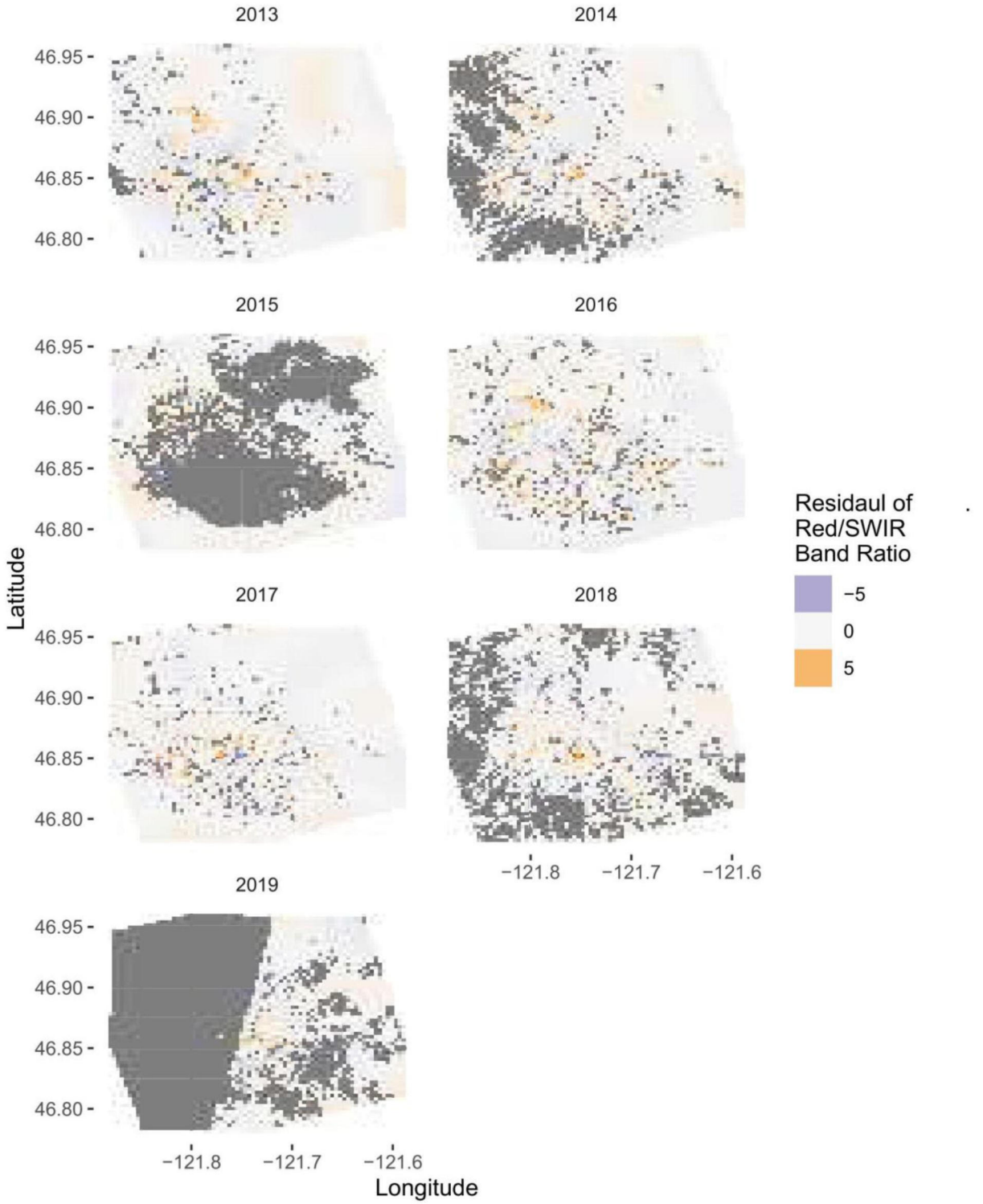


Figure 5.7: Grid of residuals of Red/SWIR Band ratio of Mount Rainier's glaciers 2013-2019. The grey indicates missing data due to clouds or cloud shadows.

