Spatio-temporal modelling of extreme temperature events on the Greenland ice sheet

Daniel Clarkson, BSc (Hons), MSc

Department of Mathematics and Statistics

Lancaster University

A thesis submitted for the degree of

Doctor of Philosophy

May, 2023

Declaration

I declare that the work presented in this thesis is, to the best of my knowledge and belief, original and my own work. The material has not been submitted, either in whole or in part, for a degree at this, or any other university. This thesis does not exceed the maximum permitted word length of 80,000 words including appendices and footnotes, but excluding the bibliography.

Daniel Clarkson

Acknowledgements

Firstly, I'd like to hugely thank my supervisors Emma Eastoe and Amber Leeson for their guidance, encouragement, and support throughout the PhD. From initial meetings before applying for the position to the last days of the PhD, I have felt fully supported and encouraged to pursue new ideas and have thoroughly enjoyed the entire process in large part due to their encouragement and positivity. I'm especially grateful to Emma for having to read through so many of my writing attempts and to Amber for her ever-present enthusiasm for new research ideas, both of whom managed the collaboration between topics extremely well.

I'd like to particularly thank Emma for her personal support of my work with Morecambe Bay Foodbank, which has been a consistent feature alongside the PhD. The work has been a huge learning experience for me evolving from a simple request for one line graph to however I can even describe the work now, and I'll always appreciate the support despite it being neither of our areas of expertise and completely separate from this thesis. I'd also like to thank Annette Smith for her faith in me when beginning the work, and Morecambe Bay Foodbank for their continued support of my work and all that they do for the local community.

I'm grateful to David Leslie and the DSNE project for allowing me to have such an enjoyable time doing research and providing a great working environment to do it in. Having multiple areas of science on the same project helped me learn a lot more than I would have in isolation, and the project supported me and the other students well throughout our time.

For my time outside of work, I thank LUBDS and everyone who has been in the society over the years. I never quite imagined I'd be in it for so long when I first joined, but my passion for Ballroom is as strong as ever and I'll always treasure my time in the society. Special thanks to Eloise, Gary, Ethan, Shane, Jade, Izzy, Laura, Ellie, Anna, and Aldo for their support in so many ways.

Lastly, thank you to my family and friends for their support and encouragement

throughout my surprisingly long time as a student.

Spatio-temporal modelling of extreme temperature events on the Greenland ice sheet

Daniel Clarkson, BSc (Hons), MSc.

Department of Mathematics and Statistics, Lancaster University

A thesis submitted for the degree of *Doctor of Philosophy*. May, 2023.

Abstract

The Greenland ice sheet has experienced significant melt over the past six decades, with rare extreme melt events covering large areas of the ice sheet. Melt events are typically analysed using summary statistics from satellite data, but the nature and characteristics of the events themselves are less frequently analysed. In this thesis, we take MODIS satellite temperature data and develop a series of models to build a detailed understanding of temperature, melt, and extreme temperature events on the ice sheet. A core aim of the modelling work is to create and use models that are statistically robust that also strongly consider the scientific context of the variables and processes being modelled. We first develop a statistical model for temperatures at a single location on the ice sheet. We define a novel method of identifying melt observations using a Gaussian mixture model to capture the distribution of temperatures across the ice sheet in a consistent format. In the next chapter, we begin to examine the spatial trends in the data by examining the mixture model's parameters in a spatial setting. We use a regression model to predict the mixture model parameters for a given location based only on geographic spatial variables, allowing us to estimate the distribution of temperatures for any location using only a set of coordinates and information derived from them. We then examine spatial dependence between locations using a Gaussian process. Using the mixture model as a marginal model and insights from the regression model, we quantify the spatial dependence in the data and simulate temperature realisations for the entire ice sheet. Finally, we use the spatial conditional extremes model to model extreme

temperature events. Using the model, we can describe the characteristics of extreme temperature events and simulate and predict them.

Publications

Two publications have been created directly from the thesis, which have together formed the majority of Section 2:

Clarkson, D., Eastoe, E., and Leeson, A.: Melt probabilities and surface temperature trends on the Greenland ice sheet using a Gaussian mixture model, The Cryosphere, 16, 1597–1607, https://doi.org/10.5194/tc-16-1597-2022, 2022

Clarkson, Daniel and Eastoe, Emma and Leeson, Amber (2022) The importance of context in extreme value analysis with application to extreme temperatures in the USA and Greenland. Journal of the Royal Statistical Society: Series C (Applied Statistics). ISSN 0035-9254 (In Press)

Contents

1 Introduction

In a globally changing environment driven by climate change, there are an abundance of questions to be answered and sources of data with which to answer them. The increase in availability and resolution of data sources relating to climate changes presents challenges and opportunities within a modelling and analysis setting. Satellites, Automatic Weather Stations (AWSs), and climate models amongst others allow more sophisticated statistical models to be used, both from applying previous models on larger amounts of data and allowing for the development of new modelling techniques. Developing these new models is of particular interest, especially when that involves consideration of the context of the scientific field in the model's structure. As the needs for scientific expertise and more advanced statistical models both increase, so does the need to balance considerations and insight from multiple fields within the same analysis to give the best insight into the challenges at hand.

In addition to the overall trends and effects observed from climate change, of additional interest is the behaviour and associated risk from extreme climate-related events. Many fields have interest in rare events that typically have few examples from data sets but the largest impacts on people, the environment and society, which naturally leads to extreme value analysis methods. The roots of the statistical modelling of extreme events lie in the theoretical work of Von Mises (1936) and Gnedenko (1943) who sought to understand the asymptotic behaviour of sample maxima. From the 1950s, this probability theory began to be translated into the statistical modelling framework now known as the block maxima method (Gumbel, 1958). Both theory and modelling framework were later extended to a peaks over threshold approach (Pickands, 1975; Davison and Smith, 1990a). It was recognised early on that both block maxima and peaks over threshold frameworks could make major contributions to the study of natural hazards, leading to a long-standing relationship with the environmental sciences and civil engineering. Applications include hydrology (Katz et al., 2002; Jonathan and Ewans, 2013), air

11

pollution (Eastoe and Tawn, 2009; Reich et al., 2013; Gouldsborough et al., 2021), temperatures (Acero et al., 2014; Winter et al., 2016), precipitation (Katz, 1999), drought (Burke et al., 2010), wind speeds (Fawcett and Walshaw, 2006), statistical downscaling (Friederichs, 2010; Towe et al., 2017), space weather (Thomson et al., 2011; Rogers et al., 2020) and climate change problems (Sterl et al., 2008; Cooley, 2009; Cheng et al., 2014). In each case, notions of risk and analysing rare, high magnitude events are of central importance, particularly in light of a changing global environment.

One impact of the rise in global temperatures caused by climate change has been an increased volume of melt water generated by the world's ice sheets (Hanna et al., 2013). This has led to a sea level rise that has already increased the risk of flooding in low-lying areas and will continue to do so whilst temperatures continue to rise (Nicholls and Cazenave, 2010). The two largest ice sheets, Greenland and the Antarctic, have had an accelerating contribution to sea level rise since the early 1990's (Rignot et al., 2019), and are likely to continue to play a key role in the near future (Rahmstorf et al., 2017). In this context, both individual instances of melt on the ice sheet and large spatial melt events can both be considered as extreme events from a modelling perspective. Understanding the spatial extent, frequency, magnitude and trends of such melt events is vital to quantify risks caused by climate change to sea levels and to the ice sheet itself, both now and in the future. This is not least because the consequences of increased melt are likely to persist for years to come (Kulp and Strauss, 2019).

The Greenland ice sheet has experienced significant melt over the past 6 decades (Fettweis et al., 2011), with an accelerating contribution to sea level rise (Rignot et al., 2011). Wide regions of the ice sheet have lost mass over the last 2 decades, and the total surface melt area has also increased (IPCC, 2013). Combined with glacial melt, melt from other ice sheets and thermal expansion of the oceans, total sea level rise has been far above the historical rate of sea level rise during this time period (Cazenave and Nerem, 2004). Understanding how frequently, where and when melt

occurs on the Greenland ice sheet is a key part of understanding its role in sea level rise and how we can expect it to change in the future.

Melt can analysed in two main forms - direct observations or via proxy by temperature. Direct observations - often from passive microwave measurements (Fettweis et al., 2011) - give a binary indication of melt or no melt. This is useful for the identification of melt, but lacks nuance when more consideration is placed on the conditions in which melt occurs and how melt might change as a result of climate change. Temperature values offer a more detailed image of the conditions on the ice sheet, from which melt can also be derived. This enables more information about the conditions of the ice sheet to be represented by a statistical model while maintaining inference for the main variable of interest of melt. Furthermore, due to the nature of a continuous variable such as temperature compared to a binary melt variable, it allows extreme value methods to be more easily applied to the data which thereby enables a broad range of challenges to be addressed by statistical modelling.

Due to the strong links between temperature and ice melt (Vermeer and Rahmstorf, 2009), temperature data are often used to analyse melt, with temperatures exceeding 0°C or a similar threshold being interpreted as evidence of melt. There are many ways to measure the temperature of the ice sheet, including satellites (Zhengming and Dozier, 1989), Global Climate Models (GCMs) (Smith et al., 2007), and AWSs (Tedesco et al., 2013). Each data source has advantages and limitations in terms of accuracy and coverage. Comparisons between them can give greater confidence in their respective ability to represent the true surface temperature, and by choosing data sets with a high resolution in both space and time, we can examine the past 20 years of Greenland temperatures with reasonable confidence and depth.

One of the many benefits of having such high resolution data sets is that extreme melt events in both space and time can be identified and examined in greater detail. In 2012, anomalous melting was observed during mid-July, with 98.6% of the ice-sheet experiencing melt (Hanna et al., 2014; Nghiem et al., 2012). Since 2012,

particular attention has been paid to these extreme melt events, in which large areas of the ice sheet simultaneously experience melt. The existing literature mostly considers summary statistics, whereby melt is analysed using sum totals (Mernild et al., 2011) or extreme instances of melt are observed from the data (Nghiem et al., 2012). Although such measures are helpful in giving a broad view of the total ice melt in a year, examining extreme events gives us a much more detailed understanding of how melt occurs on the ice sheet. Understanding the conditions in which these events occur and the times and locations that they occur can give us a deeper understanding of the melt process when combined with traditional approaches.

When identifying both melt at an individual location and wider spread melt events, it is important to think about how to accurately define melt. At what temperature can we most accurately say melt occurs? How much of an area needs to have experienced melt for us to classify it as an extreme melt event? Do the properties of melt and melt events vary over space and time? We can attempt to answer these questions using statistical techniques that account for the clear uncertainty of the answers we're looking for. Particularly for extreme melt events, extreme value theory is an ideal branch of statistics to apply to the topic and can give us good insight into extreme melt on the ice sheet.

1.1   Structure

In Section 1.2 we give an overview of Extreme value analysis. We describe the set of models and principles that will feature heavily throughout the thesis, either in approach or in the specific models applied to the data set. This includes a definition and background of extreme value analysis; introduction to the peaks over threshold approach; introduction to conditional extreme value theory; and a definition of asymptotic dependence. This sets the background for our approach to analysing extreme temperature events and gives an indication of how and why we use the methods that we use to conduct our analysis.

In Section 2, we aim to gain a thorough understanding of temperature across the

14

Greenland ice sheet. We assess possible data sets to use for temperature analysis and the types of measurement that can be taken; conduct a simple melt extent analysis using satellite data from the last 20 years; and model the temperature distribution at each observed location using a Gaussian mixture model. From this, a clear picture of overall temperature and melt trends across the ice sheet can be drawn, as well as an initial understanding of the spatial variation of extreme temperatures.

In Section 3, we explore the spatial trends of the mixture model parameters from the fitted models in Section 2. We aim to identify key spatial covariates that are related to temperatures, specifically through relationships with the parameters of the Gaussian mixture model rather than directly with temperature observations. By assessing these relationships in a regression framework, we can predict the temperature distribution via its mixture model parameters for any given location on the ice sheet. This provides evidence of the most useful spatial covariates to use in future modelling attempts, and gives us a deeper understanding of the marginal mixture models for single site behaviour.

Section 4 moves from spatial relationships between marginal model parameters to examining the correlation between temperatures at different locations. Using the mixture model as a marginal model, we fit Gaussian process models to the data to examine the spatial structure of the data across the ice sheet. We consider a range of distance measures - informed by the regression work - to include in the covariance matrix of the Gaussian process in order to improve the model fit, as well as splitting the ice sheet into regions to account for areas with different dependence structures. The allows us to simulate temperatures across the entire ice sheet for a given day, and evaluate the probability of simultaneous melt at multiple locations rather than at a single location. It also shows the scale of dependence between locations and how closely linked temperatures on the ice sheet are.

The final section - Section 5 - focuses specifically on spatial extreme events, having built up an understanding of marginal extremes, spatial trends and spatial dependence from the previous models. Rather than modelling the dependence of

15

the entire temperature distribution, we focus on the dependence of the extreme temperatures. The Spatial conditional extremes model allows simulations of extreme temperature events, giving insight into how the structure of extreme events differ to regular temperature events.

1.2   Extreme value analysis

Extreme value analysis (EVA) focuses on the modelling and analysis of the highest or lowest observations from a data set - the tails of the distribution. An extreme value is defined within the context of a data set, and all points classified as extreme are extracted and modelled separately to the rest of the distribution using an extreme value distribution. This tail distribution can be used to examine the probability of rare events outside the range of the observed data, such as estimating an extreme value only expected to occur once in 1000 years from 20 years of data. The motivation for this approach is that often the most extreme values in a data set make up only a small part of the data set - typically $\leq 10\%$ of values are defined as extreme for the most common approaches. If interest lies solely in the distribution of the extreme values rather than the full data set, it may be inappropriate to make inference on their distribution by fitting a model to the entire data set. While fitting a distribution to the entire data set may result in a good fit for the main body of the observed distribution of data, for the tails of the distribution it may be more difficult to find an appropriate fit. For the accurate modelling of extreme values, it is be more appropriate to fit a model or distribution only to the extreme values, particularly if their behaviour is suspected to be different from that of the main body of the data set. Depending on how the extreme values are defined, specific tail (extreme value) distributions can then be applied to the extreme values. Once fitted, these distributions may allow for predictions of future extreme values and estimates of return levels (n year high value) amongst other attributes. We consider the maximum values in this report, although the results hold for minimum values in a similar fashion.

There are two main ways of defining the extreme values in a distribution - the maximum values within specific non-overlapping time periods, such as a yearly or monthly maximum, or any values that lie above a suitably high "extreme threshold". These two approaches are referred to as the block maxima and the peaks over threshold approach respectively. Each approach uses a different distribution to model the extreme values, and are both suitable for different applications depending on the nature of the data set being used. Some of the main considerations to make when choosing which approach is most appropriate are the amount of data available, whether the data is time dependent and has any seasonality, and whether the analysis focuses on absolute or relative extreme values. For absolute extremes and cyclical data, block maxima are typically taken from a block that that contains one full cycle of any trends in the data to ensure that extreme values are extreme relative to the full range of data. For example, for Ice Surface Temperature data we would take yearly maxima because smaller blocks would result in extreme values from a subset of the strong cyclical trend from the data, which would thereby give us values that are extreme relative to the a subset of the data set rather than the full data set. Since there are 20 years of data available, this would result in only 20 extreme values being identified from the entire data set, which is a huge reduction in the amount of data available to use. On the other hand, a peaks over threshold approach gives us more data to work with as we can define a threshold at the 90% quantile level to give us 10% of the data rather than < 1%. Using a set threshold also works well for absolute extreme values as long we define the threshold based on the entire data set. Therefore, we use the peaks over threshold approach in our analysis.

1.2.1 Peaks over threshold

The peaks over threshold (Davison and Smith, 1990b) approach for extreme value analysis defines extreme values based on whether they lie above or below some threshold. The threshold can be a constant value for the entire data set, or can vary

with time in the case of a step function, a cyclical function or some other smoothly varying function. The choice of the type of threshold is largely based on the nature of the data set being analysed and whether there are any consistent patterns within the data set such as yearly cycles. One of the most commonly used thresholds is the 90% quantile of the data, which is suitably high for exceedances to be considered extreme while also leaving enough data to be analysed. Once defined, any values that exceed the threshold are extracted and fitted with a tail distribution such as the Generalised Pareto Distribution (GPD). The tail distribution models how much the values exceed the threshold by, and so is typically used in conjunction with a second distribution for the main non-extreme body of the data if the entire data set is to be analysed. In this report, we only consider distributions to model threshold exceedances and do not model values lower than the threshold.

One of the most important considerations for this approach is an appropriate threshold definition. If the threshold is misspecified, then there may be insufficient data to properly fit a tail distribution to, time trends may not be considered, or the threshold may define non-extreme values as extreme and worsen the appropriateness of the tail distribution. Although a 90% quantile is often the easiest threshold to consider, in the case of seasonally varying data - such as IST measurements - other thresholds may be more accurate. As a result, for data with a strong trend a choice must be made about whether extremes are defined relative to the trend of the data set or defined in absolute terms relative to the distribution of the entire data set. If interest lies in values that are extreme relative to the time that they occur, then the highest 10% of values may simply occur in the hottest month of the year and relatively extreme values in the colder months will not be defined as exceedances. While this is certainly a consideration for future analysis, our initial investigations use a fixed 90% quantile for the entire data set, as proximity to the boundary at 0°C is currently more of a focus than relative extreme values in the context of IST data.

This approach works well for a single time series and can also be extended

to spatial data for time series data at multiple locations, as is often the case in environmental applications. The simplest method for spatial data is to assume independence between each location and fit a GPD to each location. The parameters for each model can be compared across the spatial region, and the fitted models can be used to estimate probabilities of exceeding certain values at each point. Modelling the data like this is cheaper computationally than a full spatial model as the models themselves are simpler than full spatial models, and due to the independence assumption the model for each point can be fit in parallel. Although this doesn't give any indication of spatial dependence as the timings of the extreme values aren't considered in the models, it gives a good overview of the extent of extreme observations, particularly when comparing all points and models to common threshold values.

1.2.2 Conditional extreme value theory

The next step in the spatial modelling of extremes is to relax the assumption of independence at different points and to consider when the extreme values occur. Often in environmental applications such as temperature data, co-occurrence of extreme values between sites is of interest, and the assumption of independence can become less realistic due to a small distance between observation sites. For example, when analysing satellite IST data, the distance between observation sites can be as low as 1km, making independence between sites highly unrealistic especially in regions with consistent terrain. When the assumption of independence is no longer appropriate, modelling the extremal dependence between sites becomes an additional consideration of the modelling approach.

Conditional extreme value theory (Heffernan and Tawn, 2004) aims to account for extremal dependence between sites by modelling the behaviour of one or more sites on the condition that a particular site has experienced an extreme event. In effect, the aim is to examine the behaviour of some random vector X given that a single component X_i is an extreme value. Although we still define events as being

19

extreme or not using thresholds, values that are not considered extreme are now included when fitting the model if they occur at the same time as an extreme value at a different site. By looking at extreme values from this perspective, we can infer details of how closely linked extremes are at different sites, and as a result whether or not independence would be a reasonable assumption to make.

The main assumption in conditional extremes is similar to that of univariate extremes, and we can define the asymptotic distribution in a similar manner. First, using the probability integral transformation we standardise the distributions to have exponential-tailed margins, so that the extreme values at any individual site have a standard exponential distribution i.e $P(X(s_i) > x) = \exp(-x)$ as $x \to \infty$. More generally, where we want to consider both positive and negative dependence, we also assume $P(X(s_i) < x) = \exp(-|x|)$ as $x \to -\infty$.

For points s in some domain $\mathcal{S} \in \mathbb{R}^2$, let $X(s)$ be a stationary stochastic process with margins defined as above. We assume that functions $a_{s-s_0}(s) \in \mathbb{R}$ with $a_0(x) = x$ and $b_{s-s_0}(s) > 0$ exist such that for any site $s_0 \in \mathcal{S}$, any $d \in \mathbb{N}$, and any sites $s_1, \ldots, s_d \in \mathcal{S}$,

$$\left(\left\{ \frac{X(s_i) - a_{s_i-s_0}(X(s_0))}{b_{s_i-s_0}(X(s_0))} \right\}_{i=1,\ldots,d}, X(s_0) - t \right) \bigg| X(s_0) > t \xrightarrow{d} \left(\left\{ Z^0(s_i) \right\}_{i=1,\ldots,d}, E \right) \tag{1}$$

as $t \to \infty$, where $E \sim \mathrm{Exp}(1)$ is independent of the process $Z^0(s)$ and $Z^0(s_0) = 0$ almost surely.

1.2.3 Asymptotic dependence

One of the motivations for conditional EVA is to examine the degree to which extreme values at different sites depend on each other, and as such we require a way of defining the type and strength of dependence between sites. The 2 main notions of dependence that we will refer to are asymptotic independence and asymptotic dependence. In the bivariate case, we consider a bivariate random variable (Y_1, Y_2) at spatial locations s_1 and s_2 with associated distribution functions F_1 and F_2. We

examine the probability that both components of our bivariate random variable exceed some quantile q - that is, $P(F_1(Y_1(s_1)) > q, F_2(Y_2(s_2)) > q)$. In particular, we are interested in the behaviour of this probability asymptotically as $q \to 1$, which is the limit:

$$\chi(s_1, s_2) = \lim_{q \to 1} P\left(F_1\left(Y\left(s_1\right)\right) > q, F_2\left(Y\left(s_2\right)\right) > q\right) / (1 - q).$$

If this limit exists and is positive, then we define the random variables as asymptotically dependent, and if the limit is 0 then they are asymptotically independent. That is, if the joint probability of an extreme value from both components of the bivariate random variable is non-trivial as the quantile tends to 1, then the extreme values from the two components depend on each other to some degree.

2 Marginal modelling of extreme values (Ice Surface Temperature)

2.1 Introduction

We begin our analysis of extreme temperatures on the Greenland ice sheet by examining the behaviour of Ice Surface Temperature at individual cells of satellite data. By characterising the behaviour of ISTs at a single location, we can get a baseline understanding of the distribution of ISTs and thereby the properties of the extreme temperatures. This presents a thorough and informative exploration of IST data that will inform future modelling decisions and give us a marginal model that can be used to describe temperatures at a single cell, which can then be used for inference alone or in combination with other model structures.

We propose a novel statistical approach applied to Moderate Resolution Imaging Spectroradiometer (MODIS) IST data to model the distribution of temperatures on the Greenland ice sheet at 1139 MODIS cells, with a particular interest in identifying and modelling melt temperatures. The approach is based on three key characteristics of IST data: firstly, the presence of physical bounds on the range of ice and ice-melt temperatures; secondly, the multi-modality of the distribution; and thirdly, ambiguity about whether measurements close to 0°C represent melting of the ice sheet surface. This model-based approach has several advantages over a purely empirical analysis, including allowing full characterisation of the distribution of IST and resulting properties e.g. melt threshold exceedance probabilities, quantiles, return periods and return levels, as well as allowing for out-of-sample prediction and extrapolation. Since the sample of cells used to fit the statistical model is uniformly distributed over the full ice sheet, our model is sufficiently generalisable as to be useful for cells not explicitly used to generate the model, regardless of elevation, distance from the coast, or geographical location. Finally, note that we limit our analysis to the modelling of cloud-free days. This is due to the absence

of data on days with cloud cover and the bias that would ensue if we were to assume that temperatures on clear days could be used to represent these missing values. The data can not be considered as missing at random, so there would be a bias in temperatures on cloudy days compared to clear days. We use this model to investigate time trends in the observation period and to quantify both the frequency and magnitude of temperature events that are likely to result in ice melt.

2.2 Data

The ideal data source for both ISTs and air temperatures would be to use in-situ measurements. These can be obtained from AWSs, but these have sparse spatial coverage, much missing data, and are not representative of anything near the whole ice sheet. To get representative data of the entire ice sheet, we can either use GCMs or remote sensing in the form of satellite data. Both model output and satellite data have high spatio-temporal coverage, but model output is more processed since it relies on many hidden layers of modelling assumptions. Therefore, we prefer to use the satellite data which is closer to the real in-situ measurements.

For the in-situ measurements, air temperature and IST are closely linked in their overall properties but have some key differences in the way they are measured, their availability, and their relationship with melt. Despite the lower spatio-temporal coverage of weather stations that capture air temperature, it is important to distinguish the differences in the measurement types to gain a baseline understanding of the satellite data. As a result, we will examine both direct measurement types in more detail to assess the differences between them and learn some of the key properties of remote sensing data from satellites compared to in-situ measurements.

2.2.1 Ice Surface Temperature and air temperatures

In-situ air temperature measurements are a record of the temperature of the air at a specified height above the ground, often 2m. In Greenland, this is recorded using sensors stationed on the ice sheet that record the temperature at regular time

intervals at hourly intervals. However, air temperature cannot be measured directly by remote sensing. Although there are models that can estimate air temperature using other remotely sensed variables using models (Cristóbal et al., 2008; Hooker et al., 2018), we prefer to avoid the additional hidden uncertainty that is generated by using such processed data sets. This is one of the main reasons we use in-situ air temperature measurements for our comparisons rather than modelled air temperature data.

In contrast to air temperatures, IST measurements record the temperature directly at the surface. For ISTs in Greenland, direct measurements can be made on the ice sheet similarly to air temperature measurements, with the temperature sensor being in direct contact with the ice rather than 2m above it. Alternatively, IST can be recorded remotely using thermal infrared or passive microwave radiation. Sensors that can detect these wavelengths of electromagnetic radiation can be placed on-board satellites for reasonably accurate IST measurements at high spatial and temporal resolution.

A unique feature of IST data compared to Land Surface Temperatures (LSTs) not on ice sheets is a theoretical upper limit at 0°C. For non-ice terrain, LSTs have similar trends to air temperatures since they both have cyclical annual trends and are both influenced by the same external factors such as climate and weather. However, on ice this linked behaviour begins to break down around 0°C. Since the land surface takes the form of ice, the IST is bounded above by 0°C because above this temperature the ice is no longer ice and consequently no longer the surface. Therefore, for temperatures exceeding 0°C, air temperatures will continue to rise whereas ISTs will remain at 0°C. In practice, this limit is sometimes not strict as other factors such as surface water, measurement error and water vapour from clouds can result in temperatures above 0°C, but it has a large impact on the overall distribution of IST data sets.

Some of the most important factors by which to compare in-situ air temperatures and remotely sensed ISTs are spatial coverage, temporal coverage, time-series

properties and distributional properties. The first three factors will give an indication of the scale of analysis that can be performed of extreme temperatures, while the last two will provide information that may need to be considered when defining a model to apply to the data. The main sources of data that will be compared for the two variables are AWS data for air temperatures and satellite data for IST.

For our satellite data, we use MODIS IST data from MODIS/Terra Sea Ice Extent 5-Min L2 Swath 1km, Version 6 (MOD29) contained within a multilayer Greenland MODIS-based product (Hall et al., 2018). MODIS records surface reflectance from 36 spectral bands of different wavelengths - including those used in IST - near daily for the entire Earth. The MODIS instrument is located in two separate satellites - Terra and Aqua - that produce data sets with the same specifications, so we use Terra for its earlier launch date and thereby start time for its data sets. This data set spans the period 01/03/2000 to 31/12/2019 and has a spatial resolution of .78 km × .78 km. Here we discard the first 10 months of the data set, up to 01/01/2001, in order to work only with those years for which a full annual cycle is available. To reduce the computational burden in analysing/modelling the data, we subsample the data taking 1 in every 50 cells in both x and y dimensions for a total of 1139 cells. The resulting grid consists of 1139 cells, equally spaced in latitude and longitude and covering the full range of glaciological and climatological settings across the ice sheet. Depending on the analysis and model used, this sample will be changed to match the computational cost of the model/analysis.