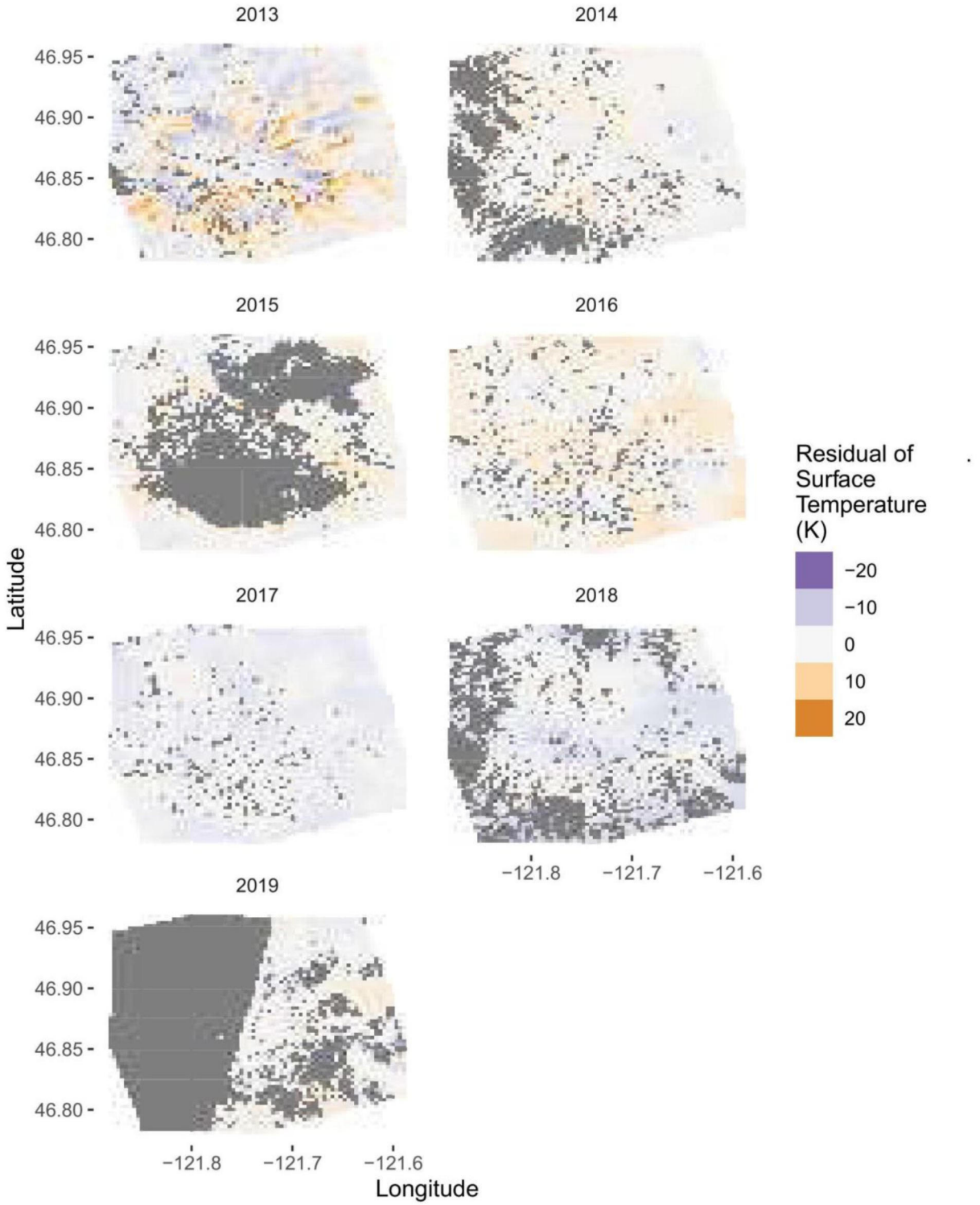


Figure 5.8: Grid of residuals of surface temperature of Mount Rainier's glaciers 2013-2019. The grey indicates missing data due to clouds or cloud shadows.

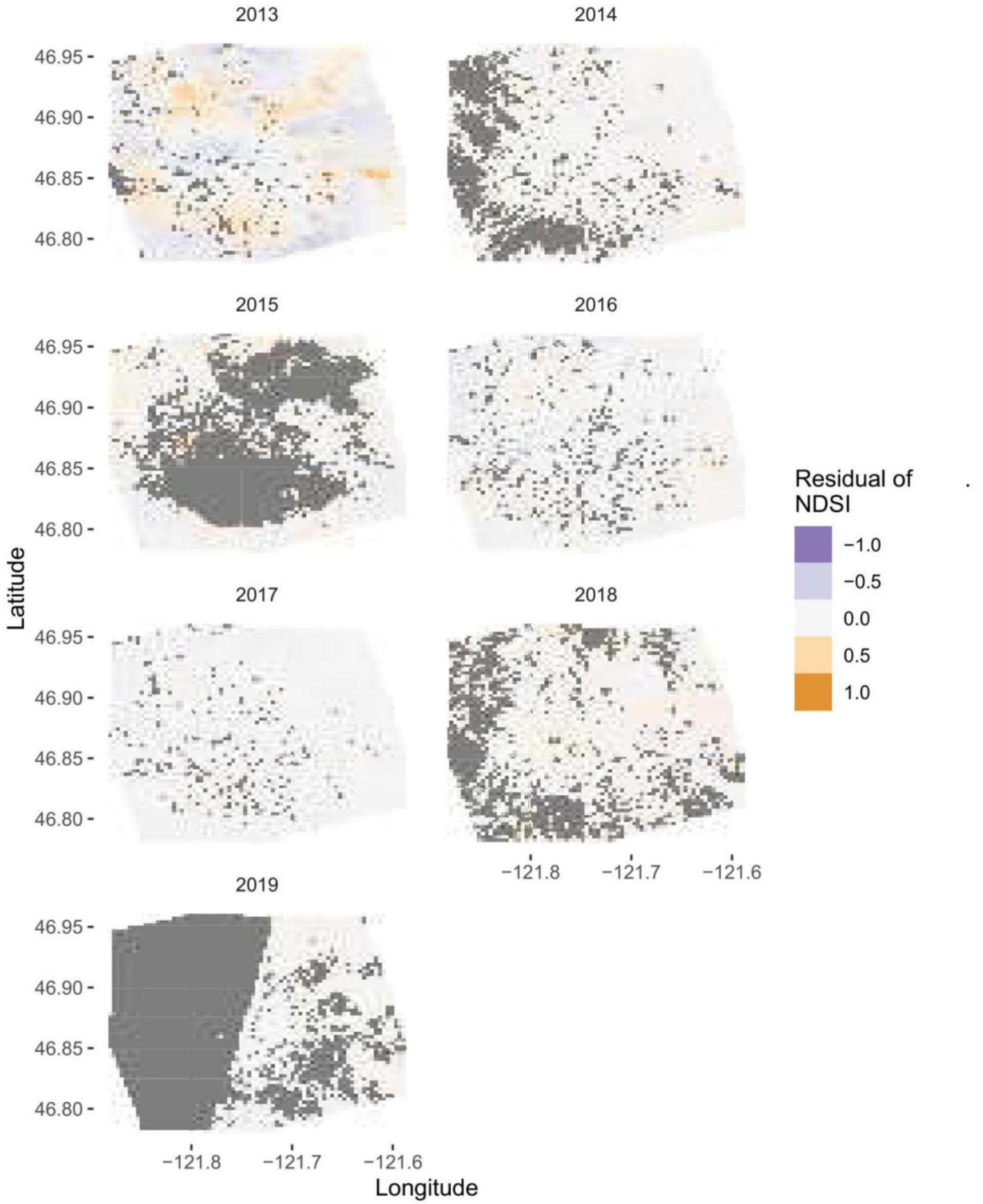


Figure 5.9: Grid of residuals of Mount Rainier's glaciers 2013-2019. The grey indicates missing data due to clouds or cloud shadows.

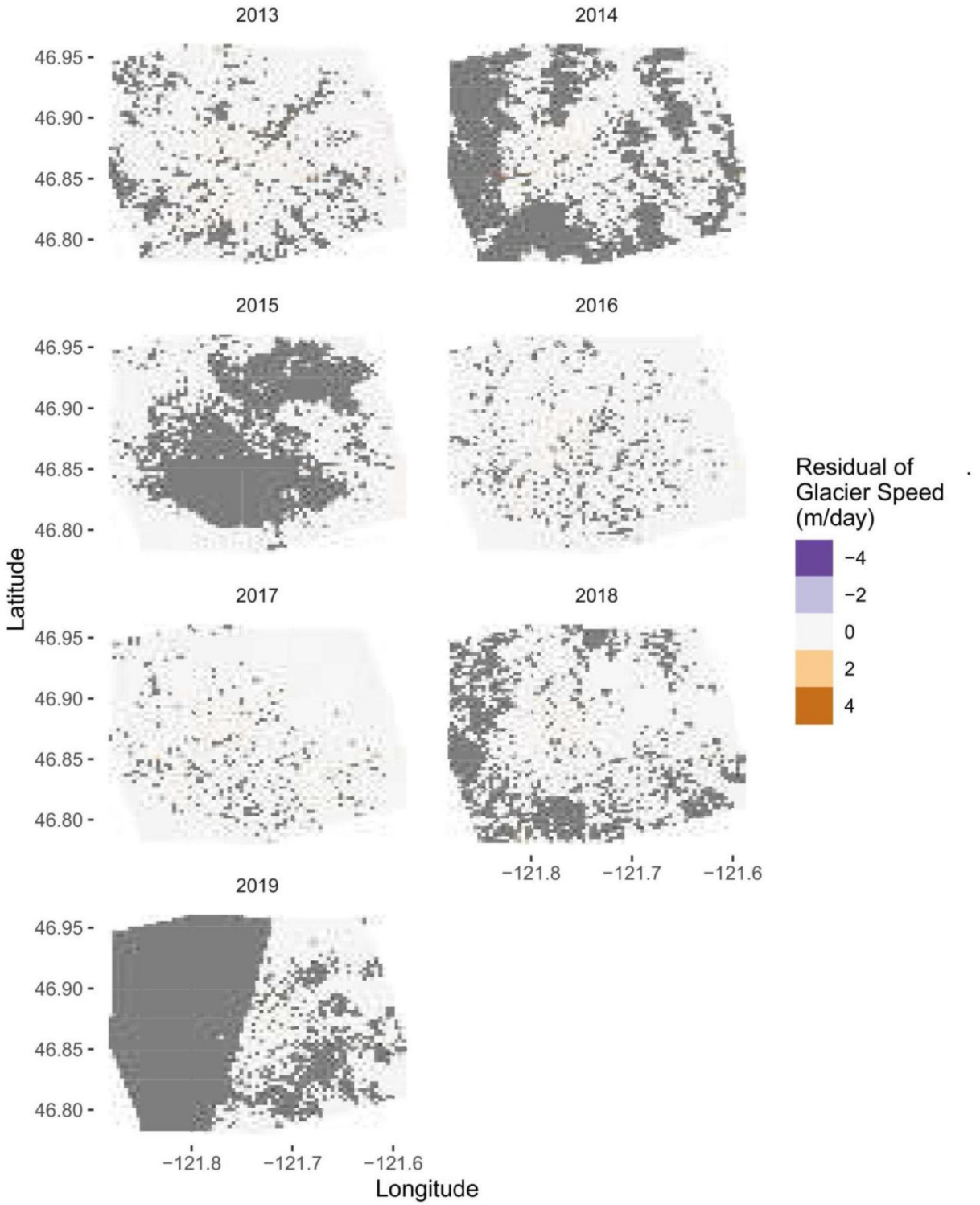


Figure 5.10: Grid of residuals of glacier speed of Mount Rainier's glaciers 2013-2019. The grey indicates missing data due to clouds or cloud shadows.

5.3 Classification

The classification was performed using the algorithms outlined in Section 4.9 to classify the glacier and non-glacier areas for 2015. The total classification was calculated as a weighted average of the glacier and non-glacier classifications with respect to the amount of pixels.

Figure 5.11 shows the classification of the final model sampled from the posterior for 2015 with an outline of the 2015 glacier mapping superimposed upon each image. Table 5.3 summarises the results for true values of glacier and surrounding area. Overall, all classification methods failed to classify the debris covered tongues of the northeastern glaciers. The Gaussian mixture model has the highest accuracy in classifying the glacier, and it does pick out some potential toes on the North edge of the glacier; however, it misclassifies these toes, showing them offset in adjacent locations. Compared to the other probabilistic classification techniques, soft K-means, Gaussian mixture model outperforms on both glacier and non-glacier accuracy. However, the Gaussian mixture model appears to classify too much glacier, and the edges of the glacier often protrude outside the red boundaries. The reason Gaussian Mixture Models still outperforms soft K-means is that the soft K-means algorithm had difficulty determining the non-glacier areas. There are large striations in the soft K-means image, and these add up and contribute to its low classification of the surrounding area.

The hard K-means and soft K-means perform similarly. Each algorithm finds the true edges of the glacier well, outside of the northern toes. However, both algorithms have difficulty finding the outline between adjacent glaciers. As pre-

Probability of Correct Classification

Glacier	Gaussian Mixture Models	Hard K-means	Hierarchical Clustering	Soft K-means	Support Vector Machines	Hard K-means Band Ratio (2016)
Yes	0.874	0.852	0.803	0.819	0.764	0.688
No	0.854	0.940	0.943	0.832	0.958	0.973
Total	0.858	0.922	0.915	0.830	0.919	0.916

Table 5.3: Summary table for the classification of Mount Rainier's glaciers in 2015. The total classification was calculated as a weighted average of the glacier and non-glacier classifications with respect to the amount of pixels.

viously noted, the large classification striations in the soft K-means image add up and contribute to its low classification accuracy of the surrounding area.

When it comes to finding the outline between adjacent glaciers, the support vector machine is the best model. While the hierarchical clustering and both K-means algorithms also did this to an extent, the support vector machine best separated adjacent glaciers. This is seen in the highest classification accuracy of the surrounding area. However, the support vector machine algorithm performed the worst in classifying the glacier itself, with the lowest accuracy of 76.4%. Support vector machines work by drawing a hyperplane to segment the data into groups. As the model predicted variables appear to underestimate the edges of the glacier, the SVM may draw the hyperplane to underlassify the glacier.

The hierarchical clustering algorithm had the second highest classification accuracy of the non-glacier area at 94.3%, just behind of the support vector machine at 95.8%. However it was least effective for the glacier itself, correctly classifying only 80.3%.

To show comparison to current work done in a univariate non-Bayesian para-

digm, a hard K-means classification was performed. As the 2016 data is much more complete than the 2015 data, it was performed on the red/SWIR band ratio on the 2016 data. When classifying the non-glacier areas, the univariate classification performed very well, this is seen in the highest non-glacier classification of 97.3%. This is clearly seen in the middle of Mount Rainer's glaciers, where only the support vector machines and the univariate K-means managed to delineate the individual glaciers. Despite this, the univariate K-means had the lowest classification of the glaciers itself as it correctly classified 68.8% of the glacier. This suggests that this multivariate and Bayesian analysis adds value.

Overall, the hard K-means algorithm appears to perform best. It does not pick up on any glacial toes; however, it does classify the easternmost isolated glacier, and does a good job of isolating adjacent glaciers, unlike the Gaussian mixture model. And unlike the soft K-means, the h K-means accurately classifies the surrounding bedrock as bedrock. While it does not have the highest accuracy of either the glacier or surrounding area, it is marginally behind the leading algorithms, and it has the highest overall accuracy of 92.2%.