For in-situ air temperature data, we use AWS data sets from Programme for Monitoring of the Greenland Ice Sheet (PROMICE) (Fausto et al., 2021), which make up around half (25) of the available AWSs. The stations record a range of meteorological conditions on the ice sheet, the most relevant of which are air temperature and surface temperature. Each station has a different range of data available, from around 20 years to only a couple of years.

For remote areas such as Greenland, recording air temperature accurately
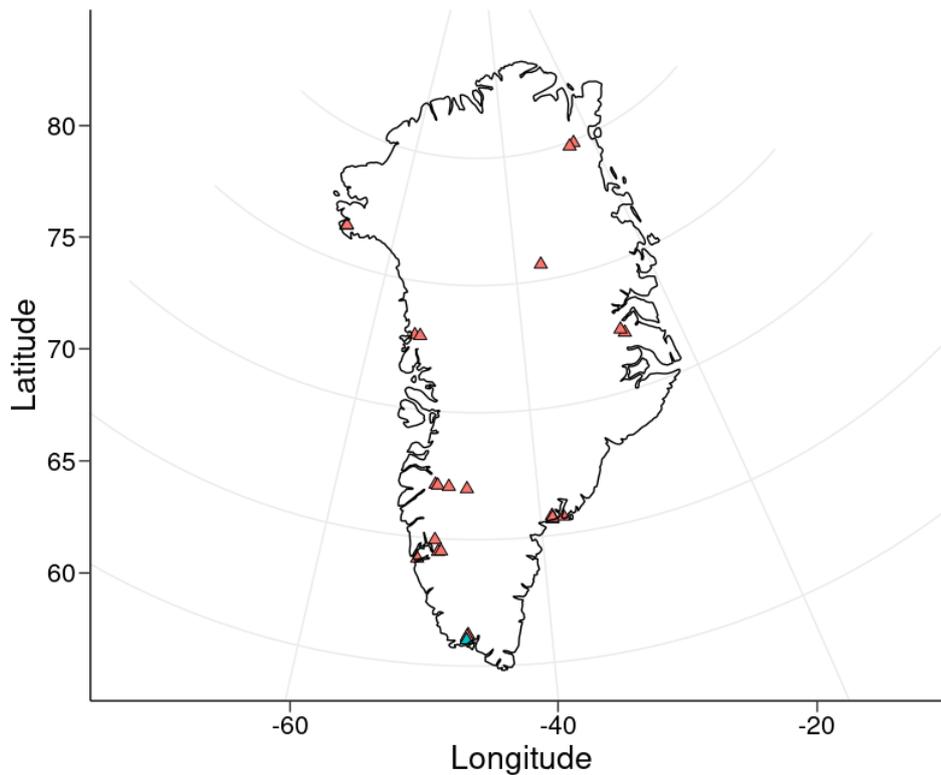
25

Figure 1: Locations of automatic weather stations from the PROMICE in Greenland that measure air temperature daily.

requires individual sensors on the ice sheet that are both expensive and time-consuming to setup. The ice sheet has a surface area of $\sim 2,200,000$ km and a maximum elevation of 3.3 km. Therefore, it would require millions of sensors to get good spatial coverage which would require high resolution sampling due to the variability in elevation, which is not at all feasible. As a result, spatial coverage of air temperatures in Greenland is limited to less than 50 sites. Figure 1 shows the locations of the AWSs. As can clearly be seen, most of these are situated close to the coast as it is more difficult to setup sensors further inland, and they are often placed in series of locations close to each other, further reducing the overall spatial coverage. By comparison to the air temperature data sets, IST data have far more comprehensive spatial coverage than air temperature data due to the use of satellite instruments. Satellite instruments determine IST by measuring infrared
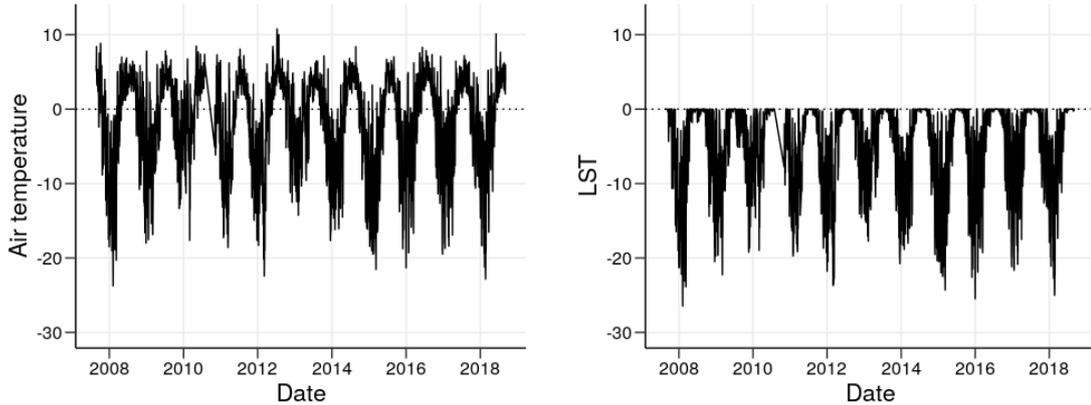
Figure 2: Time series of air temperatures (left) and IST (right) recorded at PROMICE site QASL.

radiation in the range of $10.5 - 12.5$µm which can be used to calculate IST, allowing for near-global coverage of IST data at resolutions of 1km or higher depending on the temporal resolution. This results in data sets with several million observations per day evenly distributed across the entire ice sheet, which is several orders of magnitude better than the spatial coverage of AWSs.

The difference in characteristics of air and ISTs is clear from Figure 2, which shows time-series of both sets of measurements for the PROMICE AWS site QASL. Although the overall seasonal and yearly trends in the data are fairly consistent, the cut-off point of $0°$C in the IST data blocks the full temperature seasonal cycle that can be seen in the air temperature data. It also acts as an indicator that the state of the ice surface has changed from ice to water. As a result, is it difficult to distinguish between the hottest temperatures in the IST data where temperatures reach this boundary. IST data typically reaches the upper limit for areas on the coast and further south, so although the issue is not persistent across the entire ice sheet, it is more prevalent in the areas where the AWSs are located which is where the best direct comparisons between the two variables can be made.

For both measurement types, there is missing data in late 2010 that causes the time-series to jump quite severely, which may be due to an instrument failure during that time as mentioned in the previous section. Because of the remote location of the
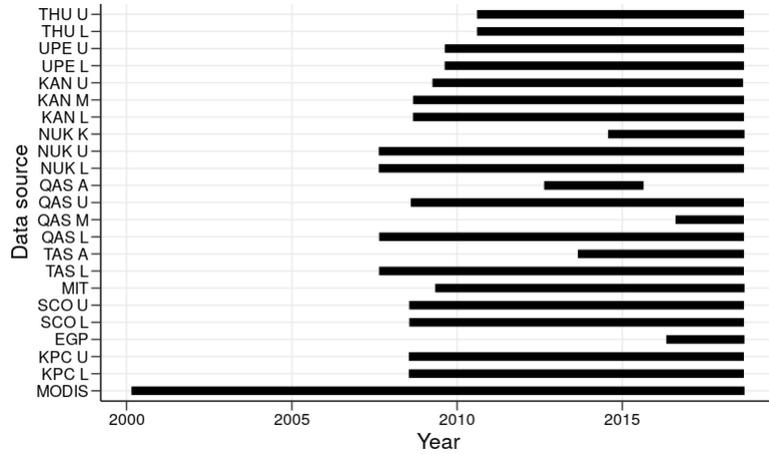
27

Figure 3: Range of data available for each of the PROMICE AWSs and for the MODIS data sets

instruments, it can take a significant amount of time to repair them if they develop faults or are damaged by severe weather. This can cause reasonably large amounts of time to have no recorded values depending on which/how many instruments develop problems. One-off missing measurements are not problematic unless there is some consistency in when or why they are missing, but large chunks of missingness can cause more issues. However, this problem can largely be avoided by using satellite remote sensing instruments that are far less prone to prolonged outages.

2.2.2 Temporal coverage

Figure 3 shows the start and end times of the data record for each of the 22 PROMICE stations for which data is available and of the MODIS Terra data sets. MODIS has nearly double the data available than some of the PROMICE stations. Further, whilst most of the AWSs have been in operation for similar time spans, there are some stations for which only a couple of years of data is available. Furthermore, some of the time ranges shown are not complete since, as mentioned previously, there can be instrument outages that can't be immediately repaired due to the stations' remote locations. This is not the case for the MODIS instrument, which operates uninterrupted and without outages for the entire duration of its operation.

28

The IST measurements represent the temperature at the surface of the ice in cloud-free conditions. Clouds (specifically water vapour) can interfere with the measurements, so a cloud mask is used in the MODIS product to remove measurements made in cloudy conditions. As a consequence, our analysis and predictions are valid for clear conditions only. Due to the generally warmer temperatures seen on cloudy days, were the analysis to be interpreted as representative of clear days also, there would be a strong likelihood of over over-estimating the magnitude and frequency of melt events (Koenig and Hall, 2010).

2.2.3   Distributional properties

Figure 4 shows the empirical distribution of both air and ISTs. It is clear from these plots that the distribution of air temperatures is uni-modeal, with no clear-cut upper bound. While air has no state changes in the typical air temperature range and therefore recorded values never reach the theoretical minimum or maximum values, the change from ice to water at 0°C results in an upper limit at 0°C for the IST data. This causes a large amount of the total density of the IST measurements to form a point mass at 0°C. This heavily skews the distribution of IST and presents a challenge for modelling the data. The air temperature data is also somewhat skewed above 0°C, although to a much lesser extent.

In the context of measuring ice melt and extreme melt events, IST data will be more appropriate for our work than air temperature data. The greater spatial resolution of the satellite data compared to the AWS in-situ measurements can give us a clearer picture of melt across the entire sheet, as opposed to only at a small number of locations in very specific parts of the ice sheet that have an AWS. IST data is also more closely linked to melt as despite the links between the two measurements, ISTs are a measurement of the exact quantity (the ice temperature) that we are interested in. Provided that our interest lies in the occurrence of melt with less focus the extent of the ice temperatures above 0°C, the upper limit of ISTs can be considered appropriately in our analysis to still allow us to accurately
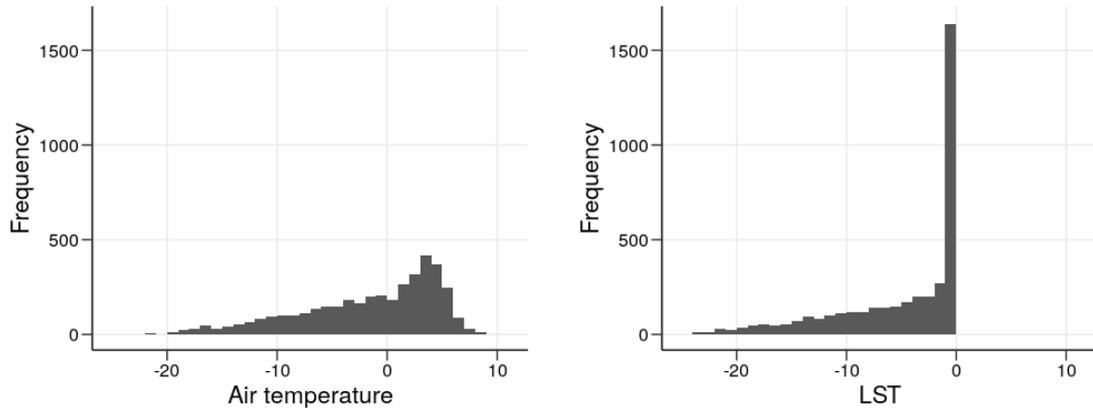
Figure 4: Histograms of air temperatures (left) and IST (right) recorded at PROMICE site QASL.

analyse melt events.

2.2.4 Cloud cover

Due to the variability in climate, areas in the north and coastal areas tend to have more cloudy days and hence a greater proportion of missing data (Figure 5. We also see that winter months have more missing data on average than summer months because of cloud cover, with a range of 65.1% of data available in December compared to 91.1% of data available in May. This is important to bear in mind when interpreting the predictions made from the statistical models, as the IST distributions will be more heavily weighted towards warmer temperatures. This shouldn't affect our inference with regards to melt, however, as melt temperatures almost exclusively occur in the summer months which have a much lower proportion of missing data.
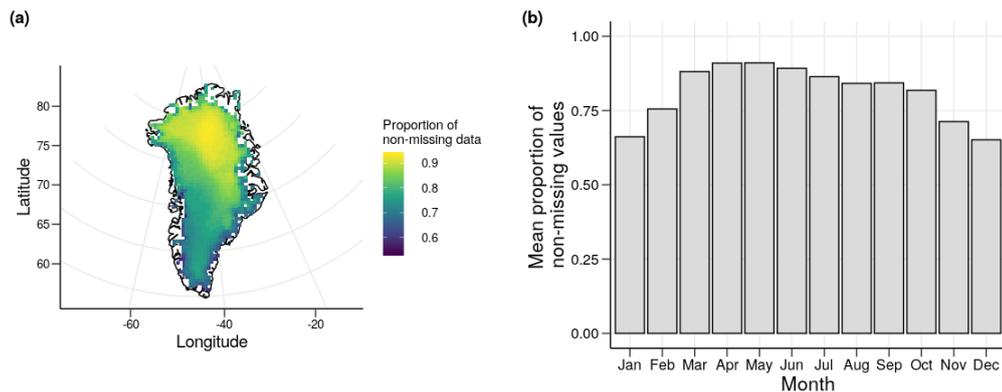

Figure 5: Proportion of available MODIS IST data (i.e not filtered by the cloud mask) at 1139 cells on the Greenland ice sheet between 2001 and 2019 (a) and mean proportion of available MODIS IST data by month between 2001 to 2019 (b).

To check for the possible presence of longer term changes in surface conditions in response to changes in cloud cover, the proportion of data missing due to cloudiness is compared between decades (Figure 6). The clear spatial trends seen in the overall proportions of missing data are not present here, though locations in the north and south of the ice sheet have slightly more data in the most recent decade in contrast to locations at mid-latitudes where the reverse is true. However, these trends are not consistent over all locations, and the magnitudes of the changes are relatively minor
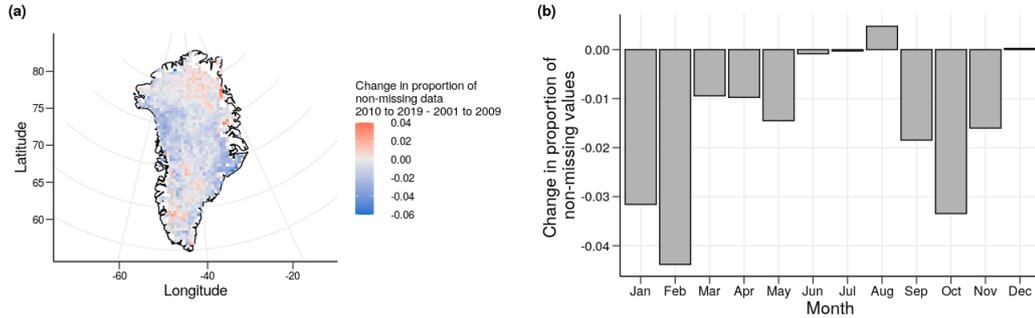
Figure 6: Change in proportion of available MODIS IST data (i.e not filtered by the cloud mask) at 1139 cells on the Greenland ice sheet from 2010 to 2019 - 2001 to 2009 (a) and change in mean proportion of available MODIS IST data by month from 2010 to 2019 - 2001 to 2009 (b).

with a maximum absolute change of 0.060 and a mean absolute change of 0.010. Seasonal differences in cloud cover and thereby data availability also vary between the two decades. January, February, and October show the largest changes, with a maximum absolute change of 0.044 in February. Ten months have an average decrease in data availability between the decades but the absolute differences are reasonably minor. Furthermore, of the four months with the lowest changes, three (June, July, and August) also see the highest average temperatures, giving further confidence that the cross-decade comparison shows no obvious inter-decadal trends and so the data can be analysed as a single sample without adjusting for long-term trends.

2.3 Methods

To create a statistical model that is parsimonious and applicable at all cells over a large and geographically-varied region, we model the IST data using statistical methods that allow us to treat melting in a probabilistic manner. Exploratory data analysis shows that there is no clear quantile in the temperature distribution that can be attributed as the onset of melting (Figure 7). As a result, we model melting ice temperatures and non-melting ice temperatures separately and estimate the probability of melt occurring over a range of temperatures. This approach allows

for some uncertainty in the observations from factors such as the precision of the data set, which has a stated uncertainty of $\pm 1°C$. We hereby refer to temperatures associated with melting ice as "melt" temperatures and temperatures associated with non-melting ice as "ice" temperatures.
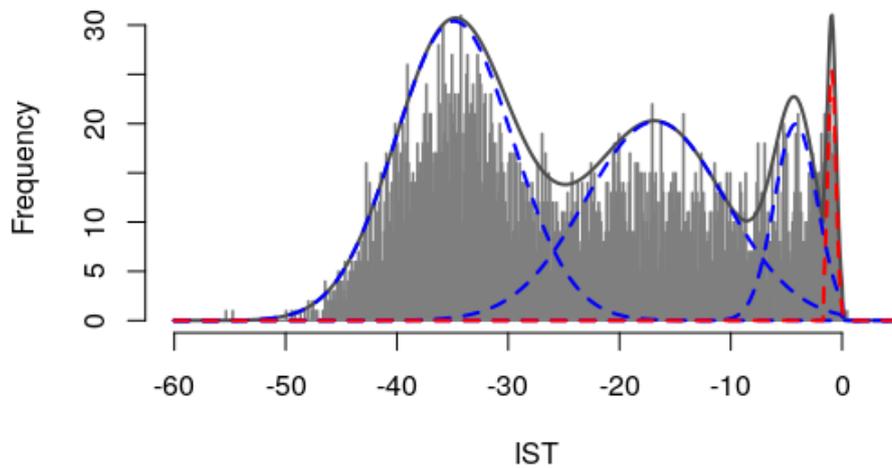

Figure 7: Frequency distribution of daily MODIS IST data from an example cell (82.47, -37.50) on the Greenland ice sheet between 2001 and 2019. Solid lines show a mixture model fit to these data where blue indicates the three 'ice' components, red indicates the 'melt' component, and black indicates the full model as the sum of the ice and melt components. Each individual component is a truncated Gaussian distribution, and the lines represent the probability density function of these on a scale to matching that of the histogram.

A key feature of the data set and a core modelling consideration is the soft upper limit at $0°C$. The melting point of the ice acts as a physical upper limit on ISTs, as once the ice exceeds this temperature it melts and may no longer form the surface of the ice sheet. Some sites have measurements above this limit, which arise due to meltwater sitting on top of the ice. However, the ice under the water places a limit on these melt temperatures, hence the distribution of positive temperatures is truncated close to $0°C$. This soft upper limit of ISTs causes a significant peak in the distribution centred at approximately $-0.5°C$, as any ISTs that would exceed $0°C$

33

are truncated to small positive values close to 0°C.

The simplest statistical model would be to fit a single distribution to the full data set, potentially after an initial transformation. This raises two issues. Firstly, a bimodal distribution is clear at all cells with cells that experience melt having the highest mode close to 0°C (Figure 7). For non-melt cells, the location of the higher mode is more variable. This does not appear to be directly attributable to seasonal differences in temperatures, as the shape of seasonal temperature distributions show as much inter-site variability as they do inter-season variability. Fits of unimodal distributions are particularly poor at the tails of the distributions, which is particularly problematic since our interest lies in melt which is directly connected to the upper tail of the temperature distribution.

Given the focus on melt, an alternative option would be to undertake an extreme value analysis of only the highest temperatures at each cell. This would allow the model to focus on the temperatures of highest interest that are the most difficult for more standard models to capture. This also proves problematic though, as in order to fit the model the temperatures must first be identified as extreme using an extreme value threshold, with temperatures above the threshold being classed as extreme and those below being classed as non-extreme. Due to the mode around 0°C, finding a consistent threshold location using quantiles, gradient analysis or a specific temperature all encounter problems due to the large variety of tail shapes at different cells. Each of the above mentioned threshold types do not work universally across the ice sheet, and in many cases provide much worse fits than a distribution applied to the whole range of temperatures.

A consistent model that can be automatically applied at cells across the ice sheet therefore requires a multi-modal distribution or a time-series model to capture seasonal behaviour. The disadvantage of the latter is that it is less able to capture the mode around 0°C and the truncation of the temperatures which is where our research interest lies. This further motivates a modelling approach that more directly considers the distribution of the specific data set and allows for multi-

modal distributions. The overall distribution shape is broadly similar between sites with the main difference being the proximity of the distribution to 0°C and thereby the amount of truncation in the data. The considerations above around the multimodality of the data set and of the nature of melt temperatures around 0°C give us a basic set of assumptions to base our modelling around that allow the model to retain the same underlying structure regardless of the absolute difference in ISTs between cells.

In order to accommodate spatial variability in the temperature distribution, we model IST using a truncated Gaussian mixture model in which components are assigned to model groups of temperatures that we assign to be either ice or melt. For n_I ice components and a single melt component, let ϕ_i be the weight associated with model component i such that for $n_c = n_I + 1$ total components, $\sum_{i=1}^{n_c} \phi_i = 1$. For each ice component i (and melt component M) let $f_i(x)$ be the probability density function of the truncated normal distribution $X \sim TN(\mu_i, \sigma_i^2, a_i, b_i)$, such that

$$f(x) = \frac{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)}{\Phi\left(\frac{b-\mu}{\sigma^2}\right) - \Phi\left(\frac{a-\mu}{\sigma^2}\right)}$$

where μ_i is the mean, σ_i is the standard deviation, a_i (b_i) is the lower (upper) truncation point, and Φ is the cumulative distribution function of a standard normal distribution. Then the probability density function of ISTs x is:

$$p(x) = \sum_{i=1}^{n_I} \phi_i f_i(x) + \phi_M f_M(x).$$

We set the upper and lower truncation points for the ice and melt components at values that bound each measurement type with relative certainty. For the ice components, $a = -\infty$ as there is no hard lower limit on the temperature of ice (aside from absolute zero), and $b = 0$ as, theoretically, ice temperature can't exceed 0°C. This means that there is no limit on how low ice temperatures can go, but they can't exceed 0°C. For the melt component, $b = \infty$ and $a = -1.65$, so that

temperatures in the melt component can't go below $-1.65°C$ but are not upper truncated. We take a bound lower than zero here to account for uncertainty in the data and any potential impurities in the ice surface. $-1.65°C$ is the theoretical minimum temperature at which saline ice can melt (Hall et al., 2004), and thus should be a conservative estimate for this lower bound. Temperatures between $-1.65°C$ and $0°C$ can be modelled by either/both the ice and melt components as there is uncertainty as to whether they are associated with melting or non-melting ice.

A mixture model was fitted using the EM algorithm for each sample cell. The algorithm alternates between two main steps: calculating the component probabilities that each observation x_i comes from model component k, and maximising the expectations of the model parameters using the component probabilities (for full details see Appendix 1). We used this method to obtain estimates of μ, σ, and ϕ for each model component at each cell.

We used Bayesian Information Criterion (BIC) (Keribin, 2000) to assess the most appropriate number of ice and melt components and found that three ice components and one melt component fit the data best. These components may be broadly interpreted as winter, autumn, spring, and the melt season for the three ice components and single melt component respectively.

When modelled with separate Gaussian components, the characteristics of the different modes of the data are much clearer (Figure 7). The melt component at each cell generally has a much lower variance than the ice components due to the soft upper limit of ISTs and the lower truncation point of the model, whereas the ice components have higher variances and more overlap between components. For the sites that experience melt regularly, a substantial proportion of the overall temperature distribution occurs in the overlap between true ice and true melt. A similar result is seen across sites located on or near the coasts, which further validates the decision to use a fixed melt threshold as the melt temperatures - and thereby the melt process - appear to have consistent characteristics across cells.

Using this model, the probability of melt occurring, which we denote by $\rho(x)$, can be quantified as the ratio of the densities of the ice and melt components. For a given IST x, n_I ice components melt component M, we have:

$$\rho(x) = \frac{\phi_M f_M(x)}{\phi_M f_M(x) + \sum_{i=1}^{n_I} \phi_i f_i(x)}.$$

Consequences of this definition are that for ISTs below $-1.65°$C, the probability of melt is 0, for ISTs above $0°$C the probability of melt is 1, and between these values the melt probability depends on relative values of the melt and ice components' densities. For cells with very few or no ISTs above $-1.65°$C, the weight of the melt component may be close to or equal to 0, in which case the probability of melt occurring is effectively zero. Note that there are discontinuities in the model-based estimate of this probability due to the censoring of the mixture components. These discontinuities occur at the edges of the range of interest ($-1.65°$C and $0°$C) and are more or less severe depending on the degree of truncation of the ice and melt components. There is also a lack of uncertainty measure from this definition of melt probability due to its simplistic nature. This does increase the risk of inaccuracy for individual probabilities, however the risks should be minimised if the probabilities are used for averages and summary statistics rather than in-depth analyses of individual cells and days.

2.4 Results

2.4.1 Melt extent comparison

Using our model, we calculate the expected number of melt days in each year at each sample cell. Let N_y be the number of melt days in year y, then $\mathrm{E}[N_y] = \sum_{i=1}^{m} \rho(x_i)$ where $\rho(x)$ is the notation introduced earlier to denote $\Pr[\text{melt}|X = x]$ and m is the number of observations in year y. The overall annual average is simply the average of the individual annual averages. We then compare our modelled estimates to a simple threshold-based approach to defining melt, i.e. the average number of days
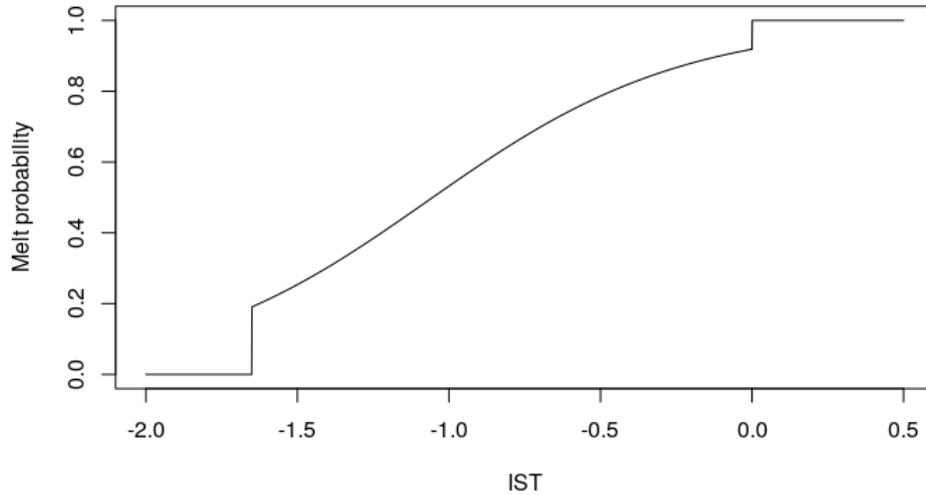
Figure 8: Melt probability estimates of a range of ISTs using the fitted mixture model at a single cell (75.37,-58.13) on the Greenland ice sheet between 2001 and 2019. Because some cells have very limited data above −1.65°C, we use a cell on the west coast with a high proportion of data above −1.65°C (22.55%), thus giving us an increased amount of information in the most pertinent temperature range.

per year with temperatures greater than or equal to −1°C (Figure 9).
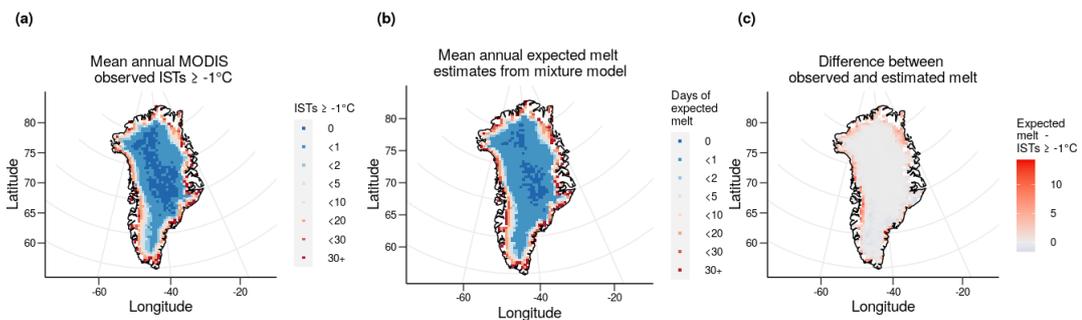

Figure 9: The mean annual number of ISTs above −1°C per year (a), the mean annual expected melt days estimated from each cell's mixture model (b), and the difference between the two variables (c).