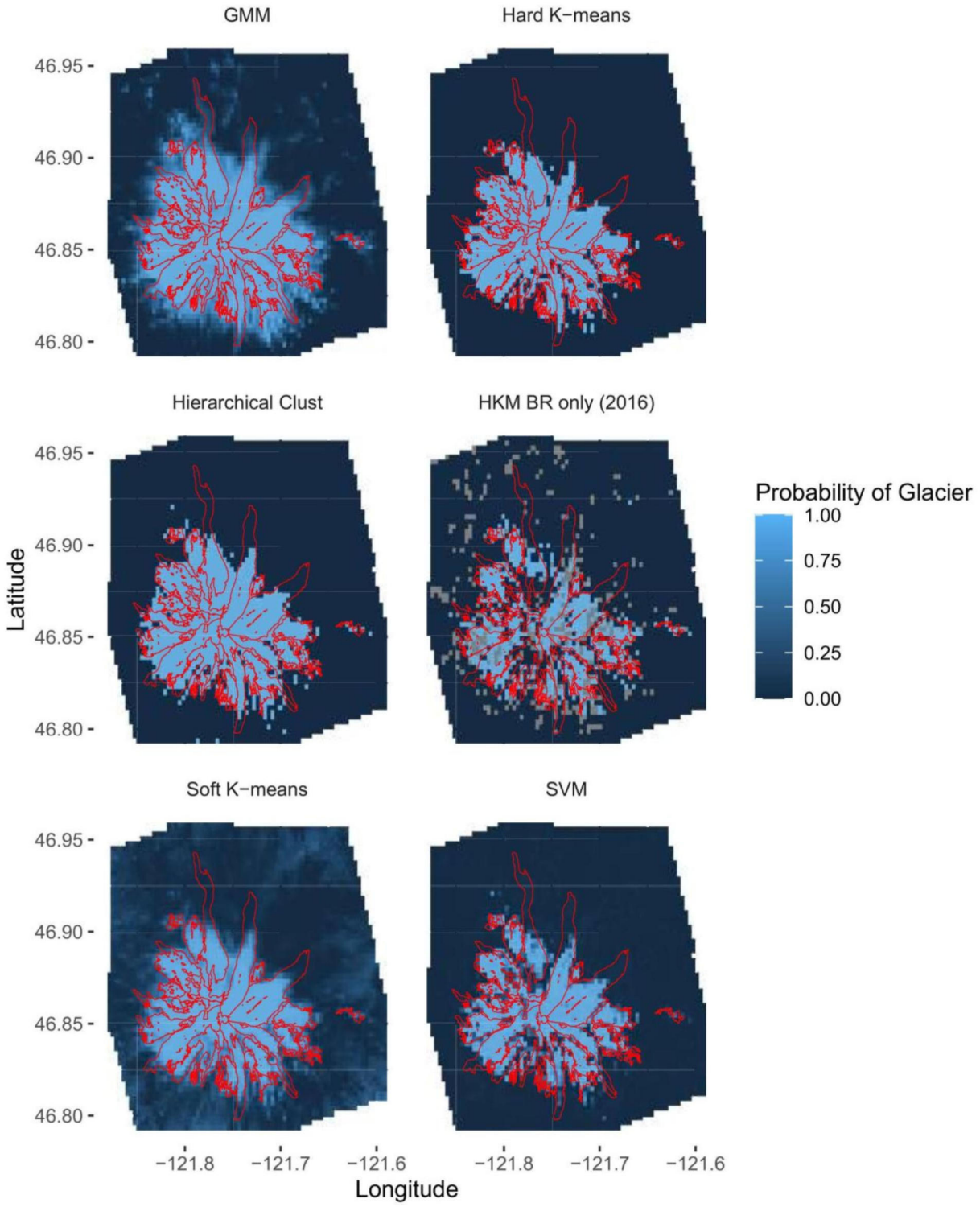


Figure 5.11: Glacier classification techniques for Mount Rainier's glaciers in 2015. The 2015 glacier extent is outlined in red.

5.4 Principal Components Analysis

In this work, the PCA was performed using the correlation matrix (or equivalently the covariance matrix with standard-normalised variables) on the model output of the four variables of interest for the 2015 data. The screeplot in figure 5.12 shows that at least 3 principal components should be used. Given that there is little difference between the eigenvalues of the third and fourth principal components, four principal components were chosen. The loadings of the principal components are shown in Table 5.4. The loadings of the first principal component suggests a contrast between the Red/SWIR band ratio and the glacier speed with the surface temperature. The second principal component is the NDSI. The third principal component is the contrast with Red/SWIR band ratio with glacier speed, and finally the fourth principal component is the contrast of Red/SWIR band ratio and surface temperature with glacier speed.

Loadings of PCA of mean estimates

	PC1	PC2	PC3	PC4
Band Ratio	0.535	-0.293	0.688	0.394
NDSI	0.206	0.955	0.190	0.099
Surface Temperature	-0.598	0.041	0.025	0.800
Glacier Speed	0.560	-0.028	-0.700	0.442

Table 5.4: Loadings of principal components analysis of the variables of interest. The screeplot in Figure 5.12 suggests all four principal components are necessary.

Given that all four principal components are used and there is no dimension reduction, this suggests that only a rotation is being applied upon the data. As Gaussian mixture models, the K-means algorithm, hierarchical clustering, and support vector machines with radial bases (as are used here) are invariant under rotation [70–72], this does not aid in the classification. Table 5.5 shows the results of the classification upon the rotated data; there is minimal change compared to Table 5.3. Note that approximate principal components are used for ease, so it is not a true orthogonal transformation, and thus minor differences are seen.

Despite not being able to aid in the classification, the PCA does add value. It suggests that one single variable (a linear combination of multiple variables) does not capture enough of variability in the model. So having a model with all four multivariate responses, as in this work, better models the spatiotemporal process compared to a univariate model.

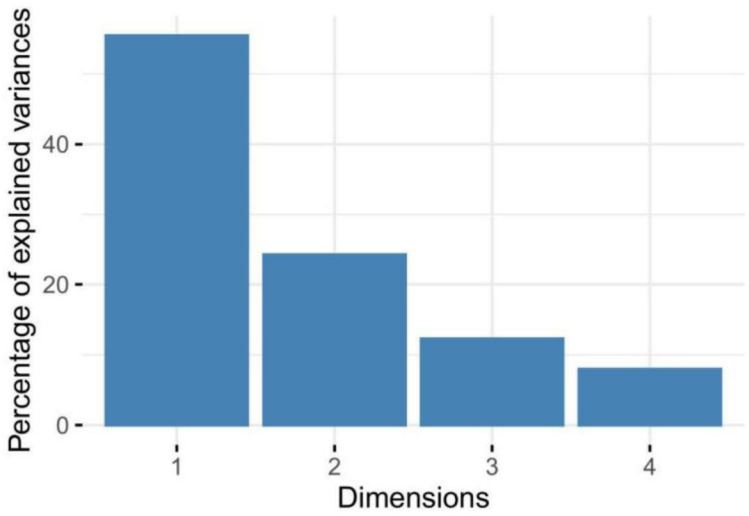


Figure 5.12: Screeplot of estimated Red/SWIR band ratio, NDSI, surface temperature, and glacial speed for Mount Rainier's glaciers.

Probability of Correct Classification of Rotated Data

Glacier	Gaussian Mixture Models	Hard K-means	Hierarchical Clustering	Soft K-means	Support Vector Machines	Hard K-means Band Ratio (2016)
Yes	0.868	0.851	0.805	0.818	0.764	0.688
No	0.859	0.940	0.945	0.836	0.958	0.973
Total	0.857	0.922	0.915	0.829	0.919	0.916

Table 5.5: Summary table of the rotated data for Mount Rainier's glaciers in 2015. The univariate classification was not rotated but added for comparison. The total classification was calculated as a weighted average of the glacier and non-glacier classifications with respect to the amount of pixels.

Chapter 6

Conclusions and Discussion

In terms of satisfying its research goals, this work successfully modelled the spatiotemporal data of the glaciers surrounding Mount Rainier, and it had mixed successes with the delineation of said glaciers. The delineation outperformed a comparable univariate analysis; however it failed to classify any of the debris covered glacier tongues. In order to do this modelling, a spatiotemporal GAM was chosen for its flexibility and computational speed, with ARMA processes and Gaussian processes respectively modelling the temporal and spatial autocorrelations. The ARMA(3,1) model with the exponential Gaussian process kernel was chosen via lowest pooled RMSE and WAIC of all four variables. Model predictions were then created by sampling from the posterior. These predictions were classified with a variety of techniques, from which the hard K-means algorithm measured by euclidean distance was determined to be the best. A further PCA was performed which showed that a reduction in the number of variables could not capture the

variability within the model and that multivariate response regression as justified. Furthermore compared to a univariate classification of the Red/SWIR band ratio only, the hard K-means multivariate classification outperformed the univariate classification by 85% to 68%.

One major limitation in this research is that there are only seven time indices. In the realm of time series analysis, this is a small sample size. The seven time indices also allowed for some assumptions in the modelling process that could not be checked by hypothesis tests. Given the nature of a glacier, each pixel is relatively stationary throughout time, until a certain point when the ice melts and the observation's expected value is no longer invariant with respect to time and the stationary assumption is violated. If these methods were applied to a longer time period, it would be reasonable to consider a change point model to accommodate nonstationarity, whereby an additional parameter can be used to separate two different time series models. Furthermore, the time and memory required, even with this restricted data, was substantial. A single ARMA model required up to 28 Gigabytes of memory for upwards of seven days of run time. Gaussian processes scale $O(n^3)$, so if double the data is used, the model requirements scale by 8. So further approximations, or different processes are needed if even a small amount of additional data is used. Since this research has been completed, additional ARD imagery has been uploaded for the years 2013-2019. Further research may use these additional images, older Landsat 7 imagery, or data from 2020 onwards for a larger dataset.

The Bayesian methods used allowed us to model the uncertainty not only in

the response, but also allows variation in the hyperparameters, unlike current methods. The issue that arose is that Bayesian computation is expensive in both time and memory demands. As a result, further approximations are needed to be made to run the model in a reasonable time. In this case, only one acquisition was used per year, and the Gaussian process length parameter was fixed at 12. This is seen in Figures 5.3, 5.4, 5.5, and 5.6 for 2013, in which the ARMA process had no support and the approximations within the Gaussian process clearly did not effectively outline the glacier. Including multiple scenes per year could help with modelling the intra-annual variation. Given the data is obtained from satellites rotating the earth, including multiple scenes per year could help mitigate sampling error. Additionally, the additional data would influence the modelling choice. Instead of modelling the spatial dependencies with Gaussian processes, conditional autoregressive models may be more suitable, as there would no longer be large amounts of missing data, and conditional autoregressive models are more sensitive to rapid changes, such as one would find at the glacier edges. However, given the large memory and time constraints on the analysis, the data was restricted to a singular image per year. While a glacier is a stationary object and is not expected to shrink more than a few meters per year, adding additional manner may help with first reconstructing a complete image on an annual basis, then modelling the spatiotemporal dynamics. This may increase the time and memory demands of the algorithm.

The methods here, as with many statistical techniques, rely on the missing data being MAR, the assumption of which can certainly be brought into ques-

tion. As previously noted, the Fmask algorithm, used to remove the clouds and cloud shadows, has difficulty distinguishing between clouds made of ice crystals and underlying snow or ice. As a result, given equal cloud cover, more glacier observations would be deemed missing than non-glacier observations. However, given the elevation range of Mount Rainier, an isolated strato-volcano, the cloud formation is not uniformly distributed, and will be more likely to form at certain elevations. This research glossed over this with a MAR assumption. I think it would be a valid extension to further investigate this by adding a missing data model parameter.

The missing data also influenced the algorithm complexity. Note that the simplest ARMA models, the AR(1) and MA(1), could be ruled out immediately by inspecting the data in Figures 3.4, 3.5, 3.6, and 3.7. The missing data in 2015 and 2019 does not allow any time series process with a single lag to make predictions for the following year. So the 2016 and future 2020 data would be not well modelled under an AR(1), MA(1), or ARMA(1,1) process. Current research avoids this problem by filling in these gaps, either through a Kriging-variant interpolation or through quantile regression [73–75] . The issues arising here are that these methods often flatten the spatiotemporal data, or they use the data twice. The first time is to fill in the gaps, and the second time is to model the spatiotemporal dynamics. A different way of approaching this would be not to use a discrete time series process, such as ARMA, but instead use a continuous time series process. The downside of this is that the sample size may grow many times in size, which will greatly increase the memory and computation time.

Glacier observations seem to be taken from several sub-populations. The first is the snow and ice on the glacier. The second is the surrounding bedrock. The third is the surrounding forest and alpine areas, and finally, speculatively, is the border area between the first two. One possible way would be to model this using a multi-modal distribution. The issue with applying this stems from the MCMC methods used to approximate the posterior distribution. The MCMC iterations happen in sequence, so if the modes of the posterior distribution are far apart then the MCMC samples have trouble transitioning from one peak to another. Given there are a finite, predetermined amount of MCMC iterations, the model may not transition in time from one peak to another for accurate modelling. The potential solution to this would be to split the data into three sections and model each independently. A difficulty with this is that as the glacier melts, pixels will transition from one state to another. This would require a dynamic spatiotemporal model with a non-separable covariance structure. This is another potential opportunity not seen in current glacier remote sensing research.

Another way to approach the debris cover question would be to involve different data sources. Given that debris covered glaciers and the surrounding bedrock are indistinguishable to the human eye, research shows little improvement with including finer resolution lidar data [76]. However, the addition of non-optical data, such as SAR data, may improve upon the classification. Given the methods themselves have no determinant connection to isolated strato-volcanos, the methods demonstrated here can easily be applied to other glaciers, ice fields, and other permanent ice structures. These methods here are also suited to other stationary

spatiotemporal data structures. Potential applications include but are not limited to mountain snow pack, algae bloom, and particulate spread from an emission source.

Bibliography

[1] USGS, "How much of the Earth's water is stored in glaciers?" https://www.usgs.gov/faqs/how-much-earths-water-stored-glaciers?qt-news_science_products=0, accessed 2019-05-01.