The majority of the ice sheet - 90.7% of cells from the expected melt from the model, 79.5% from a threshold of −1°C on the data - experiences some degree of melt on average each year, except for sites in the dry snow zone in the centre and

38

north of the ice sheet (Benson, 1960). Of the cells that experience melt, most (62.2% from the model, 57.3% from the data) sites on average see less than 2 days of melt per year, which makes up the rest of the dry snow zone and most of the percolation zone. The areas with the most melt are located around the coast and in the south and west as may be expected. The main discrepancies between the two measures are at coastal cells, particularly on the west and north coasts. Here, the model estimates a larger amount of melt, with a maximum of 14 additional melt days at 1 specific cell on the edge of the south east coast compared to the data set. However, 89% of cells have an absolute difference of less than 2 melt days, showing the broad agreement between the measures at central cells.

To place the methodology and the results within the context of the literature, we compare the estimates of melt found using our method to those estimated using data from the PROMICE AWSs (Fausto et al., 2021). The characteristics of the data sets differ with known biases between the two (Koenig and Hall, 2010), however previous validation carried out independently for both data sets suggests it is reasonable to consider both data sets as representative of the true surface temperatures. With AWS as a baseline, Figure 10 compares the difference between the AWS melt proportions with the same quantity obtained from (a) MODIS IST and (b) the model fitted to MODIS IST. Despite the relatively large distances between some of the comparison locations, there is still reasonable agreement for a substantial number of AWS locations. The differences between the estimates may also be partially due to other factors surrounding the nature of the comparison, such as the reduced amount of data available because of the need for common dates, the difference in time resolutions between the data sets in terms of the number of observations averaged to give a single daily observation, and the different measurement uncertainties that lead to different definitions of melt.
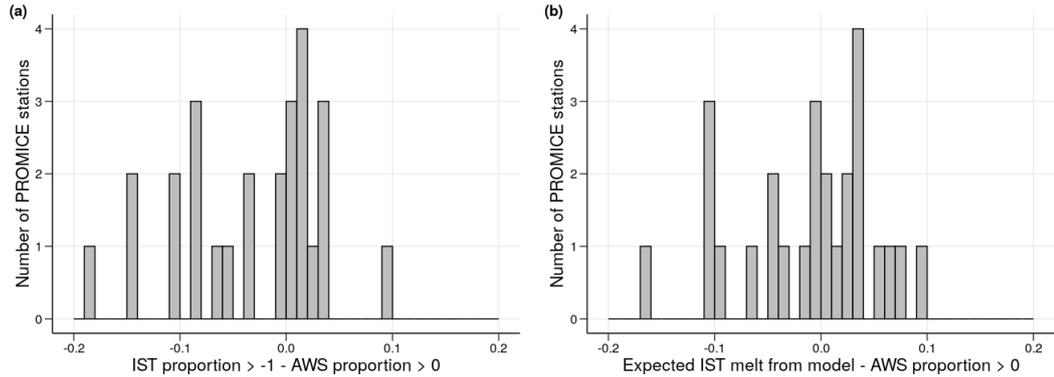
Figure 10: A comparison of the empirical melt from PROMICE AWS data (proportion of data $\geq$ 0° C) with empirical MODIS IST estimates (proportion of data $\geq$ 0° C) and mixture model based expected melt estimates. Only dates with data available from both data sets are used, meaning the above estimates are valid for cloud-free days only due to the limitations of the MODIS IST data set.

2.4.2 Temperature quantiles

We now use the model fit to calculate quantiles of the ISTs at each cell (Figure 5). This gives context to the overall temperature trends observed in the data set, before looking at melt in more detail. We calculate the 90% quantiles to examine the broader trends of high temperatures that aren't necessarily melt temperatures, as well as the 10% quantiles for temperatures that are as relatively low as the 90% quantiles are high. The estimated 10% and 90% quantiles broadly follow the same trends as elevation on the ice sheet. The 10% quantiles have a range from -53.84°C in the centre of the ice sheet to a maximum of -15.75°C at the south tip of the ice sheet. As would be expected, cells at higher elevations have a lower 10% and 90% quantile. However, of more interest are the few (30/1139) cells located on the west, east and southern coasts that have a 90% quantile above 0°C. At these cells, we would expect at least (in some cases more than) 10% of observed temperatures to be above 0°C and thereby melt temperatures.

We also calculate the 1-year return levels of each cell. This is the IST that is on average only exceeded once per year as estimated from each cell's mixture model. The return levels range from a minimum of -7.08°C in the centre to a maximum

of 7.24°C on the west coast. Although, as previously discussed, ISTs should not be seen higher than 0°C, these return levels reflect similar temperatures recorded by observations in the data set and can be plausible temperatures when considering the effect that meltwater on the surface of the ice sheet has on the observations. The rarity of melt in certain central areas can be seen more clearly, as temperatures in many cells (519/1139) on average reach −1.65°C less than once a year. The trends seen in the return levels also broadly agree with those seen in the quantiles, and are in reasonable agreement with the elevations and distance to the coast of each cell, with cells at lower altitudes and closer to the coast generally experiencing more melt.
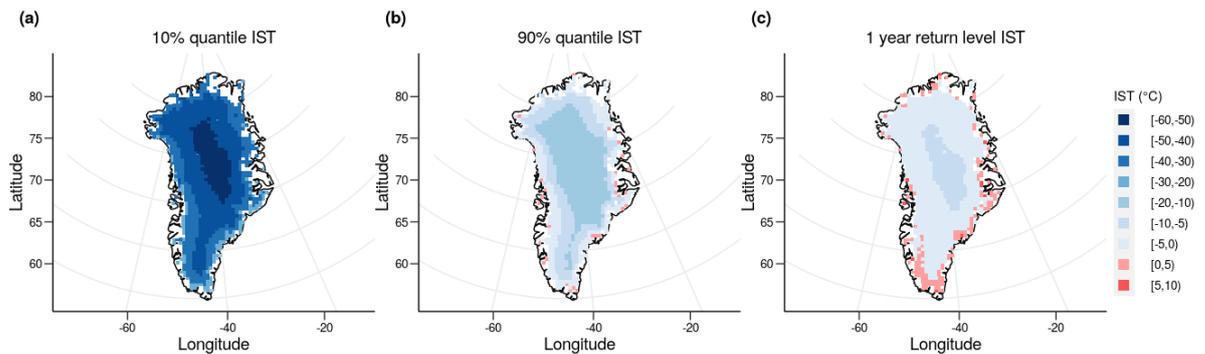

Figure 11: 10% quantile (a), 90% quantile (b), and 1 year return level (c) estimates for MODIS IST data from 2001 to 2019 at 1139 cells. Estimates are calculated from fitted mixture models at each cell.

2.4.3   Decadal fits

To examine potential changes in melt over time, we fit mixture models at each cell for data from two separate decades: 2001 to 2009 and 2010 to 2019. Analysing trends over a decade helps to smooth some of the annual variability and thus highlight any potential differences as a result of climatic change. To assess any changes in melt and high ISTs between the decades, we compare quantiles between the fitted models and the estimated melt probabilities at each cell in each decade.

Because some central areas of the ice sheet do not have many historical

41

melt observations, we examine the 95% quantiles and yearly expected maximum temperatures, both of which give an indication of overall trends in high temperatures even if these do not reach the level required for melt at some cells. We use the 95% quantile rather than the 90% quantile as used in the previous section as this quantile is more common in extreme value analysis and it focuses more closely on melt temperatures than the 90% quantile. As previously, we take the estimated quantiles from each of the fitted decadal models for each cell.

For almost all cells (1100/1139), the 95% quantile increased between the two decades. Cells in the south in particular have increased fairly consistently. The average change for all cells south of 73.41°N was 0.73°C, with 99.3% of all cells further south than this seeing an increase. The largest increases were also concentrated in the southern areas of the ice sheet, with a maximum increase of 1.78°C.

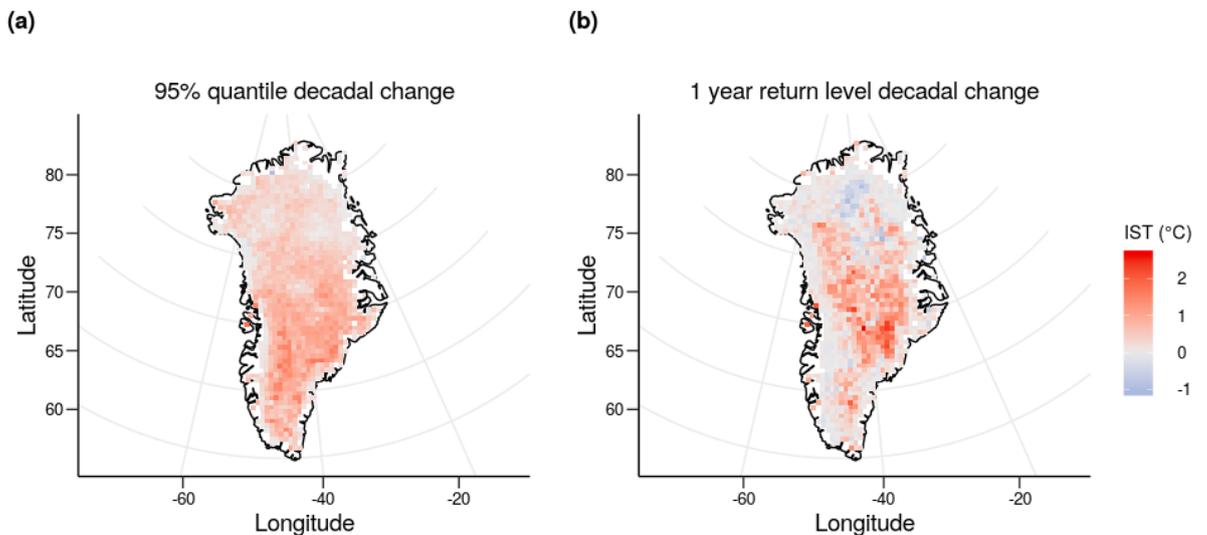Figure 12: Comparison of the change in 95% quantiles (a) and 1 year return levels (b) of mixture models fit to MODIS IST data from 2001 to 2009 and 2010 to 2019.

The 1-year return levels also generally increased, albeit slightly less consistently than the 95% quantiles (849/1139 cells). Areas in the east show the largest increases - with the largest increase being 2.66°C - however on the south west coast and

particularly the north central area of the ice sheet there are also several cells that show a slight decrease in contrast to larger increases. More clearly than in the 95% quantiles, 1-year return levels at coastal cells do not increase as much as in central cells between decades. However, it is important to note that the maxima at coastal cells are already close to or above 0°C. Because of the soft upper limit of the IST data, values already close to this limit can be partially constrained from further increases, so cells that had a 1-year return level above 0°C are less likely to show an increase than colder areas such as in the centre of the ice sheet. This makes the 1-year return level more informative for central cells than for coastal cells.

We next compare the probability that each cell experiences melt on any given day for each decade. Using the fitted models, we estimate the probability that each daily observation is a melt temperature, then take an average of all values within our defined decades. Although there is some variance between the individual years, our analysis remains focused on the broader trends between decades rather than on individual years as it is preferred to account for inter-annual variability than to examine it directly. For the purposes of interpretation, we limit our discussions to the summer months (May to September, inclusive) when considering melt probabilities, due to the almost zero probability of observing melt outside of this period.

The two decades show very similar trends in their daily melt probabilities, particularly around the coast. However, decade 2 has more cells with a non-zero probability of melt (1017) than decade 1 (853) - an increase of around 19.2% between the two decades - and 68.5% of cells saw an increased probability of melt between the decades. The cell with the single largest probability from either decade is from decade 2 (64.11,-49.93). This cell has a probability of a melt temperature on any given summer day of 0.64 - equivalent to an expected 97.92 melt days per year.

Most of central Greenland has experienced minimal change in the probability of melt between the two decades (Figure 13). This may be largely due to the probabilities being extremely small for these areas regardless of the time period chosen. Coastal cells show clearer and larger cross-decadal variation in melt

probability. South east and south west areas of the ice sheet were generally more likely to experience melt, in addition to some cells in the north east and north west areas that were less likely to experience melt, in the more recent decade. The largest increase is on the south east coast, where cells show a maximum change of 0.0351, which equates to an expected increase of 15.42 melt days each year.
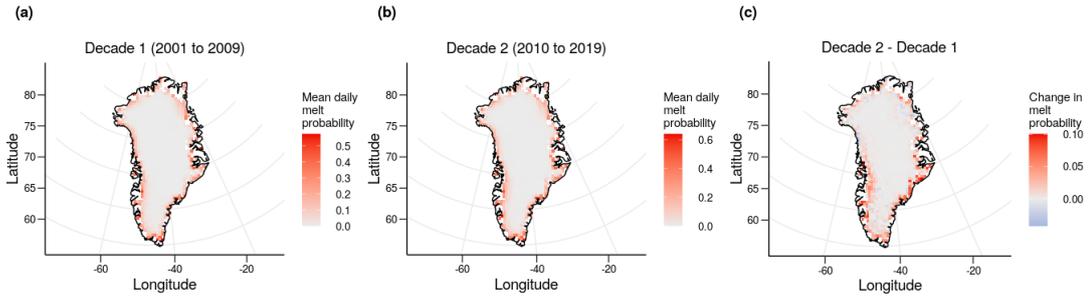

Figure 13: The average probability of a melt temperature on any given summer (May-September inclusive) day from 2001 to 2009 (a) and 2010 to 2019 (b), and the change in melt probability between 2001 to 2009 and 2010 to 2019 calculated from the fitted mixture models.

2.4.4 Variable ice components

We also test a fit of the model with a variable number of ice components. Although for consistency in the structure of future models we will use three as the optimal number of ice components, we can still infer more information about melt by allowing for different number of ice components to potentially give a more accurate fit. This approach is particularly aimed at examining the melt component in closer detail. By allowing more ice components to be considered, the main body of the distribution would have more flexibility in the shape - particularly close to 0° C - that could be represented. This could allow for more representative modelling around the crucial melt temperature range, which may also make the melt component more accurate.

To find the most appropriate number of ice components for the model at each cell, we fit the model with n_I from 3 to 6 and compare the fits using BIC. A single melt component is still kept for this approach, since melt temperatures should still

show much less cross-cell variability than the ice temperatures and maintaining a consistent number of melt components makes comparisons between cells easier.

The impact of the soft upper limit on ISTs can once again be seen from the number of ice components selected by the BIC to model the data at different cells. Figure 14 shows the number of ice components in the best fitting model to be higher at cells close to the coastline than those closer to the centre of the ice sheet. There are also trends within coastal areas, with cells on the south west coast generally having the highest number of model components and areas in the north west coast having fewer. Cells closer to the coast generally have higher average temperatures, experience more melt and therefore have more frequently truncated observations. The coastal temperature trend is consistent with local climate trends of higher temperatures occurring closer to the coast and demonstrates the impact of the upper limit from the truncated observations. This also demonstrates the difficulty mentioned previously of finding a single distribution to describe the data at all locations of the ice sheet.
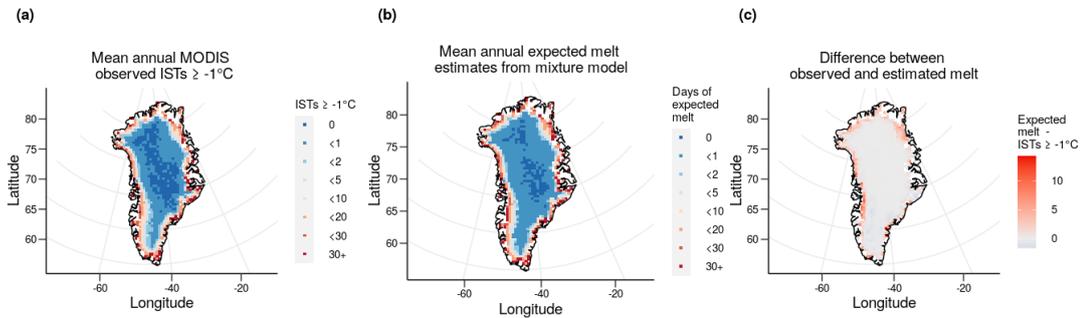

Figure 14: (a) Estimated weight $\hat{\phi}$ of the melt component of each fitted mixture model. Due to restrictions during the fitting process the minimum possible value of this estimate is 0.0005. (b) The number of ice components in the best fitting mixture model at each cell in our study area.

While there is value in examining how each parameter varies across the ice sheet, with the addition of extra ice components these comparisons are less appropriate for this set of model fits. To understand melt, it is most useful to consider variability and trends in the weight of the melt component. This gives an overview of melt

observed across the ice sheet, and reflects broad climate trends and the coastal effect. Figure 14 shows coastal areas having more of their distribution associated with melt temperatures, with cells in the south having a slightly higher weight than those in the north. Melt probability (2.3) is estimated for each observation and used to calculate the expected number of melt observations at each cell in the same manner as in Section 2.4.1. These values are still estimated with a high probability of being under-estimates since the expected melt at each cell would be higher at almost all cells if it were possible to also include cloudy days in the analysis. Figure 15 shows the majority (90.7%) of cells have a non-trivial expectation of melt (> 0.001) in an average year. Although melt is more common on the coasts, it is clear from the expected melt estimates that this is not exclusive with some inland cells also experiencing melt with relatively high frequency, and the main factors that appear to determine the melt extent are distance to the coast and elevation as seen previously.
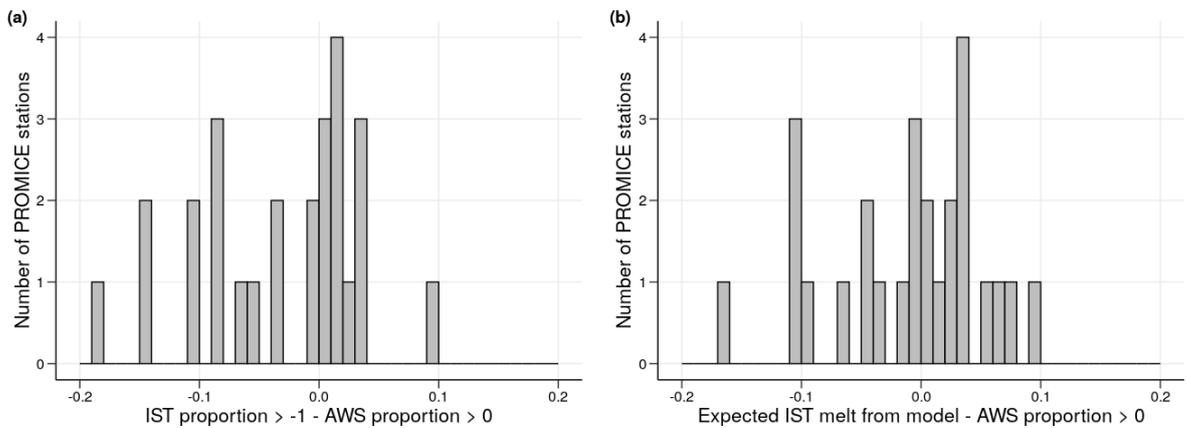

Figure 15: Expected melt calculated from the fitted mixture models. The probability of each observation representing a melt temperature is calculated then averaged to the expectation in a single year.

2.5   Discussion

Increases in ice melt in Greenland are of major concern due to the impact that it will have on sea levels (van den Broeke et al., 2016), however in-situ observations

of ice melt are sparse, spaced irregularly, and of coarse resolution. Here, we show that melting can be estimated using a relatively low-dimensional and highly flexible statistical model for IST. This enables us to assemble a record of melting that is continuous in time and space, and is sampled at high spatial (.78 km $\times$ .78 km) and temporal (daily) resolution (cloud-cover permitting) using the MODIS IST data set. In addition to the greater availability of IST data, ISTs are on a continuous scale and vary smoothly over time and space, making them better suited to statistical modelling. This is of particular interest given that, from these data, we see that there is ambiguity about whether or not temperatures below 0°C are in fact reflective of melting ice. In this paper, we have addressed this uncertainty by incorporating it into the structure of the statistical model, and thus our record of melting/not-melting is probabilistic rather than binary.

Our model gives comparable results to empirical estimates of melt obtained using a fixed threshold, while also allowing more detailed analyses of melt and the overall temperature distribution via quantile estimation, melt probabilities, and return levels. By modelling the entire temperature distribution, not only can we can gain insight into the frequency and range of melt temperatures, but also broader trends such as higher temperatures in both the high and low quantiles. Furthermore, the model allows for out-of-sample predictions and extrapolation beyond the range of observations. This is of particular interest for melt which occurs with temperatures in the upper tail of the IST distribution where there can in some cells be insufficient data to confidently make empirical estimates.

We observe that melt is much more likely at coastal cells and in the south of the ice sheet than in the centre, and that there is a non-trivial probability of melt occurring below −1°C. The spatial melt trends are in keeping with previous work examining melt using surface mass balance data (van den Broeke et al., 2016) and satellite data (Mernild et al., 2011), including MODIS data (Nghiem et al., 2012). The fitted models also show a clear link between elevation and high ISTs similarly to previous studies linking temperature to elevation (Reeh, 1991), and the

yearly expected maxima show the potential for even central areas of the ice sheet to experience melt (Nghiem et al., 2012). Trends previously observed in the south (Mote, 2007) also appeared to have continued, as all cells examined south of 75.16°N saw increases in high-temperature quantiles in the most recent decade.

One of the key considerations is the impact of cloud cover on temperatures, which will not be negligible. The data set used has a complete absence of data on cloudy days. This could be handled in three ways: analyse clear days only, impute missing values, or impute cloudy day data from a second data source. Cloudy day data are not missing at random, since the mechanism which causes the missingness is intrinsically related to the missing values themselves. Consequently the usual methods for imputation using the observed data are not valid. In particular, any such imputation of cloudy day values using the available clear day data would need to take into account the systematic differences between clear and cloudy day temperatures since, as noted, cloudy days are in general warmer than comparable clear days. Because there is a complete absence of cloudy day data, there is no way for the extent of this bias to be estimated empirically. Consequently, we would need to use external information, e.g. other sources of data, to undertake the imputation. This would open up additional problems around different levels of measurement and recording error, different spatio-temporal measurements scales, and so forth, which we believe is beyond the scope of the project.

Although some of the expected annual maxima are just below the lower censoring point of our model's melt component, melt may be possible in these areas over longer time periods. For some cells, the model fit suggests an extremely low probability of melt. This may be because these cells have few historical instances of possible melt in the data, i.e. no ISTs above the lower censoring point of the melt component. In these instances, the information in the data is insufficient to support a melt component, so only the ice components can be fit to the data, leading to an effective zero probability of melt.

The model also assumes that surface conditions remain similar over the observed

time period. Additional impurities becoming present in the ice or rocks appearing after a particularly warm summer could affect the distribution of temperatures at least in the short term and potentially in the long term, however these changes would be difficult to accurately identify using only ISTs data. A separate data set with additional information about surface conditions could be used to identify these changes, or adaptations to the current model structure could be made to allow the detection of long-term changes in surface conditions. This could take the form a regression or mixed-effects based model, which may represent the surface conditions of the ice but at the expense of being more difficult to fit and potentially interpret due to the increased number of parameters.

Given the assumptions and intuition behind some of the modelling choices, this data set could alternatively be modelled using a Bayesian framework with prior distributions that reflect these assumptions. We would expect melt to have similar distributions at different cells even if there is less evidence of melt in some cells than others. If this is the case, then a modelling framework could be established whereby the melt components of the model share information or parameters, while the ice components are independent between cells. This could be used to estimate melt probabilities even in cells where no melt temperatures have been observed, as melt components could still be estimated using information from other cells with more data resembling melt.

Fitting models to sub-decadal data sets would lead to insufficient data to fit the model; in particular, there would be many cells and time periods with an extremely low number of IST above $-1.65°$C and $0°$C, making it difficult to fit the melt component with any degree of accuracy. By separately fitting the model to data from two decades (2001 to 2009 and 2010 to 2019), the overall temperature trends were examined. South west and south east areas of the ice sheet were found to have a higher probability of melt in 2010 to 2019 compared to 2001 to 2009, and although 22.2% of cells saw a decrease in melt probability of some degree, 68.5% of cells saw an increase in melt probability and the average increase was more

than double the average decrease (-0.0044 compared to 0.011). By contrast, the 95% quantiles increased at almost all cells (1100/1139) and the 1-year return levels increased at most cells (849/1139). The overall trends of the model suggest that melt has become more frequent in the most recent decade, and temperatures more broadly are increasing in areas across the ice sheet.

3 Spatial trends of marginal parameters

3.1 Introduction

In Section 2 we developed strategies for the statistical modelling of extreme events at specific sites. Such models will be referred to as marginal models, since the model fit is based entirely on the data at the site and there is no sharing of information across sites. These models can be used to predict site specific characteristics, summary statistics, or more tailored statistics such as melt probability. Whilst such models can be useful, they are time consuming to fit since the number of parameters is large for a large number of sites. Fitting separate marginal models does not permit formal statistical quantification of spatial trends and associated uncertainty, nor does it allow pooling of information between nearby sites. For instances where there may be spatial structure in either the underlying response (and consequently in the measurements) or the marginal models, ignoring such structure is inefficient. However, we can still assess the marginal model fits for spatial trends using data visualisation tools and basic statistical summaries and hypothesis tests.