[2] R. Bhambri and T. Bolch, "Glacier mapping: a review with special reference to the Indian Himalayas," Progress in Physical Geography, vol. 33, no. 5, pp. 672–704, 2009.

[3] A. Fischer, B. Seiser, M. Stocker Waldhuber, C. Mitterer, and J. Abermann, "Tracing glacier changes in Austria from the Little Ice Age to the present using a lidar-based high-resolution glacier inventory in Austria," Cryosphere, vol. 9, no. 2, 2015.

[4] W. G. Rees and N. S. Arnold, "Mass balance and dynamics of a valley glacier measured by high-resolution LiDAR," Polar Record, vol. 43, no. 4, pp. 311–319, 2007.

[5] H. Alifu, B. A. Johnson, and R. Tateishi, "Delineation of debris-covered glaciers based on a combination of geomorphometric parameters and a TIR/NIR/SWIR Band Ratio," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 9, no. 2, pp. 781–792, 2015.

[6] A. E. Racoviteanu, M. W. Williams, and R. G. Barry, "Optical remote sensing of glacier characteristics: a review with focus on the Himalaya," Sensors, vol. 8, no. 5, pp. 3355–3383, 2008.

[7] K. Bayr, D. K. Hall, and W. M. Kovalick, "Observations on glaciers in the eastern Austrian Alps using satellite data," International Journal of Remote Sensing, vol. 15, no. 9, pp. 1733–1742, 1994.

[8] F. Paul, S. H. Winsvold, A. Kääb, T. Nagler, and G. Schwaizer, "Glacier remote sensing using Sentinel-2. Part II: Mapping glacier extents and surface facies, and comparison to Landsat 8," Remote Sensing, vol. 8, no. 7, p. 575, 2016.

[9] D. K. Hall, J. L. Foster, J. Y. Chien, and G. A. Riggs, "Determination of actual snow-covered area using Landsat TM and digital elevation model data in Glacier National Park, Montana," Polar Record, vol. 31, no. 177, pp. 191–198, 1995.

[10] R. Sidjak and R. Wheate, "Glacier mapping of the Illecillewaet icefield, British Columbia, Canada, using Landsat TM and digital elevation data," International Journal of Remote Sensing, vol. 20, no. 2, p. 273–284, 1999.

[11] M. Zhang, X. Wang, C. Shi, and D. Yan, "Automated glacier extraction index by optimization of red/SWIR and NIR/SWIR ratio index for glacier mapping using Landsat imagery," Water, vol. 11, no. 6, p. 1223, 2019.

[12] P. Rastner, T. Bolch, C. Notarnicola, and F. Paul, "A comparison of pixel- and object-based glacier classification With optical satellite Images," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 7, no. 3, pp. 853–862, 2014.

[13] T. Strozzi, A. Luckman, T. Murray, U. Wegmuller, and C. L. Werner, "Glacier motion estimation using SAR offset-tracking procedures," IEEE Transactions on Geoscience and Remote Sensing, vol. 40, no. 11, pp. 2384–2391, 2002.

[14] N. Short and A. Gray, "Potential for RADARSAT-2 interferometry: glacier monitoring using speckle tracking," Canadian Journal of Remote Sensing, vol. 30, no. 3, pp. 504–509, 2004.

[15] H. Hindberg, E. Malnes, and K. A. Høgda, "A monitoring system for glaciers on Svalbard based on Envisat ASAR wide swath data," in 2012 IEEE International Geoscience and Remote Sensing Symposium, pp. 1856–1859, 2012.

[16] NSIDC, "Global Land Ice Velocity Extraction from Landsat 8 (GoLIVE)." `https://nsidc.org/data/golive`, accessed 2019-05-01.

[17] B. Altena, T. Scambos, M. Fahnestock, and A. Kääb, "Extracting recent short-term glacier velocity evolution over southern Alaska and the Yukon from a large collection of Landsat data," The Cryosphere, vol. 13, no. 3, pp. 795–814, 2019.

[18] L. Sam, A. Bhardwaj, R. Kumar, M. Buchroithner, and F. J. Martín-Torres, "Heterogeneity in topographic control on velocities of Western Himalayan glaciers," Scientific Reports, vol. 8, 08 2018.

[19] C. Nuth, J. Kohler, M. König, A. von Deschwanden, J. O. Hagen, A. Kääb, G. Moholdt, and R. Pettersson, "Decadal changes from a multi-temporal glacier inventory of Svalbard," The Cryosphere, vol. 7, no. 5, pp. 1603–1621, 2013.

[20] B. Medley, I. Joughin, S. B. Das, E. J. Steig, H. Conway, S. Gogineni, A. S. Criscitiello, J. R. McConnell, B. E. Smith, M. R. van den Broeke, J. T. M. Lenaerts, D. H. Bromwich, and J. P. Nicolas, "Airborne-radar and ice-core observations of annual snow accumulation over Thwaites Glacier, West Antarctica confirm the spatiotemporal variability of global and regional atmospheric models," Geophysical Research Letters, vol. 40, no. 14, pp. 3649–3654, 2013.

[21] D. J. Wingham, D. W. Wallis, and A. Shepherd, "Spatial and temporal evolution of Pine Island Glacier thinning, 1995–2006," Geophysical Research Letters, vol. 36, no. 17, 2009.

[22] A. Greene, Response of Mountain Glaciers to Climate Forcing: Analysis and Applications. PhD thesis, Columbia University, 2002.

[23] Á. Moreno, F. J. García-Haro, B. Martínez, and M. A. Gilabert, "Noise reduction and gap filling of fAPAR time series using an adapted local regression filter," Remote Sensing, vol. 6, no. 9, pp. 8238–8260, 2014.

[24] M. Neteler, "Estimating daily land surface temperatures in mountainous environments by reconstructed MODIS LST data," Remote Sensing, vol. 2, no. 1, pp. 333–351, 2010.

[25] N. E. Barrand and T. Murray, "Multivariate controls on the incidence of glacier surging in the Karakoram Himalaya," Arctic, Antarctic, and Alpine Research, vol. 38, no. 4, pp. 489–498, 2006.

[26] A. Dehecq, N. Gourmelen, and E. Trouvé, "Deriving large-scale glacier velocities from a complete satellite archive: Application to the Pamir–Karakoram–Himalaya," Remote Sensing of Environment, vol. 162, pp. 55–66, 2015.

[27] P. Kraaijenbrink, J. Shea, F. Pellicciotti, S. De Jong, and W. Immerzeel, "Object-based analysis of unmanned aerial vehicle imagery to map and characterise surface features on a debris-covered glacier," Remote Sensing of Environment, vol. 186, pp. 581–595, 2016.

[28] V. Ayma, C. Beltrán, P. Happ, G. Costa, and R. Feitosa, "Mapping glacier changes using clustering techniques on cloud computing infrastructure," International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, 2019.

[29] S. Z. Gilani and N. I. Rao, "A clustering based automated glacier segmentation scheme using digital elevation model," in 2009 Digital Image Computing: Techniques and Applications, pp. 277–284, IEEE, 2009.