In terms of model structure, we consider three different model structures for extreme events: marginal models, regression models, and spatial models. Let Y_{ij} be the jth observation at site i, and let x_i be a vector of site-specific covariates for site i:

$$\text{Marginal models:} \qquad Y_{ij} \overset{IID}{\sim} F(\theta_i);$$

$$\text{Regression models:} \qquad Y_{ij} \overset{IID}{\sim} F(\theta_i)$$

$$\theta_i = \beta' x_i + \epsilon_i$$

$$\text{Spatial models:} \qquad Y_j \overset{IID}{\sim} F_d(\theta)$$

where $Y_j = (Y_{1j}, ..., Y_{dj})$, d is the number of sites and F_d is a d-dimensional distribution. In this section, we move from a marginal model to a regression model to begin to incorporate spatial information in our model.

One of the main motivations for considering a spatial modelling approach with regards to MODIS data is the natural spatial structure of the data set. ISTs on individual days, summary statistics, and model parameters all show clear spatial trends. While these trends can be observed from the marginal models, we aim to quantify the trends in the properties of marginal extreme events. This enables spatial interpolation to sites at which observations are unavailable as well as an improved understanding of factors that drive the frequency and magnitude of extremes events. Therefore, we aim to fit a model that capitalises on this spatial structure to better represent the data set.

An additional, more practical motivation of a spatial modelling approach is to better deal with large amounts of data and/or sites. The MODIS IST data set used in Section 2, there are 2,868,630 cells identified by the ice mask as having an ice surface on the Greenland Ice sheet, and the high spatial resolution makes it unrealistic to assume that these all behave independently of each other. The volume of data presents both opportunities and challenges. Although there is a lot of information at each cell and of nearby cells to describe the behaviour of ISTs, even for the independent mixture models only a small fraction of the total data could be modelled due to the computational cost of applying the models. The desire to use all of the information available can be restricted by the computational or time cost of using it all, which motivates quicker methods of accurately summarising or analysing more of the data set.

To bridge the gap between a spatial model with dependence between locations and fully independent models at each location, we first examine the parameters of the independent marginal models spatially and aim to model those rather than the data at each location. In essence, we aim to find the relationship between the parameter estimates of our independent models rather than between all of the observations. This reduces the computational cost as each location is represented by a small number of parameters rather than thousands of data points, and can potentially extend the marginal models to new locations if the spatial relationships

are consistent enough. This approach would also be particularly useful for data sets with less consistent and high resolution spatial resolution.

Furthermore, more formal identification of spatial correlations provides further motivation for spatial models that don't assume independence between locations. Although it would appear that certain spatial covariates are associated with ISTs and that there are spatial correlations in the data, modelling these gives a more formal recognition of their magnitude and form that will inform any future model structures. The covariates are particularly important to examine when considering spatial dependence models that use a distance metric as a measure of similarity of locations. A distance measure could be as simple as a great-circle distance, but if there are other associated spatial variables then other measures such as elevation could also be included. Testing their association in a modelling framework gives us a clearer idea of whether this could be a sensible modelling consideration for spatial models.

In this section, we use a regression modelling framework to examine the spatial variability and trends in the marginal model parameters relationships between these parameter estimates and spatial covariates. By examining the model parameters, we aim to identify key spatial covariates that could be used in spatial dependence models and to get a more detailed perspective of the spatial trends of the data set. Furthermore, the regression model allows us to estimate marginal model parameters for new locations. In effect, we aim to estimate model parameters using only a handful of spatial covariate values rather than the full data set at each location, permitting estimation of marginal behaviour at sites where data is unavailable.

We use and build upon a linear regression model as the basis for our marginal distribution parameter predictions. For marginal model parameters $(\underline{u}_i, \underline{\sigma}_i, \underline{\phi}_i)$ at site i, we have p associated covariates $\underline{x}_i = (x_{i1}, \ldots, x_{ip})$. We aim to estimate each parameter individually (i.e each of $(\underline{u}_i, \underline{\sigma}_i, \underline{\phi}_i)$) as a function of our covariates $\underline{x}_i$. The covariates include variables such as latitude, longitude, elevation, drainage basin, and the distance to the coast amongst others. Each of these covariates is

defined for each individual cell, either as a continuous variable, as a binary indicator in the case of an elevation cutoff point, or as a factor for the drainage basins (1).

3.2 Data

3.2.1 ISTs

Once we have fitted the mixture models, we assess the goodness of fit using the Kolmogorov-Smirnov (KS) test to compare the distributions of the data and of the mixture model using maximum likelihood estimation. Let F_n be the empirical distribution function of the data, and $F(x)$ be the cumulative distribution function of the fitted mixture model. For each cell of our sample data, the KS test statistic D_n is defined as:

$$D_n = \sup_x |F_n(x) - F(x)|,$$

which is the maximum distance between the cumulative distribution function and empirical distribution function evaluated over the interval $[\min_i(x_i), \max_i(x_i)]$.

First, we establish more clearly what the KS statistic means in terms of distributional fit. Figure 16 shows predicted mixture model fits for cells at which the KS statistic was closest to 0.01, 0.05, and 0.1. For each cell, we plot the probability density function with the individual model components and a histogram of the data, as well as a probability-probability (PP) plot.

The lowest KS statistic examined is 0.01, which is a reasonably common value for the fitted mixture model as can be seen from Figure 17. Figure 16 shows that the overall distributional fit matches very well with the distribution of the data. The three ice components contribute to the different peaks of the multi-modal distribution, and the melt component is very small but still identifies a small number of possible melt temperatures from the data. The goodness of fit is confirmed in the PP plot, as there are no significant deviations from the centre line. This suggests that the model fits the data well across all quantiles consistently.

For a KS statistic of approximately 0.05, the model fit is still relatively good
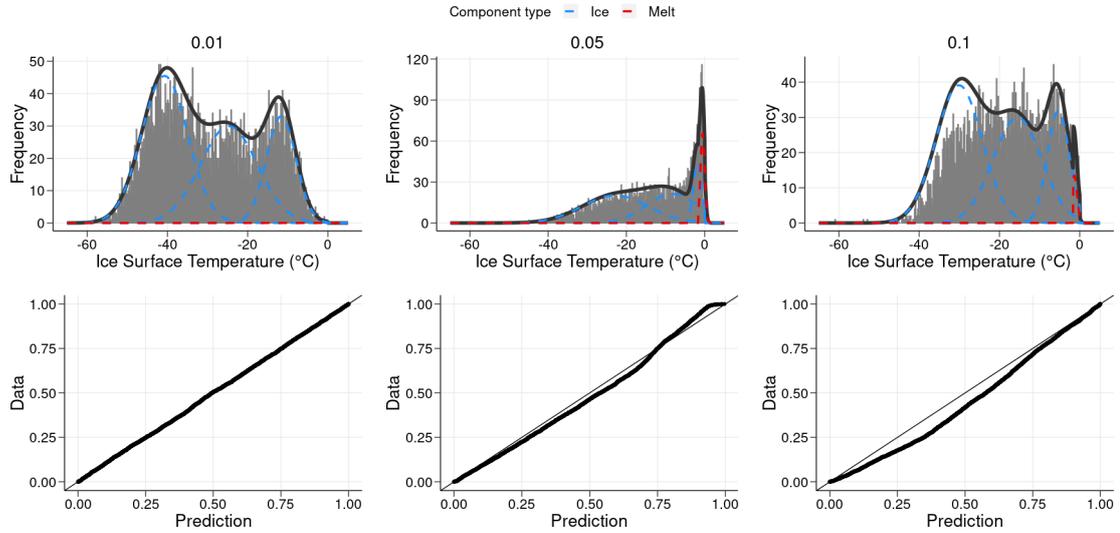
Figure 16: Top row: density functions of predicted mixture models overlayed onto the data at the same cell. Density functions are scaled in height to match the data histogram. Bottom row: PP plots of the predicted cumulative density functions against the empirical density function of the data evaluated for each observation. Examples represent the predicted distributions with KS statistics closest to 0.01 (left), 0.05 (middle), and 0.1 (right).

overall but with some noticeable inaccuracies. The overall shape of the distribution at lower temperatures - below around $-10°$ C - has been captured well by the model, particularly the lower tail for the coldest temperatures. However, for temperatures above $-10°$ C the accuracy decreases slightly. The predicted distribution has a more pronounced dip around $-5°$ C than the data, and the predicted peak of the temperatures just below $0°$ C is slightly offset from the true peak. This is reflected in the PP plot too, with a clear deviation from the central line for probabilities above 0.75. The overall fit is still reasonable though as the relative size of peaks and overall distributional shape have been captured with a reasonable degree of accuracy by the predicted density.

The third example is a case when the distribution has not been captured well by the predicted density. For a KS statistic of around 0.1, the predicted density has more significant inaccuracies from the distribution of the data. The lower temperatures are assigned more weight than the data would suggest due to a peak

in the predicted density that is not present in the data. The melt peak around 0°C is represented well, however this is not consistent for KS statistics around 0.1 as this region of the distribution can sometimes be offset from the true melt peak. The inaccuracy can also be seen in the PP plot with a large consistent curve away from the central line for quantiles between 0 and 0.75. This isn't necessarily unexpected from the previous inaccuracies seen in the regression modelling for component 4. This diminishes our confidence in models with KS statistics around this value, and demonstrates the possible level of inaccuracy that can be seen from the predicted distributions.

From Figure 17, we can see that much of the ice sheet does not display significant trends in the value of the KS-test statistic. This is particularly true in the south of the ice sheet. However, there is an area of the ice sheet on the north west coast that consistently displays higher than average values reasonably, suggesting that there may be some geographic covariate or common factor that causes a worse fitting model. Cells in the centre of the ice sheet generally have a slightly better fit than those near the coast, although this is expected since the cells near the centre of the ice sheet are likely to have a melt component with a weight close to 0, which simplifies their distribution and hence the model fit.
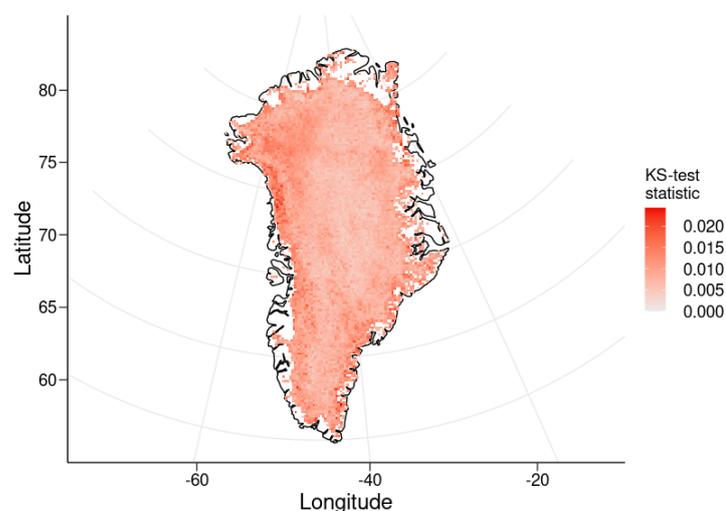

Figure 17: Cell-wise KS-test statistic to assess fit of the mixture model to the data.

In conclusion, the model appears able to capture the distribution of the data consistently across the majority of cells on the ice sheet. This gives us confidence that we can trust the ice mask and its selection of cells that contain ice. It also validates our decision not to discard ISTs above 5° C as the model fit does not seem to be significantly negatively impacted by them, and our extra model considerations seemed to have allowed the model to more appropriately consider these higher values.

3.2.2 Covariates

For some of the main spatial covariates such as latitude, longitude and elevation, the Greenland ice sheet has smoothly varying geographic characteristics. For temperatures, there are two main trends. Locations in the north at higher latitudes are generally colder than those at lower latitudes, and temperatures decrease as the distance to the coast increases. This also coincides with elevation, as areas closer to the coast generally have lower elevations. The highest elevations are found close to the centre of the ice sheet which occurs in the mid-range of latitudes and longitudes. Longitude varies the most with temperature as there are coasts and areas with high elevation across the range of longitude values. Latitude and elevation are also not as strongly correlated due to low elevation areas at both high and low latitudes since coasts occur at all latitude values.

In order to be useful as regression covariates the spatial covariates need to be either easily interpolated between sample locations or at a higher spatial resolution than the response variable. Latitude and longitude are natural choices. Elevation is also easily accessible at a high resolution and can be interpolated reasonably simply if even higher resolutions are required. Distance to the coast can be calculated for any latitude/longitude using a shapefile of the coastal boundary. This is also true for drainage basins, which are also defined in shapefiles. These covariates therefore form the basis of our regression model.

A final covariate used in the regression models is an indicator variable based on elevation being above or below 1650 m. Although this may seem arbitrary,

for the mean estimates there is a difference in distribution of the estimates once elevation falls below a certain level. We tested regression fits with different elevation thresholds for this variable by fitting models with an elevation indicator covariate at intervals of 25m. By comparing the adjusted R^2 values of each model, we could compare between model fits to find the optimal value of the change point within a consistent framework of the rest of the model and data. In doing so, 1650 m was found to be the optimal change point.

A consideration of using the above covariates is that, particularly for elevation and distance to the coast, there is potential for multi-collinearity if both are used in a single model. Figure 18 shows the extent of this relationship, with a clear non-linear relationship between the two variables. This relationship would be a clear issue if we were primarily interested in the interpretation of the regression parameters estimates rather than the predictions of the regression model. Since the main goal of the regression modelling is to get the most accurate predictions of the overall IST distribution using the minimum amount of data possible, interpretation of individual parameter estimates is of low interest. Consequently, where two correlated variables each have an association with the mixture model parameters then we include them both.
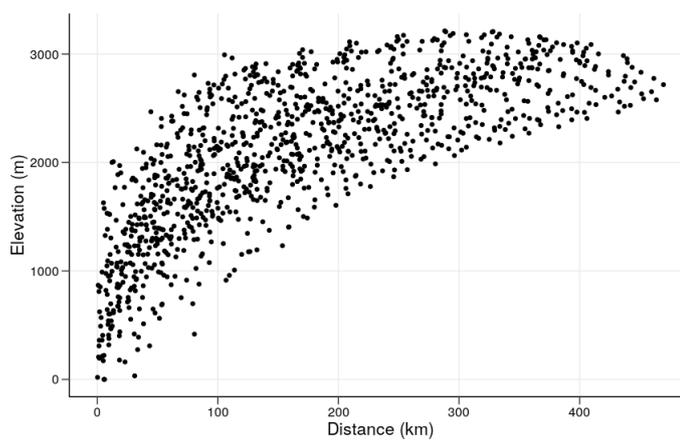

Figure 18: Elevation against distance to the coast of all cells in our sample.

Figure 19 (b) suggests some interaction between the effects of elevation and

58

latitude on the estimated means. Locations at higher latitudes generally have a lower estimated mean and conversely for the locations at lower latitudes (ie further south). However, the points at the highest latitudes also generally have higher estimated means than those at the mid latitude values (around 70-75), showing that this is not a linear trend. The trends with latitude follow the two main temperature trends discussed earlier - higher means generally indicate locations closer to the coast and/or further south. For a given elevation, sites with lower latitudes have a higher mean. This is true for all ice components, though the difference decreases with order of component and increases in elevation. The latter is likely to be confounded by the reduced range in latitude at higher elevations.
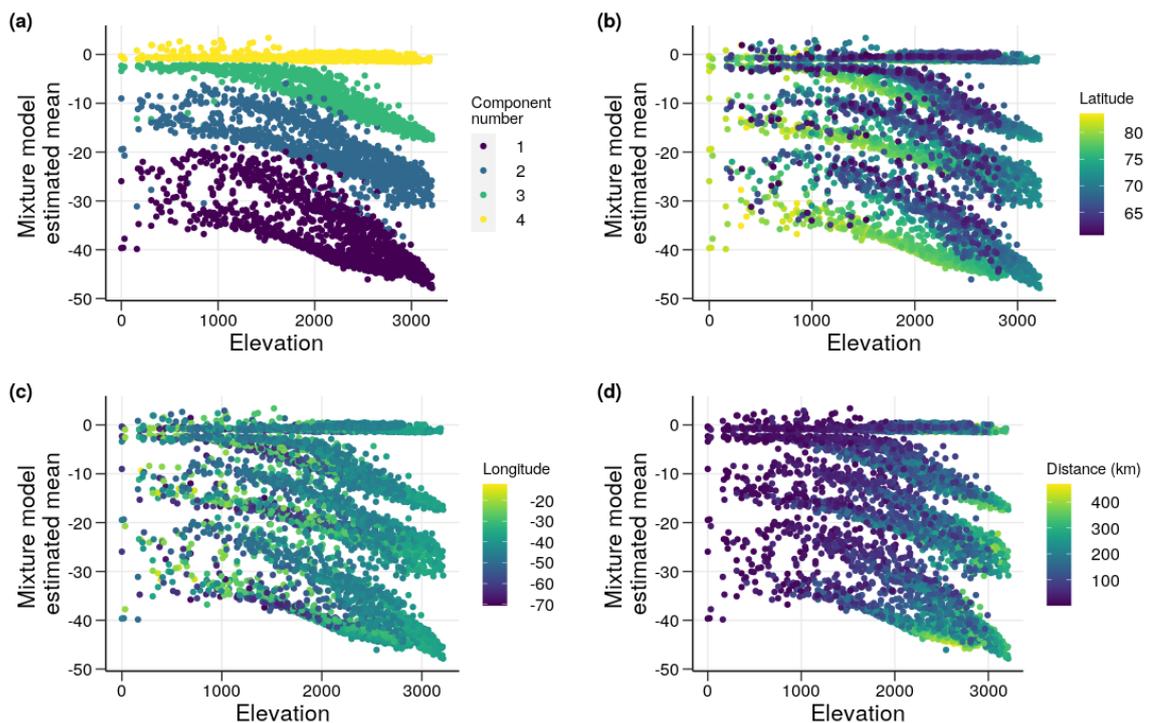

Figure 19: Plots of the estimated mixture model means for each location against the location's elevation, coloured by the component number, latitude, longitude, and distance to the coast.

3.3 Methods

To have confidence in the regression modelling and of inference from the resulting mixture models, we check how accurately the models represent our data and whether any additional modelling considerations need to be made. Because the regression model is fit to the mixture model parameters, there are two levels of model diagnostic that we need to perform. In addition to standard regression diagnostics, we also need to check the fit of the marginal mixture models using parameter estimates from the regression model. This is because the mixture model has 12 parameters, so even if the individual estimates seem accurate, small inaccuracies in several parameters can contribute to a poorly fitting distribution when combined.

Further to the model definition of the mixture model in Section 2, we make some additional modelling considerations for our second application of the model. Due to the melt boundary and the resulting upper limit on IST measurements, we define IST observations > 5°C as outliers as the surface of the ice sheet should not feasibly be able to reach these temperatures; 5° C is chosen so that we have confidence that any values above this are either measurements of a non-ice surface type or are a result of measurement error. This issue is mostly isolated to coastal areas, as cells nearer to the centre of the ice sheet are almost certain to be ice, and their temperatures do not go close enough to the upper IST to cause any issues for the model.

The simplest solution would be to simply remove any ISTs above 5° C to remove all potentially erroneous observations. However, if the highest values are indeed from ice surface conditions and are measurement errors of high temperatures, then they still indicate that melt has occurred even if the precise temperatures are not as for accurate as other observations. Removing them from the data set entirely would result in the mixture model under-estimating the proportion of melt observations and consequently underestimating melt probabilities.

Alternatively, we could remove any cells that have ISTs over a certain threshold to remove any cells that could have non-ice surface conditions. This would ensure

that there are no erroneous observations and that the surface conditions at each cell were consistent for the observed time period of the data set. However, this would also eliminate some of the more interesting cells in terms of melt, ie. cells that are closer to the coast. Whether they are removed or not may depend on whether these cells can be accurately modelled by the mixture model, or whether the fits in such cells are consistently inaccurate. If so, it would be better to remove them from the parameter estimates used in the regression models to keep more consistent and accurate trends in the estimates. Therefore, we need to examine whether the values are the result of unrecorded measurement error or errors in the ice mask.

If the ice mask was assumed to be incorrect, then some of the cells in our sample are not ice and would not suffer the same truncation issue as ISTs. The model assumption of a peak in ISTs just below 0° C would therefore be incorrect and the truncation points would have no interpretation. We would also expect a non-trivial number of temperatures exceeding the normal range of non-ice temperatures due to the lack of upper limits on surface temperatures. This would then mean that the mixture model would not necessarily fit the data as well as it would for ice cells.

To examine whether it is better to trust the observations or the ice mask, we fit mixture models to a larger sample of cells. We increased the chosen sampling resolution of cells in both x and y dimensions from the modelling work in Section 2 from 1 in 50 to 1 in 20 cells, with an additional requirement that the cells needed at least 500 data points. This resulted in 7160 cells being chosen after 6 were removed due to insufficient data. We then checked all model fits in which the ice component had a very small estimated weight. This resulted in some distributions having two main ice components and a third ice component with a very small weight. This small component made very little difference to the fit, but emerges as an outlier in the regression modelling as its mean and standard deviation were no longer linked to the main body of the distribution. Therefore, 4 cells with an ice component weight below 0.005 were removed for a final total of 7156 cells.

For ISTs above 5° C, we assume that the ice mask is correct and the observations

are incorrect to check our assumptions of these features. To do this, we treat the ISTs above 5° C as censored. This means that we assume that the temperature is at least 5° C, but we can't know exactly what the temperature is. This allows the data to still be considered by the model, but not to suggest that ISTs can reach unrealistic temperatures such as 15° C.

To treat these observations as censored requires a change to the model fitting method from the EM algorithm to maximum likelihood estimation. In the likelihood framework, a non-censored observation x with IST t contributes its corresponding probability from the mixture model to the likelihood by using the probability density function of the model $P(x = t)$. For observations above $5°C$ the likelihood contribution is given by the survival function $P(x \geq 5) = 1 - P(x < 5)$. This allows us to still consider them in the likelihood function whilst not incorporating the exact measurement.

Let n_{IST} be the set of ISTs x such that $x < 5°$ C, $n_{outlier}$ be the set of x such that $x \geq 5°$, and θ be the mixture model parameters. Then we define the likelihood function as:

$$L(\theta) = \prod_{i \in n_{IST}} f(x_i) \prod_{j \in n_{outlier}} 1 - F(5).$$

Figure 20 shows that only a small proportion of cells are affected by the changes to the likelihood, with only 184 cells having any ISTs above 5° C. Of these 184 cells, the mean maximum IST is 8.73° C and the highest observed IST from all cells is 23.36° C. Further, the mean number of exceedances of 5° C is 39.48 amongst these 184 cells, with a range of 1 to 633. This demonstrates the importance of establishing whether the ice mask or observations are correct, as some of these cells have a particularly high number of melt temperatures and would influence any overall summary statistics if they are included in our analysis.

Having established that the mixture models have a good fit to the data, we fit a series of linear regression models to the mixture model parameters. Each of the regression models are fit separately to each model parameter for a total of 12
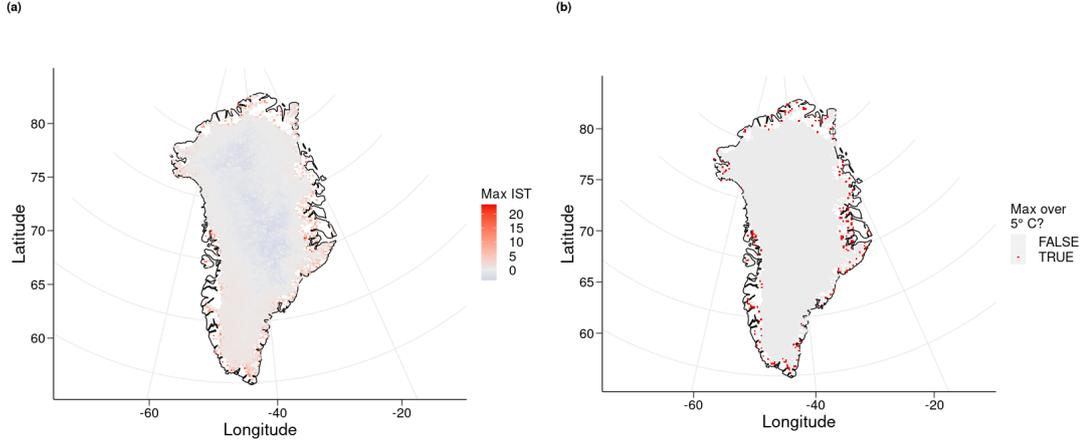
Figure 20: (a) Maximum observed IST from 2000 to 2019, and (b) locations at which the maximum observed IST is greater than 5° C

models - four mean parameters, four standard deviation parameters, and four weight parameters. We use six covariates as discussed in Section 3.2.2 to estimate the parameters using only spatial information - latitude, longitude, elevation, distance to the coast, elevation above 1650m, and drainage basin (specific details of each covariate can be seen in Table 1). With the model structure described at the beginning of Section 3, we define linear models for each mixture model parameter θ_j for $j \in \{1, \ldots, 12\}$.

Model 1 uses only first-order covariate terms, such that:

$$\text{Model 1} \quad \theta_j = \beta_0 + \beta_{1,j} x_1 + \cdots + \beta_{14,j} x_{6,9} + \epsilon_j.$$

Model 2 includes the same terms as Model 1 with the addition of all second-order interaction terms, such that:

$$\text{Model 2} \quad \theta_j = \beta_0 + \beta_{1,j} x_1 + \cdots + \beta_{14,j} x_{6,9} + \beta_{15,j} x_1 x_2 + \cdots + \beta_{63,j} x_5 x_{6,9} + \epsilon_j.$$

Model 3 includes the same terms as Model 2 with the further addition of all

63

Label	Covariate	Data type	Covariate levels
x_1	Latitude (°)	Continuous	N/A
x_2	Longitude (°)	Continuous	N/A
x_3	Elevation (m)	Continuous	N/A
x_4	Distance to the coast (km)	Continuous	N/A
x_5	Elevation above 1650m	Binary	2
$x_{6,1:9}$	Drainage basin	Factor	9

Table 1: Details of the covariates used in the regression models.

third-order interaction terms between all covariates, such that:

$$\text{Model 3} \quad \theta_j = \beta_0 + \beta_{1,j}x_1 + \cdots + \beta_{14,j}x_{6,9} + \beta_{15,j}x_1x_2 + \cdots + \beta_{63,j}x_5x_{6,9} +$$
$$\beta_{64,j}x_1x_2x_3 + \cdots + \beta_{153,j}x_4x_5x_{6,9} + \epsilon_j.$$

Interaction terms are considered between all covariates, and stepwise selection in both directions is used to select the best set of covariates to use for the model.

Although for a linear regression model we would normally want independent variables as our covariates, our aim here is less focused on interpretation and more on the accuracy of our predictions based on limited data. This model aims for accurate predictions by sacrificing some of the potential interpretation of the previous models. In terms of directing our subsequent spatial models, the simpler regression models help to inform the relative importance of the spatial covariates, whereas the interaction model gives us the best prediction of our marginal model parameters without having to process data at a new location. Both models provide information for our future models, either in predictions or covariate structure.

The only restriction placed on the predicted parameter values was on the model weights, which have to sum to one and cannot be negative for the model to have a valid probability density function. This could have been implemented in the model framework as an extra condition in the regression fit but we chose to scale the weights after estimation to ensure these conditions were held. Any weights estimated to be less than 0.0001, which happened only for the melt component, were set to 0.0001

to give such cells a very small probability of melt - less than 1 melt observation in 20 years. Then, the weights for each model were scaled to equal one once they were all positive.

Rather than use all of the fitted mixture model parameter estimates to fit the regression model we define training and test data sets to allow for model validation after fitting. To do this, we randomly assigned 70% of the cells as the training set and 30% as the test set using a consistent seed. All regression models are thereby fit to the training set of 4960 cells, leaving 2196 cells for testing.

We also create an order in which parameters are estimated to allow earlier parameter estimates to inform further estimates. We tested all combinations of orders of the three parameters types, and found the most effective ordering of the parameters was to first estimate the means, then the standard deviations, then the weights. To make the component parameters more distinct and trends more identifiable, estimates of the same category (e.g means or standard deviations) are indexed based on the ordering of the means, with 1 indicating the lowest mean and 4 indicating the highest mean i.e the melt component. For each component, the standard deviation includes the predicted mean value as a covariate, and the predicted weight models use both the predicted mean and standard deviation values as covariates. This adds information for the standard deviation and weight models to use without having to know any more spatial information or adding any interaction terms.

3.4   Results

3.4.1   Model 1

The first application of the linear regression model has progressively worse estimates for the means, weights, and standard deviations respectively, as can be seen in Figure 21. The mean estimates are not too inaccurate, but the predictions for the standard deviations and weights differ considerably from their expected values and seem to
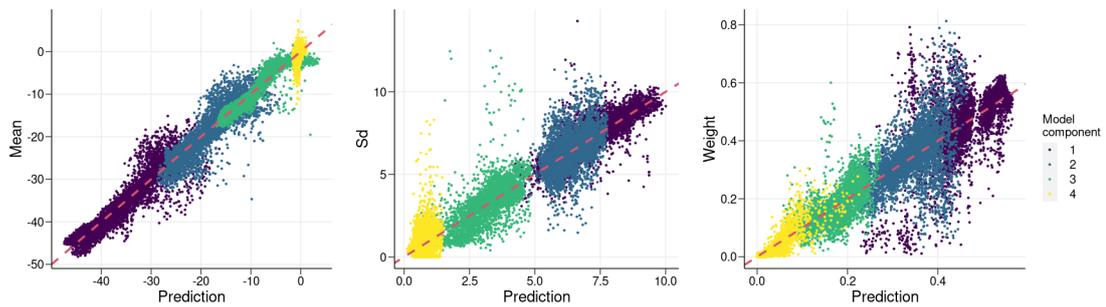
Figure 21: Predicted parameter estimates for the means, weights, and standard deviations from the basic linear regression model plotted against the values from the previously fitted values from the mixture model. Red dashed line represents $y = x$.

be missing some information as to what values they should be. There also appears to be some non-linearity in the predicted values for the means, as the predicted values for components 3 and 4 show non-linear trends not seen in components 1 and 2. However, despite only a limited amount of information available to the model, there are clearly already some trends that have been found in the data that motivate further developments of this model.

3.4.2   Model 2

The second application of the regression model adds interaction terms between the spatial covariates. This allows us to consider how the effects of our covariates vary depending on the values of the other covariates, and gives us additional structure without requiring any more data. We have already seen that there are some relationships between some of the spatial covariates - such as in Figure 18 - so adding interaction terms to the model may help us to capture how the effects of the covariates change relative to the values of other covariates. For this model, we consider all possible second order interaction terms and use the same stepwise variable selection in both directions as previous models to select the most relevant variables.

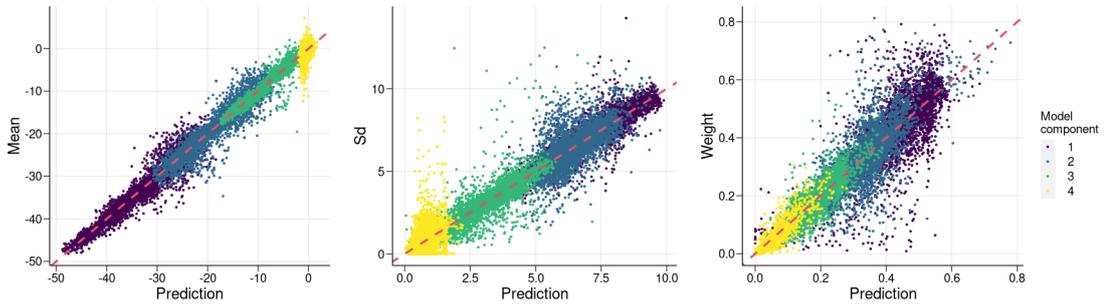The improvement in all three parameter types can be seen in Figure 22. The

Figure 22: Predicted parameter estimates for the means, weights, and standard deviations from the ordered linear regression model with second order interaction terms plotted against the values from the previously fitted values from the mixture model. Red dashed line represents $y = x$.

mean estimates from the ice components are all on average closer to their true values than in the previous models. In the previous model, the main lack of fit in the mean estimates was for component 3, which had a drift away from the line for values from $[-5, 0]$. This has now been straightened towards the central line aside from a small number of points that are over-predicted. The main area for improvement for the mean estimates is for component 4. The average of these estimates is correct, but the spread of the true mean values is much larger than the spread of the predicted values. This may be due to the nature of the ice component compared to the other three components, as by definition it is less likely to vary spatially. Although the weight of the component does vary depending on location, a higher weight can be inferred as more evidence of the same melt process, the mean of which should still be the same. This may explain why the mean estimates for this component are worse than the weight estimates, and could potentially motivate more globally defined melt component parameter estimates based on the idea of the process being the same everywhere.

For the standard deviation estimates, all four components show an improvement in predictive accuracy over the previous models. For components 1, 2, and 3 there are a small number of values that could be classed as outlier predictions given their lack of accuracy, however this could be influenced by the specific shape of the model

components at each cell. In most cases, the ice components form a similar pattern in terms of weights, with a general trend of component 1 having the highest weight ordered down to component 4 with the lowest weight. This is not always the case though, with some very low weights for component 1. In some instances, some ice components have a higher standard deviation and lower weight, so that they add to the overall shape of the distribution rather than modelling a specific peak in the data set. These flatter component 1 curves may be the cause of some of the less accurate predictions, as their form is different to the majority of other component 1 curves. Component 4 again shows some inaccuracy in its estimated values, however this may be attributable to the same cause as the mean estimates of component 4 as the variance of melt temperatures may be similar globally if the process is expectedly consistent.

The estimated weights arguably show the greatest improvement. All four components show increased accuracy in their predicted values compared to the previous regression predictions. Component 4 has potentially the most accurate set of predicted values, which gives us additional confidence in any predicted model estimates if we aim to use them to infer melt probabilities. The main trend in the weight estimates is the larger variance in true weights than the predicted weights for a given weight and component number. The weight values from the mixture model appear to have a large range for a given parameter value such as for component 1 and predicted values around 0.4, whereas the predicted weights vary much less. This may suggest that there is another feature that the model is not considering, however the main trend appears to have been captured.

3.4.3   Model 3

For the final model, we see that the inclusion of each additional order of interaction terms decreases the gain in model accuracy and the time required to fit the models increases at a rate similar to the number of regression terms. Third order terms compromise between fitting time, model complexity and model accuracy, and hence
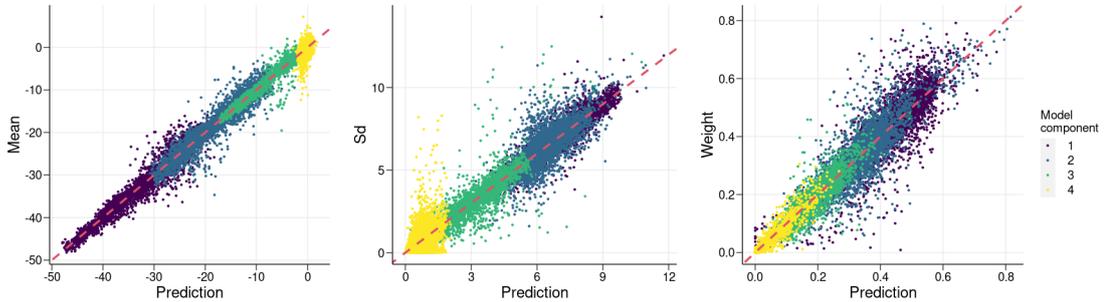
are the end-point for our model.

Figure 23: Predicted parameter estimates for the means, weights, and standard deviations from the ordered linear regression model with 3rd order interaction terms plotted against the values from the previously fitted values from the mixture model. Red dashed line represents $y = x$.

Figure 23 shows minor improvements across nearly all parameters. The means remain the most accurately predicted parameters, with the majority of predicted values lying very close to the actual values and very few clear outliers. A significant portion of the weight and standard deviation estimates are also similarly accurate albeit with a larger number of potential outlier values and slightly more variance in the predictions. The weights demonstrate the clearest improvement from the second order interaction model with all four components' estimates lying closer to the line of perfect fit. Component 4 still has some inaccuracy in the mean and standard deviation estimates as previously discussed for the second order interaction model. The addition of the extra interaction terms is unlikely to provide the model with information relevant to these parameters if, as previously suggested, they are more globally defined parameters that do not depend on location.

3.4.4 Accuracy of predicted densities

To assess the accuracy of the regression model for predicting overall mixture model distributions rather than individual parameters, we re-calculate the KS statistic for the estimated distributions. This gives us a clearer indication of the overall mixture model fits than obtained by considering the parameter estimates alone.

Small inaccuracies in several parameters could contribute to an overall poor fit of the distribution.

Figure 25 shows KS statistics for both the training set and the test set. For the cells in the training set, the fits are consistently good; 10.99% of cells in the training set have a KS statistic less than 0.01, and 50.77% of values are below 0.02. Although this may be expected for cells included in the fitting of the regression model, it acts as a strong indication that the regression model structure has sufficient information to accurately recreate cell-specific distributions in addition to individual parameter values.
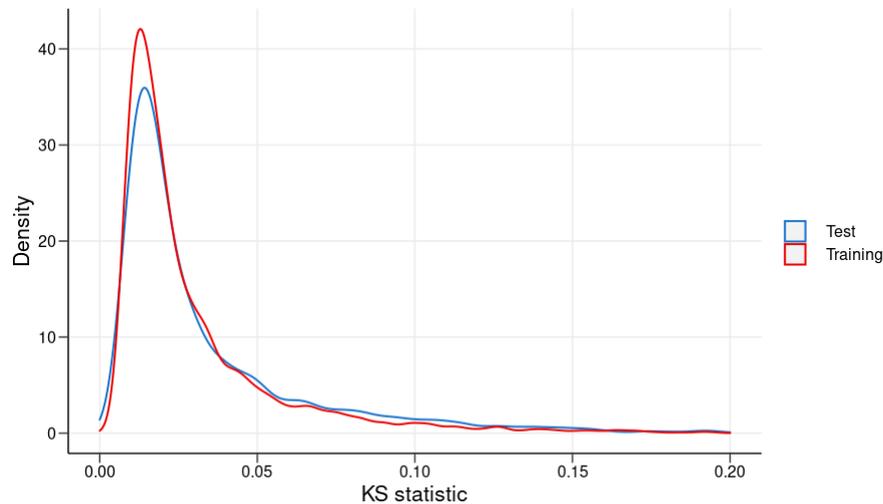

Figure 24: Densities of the KS statistic for the predicted distributions from the final regression model for the training (red) and test (blue) sets.

The test set has on average less accurate predicted distributions, but a large amount of cells still have substantially accurate predicted distributions. 9.65% of cells have a KS statistic less 0.01, 47.91% less than 0.02 and 64.62% less than 0.03. In relation to the examples in Figure 16, 80.59% of cells in the test set have a value less than 0.05, indicating that a considerable proportion of cells have at least a reasonably acceptable level of accuracy and that predictions for unseen locations can be made from the model with some degree of accuracy. The main difference between the training and test set values is the variance in accuracy. While the training set

estimates are unsurprisingly very consistent, some of the estimated distributions for the test set are so inaccurate that they would unusable. To illustrate this, 6.47% of KS statistics are above 0.1, with a maximum of 0.42.

Having established how to interpret the KS statistic, we estimate the KS-statistic for all cells and examine whether there is a consistent difference in fit between the cells with ISTs above 5° C and those without. Figure 25 shows that the distribution of KS-test statistic values between the two groups is reasonably similar in shape, with both groups having a mode around 0.008. Cells with a maximum IST below 5° C have a right skew in their errors, and also proportionally more models with a higher test statistic than the other group. Most importantly, the cells with a maximum IST above 5° C do not have a significant or consistent worse fit than the cells without, suggesting that the mixture model is equally capable of modelling both groups of cells. This indicates that we can include these cells in our analysis.
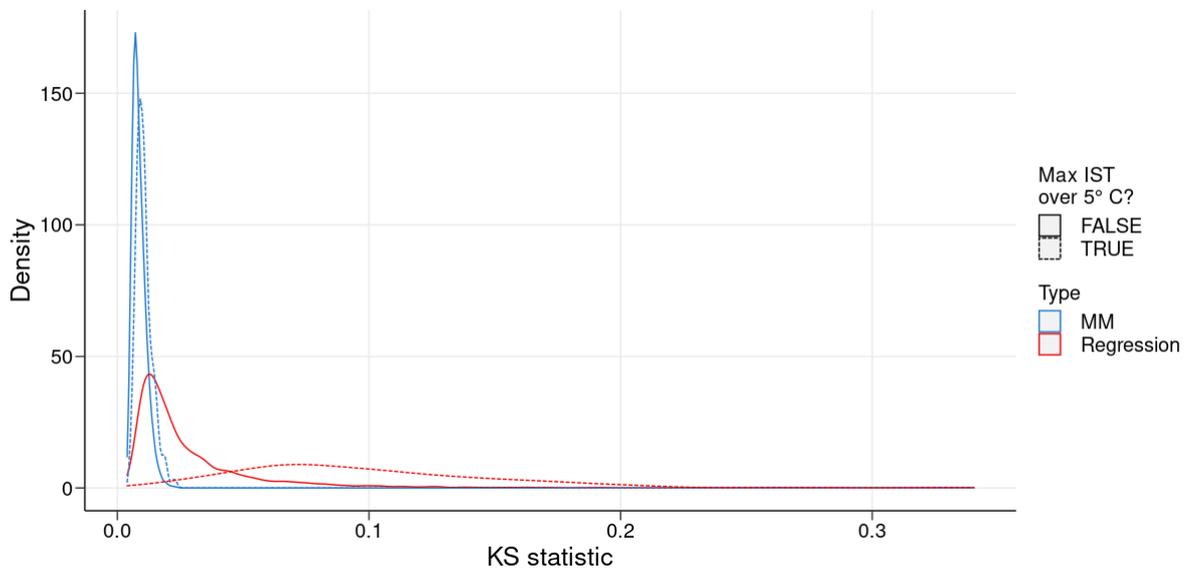

Figure 25: Densities of the KS test statistic values for models fit directly to the data (rethe cells with (red) or without (blue) ISTs above 5° C.

To assess which cells have less accurate distributions, we examine the KS statistic spatially. Figure 26 (a) shows that most of the cells with higher KS values are close to the coast. These cells are found along most of the coastline aside from a small

gap in north west. This area seems to coincide with an area of higher KS values from the models fit directly to the data (Figure 17), suggesting that while model fits in that area are less accurate than average, predictions are more consistent than average. There are also more KS values over 0.1 on the east coast than the west, suggesting that local behaviour on the east coast is more difficult to predict. The trend with distance to the coast is corroborated in Figure 26 (b), where all but six cells with KS values greater than 0.1 are less than 100km from the coast. Higher local variability in geophysical and meteorological conditions, higher probability of melt and relatively fewer cells on the coast than the centre may be contributing factors.
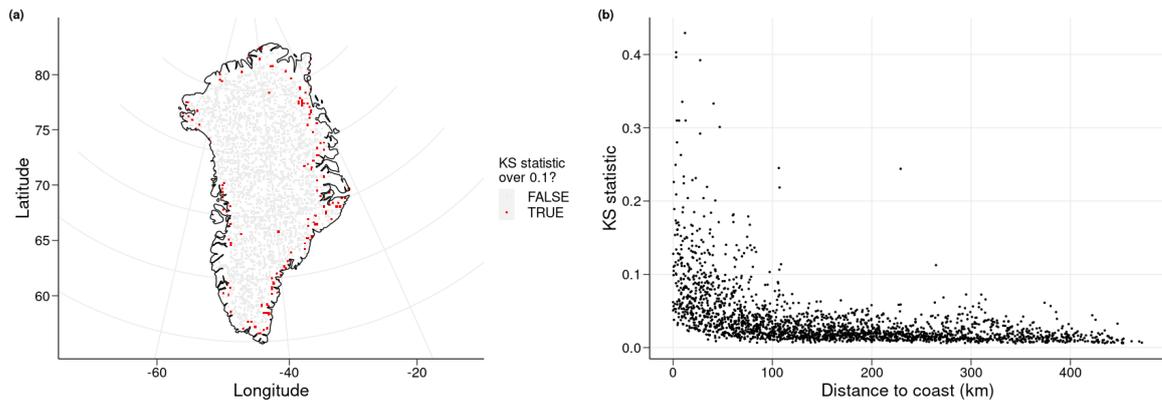

Figure 26: The KS statistic for the predicted distribution of cells in the test set plotted spatially (a) and against the distance to the coast (km) of each cell (b).

3.5   Discussion

In this section we have demonstrated that even simple predictive models with limited information can be used to accurately and reasonably reliably predict marginal distribution parameters. The final regression model is capable of predicting a range of complexly shaped distributions using basic spatial information. The model predicts the distribution by accurately estimating each of the 12 parameters in the Gaussian mixture model structure we previously defined. For any location on the ice sheet, all other covariates used by the model can be calculated using the coordinates,

allowing distributions to be analysed without the need to handle large amounts of satellite data. This means that, with the relevant shape files and elevation map for calculating the covariates, the distribution at a particular cell can be estimated using only a set of latitude and longitude coordinates. Although it could be said that using these still represents some data processing, evaluating a single cell's location and elevation is much more efficient than extracting 20 years of data from a large satellite data set then fitting a statistical model to it. When interest lies in many cells, the benefits become even clearer as there is no need to store or process very large satellite data files, yet thousands of distributions can still be estimated.

This approach can be applied to other spatial data sets with a similar need for estimation of marginal distributions. The overall approach is to find a consistent marginal model structure, then fit a regression model to the marginal model parameters using only spatial information as the covariates. In the context of ISTs, the similarity of surface conditions across the ice sheet, high spatial resolution of the data and high resolution of elevation data all aid the effectiveness of the regression model and predictions from it. Surface conditions may not be as uniform in other areas of the world, but if there is data describing the changes in conditions such as vegetation indices or land cover data then these can be included in the regression model in addition to the other covariates. Furthermore, the model used for IST has a relatively high number of parameters due to the multi-modality of the data and variance in distributional shape. Both of these properties may not be present in other data sets with more regular marginal distributions, potentially giving clearer trends in marginal parameters over space.

For our work, the main use of this approach is to quickly estimate marginal distributions for use in spatial models. The spatial models considered in subsequent sections consider the dependence between different locations by considering data at each location on the same scale. This requires a marginal model at each location to translate the data onto the same scale as required by the spatial model. Once on the same scale, the spatial models use only the spatial location data to define

dependence and do not require the underlying marginal data. Because of this, we can extend the spatial model to new locations. We estimate its spatial dependence with other locations using the spatial location data, then estimate the marginal distribution using the above approach or similar to translate any predictions back to the original marginal scale. This idea will be expanded upon in Section 5.

4 Modelling spatial dependence with Gaussian Processes

4.1 Introduction

Our work so far has focused on the distribution of IST at a single cell. This allows us to make inference of single cell behaviour of ISTs as well as examine spatial trends in the marginal distributions. Due to the clear spatial trends observed in both site-wise summary statistics (Section 2) and marginal model parameters (Section 3), we now investigate the spatial dependence structure given the clear trends in the marginal behaviour.

There are many motivating reasons to use spatial models to study melt. The most natural of these is to describe additional aspects of melt events. Using spatial models, we can examine the size and spread of melt events, including whether this changes over the ice sheet. This gives a more nuanced view of spatial melt events as the spatial models can consider the dependence between temperatures that occur at the same time rather than in isolation from each other.

More specifically, a Gaussian process can be used to achieve these goals and to view the data's spatial dependence. The model treats the cells as a single temperature surface rather than as a series of independent temperature observations. As a result, it can consider the dependence between cells, allowing for surface wide inference. This also presents a useful first modelling framework for an initial spatial model, as there is no requirement for the identification of extreme temperatures. The Gaussian process can model all temperatures at all sites, making it simple in both implementation and interpretation.

We first examine the coincidence of ISTs at different cells at the same time point in light of missing data, further to the previous analysis of missing data in Section 2.2.4. This validates the decision to use a spatial model further as the spatial model considers the correlation of observations at the same time point, and gives us a

framework with which to assess empirical correlation in the data set. Using the mixture models from Section 3 as marginal distributions, we transform the data to standard normal margins and fit a Gaussian Process to the data. This model is used to assess the dependence between data across different cells and gives us an initial understanding of the spatial dependence in the data. We then progress to fitting Gaussian Process models to specific quantiles of the data set in order to assess whether dependence is the same or different at all quantile levels, and thereby whether a spatial extremes model could be more suitable to model spatial melt events.

4.2 Data

4.2.1 Missing data

The marginal mixture models do not have to explicitly deal with the missing data - aside from limiting analysis to cloud free days - as the models do not take into consideration the time at which the observations were made. After filtering for cloud free days, the data are assumed to be missing at random, and since the model assumes identically distributed data at each site, there is no need for them to take further account of missingness. As a result, the date of the observations is not included in the mixture models as their purpose is to represent the entire distribution of the data. Inference for specific time periods outside of the observation period used for the model fit can not be made as there is no information to distinguish when each observation used in the model has occurred. However, the spatial models consider the relationship between observations that occur at the same time point, whereby a single observation as considered by the model is a spatial surface at a given time point. For a time point t, we consider the behaviour of surface $\underline{X}_t = (X_t(s_1), \ldots, X_t(s_d))$ for d cells. This is to eliminate other possible sources of variation such as varying weather conditions on different days and the broader yearly temperature cycle.
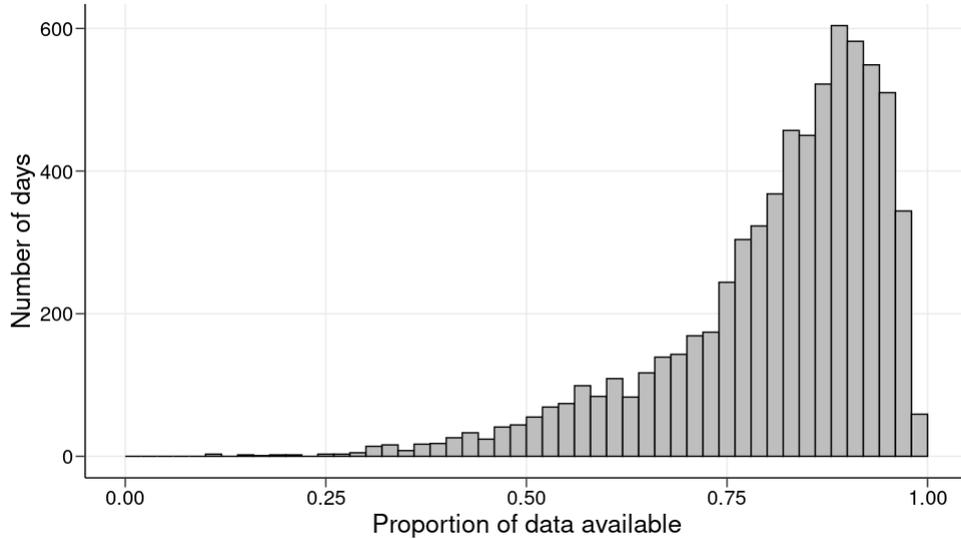
Figure 27: Frequency of the proportions of available MODIS IST data at 7156 sample cells each day between 01/01/01 and 31/12/19.

As a result, we examine the daily proportion of missing data as opposed to the previous missing data analysis that focused on larger time scales (Section 2.2.4). From Figure 27 we see that the majority of days in the chosen observation period have a high proportion of available data. The median proportion is 0.85 and 75.45% of dates have at least 75% of cells with a valid observation and only 3.79% of dates having less than 50% available data. This confirms our early conclusions that the extent of missingness is sufficiently low to continue with our analysis, i.e there is sufficient data at each time point for spatial dependence to be thoroughly examined.

4.2.2 Coincidence of missingness

In addition to the proportion of available data, it is important to examine the spatial trends in missing data too to check that there are no regions or between-cell distances that are disproportionately affected by missingness. Based on the trends seen from the mixture models and our understanding of the climate and meteorology of Greenland, we expect there to be greater correlation between ISTs at cells closer

together than those further apart. Although we can see that the proportion of missing data on each day is relatively low, in order to identify spatial correlation and dependence there also needs to be coincident data at cells close together based on this expectation. If there is insufficient data at closely related cells, then identifying smaller scale correlations or dependences will be difficult.

To check the scale of missingness over different distances, we first calculate for each location an indicator variable denoting whether, for each time point, we have a valid observation (true) or a missing value (false). This results in a vector at each location with length equal to the total number of days in our sample. For our two chosen locations, we then calculate the correlation between this missing data indicator with the same metric at every other sample location. If the missing data between two sites occurs on the same days, then we would expect the correlation of the two locations indicators to be close to 1, and if not then we'd expect it to be closer to 0.

To look in more detail at how the correlation of missing data (and later of ISTs) varies spatially, we select three locations with which to compare the rest of our sample locations. Because of the already observed differences in the amounts of missing data in the centre of the ice sheet compared to the coast, we select one point in the centre and two on opposite coasts in the north west and south east. Furthermore, in order to investigate the effect of elevation, we choose the locations with the highest (central location) and lowest (north west location) elevations from the data set. The three locations are shown in Figure 28: Point 1 is the central location with an elevation of 3213m, Point 2 is the north west location with an elevation of 0m, and Point 3 is the south east location with an elevation of 2451m. This highlights that elevation cannot be used as a proxy for distance to the coast, as Points 2 and 3 are both close to the coast but with very different elevations. This is due to the specific surface characteristics of each Point, and more generally due to coastal areas having more variable surface conditions than the centre of the ice sheet. Additionally, the locations have slightly different amounts of missing data,

78

with 13.39% missing for Point 1, 27.00% missing for Point 2, and 29.16% missing for Point 3.
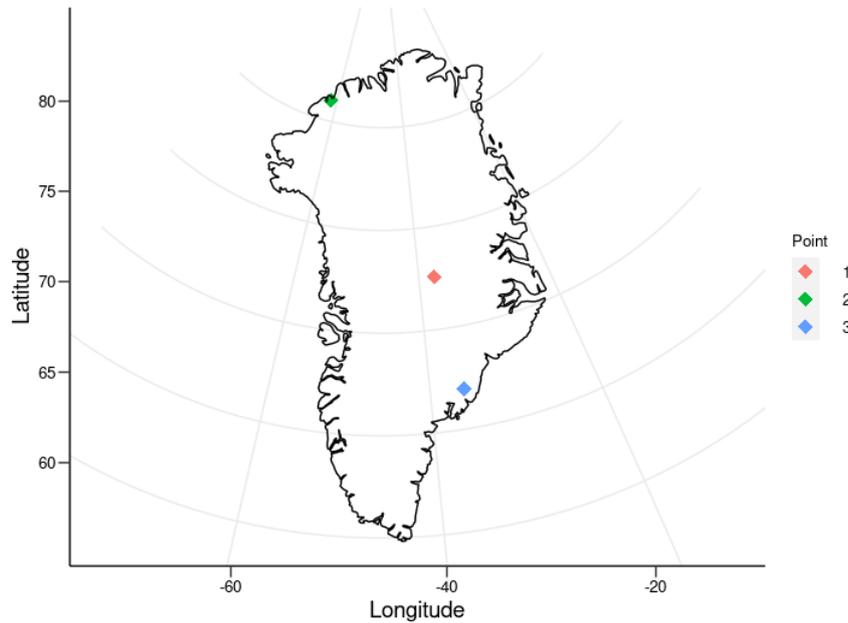

Figure 28: Locations of the three example points used to calculate the correlations in this section. Point 1 is the highest location in our sample (3213m) on the north west coast, Point 2 is the central location and is the lowest (0m), and Point 3 is the south east location at an intermediate elevation (2452m).