[30] N. Karimi, A. Farokhnia, L. Karimi, M. Eftekhari, and H. Ghalkhani, "Combining optical and thermal remote sensing data for mapping debris-covered glaciers (Alamkouh Glaciers, Iran)," Cold Regions Science and Technology, vol. 71, pp. 73 – 83, 2012.

[31] C. E. Rasmussen, "The infinite Gaussian mixture model," in Advances in Neural Information Processing Systems, pp. 554–560, 2000.

[32] F. Paul, "Evaluation of different methods for glacier mapping using landsat tm," EARSeL eProceedings, vol. 1, pp. 239–245, 2000.

[33] J. L. Dwyer, D. P. Roy, B. Sauer, C. B. Jenkerson, H. K. Zhang, and L. Lymburner, "Analysis ready data: enabling analysis of the Landsat archive," Remote Sensing, vol. 10, no. 9, p. 1363, 2018.

[34] S. Foga, P. L. Scaramuzza, S. Guo, Z. Zhu, R. D. Dilley Jr, T. Beckmann, G. L. Schmidt, J. L. Dwyer, M. J. Hughes, and B. Laue, "Cloud detection algorithm comparison and validation for operational Landsat data products," Remote Sensing of Environment, vol. 194, pp. 379–390, 2017.

[35] T. Bolch and U. Kamp, "Glacier mapping in high mountains using DEMs, Landsat and ASTER data," Zurich Open Repository and Archive, 2005.

[36] A. Bhardwaj, "Applicability of Landsat 8 data for characterising glacier facies and supraglacial debris," International Journal of Applied Earth Observation and Geoinformation, vol. 38, 12 2014.

[37] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, Bayesian Data Analysis. CRC press, 2013.

[38] R. P. Bukata, Satellite Monitoring of Inland and Coastal Water Quality. CRC Press, 2005.

[39] N. Cressie and C. K. Wikle, Statistics for Spatio-Temporal Data. John Wiley & Sons, 2015.

[40] G. C. Tiao and G. E. Box, "Some comments on 'Bayes' estimators," The American Statistician, vol. 27, no. 1, pp. 12–14, 1973.

[41] D. J. MacKay, "Bayesian interpolation," Neural Computation, vol. 4, no. 3, pp. 415–447, 1992.

[42] B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell, "Stan: A probabilistic programming language," Journal of Statistical Software, vol. 76, no. 1, 2017.

[43] L. Fahrmeir and S. Lang, "Bayesian inference for generalized additive mixed models based on Markov random field priors," Journal of the Royal Statistical Society: Series C (Applied Statistics), vol. 50, no. 2, pp. 201–220, 2001.

[44] R. Murray-Smith and A. Girard, "Gaussian process priors with ARMA noise models," in Irish Signals and Systems Conference, Maynooth, pp. 147–152, 2001.

[45] T. Hengl, G. B. Heuvelink, M. P. Tadić, and E. J. Pebesma, "Spatio-temporal prediction of daily temperatures using time-series of MODIS LST images," Theoretical and Applied Climatology, vol. 107, no. 1-2, pp. 265–277, 2012.

[46] Q. Yao and P. J. Brockwell, "Gaussian maximum likelihood estimation for ARMA models II: spatial processes," Bernoulli, vol. 12, no. 3, pp. 403–429, 2006.

[47] C. J. Verzilli, N. Stallard, and J. C. Whittaker, "Bayesian modelling of multivariate quantitative traits using seemingly unrelated regressions," Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society, vol. 28, no. 4, pp. 313–325, 2005.

[48] A. Zellner and T. Ando, "A direct Monte Carlo approach for Bayesian analysis of the seemingly unrelated regression model," Journal of Econometrics, vol. 159, no. 1, pp. 33–45, 2010.

[49] B. Liquet, K. Mengersen, A. Pettitt, and M. Sutton, "Bayesian variable selection regression of multivariate responses for group data," Bayesian Analysis, vol. 12, no. 4, pp. 1039–1067, 2017.

[50] M. A. Zapala and N. J. Schork, "Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables," Proceedings of the National Academy of Sciences, vol. 103, no. 51, pp. 19430–19435, 2006.

[51] A. Schuster, "On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena," Terrestrial Magnetism, vol. 3, no. 1, pp. 13–41, 1898.

[52] W. M. Persons, "Correlation of time series," Journal of the American Statistical Association, vol. 18, no. 142, pp. 713–726, 1923.

[53] G. Box, "Time Series Analysis, Forecasting, and Control," Francisco Holden-Day, 1970.

[54] H. Wold, A study in the Analysis of Stationary Time Series. PhD thesis, Almqvist & Wiksell, 1938.

[55] J. F. Monahan, "Fully Bayesian analysis of ARMA time series models," Journal of Econometrics, vol. 21, no. 3, pp. 307–331, 1983.

[56] K. A. Hoover and M. G. Wolman, "Beyond the semivariogram: Patterns, scale, and hydrology in a semi-arid landscape," Advances in Water Resources, vol. 28, no. 9, pp. 885 – 898, 2005.

[57] F. Pellicciotti, S. Ragettli, M. Carenzo, and J. McPhee, "Changes of glaciers in the Andes of Chile and priorities for future work," Science of The Total Environment, vol. 493, pp. 1197 – 1210, 2014.

[58] M. S. Handcock and M. L. Stein, "A Bayesian analysis of Kriging," Technometrics, vol. 35, no. 4, pp. 403–410, 1993.

[59] D. G. Krige and D. Krige, Lognormal-de Wijsian Geostatistics for Ore Evaluation. South African Institute of Mining and Metallurgy Johannesburg, 1981.

[60] D. J. Mackay, "Gaussian processes," A replacement for neural networks, NIPS Tutorial, pp. 514–520, 1997.

[61] J. G. Anderson and J. N. Brune, "Probabilistic seismic hazard analysis without the ergodic assumption," Seismological Research Letters, vol. 70, no. 1, pp. 19–28, 1999.

[62] Stan Development Team, "Stan modeling language users guide and reference manual, version 2.18.0," 2018.

[63] P. C. Phillips, "Bayesian model selection and prediction with empirical applications," Journal of Econometrics, vol. 69, no. 1, pp. 289–331, 1995.

[64] R. Salakhutdinov and A. Mnih, "Bayesian probabilistic matrix factorization using Markov chain Monte Carlo," in Proceedings of the 25th International Conference on Machine Learning, pp. 880–887, 2008.

[65] A. Vehtari, A. Gelman, and J. Gabry, "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC," Statistics and Computing, vol. 27, no. 5, pp. 1413–1432, 2017.

[66] T. M. Ludden, S. L. Beal, and L. B. Sheiner, "Comparison of the Akaike Information Criterion, the Schwarz criterion and the F test as guides to model selection," Journal of Pharmacokinetics and Biopharmaceutics, vol. 22, no. 5, pp. 431–445, 1994.