For the three locations, we calculate the correlation of the missing data using the indicator variable defined above (results are shown in Figure 29). Point 1 has the highest mean (0.149) and median (0.131) correlation with all other cells. The correlation appears to decrease at a similar rate in all directions with distance. The high rate of coincident missing data could be due to the larger number of cells close by due to its location in the centre of the ice sheet rather than on a coast. Previous missing data trends suggested that cloud conditions (and thereby the amount of missing data) are more localised and intermittent at coastal locations, which could be further linked to elevation and to a lesser extent distance to the coast from this analysis. This would suggest that cloud patterns and thereby missing data depend more on elevation changes and distance to the coast than they do the absolute temperatures of the surface at a particular location.

Point 2 in the north west shows very different trends to Point 1, in addition to having a much lower mean (0.109) and median (0.0824) correlation. Rather than the correlation of missing data decreasing with distance, correlations appear to decrease with changes in distance to the coast. Somewhat similar trends to Point 1 can be seen in the immediate vicinity of Point 2 as cells within a relatively short distance show the highest correlation. However, the higher correlation also extends down the coastline, especially towards the east side of the ice sheet. The distinction between distance and distance to the coast is clearest on the east coast at mid-latitudes. There is a distinct difference in correlation between cells on the coast and further inland, despite the inland cells being closer to Point 2. This suggests that weather patterns may be more linked to distance to the coast - and thereby elevation due to their similarities - than absolute distance, at least for the purposes of cloud cover and missing data.

Point 3 on the south east coast displays some of the trends observed at both Points 1 and 2. It has a similar mean correlation to Point 2 (0.108) but a higher median value (0.980). The trends over distance seen at Point 1 can be seen again at Point 3, with a reasonably similar rate of decrease in all directions and almost no correlation with cells at the very north of the ice sheet. The exception to this is some evidence of the coastal trends observed at Point 2 too. Coastal cells south of Point 3 maintain relatively high correlation values towards the southern most tip of the ice sheet, over greater distances than can be seen by moving directly inland from the point. The coastal trend can be most clearly seen at cells with a latitude around 75. Inland cells closer to Point 3 have a noticeably lower correlation of missing data than those further away on the west coast.

We examine the distance and elevation trends further in Figure 30. For Point 1, there is a clear trend of lower correlation as separation distance increases. The trend with distance is particularly clear because the geophysical surface conditions do not drastically change as distance increases in any direction, removing confounding factors. The difference in elevation also shows clear trends with missing data
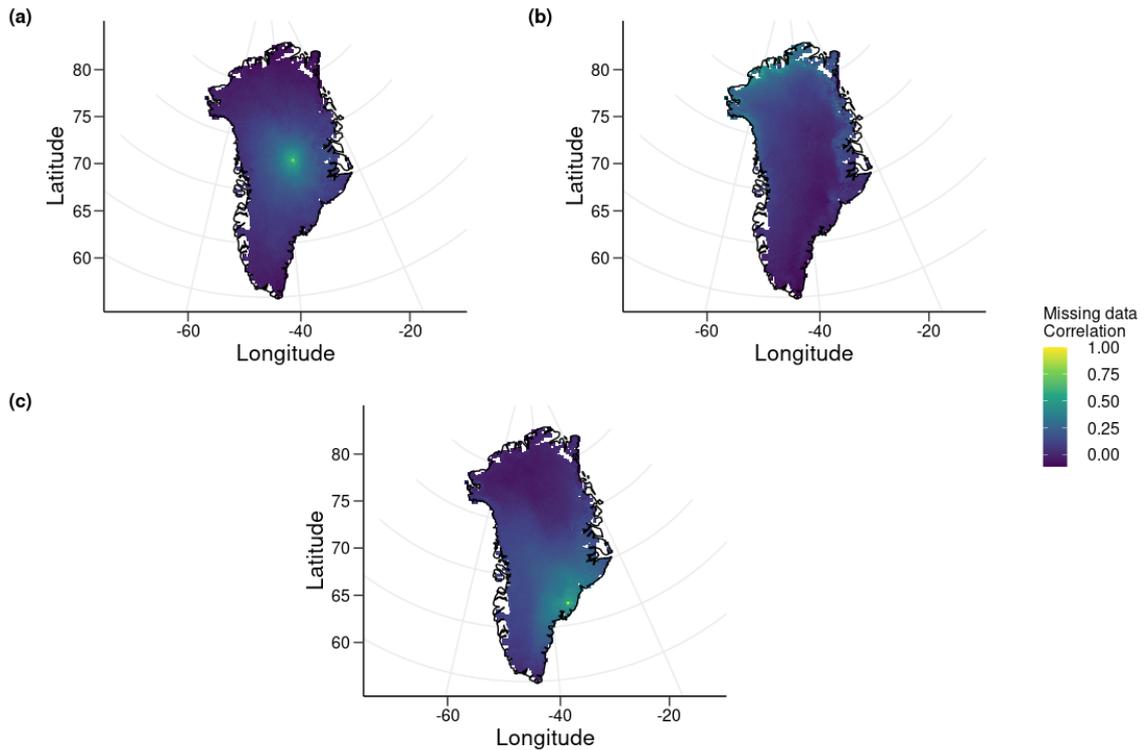
Figure 29: Correlation of missing data indicator variable at sample Points 1 (a), 2 (b), and 3 (c). Correlations are calculated between the three points defined above and all other locations in our sample.

correlation, as cells with a greater elevation difference in general have a lower correlation.

The trend between elevation difference and missing data correlation can be more clearly seen for Point 2. The decrease in correlation as distance increases is still evident although with more variability. Further, across the range of distances, cells at the same distance but with a greater difference in elevation have a lower correlation than those at a more similar elevation. The rate of decrease in correlation for low distances is similar to that of Point 1, as the points form a similar curve for distances below 1000km albeit with the extra variance for Point 2.

Point 3 confirms the trends between distance and correlation seen in the other two points. Again, there is a similar rate of decrease in correlation as distance increases, as well as a slight extension of this trend as some cells have slightly

81

negative correlations for distances between 1000 - 1800. The trends with elevation difference are less clear here than for Point 1 and Point 2. The trends seen spatially in Figure 29 are more nuanced for Point 3 than Point 1 or Point 2, which may make them more difficult to analyse from this perspective.
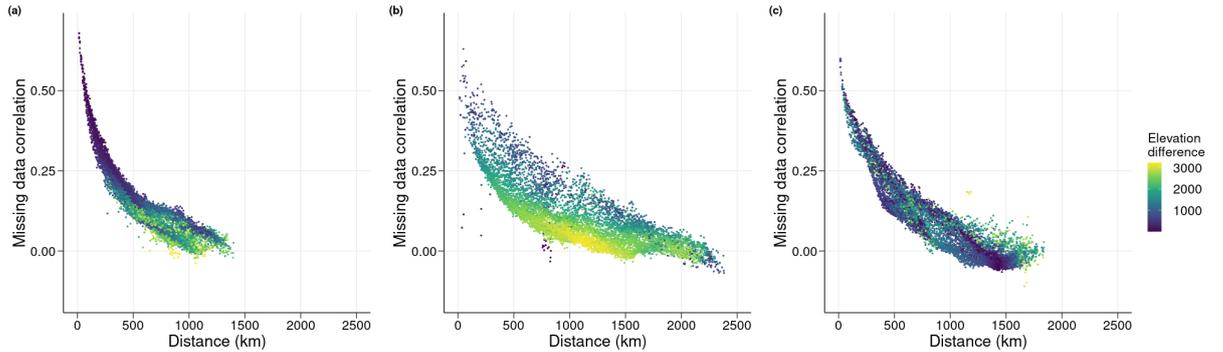

Figure 30: Correlation of missing data indicator variable at sample (a) Points 1 , (b) 2, and (c) 3 plotted against the distance between the point and the cell being compared. Colour represents the absolute difference in elevation between the two cells being compared.

From this analysis, we continue to limit our future modelling to only cloud-free days rather than to incorporate missing data into the model structure. There are clearly spatial trends present in the missing data and thereby in the cloud cover across the ice sheet, however our main aim is to explore and model melt and particularly spatial melt events rather than to model the weather system of the ice sheet. The missing data should not significantly impact our analysis of spatial extreme temperatures, and acts more as an indication of spatial correlation and weather trends than of the magnitude of temperatures observed in the data set.

4.2.3   IST correlation

With an improved understanding of the missing data patterns both on individual days and spatially, we now examine the correlation of the ISTs themselves. We use the same three example locations as in Section 4.2.2 as our comparison points. To compare the relative temperatures between cells, we transform the data onto standard normal margins (Section 4.3.1) and then calculate the correlation.

We compare relative temperatures instead of absolute temperatures because this removes differences in measurement scale between cells, and as a result is the scale on which the data will be used by subsequent spatial models. We also limit the correlation calculations to only days on which both cells have an observation to mitigate the effects of daily weather variability and yearly temperature trends.

From Figure 31, we see that Point 1 has reasonably high correlation with most of the ice sheet, with a mean correlation of 0.823 and a median of 0.828, both the highest of all three points. The correlation decreases as the distance from Point 1 increases, although the decrease is relatively slow. The only area that deviates from this trend is the south east coast. In the south east, there is a clear and more sudden decrease in correlation at cells close to the coast, more so than would be expected from the trends seen elsewhere. The south east coast does not appear to have a sharper elevation change compared to other coastal areas or distinctly different surface conditions. This may suggest that the climate in the south east is more distinct from the rest of the ice sheet.

Points 2 and 3 show reasonably similar overall trends, although Point 2 has a higher mean (0.781 compared to 0.729) and median (0.785 compared to 0.728) than Point 3. The correlation decays more rapidly for both points than for Point 1. This may be due to surface conditions being more varied close to the coasts than in the centre of the ice sheet. Both points also have low correlation with cells on the opposite coasts, with Point 2 having low correlation with the south east coast and Point 3 with the north west coast. However, both points have a low correlation with the south coast irrespective of their distances away from it, and the area at the southern most tip of the ice sheet shows low correlation with all three points. This suggests that the climate in this area may be different to the rest of the ice sheet.

In Figure 32, we also examine the relationship between IST correlation and the distance between the two cells being compared. Point 1 shows a clear trend with a close to linear association for the nearby cells and a higher variability in correlations for distances over 500km. The higher variability most likely reflects
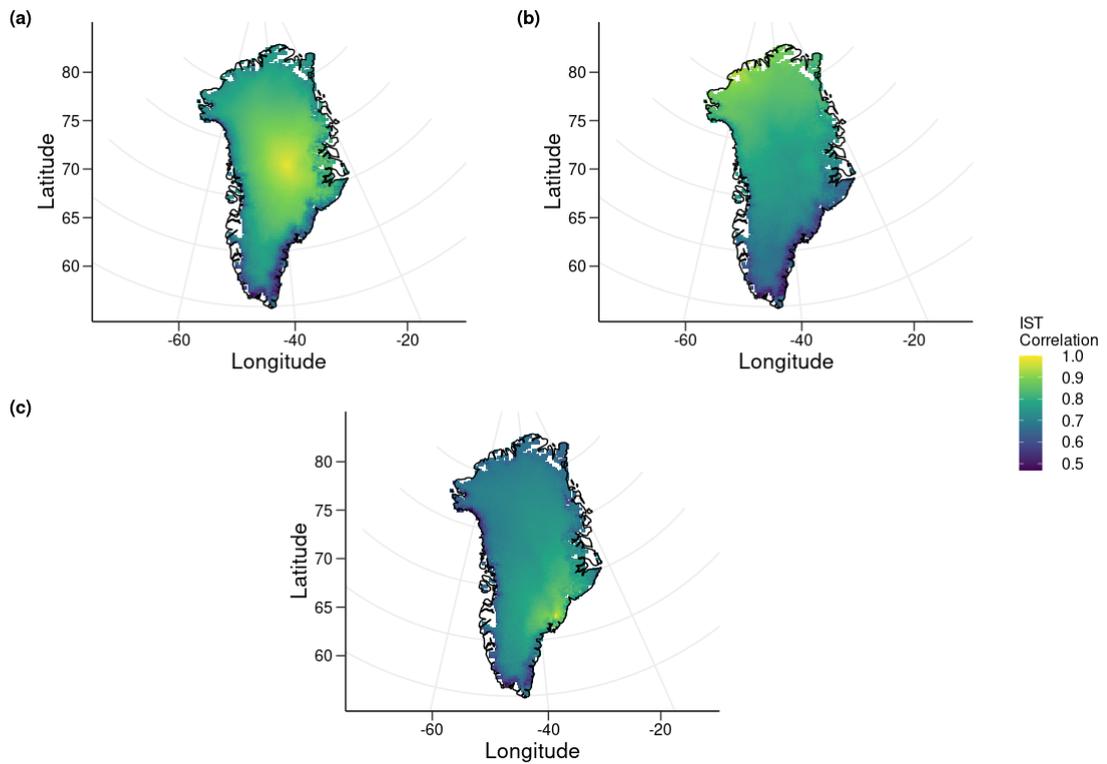
Figure 31: Empirical spatial correlation of transformed IST at sample (a) Points 1, (b) 2, and (c) 3.

increased differences in surface conditions across sites that are equidistant from Point 1 as the separation distance increases. As expected, we also see cells with considerably different elevations having a lower correlation.

Point 2 has a similar trend with distance, although the trend is more noisy. There is less variance in the correlations at distances below 1500km than for Point 1, but a similar overall range of correlations. The relationship between elevation difference and distance is also less clear. Conditional on separation distance, the cells with the lowest correlation also have a very low elevation distance, suggesting that other coastal or low elevation areas are the most distinct from Point 2. The only exception to this is for cells at a distance of less than 1000km, for which distance is a more important factor than elevation.

Point 3 may be expected to show similar trends to Point 2 given both are coastal, however the small difference in distance to the coast and subsequent difference in

elevation and locations on opposite sides of the ice sheet result in very different trends. Correlations here decrease at a faster rate than at either Point 1 or 2, suggesting that the conditions at Point 3 are most distinct from the surrounding cells. This may be due to proximity to the south east region. The elevation difference maintains a reasonably consistent trend as higher differences generally indicate a lower correlation, in agreement with the trends seen from Points 1 and 2.
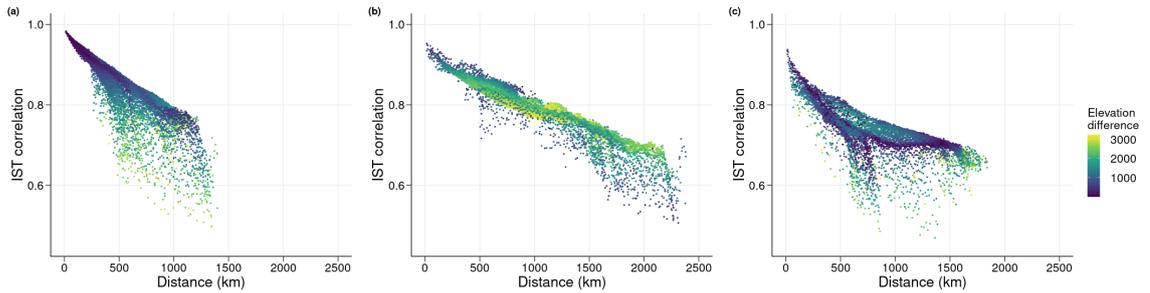

Figure 32: Correlation of IST at sample (a) Points 1, (b) 2, and (c) 3 plotted as a function of separation distance. Colour represents the absolute difference in elevation between the two cells being compared.

4.3 Methods

We now aim to quantify the spatial structure across the entire ice sheet. A suitable model for this is a spatial GP. This model captures spatial dependence between locations through a covariance function, which defines the variance of each location and its covariance with other locations as a function of distance. The model gives an interpretable structure that we can use to understand the spatial correlation of the data.

4.3.1 GP assumptions

The mixture models (Chapter 2) are used to transform the data onto standard normal margins before fitting the Gaussian Process. To undertake the transformation, we use the Probability Integral transform. For each cell, denote the marginal mixture model Cumulative Density Function (CDF) as $F_M()$, and transform the

data y such that our transformed data x_N is

$$x = \Phi^{-1}\left(F_M\left(y\right)\right)$$

where $\Phi()$ is the CDF of the standard normal distribution. This gives us data at each cell with a standard normal distribution. The reverse transformation is applied to simulations from the GP, so that we have

$$x_s = F_M^{-1}\left(\Phi\left(x\right)\right)$$

for our final simulated data x_s.

For the GP model, let $\underline{X}_t = (X(\underline{s}_1), \ldots, X(\underline{s}_t))_t$ be the d_t dimensional vector of observations at all sites on day t. Then,

$$\underline{X}_t \overset{IID}{\sim} \mathrm{MVN}_{d_t}(0, \Sigma)$$

where Σ is the covariance matrix with dimension d_t.

4.3.2 The GP kernel

When defining the form of the kernel used in the GP, it is important to consider how the dependence changes as the distance between cells increases. The exploratory analysis in Figure 31 shows a clear negative relationship with distance. However, the dependence also appears to be affected by specific areas of the ice sheet and other geographic covariates. Consequently, standard stationary kernels which define the correlation matrix Σ only by distance between cells are unlikely to fully capture the dependence structure.

One possible solution to this problem would be to consider non-stationary kernels (Heinonen et al., 2016; Pandita et al., 2021). This would enable the GP to capture more complicated dependence structure at the expense of greater computational cost. Our preferred approach is to include additional variables e.g. elevation in a

stationary kernel to determine separation data. The amount of relevant and useful additional data - as seen from Section 3 - suggests that alternative distance measures could be of use for dependence without having to incorporate non-stationarity. This maintains a simple model structure while allowing greater expression from the GP.

To include other measures of distance and other continuous variables in the kernel, we include each variable in its own kernel function then take the product of these functions for the overall function. We choose a rational quadratic kernel for all distance variables. This kernel balances flexibility in the shape of dependence function with a low number of parameters for relatively easy fitting. Recalling that we assume the variance of the GP to be 1, we define the kernel for a continuous variable x as:

$$k(x, x') = \left(1 + \frac{\|x - x'\|^2}{2\alpha l^2}\right)^{-\alpha},$$

for parameters l and $\alpha > 0$ and for all combinations of cells x, x'. Here $\|x - x'\|$ represents the Euclidean distance between two cells in the case of latitude/longitude distance being used as the kernel's distance measure, or the absolute difference between values in the case of other continuous variables.

We can extend our kernel definition to include discrete categorical variables. For these variables, we aim for each level of the variable to have a separate set of model parameters to reflect their differing dependence structures. We calculate between and within group kernels separately to form the overall covariance matrix, such that for the covariance of points in group R we have:

$$k_R(X, X', c, c') = \left(1 + \frac{\|x - x'\|^2}{2\alpha_{1R} l_{1R}^2}\right)^{-\alpha_{1R}} \left(1 + \frac{|c - c'|^2}{2\alpha_{2R} l_{2R}^2}\right)^{-\alpha_{2R}}$$

where $\|x - x'\|$ and $|c - c'|$ are the separation distance and the difference in distance to the coast between all combinations of cells in groups R, with parameters α_{1R}, l_{R1} associated with distance and α_{2R}, l_{2R} associated with distance to the coast. We

define group R as

$$
R = \begin{cases} 1 & \text{if both cells in Group 1} \\ 2 & \text{if both cells in Group 2} \\ 0 & \text{if one cell in Group 1 and one in Group 2} \end{cases}
$$

This kernel is calculated for Group 1 with itself, Group 2 with itself and between Group 1 and 2 for a total of 12 parameters. The overall covariance matrix for cells in Group 1 (x_1) and in Group 2 (x_2) is defined as:

$$
\Sigma = \begin{pmatrix} k_1(x_1, x_1) & k_0(x_1, x_2) \\ k_0(x_2, x_1) & k_2(x_2, x_2) \end{pmatrix}
$$

This kernel is still stationary as distance to the coast differences are considered without consideration of the relative coastal distances individually, and allows both features to be considered for the dependence structure of the model.

The Gaussian process models are fitted using maximum likelihood estimation. For data $\mathbf{x}$ from n days with k_i cells of data available at time point i and observations $\mathbf{x_i} = (x_1, \ldots, x_{k_i})$ for day i, we have

$$
L(\mathbf{x}) = \prod_{i=1}^{n} \frac{\exp\left(-\frac{1}{2}\left(\mathbf{x_i}^T \Sigma_i^{-1} \mathbf{x_i}\right)\right)}{\sqrt{(2\pi)^{k_i} |\Sigma_i|}},
$$

where $|\Sigma_i|$ is the matrix determinant of the covariance matrix Σ_i.

The other variables used to define distance were chosen based on previous results: the absolute difference in latitude, elevation, and distance to the coast. These variables are not completely independent, but previous work has showed that each variable can add extra information to the model even if they are collinear. Previous work also demonstrated that considering interactions between spatial variables can result in a model that better represents the data, so rather than prioritise any single measure, we consider them all both alone and in combinations with each other.
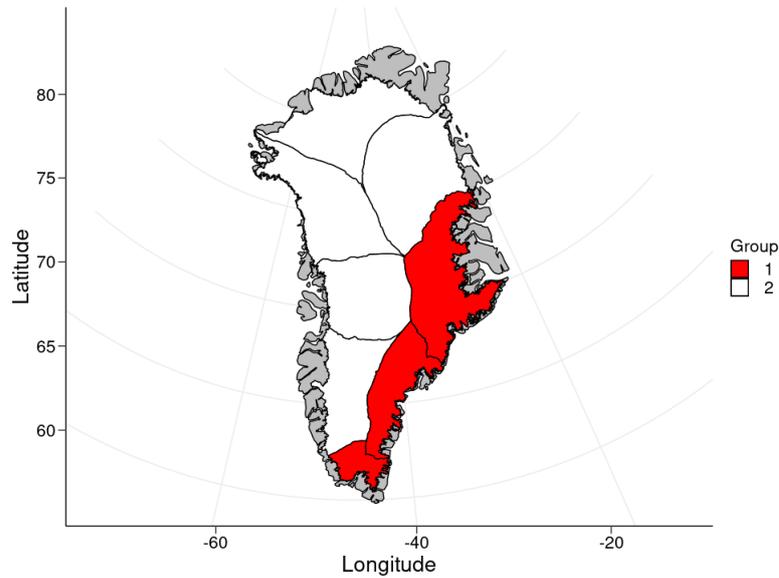
Figure 33: Drainage basins as derived from ICESat split into two groups for the Gaussian process. Groups were chosen based on the different characteristics seen in the south east from previous modelling and analysis.

An additional variable that is considered for use in the kernel that is not continuous is the drainage basin that each cell is located in. Rather than a separate kernel for all eight drainage basins we split the ice sheet into two regions that show different climate properties. This simplifies incorporating the variable in the kernel and allows us to target any area that shows distinctly different trends to the rest of the ice sheet. From exploratory analysis and particularly Figure 31, the south east coast of the ice sheet appears to have the most distinct trends seen on the ice sheet. ISTs there do not appear to correlate with other areas of the ice sheet such as the centre as much as other areas. In both plots 31 (a) and (b), the south east has a clear drop in correlation that is not seen for cells also a similar distance away from the comparison point. To capture this area from the drainage basin data, we group three drainage basins together to cover this area, as can be seen in Figure 33. This gives the Gaussian process the ability to express dependence in terms of the spatial variables differently for the two defined areas.

For the sample cells, we tested fits of the Gaussian process using combinations

89

of the spatial variables in the covariance matrix. Each covariance matrix consisted of the product of each variable's individual kernel as previously discussed. Distance defined by latitude and longitude was included in all covariance matrices as the trends between distance and IST correlation were clearest and it provides a baseline description of the data's dependence that can potentially be improved by adding additional variables. The variations tested were distance with elevation, latitude, or distance to the coast; the same models with the data grouped by drainage basins; and distance with two of elevation, latitude, and distance to the coast. We compared the negative log-likelihood values of each of the fitted models to determine which of the models had the best fit. We first assessed the grouping and additional parameter fits to see whether, in general, additional parameters had a clear impact on the model fit compared to the fits with fewer parameters. Having shown noticeable improvement with additional parameters, models were compared to other models with the same number of parameters. On this basis, the model with the best fit used a kernel with the groups of drainage basins, distance, and the absolute difference in distance to the coast.

4.4   Results

Due to the computational cost of the Gaussian Process model compared to the independent mixture models, we reduce the number of cells that we fit the model to. We change the sampling resolution of cells in both x and y dimensions from the regression modelling in Section 3 from one in 20 (7156 cells) to one in 120 cells (206 cells). All kernels and models discussed in the rest of this section use this same set of sample cells.

As a starting point, we fit a model in which the covariance of the data is a function of only the distance between cells. Figure 34 shows how under this model the covariance between locations decreases as the distance between them increases. The form of the relationship most closely resembles the empirical correlation of Point 3, which may be expected as this point has a more typical elevation and therefore

typical distribution and is therefore more likely to be representative of the majority dependence structure. The lowest covariance value of just over 0.7 for cells 2500 km apart also agrees with the exploratory data analysis on three case study sample points, as the mean of the correlations at the maximum distance from each point also appears to be around this value.
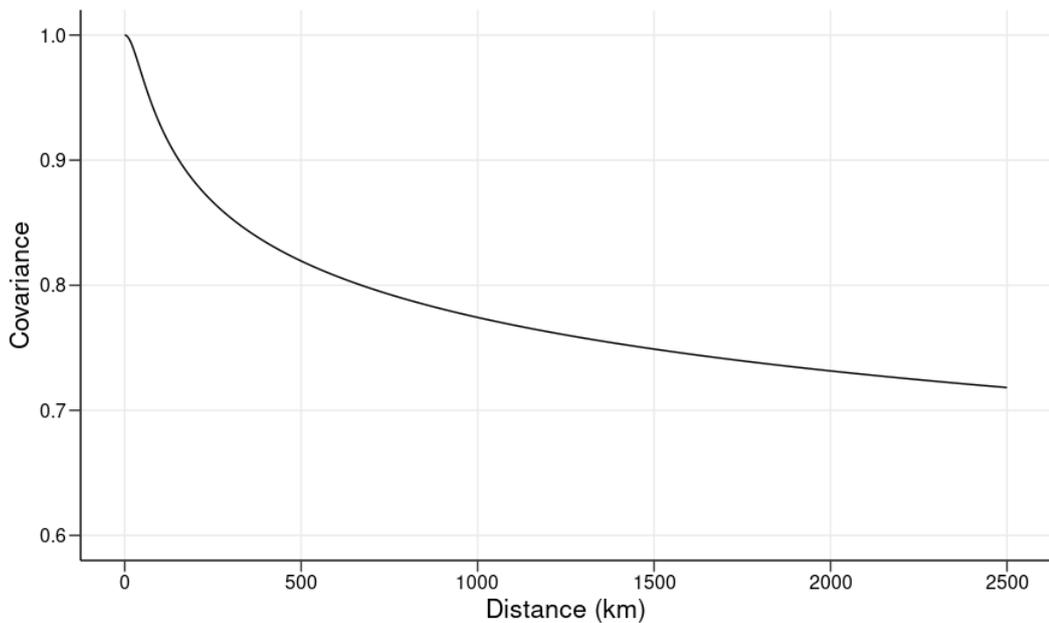

Figure 34: Fitted kernel of the Euclidean distance-only Gaussian process.

For the main Gaussian process model we include grouping, distance and distance to the coast in the covariance matrix. We first examine the fitted covariance matrix evaluated at the sample cells. Since the model is fitted to a different sample of cells to those used in the previous empirical correlation analysis, we select four new comparison points: the cells with the highest and lowest distances to the coast, a cell in the east of drainage basin Group 1 of the drainage basin groups, and a cell in the south of the ice sheet.

Figure 35 shows a map of the covariance function for the four comparison points. Point 1 shows most clearly the impact of grouping by drainage basins: with a difference in the dependence model-bases estimates between the two groups as previously suggested by empirical IST correlations. The distinction between distance

and the change in distance to the coast can clearly be seen by considering the line of latitude through the centre of the ice sheet. Cells in the south central areas have a higher covariance than cells on the northern coasts despite being further away. This demonstrates the model's functionality - covariance still decreases as both distance and the difference in distance to the coast increases, but the rate of decrease changes depending on how similar both of these values are. Cells a similar distance away from the coast have a higher covariance than would otherwise be possible if only the distance between them was considered.

As the closest cell to the coast, point 2 has higher covariance with other cells on the coast than with inland cells at similar distances. The most noticeable impact is on the cells on the west coast. Despite large distances between these cells and point 2, their covariance is still relatively high given the distance between them. This could suggest that the similarity in relation to the coast has a greater impact on the similarity of temperatures than the distance between cells. The change in covariance between the two drainage basin groups is less apparent here than for point 1, although there is on average still lower covariance with points in group 1 than in group 2.

Point 3 is situated in group 1 but is on the border between the two groups. Despite this it shows even less distinction between the two groups than either of the previous points. It has a reasonably smooth covariance across the ice sheet, aside from two notable areas. But, it has a lower covariance with cells in the centre of the ice sheet that are the furthest distance away from the coast. This again shows the strong effect of distance to the coast, and thereby also on the similarity of temperatures for cells the same distance from the coast, even if they are very far apart. The other area of interest is on the south east coast, where covariances are noticeably lower than other regions a similar distance away from the reference point. These cells are in the same drainage basin group, acting as an illustration of the different relationship the two groups have with distance. The cells on the south coast are a similar distance to the coast as point 3, suggesting that the covariance

is lower due to the effect of distance instead and that group 1 is more impacted by distance than group 2.

The importance of distance for group 1 is even clearer at the final point. Point 4 is again located in group 1, but is located closer to the south end of the group, meaning that larger distances between cells in the same group are available. The overall covariance trends for cells in group 2 are largely the same as point 3, however the decrease in covariance with cells in group 1 can be clearly seen along the east coast. These cells show comparably low covariances to the cells in the centre of the ice sheet, again showing the different dependence behaviour of cells in group 1. Finally, point 4 is also the only point with relatively high covariance with the cells at the very south of the ice sheet, most likely due to their proximity.

4.4.1 Simulations

We next examine simulations from the model to learn more about both melt and general temperature trends. The crucial difference between the mixture model and Gaussian process is that we can simulate from all cells simultaneously rather than making inference only at individual cells, allowing us to make a fully spatial inference.

We simulate 1,000,000 realisations of the Gaussian process using the fitted parameters. For each simulation and each cell we calculate the probability of occurrence of a melt temperature using the marginal mixture model, as in Section 2.4.1. This allows us to make inference with regards to both absolute temperatures and their associated melt probabilities.

Figure 36 shows the expected melt estimated from the GP-based melt probabilities compared to estimates from the marginal models only. When compared to the mixture model, the Gaussian process captures the overall melt characteristics of the ice sheet well. Expected melt in the centre of the ice sheet is marginally higher under the Gaussian process model than under the mixture model. This result occurs because the Gaussian process gives more central cells a non-zero probability
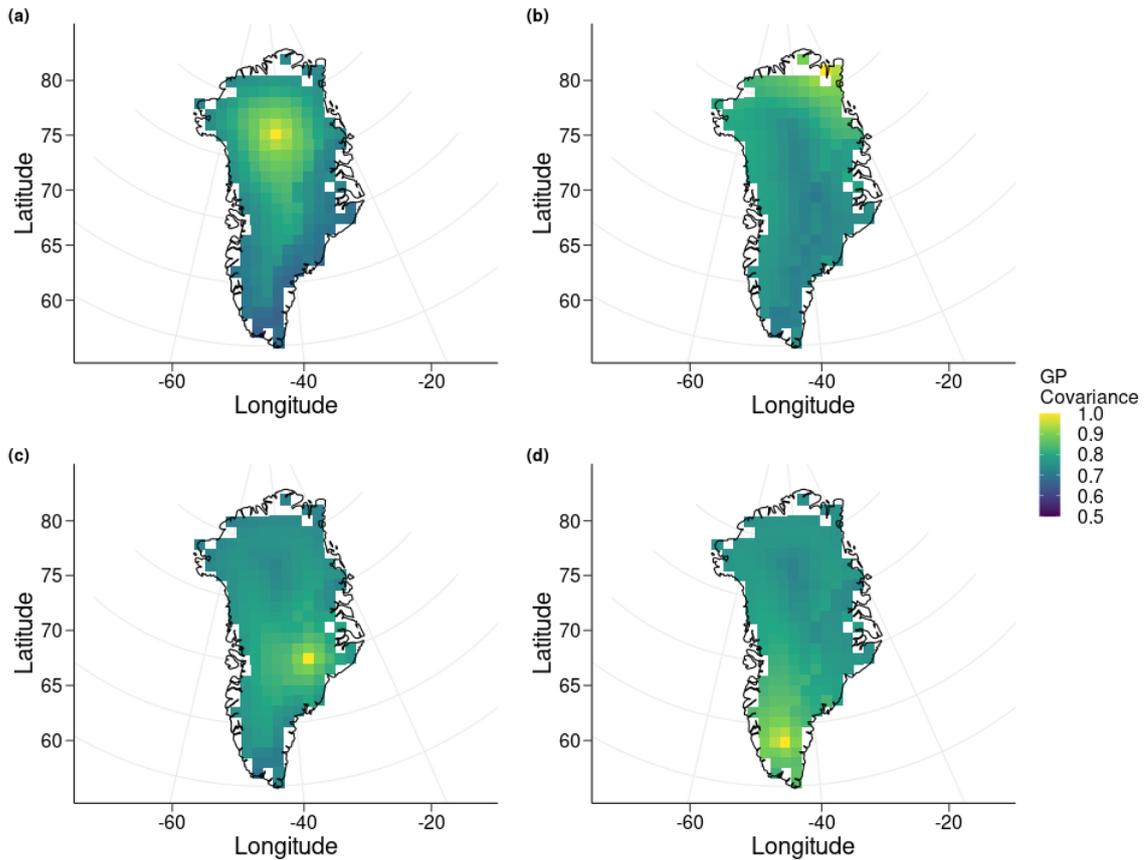
Figure 35: Covariance matrix of the fitted Gaussian process with drainage basin groups, distance, and distance to the coast included in the covariance function evaluated. The four comparison points used to calculate the covariances can be seen in yellow in each plot.

of melt than is observed in the data. This may reflect the difference in sample sizes between the observed and simulated data. Most of the coastal cells show similar melt estimates to the mixture model, and importantly the cells in group 1 are consistently close to the marginal model estimates. This provides further confidence in the model structure and that there is indeed different behaviour in the different regions of the ice sheet.

We can look in more detail at the melt estimates from the Gaussian process by examining the mean probability of high temperatures and specifically melt temperatures. In Figure 37, we see that the probability of an IST exceeding 5° C follows some of the broad temperature trends previously seen of temperatures being
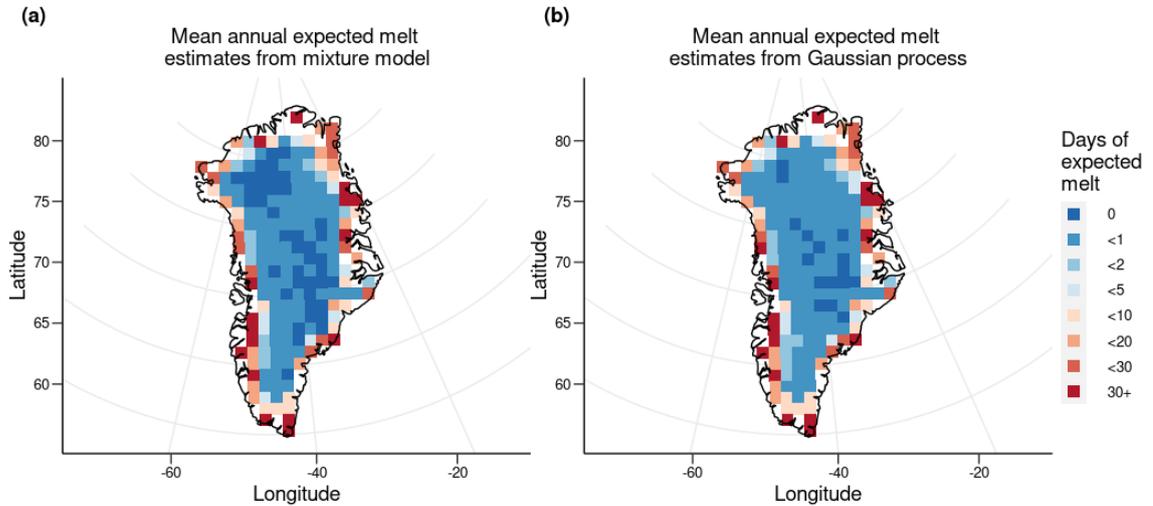
Figure 36: Expected melt estimates per year for (a) the mixture model and (b) the Gaussian process.

higher near the coast. However, the probabilities extend slightly further towards the centre of the ice sheet than the expected melt estimates, particularly at lower latitudes. The coastal estimates are also relatively high when considering the context of daily probabilities, with a maximum estimate of 0.435 in the north east of the ice sheet. When extended to a year, this represents a large number of temperatures close to the melt threshold for a typical summer.

The daily melt probabilities are higher only for cells on the very edge of the coastline. The highest probabilities are found on the south, south west and north east coasts, whereas the centre of the ice sheet has very low values, several orders of magnitude lower than the coasts. The highest value of 0.256 is on the south west coast, demonstrating that having a high probability of high temperatures does not necessarily result in a correspondingly high melt probability. These trends closely resemble those found in Section 2.4 from the independent mixture models. The main difference is the lower maximum melt probability. This may be a coverage issue, due to the larger grid in the previous section including more coastal cells for which the higher probability of melt is higher.

In addition to the pointwise estimates of melt taken from the spatial simulations,
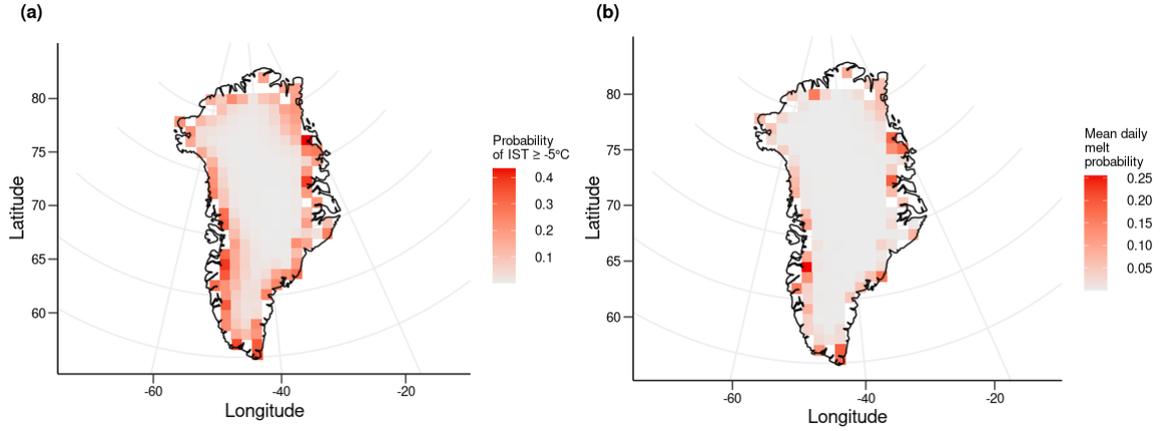
Figure 37: Mean daily probability of (a) an IST exceeding 5°C and (b) the mean daily probability of a cell experiencing melt estimated from the Gaussian process simulations.

we can also assess the simulations spatially by examining simultaneous melt at different cells on the same day/simulation. To do this, we calculate the expected melt proportion for each simulation, then calculate quantiles of increasing probability to estimate the return levels as:

$$z_p = \frac{1}{1 - F(p)},$$

where z_p is the return level with return period $1/p$. That is, the return period is the expected time until the return level is exceeded. By calculating a range of return periods, we can examine how rare melt events of a certain size are. Figure 38 shows the return levels for return periods of up to 100 years from the Gaussian process simulations. Melt events that have an expected melt proportion of at least 0.2 have a return period of 0.44 years, and for an expected melt proportion of 0.25 the return period is 1.54 years, but above this the return periods begin to increase at a much quicker rate. For an expected melt proportion of at least 0.4 the return period is 18.9 years, and for 0.5 it is 75.4 years. As the return period increases the melt proportion increases at a slower rate, suggesting that there is potentially a finite end point for expected melt proportion. However, given the number of samples from the Gaussian process and the rarity of events for high return periods, we do not extrapolate past
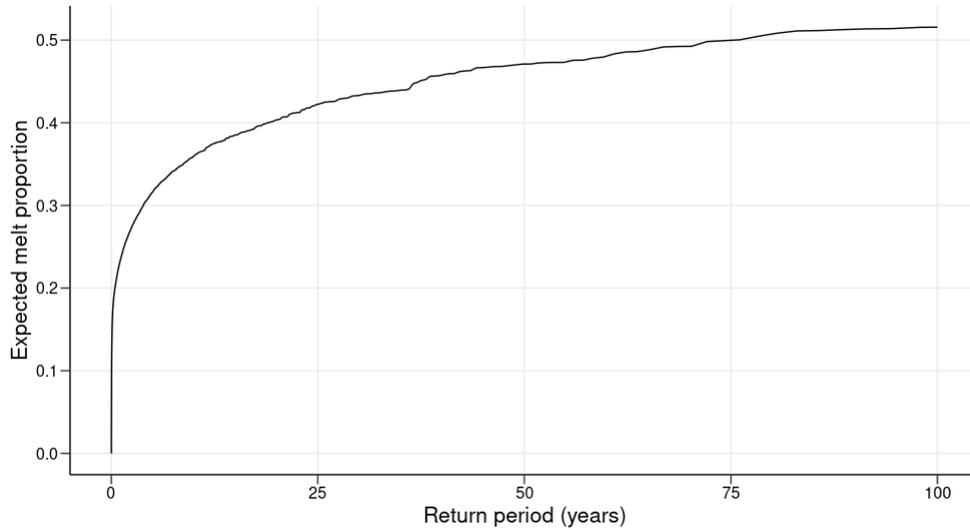
the 100 year return level.

Figure 38: Estimated return periods of the expected daily melt of the sample cells.

4.5 Discussion

In this chapter we have extended our models of ISTs on the Greenland ice sheet to quantify spatial dependence between cells using a Gaussian process. The model retains the accurate representation of the marginal mixture models while providing extra functionality for inference. By using a spatial model, we can examine the spatial dependence between cells and identify the covariates - and spatial groups - that have the strongest relationship with ISTs. The dependence function is informed by previous spatial examinations of the data using similar covariates that were previously seen to have an association with ISTs. The model allows us to simulate all cells simultaneously to give simulated days of ISTs rather than separate independent models and simulations. The simulations can then be used to make spatial inference including return periods for the amount of melt observed.

The main difference between the Gaussian process and the previous regression and independent mixture models is the inclusion of spatial structure. In previous sections we have examined spatial trends either from viewing independent measures

97

spatially or from numerical relationships between spatial covariates and model parameters. Using the Gaussian process, we can directly examine the dependence between cells and give form to the relationships via the covariance function. This results in more precise definitions of the spatial relationships seen in the data than can be obtained from the previous models.

It also allows for inference of simultaneous observations as opposed to independent observations. The mixture models on their own demonstrate the behaviour of ISTs at a particular cell, but even if the summary statistics or parameter values have spatial trends, we can't make inference about the behaviour of multiple cells at a single time point. The Gaussian process considers all observations in each time point in its likelihood rather than observations at independent cells separately. The model treats all cells as part of a single structure, both for its simulations and the covariance function. The joint consideration of all cells allows characteristics of individual time points to be made that aren't possible from independent single cell models, such as the proportion of the ice sheet that simultaneously melts.

Despite the best fitting model featuring distance, difference in distance to the coast and the drainage basin grouping, the previous models and the exploratory analysis show that there are several covariates that have strong relationships with IST. One of the main aims of each model we use is parsimony, and this is evident from our approach towards covariates. The covariates used alongside IST are not independent in almost all cases due to the geographic characteristics of the ice sheets. For example, cells at a low elevation are also very likely to be close to the coast because of the overall shape of the ice sheet. Because of the links between the covariates, the effect of adding each subsequent covariate is lower even though each individual covariate has strong correlation with IST. In this sense, we aim to focus on the covariates that have the strongest relationships with IST in a simpler model structure, rather than use every available covariate for only a small gain in model fit. This gives us a simpler model structure for fitting purposes, simplifies inference into the model structure, and helps to distinguish between the effects of different

covariates. The exception to this approach is the regression modelling. This is due to the fact that including interaction terms makes the effects of each covariate harder to assess, so the addition of further terms does not cost anything in interpretation.

In addition to using the Gaussian process for the fitted cells, the covariance structure can be extended to cells outside of the sample. The covariance function has the same structure regardless of where cells are, so as long as new cells are on the ice sheet we can evaluate the covariance function between them and the cells used in the model. As in the regression modelling, we can similarly derive the distance to the coast and drainage basin using only the latitude longitude coordinates of a new cell, and can estimate its marginal distribution from the regression models themselves. With a marginal distribution and appropriate distance measures/drainage basin information, new cells can then be simulated for as part of the Gaussian process. The main challenge of doing this is assessing whether the new cells are well represented within the sample of cells used to fit the model. If the new cells are very close to the coast - typically where models so far have had the most difficulty representing - then extending the model to these cells may be less accurate than cells closer to the centre of the ice sheet. However, using the model this way does allow for a larger size application of the model without having to fit the model to more cells. This saves computational time of the fitting process, potentially with the cost of lower confidence in the simulated values.

5 Spatial extreme value analysis

5.1 Introduction

Having modelled ISTs spatially using the Gaussian process models, we now aim to model extreme spatial temperature events and thereby spatial melt events. These often large scale events can cause melt across high proportions of the ice sheet, as clearly seen in 2012 (Nghiem et al., 2012). The models used so far for ISTs have focused on modelling the whole distribution of temperatures either at a single site or across the ice sheet. While we can make inference with regards to melt from both of these approaches, spatial melt events have not been specifically considered in any model structures yet. In order to capture these events most accurately, a more accurate approach is to model just the extreme values so that the entirety of the model fit is based upon the values of highest interest.

Our inference of spatial melt events has so far been focused on simulations from the Gaussian process models. By simulating many instances of the Gaussian process, we can obtain a catalogue of melt events. Consequently, our understanding is based upon a small proportion of simulations rather than from the model structure itself. Instead of subsetting simulations from a general model, we now model only extreme spatial events. This could be achieved by putting restrictions on the Gaussian process simulations, but a model specifically designed to represent and simulate only extreme events is a more appropriate tool for this task.

In this section we motivate the case for modelling extreme temperature events spatially. We examine the distributional form and dependence of the higher temperatures in the data set to inform choices regarding model structure. Using this insight, we modify and apply the Spatial Conditional extremes model (Wadsworth and Tawn, 2018) to MODIS IST data. The model allows us to examine spatial extreme temperature events by conditioning on a single cell having an extreme value. This can provide insight into extreme temperature behaviour at a single cell, or across the ice sheet using an importance sampling framework. This allows us

to describe spatial extreme temperature events across the ice sheet and link these temperatures to melt using marginal melt definitions, as previously seen from the Gaussian mixture models.

5.2   Exploratory analysis

We first aim to understand the marginal distribution of extreme temperatures at individual cells. For the Gaussian process models, we transformed data at each cell onto standard normal margins before modelling the spatial structure. We use a similar approach for the spatial extremes model to examine the relative magnitude of temperatures rather than their absolute values, transforming onto standard Laplace margins rather than standard normal margins. This was made possible in the Gaussian processes due to the accuracy of the marginal mixture models, which made the assumptions of normality at each cell accurate after transformation. Although this same assumption can be made when transforming to standard Laplace margins, this only gives us confidence in the form of the entire marginal distribution and doesn't describe the distribution of extreme temperatures alone.

5.2.1   Marginal distributions of exceedances

For our exploratory work, we use the same subset of 206 cells used in Section 4.3.1, and the same mixture model fits are used to transform the data at each cell to standard Laplace margins. In order to compare extreme temperatures, we first choose how to define an extreme temperature. Typical extreme value analyses choose a quantile based threshold that is specific to each location/cell in the data. This requires a balance between a high enough threshold to satisfy asymptotic assumptions, while also ensuring that there is sufficient data at each cell to get a representative model fit. Since the data has been transformed prior to our examination of the distribution of the tail, we examine a range of thresholds rather than immediately choose one. This gives us a clearer understanding of the tail behaviour of the data before we define a marginal model for the extreme

temperatures. For this purpose, we will examine exceedances at four commonly used quantile levels - 0.9, 0.95, 0.975, and 0.99.

The compromise between the amount of data and asymptotic results is demonstrated by the empirical densities in Figure 39. The densities have more consistent forms for the lower quantile thresholds, due at least in part to the higher amount of exceedances being examined. This results in a wider variety of distributional shapes for the 0.99 quantile compared with the other quantile levels. The difference between the upper tails is of importance for modelling considerations, with the higher quantiles having less weight in the tails and more weight closer to 0. This isn't unexpected given that a higher threshold will naturally make exceedances less frequent, but does illustrate the impact that threshold choice has on the distributional shape of exceedances. We choose to use the 0.95 quantile in further exploratory analysis.

Considering our previous definition of melt from the mixture models, a threshold based on absolute temperatures could be considered rather than based on a particular quantile. For example, the extreme value threshold could be set at $0°$ C or $-1.65°$ C to align with the values associated with melt by the mixture model. The benefits of this are that exceedances would have a direct interpretation in terms of melt, making the model capable of modelling specifically melt events rather than just extreme temperature events. However, the cost of this would be the exclusion of all data for certain cells that do not have many or any values that exceed these thresholds. This would make modelling temperatures at the centre of the ice sheet difficult despite the benefits to interpretation. We continue with a quantile based threshold for the current model setup, although this may become a more important consideration when melt is examined by the model in more detail.

5.2.2 Co-occurring extremes

One of the key considerations of the Gaussian process models was that observations considered for dependence had to occur on the same day. Having established
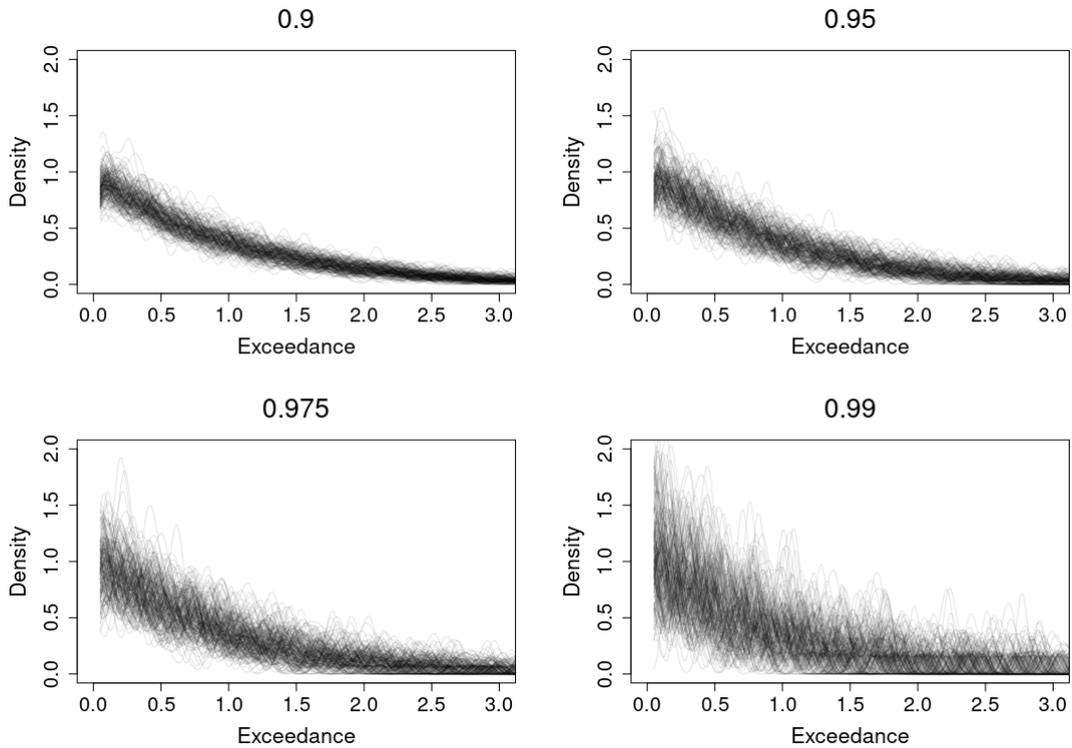
Figure 39: Empirical density curves of exceedance data at each cell using exceedance thresholds of 0.9, 0.95, 0.975, and 0.99 quantiles. Data are on standard normal margins.

the form of the marginal densities of threshold exceedances, we now examine the distribution of exceedances at the same time points. Comparing the exceedances at different cells at the same time point shows how exceedances co-occur and gives insight into the dependence structure of the exceedances in a way that is consistent with how the exceedances will be considered in the spatial extremes model.