[67] D. Anderson, K. Burnham, and G. White, "Comparison of Akaike information criterion and consistent Akaike information criterion for model selection and statistical inference from capture-recapture studies," Journal of Applied Statistics, vol. 25, no. 2, pp. 263–282, 1998.

[68] S. Watanabe and M. Opper, "Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory," Journal of Machine Learning Research, vol. 11, no. 12, 2010.

[69] B. Raup, A. Racoviteanu, S. J. S. Khalsa, C. Helm, R. Armstrong, and Y. Arnaud, "The GLIMS geospatial glacier database: a new tool for studying glacier change," Global and Planetary Change, vol. 56, no. 1-2, pp. 101–110, 2007.

[70] S. Abe, "On invariance of support vector machines," in Proceedings of the 4th International Conference on Intelligent Data Engineering and Automated Learning, 2003.

[71] S. Luo, W. Mou, M. Li, K. Althoefer, and H. Liu, "Rotation and translation invariant object recognition with a tactile sensor," in SENSORS, 2014 IEEE, pp. 1030–1033, IEEE, 2014.

[72] D. Adametz, Invariances for Gaussian models. PhD thesis, University_of_Basel, 2015.

[73] C. Zhang, W. Li, and D. Travis, "Gaps-fill of SLC-off Landsat ETM+ satellite image using a geostatistical approach," International Journal of Remote Sensing, vol. 28, no. 22, pp. 5103–5122, 2007.

[74] S. M. Howard and J. M. Lacasse, "An evaluation of gap-filled Landsat SLC-off imagery for wildland fire burn severity mapping," Photogrammetric Engineering and Remote Sensing, vol. 70, no. 8, pp. 877–880, 2004.

[75] F. Gerber, R. de Jong, M. E. Schaepman, G. Schaepman-Strub, and R. Furrer, "Predicting missing values in spatio-temporal remote sensing data," IEEE Transactions on Geoscience and Remote Sensing, pp. 1–13, 2018.

[76] T. Reid, M. Carenzo, F. Pellicciotti, and B. Brock, "Including debris cover effects in a distributed model of glacier ablation," Journal of Geophysical Research: Atmospheres, vol. 117, no. D18, 2012.

[77] S. H. Winsvold, A. Kääb, and C. Nuth, "Regional glacier mapping using optical satellite data time series," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 9, no. 8, pp. 3698–3711, 2016.

[78] R. Frauenfelder and A. Kääb, "Glacier mapping from multi-temporal optical remote sensing data within the Brahmaputra River Basin," pp. 922–925, 01 2009.

[79] M. König, J.-G. Winther, J. Kohler, and F. König, "Two methods for firn-area and mass-balance monitoring of Svalbard glaciers with SAR satellite images," Journal of Glaciology, vol. 50, no. 168, pp. 116–128, 2004.

[80] A. Brenning, "Benchmarking classifiers to optimally integrate terrain analysis and multispectral remote sensing in automatic rock glacier detection," Remote Sensing of Environment, vol. 113, pp. 239–247, 01 2009.

[81] D. Farinotti, "On the effect of short-term climate variability on mountain glaciers: insights from a case study," Journal of Glaciology, vol. 59, no. 217, p. 992–1006, 2013.

[82] U. Naeem, M. Shamim, N. Ejaz, D. H. Rehman, U. Mustafa, H. Hashmi, and A. R. Ghumman, "Investigation of temporal change in glacial extent of Chitral watershed using Landsat data," Environmental Monitoring and Assessment, vol. 188, pp. 1–13, 12 2016.

[83] C. I. Millar, R. D. Westfall, and D. L. Delany, "Thermal and hydrologic attributes of rock glaciers and periglacial talus landforms: Sierra Nevada, California, USA," Quaternary International, vol. 310, pp. 169 – 180, 2013. PACLIM: Proceedings of the 25th Pacific Climate Workshop, 2011.

[84] Y. Liu, J. Wu, Y. Liu, B. X. Hu, Y. Hao, X. Huo, Y. Fan, T. J. Yeh, and Z.-L. Wang, "Analyzing effects of climate change on streamflow in a glacier mountain catchment using an ARMA model," Quaternary International, vol. 358, pp. 137 – 145, 2015.

[85] USGS, "SLC Gap-Fill Methodology." `https://www.usgs.gov/media/files/landsat-7-slc-gap-filled-products-phase-one-methodology`, accessed 2019-05-01.

[86] USGS, "SLC-off Gap-Filled Products Gap-Fill Algorithm Methodology." `https://prd-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/atoms/files/L7SLCGapFilledMethod.pdf`, accessed 2019-05-01.

[87] R. E. Rossi, J. L. Dungan, and L. R. Beck, "Kriging in the shadows: geostatistical interpolation for remote sensing," Remote Sensing of Environment, vol. 49, no. 1, pp. 32–40, 1994.

[88] E. Addink, "A comparison of conventional and geostatistical methods to replace clouded pixels in NOAA-AVHRR images," International Journal of Remote Sensing, vol. 20, no. 5, pp. 961–977, 1999.

[89] Y. Zhu, E. L. Kang, Y. Bo, Q. Tang, J. Cheng, and Y. He, "A Robust fixed rank Kriging method for improving the spatial completeness and accuracy of satellite SST products," IEEE Transactions on Geoscience and Remote Sensing, vol. 53, no. 9, pp. 5021–5035, 2015.

[90] L. Guo, L. Lei, Z. Zeng, P. Zou, D. Liu, and B. Zhang, "Evaluation of spatio-temporal variogram models for mapping xco 2 using satellite observations: A case study in China," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 8, no. 1, pp. 376–385, 2014.

[91] R. Guhaniyogi and S. Banerjee, "Multivariate spatial meta kriging," Statistics & Probability Letters, vol. 144, pp. 3–8, 2019.

[92] P. Scaramuzza and J. Barsi, "Landsat 7 scan line corrector-off gap-filled product development," in Proceeding of Pecora, vol. 16, pp. 23–27, 2005.

[93] R. Hodgkins, "Glacier hydrology in Svalbard, Norwegian high Arctic," Quaternary Science Reviews, vol. 16, no. 9, pp. 957–973, 1997.

[94] S. Lang, S. B. Adebayo, L. Fahrmeir, and W. J. Steiner, "Bayesian geoadditive seemingly unrelated regression," Computational Statistics, vol. 18, no. 2, pp. 263–292, 2003.

[95] M. Smith and R. Kohn, "Nonparametric seemingly unrelated regression," Journal of Econometrics, vol. 98, no. 2, pp. 257–281, 2000.

[96] D. K. Hammond and E. P. Simoncelli, "Image modeling and denoising with orientation-adapted Gaussian scale mixtures," IEEE Transactions on Image Processing, vol. 17, no. 11, pp. 2089–2101, 2008.