Figure 40 shows that most of the days with an exceedances at at least one site had a low proportion of cells with exceedances simultaneously. On average, only 14.44% of cells have an exceedance simultaneously, with several days only having a single cell with an exceedance. The low frequency of extreme temperatures occurring on the same days suggests that the temperature events occur locally when considered across the whole ice sheet. If larger events were more common, then it would suggest that modelling the entire ice sheet would be the best approach as the larger scale

would be required to fully capture the events. However, because on average only a small number of extreme temperatures occur on the same days, examination of events on a finer spatial scale may be helpful.
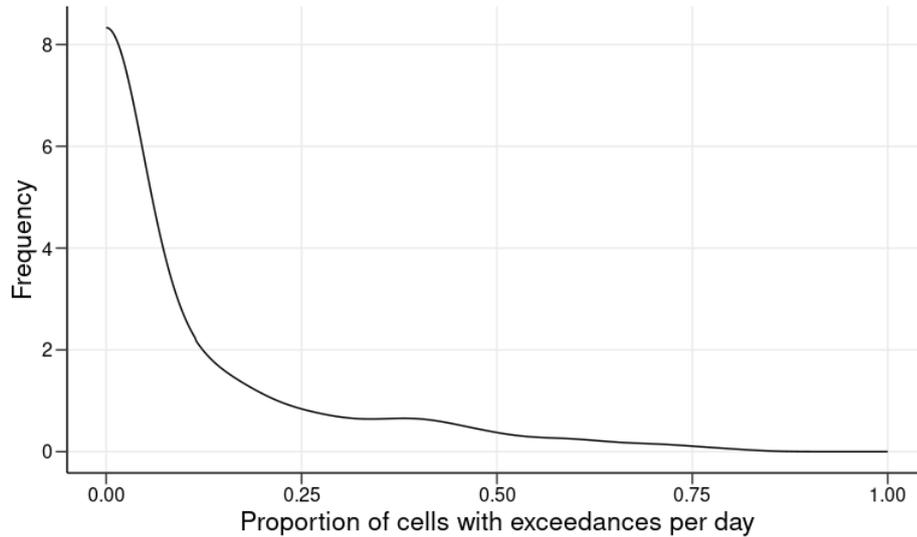

Figure 40: Density plot of the proportion of cells that have an exceedance for each day with at least one exceedance.

5.2.3 Extremal dependence

Intrinsically linked to the spatial scale of events is the distance between the locations at which extremes co-occur. We have observed from the Gaussian process models that distance and other spatial variables impact the correlation between locations, and we aim to do the same for extreme temperatures. In the previous model framework, dependence was observed and calculated between all temperatures at two different locations. The key difference when examining extreme temperatures is that if a temperature is extreme at one location, it does not imply that the temperatures at each other location will also be extreme or even that they will be high. It is this behaviour that we capture with the spatial extremes model. That is, conditional on observing an extreme temperature at one location, what do temperatures at other locations look like, and how does this change with increasing
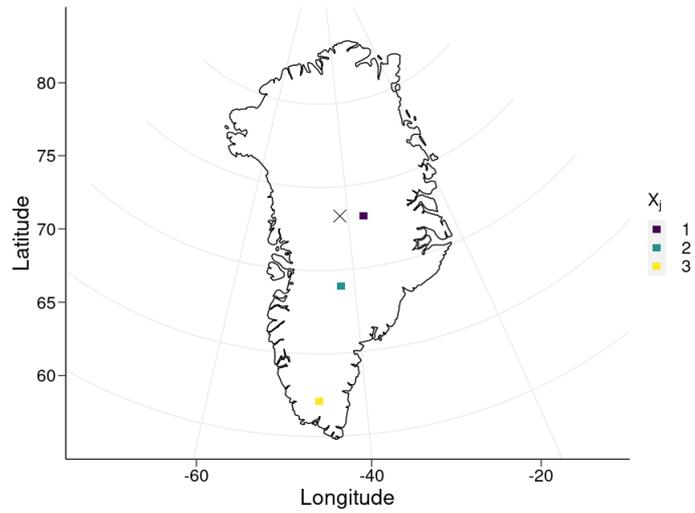
104

Figure 41: Locations of the (cross) conditioning cell c and (squares) other cells (left) 1, (middle) 2 and (right) 3.

distance?

To motivate the conditional extremes approach, we first illustrate the concept of conditional extremes. For data from 2001 to 2019, we choose a single cell as our conditioning cell c at which we condition on observing an extreme temperature then examine the distribution of X_c compared to the distribution of temperatures X_j at other cells j that are located at increasing distances from our conditioning site. All data used is transformed onto standard Laplace margins using the previously fitted mixture models, and the locations of the conditioning cell and other cells can be seen in Figure 41.

For a conditioning cell, in the centre of the ice sheet, we observe the two types of extremal behaviour present in the data. Figure 42 shows that for site 1 which is closest to the conditioning cell - around 157 km away - there is clearly some relationship with site c. For temperature exceedances at site c above our threshold μ_c, there is a positive relationship with temperature exceedances at site 1. Even for the highest observations from both cells, this trend still holds as can be seen from the points in the upper right corner of the scatter plot. In addition, there appears to be a change in variance of temperatures at site 1 for a given exceedance
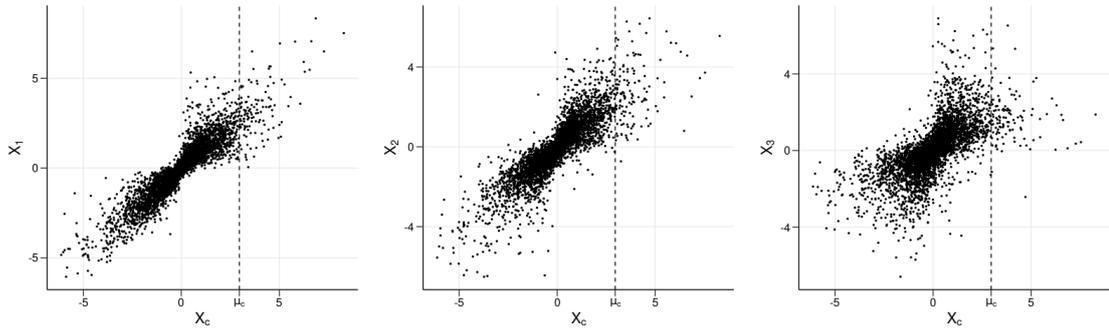
105

Figure 42: Scatter plots of data on Laplace margins from comparison cells (left) 1, (middle) 2 and (right) 3 plotted against conditioning cell c.

at site c. From this, for a known exceedance at site c, we can learn about the overall distribution of temperatures at site 1. This is an example of two sites which are asymptotically dependent.

Sites 2 and 3 are situated around 469 km and 1247 km away from the conditioning cell respectively. For site 2, there is still some dependence after conditioning on site c, however whilst the trends in the mean and variance are weaker than at site 1, knowing the temperature at the conditioning site still provides some information regarding the distribution at site 2, so there is still asymptotic dependence. However, at site 3 the trends are weaker still. At this distance, there appears to be a much weaker relationship between site 3 and the conditioning cell. For a given exceedance at site c, temperatures at site 3 appear to have a fixed mean rather than having higher average values as the exceedance gets larger. The variance or overall distribution of the temperatures does still appear to change according to the value of the exceedance, but the overall distribution of temperatures at site 3 is different than to sites 1 and 2. Site 3 shows the initially assumed overall trend between distance and dependence, that as the cells get further apart there is less dependence between exceedances.

106

5.3 Methods

We apply two types of extreme value models to the data - the bivariate conditional extremes model and the spatial conditional extremes model. We apply the bivariate model pairwise to the data first to learn about the pairwise extremal dependence properties of the data. This gives an intuition into the extremal properties of the data, as well as being useful in informing the structure of the spatial model.

5.3.1 Bivariate conditional extremes model

The bivariate conditional extremes model (Heffernan and Tawn, 2004; Keef et al., 2009)) allows us to examine the dependence of data from one cell conditional on an exceedance being observed at another cell. Given a conditioning cell c and other cells $j \in \mathcal{S}$, we consider the distribution of $X_j \mid X_c = x_c$ where $x_c > \mu_c$, such that:

$$X_j \mid X_c = x_c \sim N(\alpha x_c + \mu x_c^\beta, \sigma x_c^\beta), \qquad (2)$$

where $\beta > 0$, and so μ and σ are the mean and variance of residuals

$$Z = \frac{x_j - \alpha x_c}{x_c^\beta}$$

that are assumed to be normally distributed. Data at each cell are assumed to be on standard Laplace margins. If α and β are close to 0, then the observations at x_j are a normal sample independent of the exceedances at x_c with mean μ and variance σ. If either α or β is non-zero, then the observations at x_j are related to the observations at x_c as described by Equation 2.

Our inference focuses on the estimates of these parameters and how these estimates vary with distance between the two cells being modelled. These parameter estimates can be used to examine the range of dependence across the ice sheet, the conditional distribution after conditioning on a given value at the conditioning cell, and the strength of dependence between cells. Each of these is not only useful in

its own right, but will also help to inform the structure of the spatial conditional extremes model.

5.3.2   Spatial conditional extremes model

The spatial conditional extremes model gives us a framework upon which to define spatial extreme events by ensuring that at least one cell in our sample has an extreme observation. We examine the dependence between extreme observations at the conditioning cell and observations at other cells on the same day, which may or may not be extreme themselves. That is, for a spatial process $X(s)$ evaluated at cells s in sample $\mathcal{S}$, with conditioning cell s_0 and threshold μ, we consider the distribution of:

$$X(s) \mid X(s_0) > \mu : s \in \mathcal{S}.$$

This broadens the pairwise model by incorporating all cells in the dependence functions.

We define the spatial conditional extremes model as presented in Wadsworth and Tawn (2018). For sites s in our spatial domain $\mathcal{S}$, let X be our conditional extreme process with associated threshold μ and s_0 be the site to be conditioned on, with functions $a_{s-s_0}(x)$ and $b_{s-s_0}(x)$ such that $a_{s-s_0}(x) \in \mathbb{R}$ and $b_{s-s_0}(x) > 0$ for all x. For some residual process Z^0 we have:

$$\{X(s)|X(s_0) > u : s \in \mathcal{S}\} \stackrel{d}{\approx} \left\{a_{s-s_0}(X(s_0)) + b_{s-s_0}(X(s_0)) Z^0(s) : s \in \mathcal{S}\right\},$$

with $X(s_0) - u|X(s_0) > u \sim \text{Exp}(1)$ independent of Z^0. The functions a_{s-s_0} and b_{s-s_0} reflect the spatial dependence between the conditioning site and nearby sites, while the residual process Z^0 models overall trends of the spatial domain.

5.3.3   Residual process $Z^0(s)$

For the underlying residual process, we require a process defined for all $s \in \mathcal{S}$ with the condition that $Z^0(s_0) = 0$. That is, our process is equal to 0 at the conditioning

site. We start with an arbitrary stationary Gaussian process

$$Z_G \sim \text{MVN}_{|S|}(0, \Sigma)$$

with dimension $|S|$ equal to the number of sites in $\mathcal{S}$, and with a covariance matrix Σ such that:

$$\Sigma = \sigma^2 \exp\left(-\left(\frac{\|x - x'\|}{\phi}\right)^\nu\right).$$

This model captures the dependence between cells once they have been transformed to standard normal margins using the marginal distributions, such that the marginal distributions for each cell of Z_G are normally distributed.

In order to satisfy $Z^0(s_0) = 0$, we then condition upon $Z^0(s_0) = 0$ such that for points s_0 and all other points s can we rearrange x, μ and Σ as:

$$x = \begin{bmatrix} x_s \\ x_{s_0} \end{bmatrix}, \mu = \begin{bmatrix} \mu_s \\ \mu_{s_0} \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma_{s,s} & \Sigma_{s,s_0} \\ \Sigma_{s_0,s} & \Sigma_{s_0,s_0} \end{bmatrix}$$

resulting in the following mean function and covariance matrix:

$$\overline{\mu} = \mu - \mu\Sigma_{s,s_0}\Sigma_{s_0,s_0}^{-1}$$
$$\overline{\Sigma} = \Sigma_{s_0,s_0} - \Sigma_{s,s_0}\Sigma_{s_0,s_0}^{-1}\Sigma_{s_0,s}$$

This gives us a conditional Gaussian distribution such that:

$$Z^0(s) \sim Z_G(s)|Z_G(s_0) = 0, \tag{3}$$

which maintains the property that the marginal distributions are Gaussian for each site.

Furthermore, we transform the marginal distributions to a generalised normal distribution to allow for a broader range of distributional shapes. A random variable X with a generalised normal distribution has probability density function g such

that:

$$g(x) = \frac{\delta}{2\alpha\Gamma(1/\delta)} \exp\left\{-\left|\frac{x-l}{\alpha}\right|^{\delta}\right\},$$

where $\alpha > 0$ and l is the location parameter. This distribution encompasses both the normal ($\delta = 2$) and Laplace ($\delta = 1$) distributions whilst also allowing shapes between these distributions and more/less heavy tailed than them.

The generalised normal distribution is a broadening of the delta-Laplace distribution used in previous applications of the Spatial Conditional Extremes model (Wadsworth and Tawn, 2018). The delta-Laplace distribution has the same probability density function as the generalised normal distribution, however previous applications of the distribution consider values of δ between 1 and 2 due to the theoretical expectation that the conditional distribution will vary from a normal to a Laplace distribution as the distance between cells increases. From our exploratory analysis, we allow the model to consider values of δ outside this range for a more flexible model structure.

Rather than fit a fixed value of δ, we allow δ to vary with distance such that:

$$\delta(s - s_0) = 1 + \delta_3 \exp\left(-\left(\frac{\|s - s_0\|}{\delta_1}\right)^{\delta_2}\right)$$

which results in $\delta \to 1$ as $\|x - x'\| \to \infty$, going from a normal distribution to a Laplace distribution to match the marginal distributions of dependent or independent cells respectively.

In defining a model for a_{s-s_0}, we allow for cases of both asymptotic dependence and asymptotic independence, since closer locations are likely to be asymptotically dependent and further ones asymptotically independent. We consider dependence to be isotropic such that the distance and not direction is the only information that is considered when establishing dependence. The general form of a_{s-s_0} is:

$$a_{s-s_0}(x) = x\alpha(s - s_0) = \begin{cases} x & \|s - s_0\| < \Delta \\ x \exp\left\{-\left[(\|s - s_0\| - \Delta)/\lambda\right]^{\kappa}\right\}, & \|s - s_0\| \geq \Delta \end{cases}$$

110

where λ is a scale parameter, κ is a shape parameter and Δ is the maximum distance at which sites are asymptotically dependent. For $\|s - s_0\| \geq \Delta$, as the distance between the points increases the function a_{s-s_0} decreases with a lower limit of 0. If $\Delta = 0$, then there is asymptotic independence everywhere, and for positive values of Δ lower than the maximum distance between sites we have varying levels/areas of dependence between sites. The parameters λ and κ can be estimated from the data using likelihood methods.

For this analysis, we set $\Delta = 0$ as for the purposes of IST data, we would expect there to be some distance at which temperatures at 2 sites would be independent from each other, and therefore we would expect the same independence from the extreme temperatures. If needed, Δ could be estimated from the data, which would allow asymptotic dependence for sites close together and asymptotic independence for sites further away from each other.

We define b_{s-s_0} in terms of a_{s-s_0}, such that:

$$b_{s-s_0}(x) = 1 + a_{s-s_0}(x)^\beta, \tag{4}$$

where β is an additional parameter to be estimated. For $\beta > 0$, if $a_{s-s_0}(x) \to 0$ as $\|s - s_0\| \to \infty$, then $b_{s-s_0}(x) \to 1$. This means that for sites sufficiently far apart, conditioning on a site simply results in a Gaussian process and effectively has no impact on the observation at the other site. Two other functions for b_{s-s_0} were provided in the literature (Wadsworth and Tawn, 2018) and have been tested in our simulations, however for the purposes of our work we have chosen Equation 4 as it allows b_{s-s_0} to vary spatially unlike the other functions.

When fitting the model, we consider all extreme values observed at all sites in the domain. The model is fitted on a site-by-site basis, considering the extreme values at each site along with values from all other sites at the same time as each extreme value using a composite likelihood. For each exceedance i at each site $j \in \{1, \ldots, d\}$, we evaluate the conditional density function of all other sites $k \in \{1, \ldots, d\}/j$ given

that an exceedance occurred at site j. If there are multiple extreme observations that occur simultaneously at a single time point, each observation will be considered in the full likelihood multiple times with only the conditioning site changing. Due to the transformation of x as in Equation (1), we also require a Jacobian term so that each term in the likelihood has the form:

$$f(x) = f(z_0) \left| \frac{dz_0}{dx} \right|,$$

resulting in a composite likelihood function where:

$$L_j(\boldsymbol{\theta}) = \prod_{i=1}^{n_j} f_{Z^j} \left(\left\{ \frac{\left[x_k^i - a_{s_k-s_j}\left(x_j^i; \boldsymbol{\theta}_a \right) \right]}{b_{s_k-s_j}\left(x_j^i; \boldsymbol{\theta}_b \right)} \right\}_{k \in \{1,\dots,d\}\backslash j} ; \boldsymbol{\theta}_Z \right)$$
$$\cdot \prod_{k \in \{1,\dots,d\}\backslash j} b_{s_k-s_j}\left(x_j^i; \boldsymbol{\theta}_b \right)^{-1},$$

such that f_{Z_j} is the density of Z_0 conditional when $s_j = s_0$. This gives an overall likelihood function of:

$$L(\boldsymbol{\theta}) = \prod_{j \in \{1,\dots,d\}} \left(\prod_{i=1}^{n_j} \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \exp \left(\frac{-1}{2}(x_i^* - \mu)^T \Sigma^{-1}(x_i^* - \mu) \right) \right.$$
$$\left. \cdot \prod_{k \in \{1,\dots,d\}/j} \frac{1}{\left(1 + a_{s_k-s_j}(x_j^i) \right)^\beta} \right)$$

for parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_a, \boldsymbol{\theta}_b, \boldsymbol{\theta}_Z)$, exceedance threshold μ, and

$$x_i^* = \frac{x_k^i - a_{s_k-s_j}(x_j^i; \boldsymbol{\theta}_a)}{b_{s_k-s_j}(x_j^i; \boldsymbol{\theta}_b)}$$

5.3.4 Simulations

The model allows for simulations of all sites in the domain conditional either on a specific site observing an extreme value, or on there being an extreme value at some

point on the domain. We simulate values conditional on a selected (or sampled) conditional site. First, we generate the value of this exceedance $X(s_0)|X(s_0) > v \sim \text{Exp}(1)$ from the exponential tailed margins. We then generate $Z^0(s)$ independently of $X(s_0)$, by drawing from a Gaussian process of size $|\mathcal{S}|$ given $Z^G(s_0) = 0$ as described in Equation (3). Finally, we simulate the values for all other points $s \in \mathcal{S}$ by setting:

$$\{X(s)|X(s_0) > v\} = a_{s-s_0}(X(s_0)) + b_{s-s_0}(X(s_0))Z^0(s).$$

5.4 Pairwise bivariate fits

We fit the bivariate conditional extremes model to a range of samples to explore different properties of extremal dependence in the data. In each case, we examine the estimated parameters as a function of distance from the conditioning cell. This allows us to investigate how dependence varies across the ice sheet, at least from a pairwise perspective, before moving to a spatial view.

5.4.1 Full ice sheet

We first fit the bivariate conditional extremes model for five conditional cells in our previous sample of 206 cells. Although the previous exploratory analysis suggested that extreme dependence may only be seen at finer scales than in this sample, we use this sample as an example application of the model fit and to examine this trend more fully. The conditioning cells (Figure 43) are chosen to represent a range of locations and covariate values across the ice sheet, including the maximum and minimum distances to the coast (cells 1 and 2). For each of these cells, we fit the conditional models using maximum likelihood estimation to data from 2001 to 2019.

At the scale of the entire ice sheet, the fitted models show some dependence between the conditioning cells and other cells, but only for short distances (Figure 44). As the distance from the conditioning cell increases, the estimates of α begin
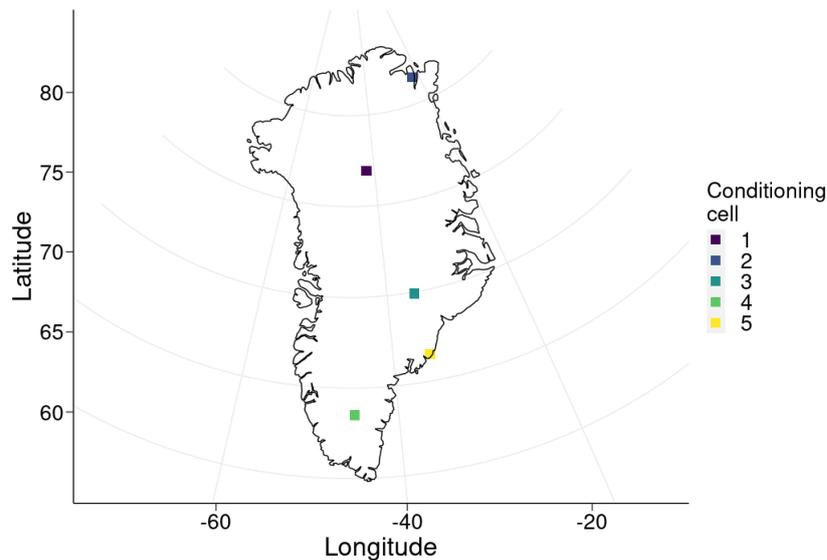
Figure 43: Locations of the conditioning cells for the fitting of the conditional extremes distribution.

to converge on 0 until the mode of the estimates for cells more than 1500km away is centred around 0. This suggests that, for small distances, there is some extremal dependence being captured by the model, but this decays over distance. This isn't unexpected given the range of distances: cells nearly 2000km apart are unlikely to show dependence in their extreme values.

The estimates for β show slightly weaker trends with distance than those for α. The estimates are closer to 0 with a similar but much weaker trend of higher values for cells a shorter distance from the conditioning cell. Although the trend for β is weaker than for α, the range of the estimates is also much smaller. This suggests that, although a linear or similar trend with distance may be difficult to model, having a single estimate for β or an estimate linked to α may be sufficient to capture the variance of this parameter.

The estimates of μ and σ also show trends with distance, although these differ from those seen in α and β. The trend with distance in μ is fairly linear but reasonably consistent. The trend with distance in σ is not as linear, estimates appear to increase as distance increases but then decrease for distances over 1000km. This
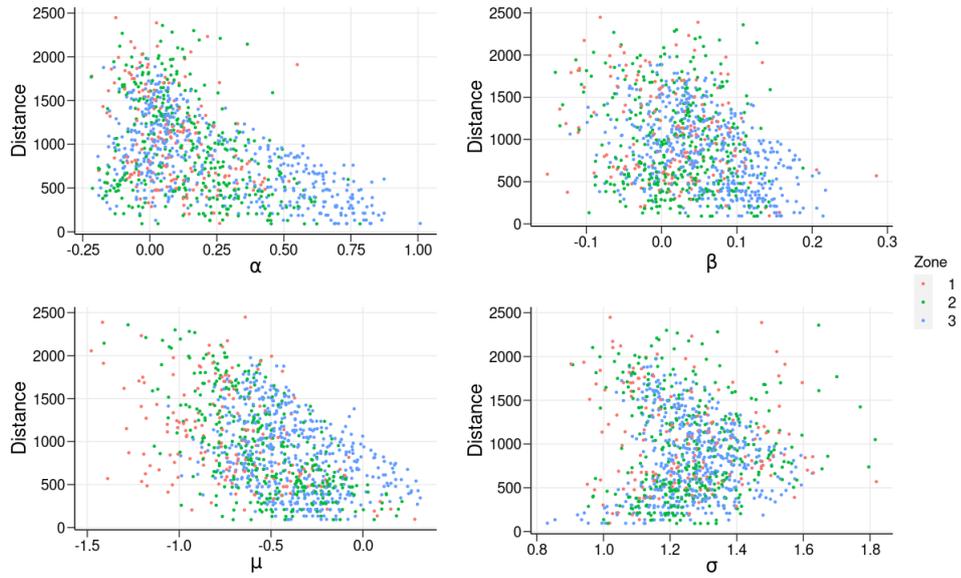
114

Figure 44: Parameters estimates for the extreme conditional distributions fit for the five conditioning cells shown in Figure 43 plotted against the distance away from the conditioning cell. Colour indicates the glaciological zone of the non-conditioning cell - 1 for ablation zone, 2 for accumulation zone, 3 for dry snow zone.

distance may be the maximum distance at which we can observe dependence, as the estimates for α also display a change in behaviour (closer to 0) for distances greater than 1000 km. The trend in the estimates of μ also slightly resembles the trend in α (Figure 45). This may be an artificial consequence of the functional form of the conditional distribution, since both parameters describe the mean of the conditional distribution.

The estimates also display some trends across the different zones in which the non-conditioning cells are located. Estimates of α are highest for cells in zone 3 (dry snow zone), however this may be primarily due to the conditioning cells chosen. The highest estimates for α are mainly for conditioning cells 1 and 3, which are the furthest cells from the coast and have the coldest overall climates due to their latitude and elevation. Although this does not extend to all conditioning cells, it may suggest that there are differences between the zones in terms of dependence, as even for very short distances estimates of α do not typically exceed 0.5 for other conditioning cells.
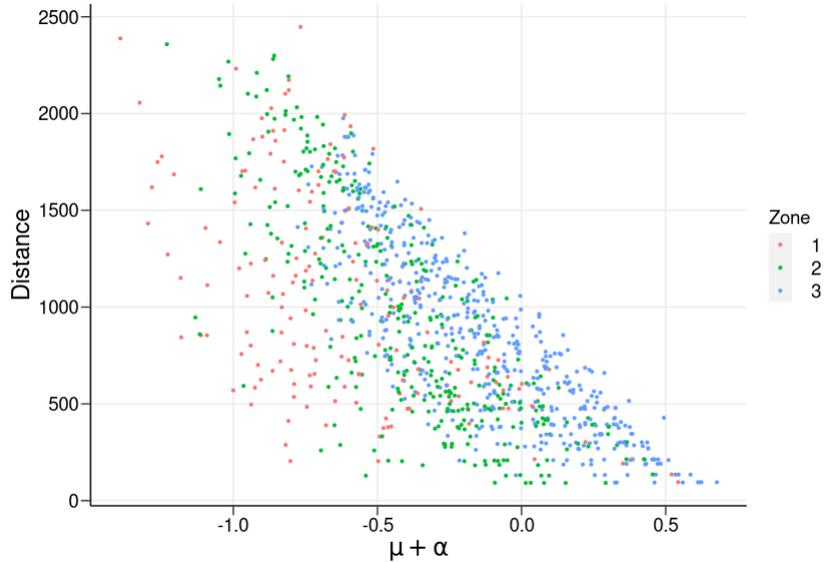
Figure 45: Sum of the parameter estimates of α and μ for the extreme conditional distribution of five conditioning cells and all over cells, plotted against the distance away from the conditioning cell.

Another trend also seen in previous analysis can be seen by examining conditioning cell 5 in more detail. It is situated on the south east coast of the ice sheet, in an area that we have previously seen different dependence behaviour to other parts of the ice sheet with comparable latitudes, elevations and distances to the coast. This can be seen more clearly in Figure 46, as the estimates for α are consistently centred closer to 0. Conditioning cell 5 has a maximum estimate of α that is less than 0.2, whereas all other conditioning cells display estimates at least exceeding 0.5. The proximity of the estimates close to 0 suggests that this conditioning cell shows low extremal dependence with other cells regardless of the distance between them, suggesting that extremal behaviour in the area is more distinct than at other conditioning cells.

5.4.2 Central area fit

Even at the scale of the entire ice sheet, there are clear trends between the parameter estimates of the conditional distribution and distance. The trends are not
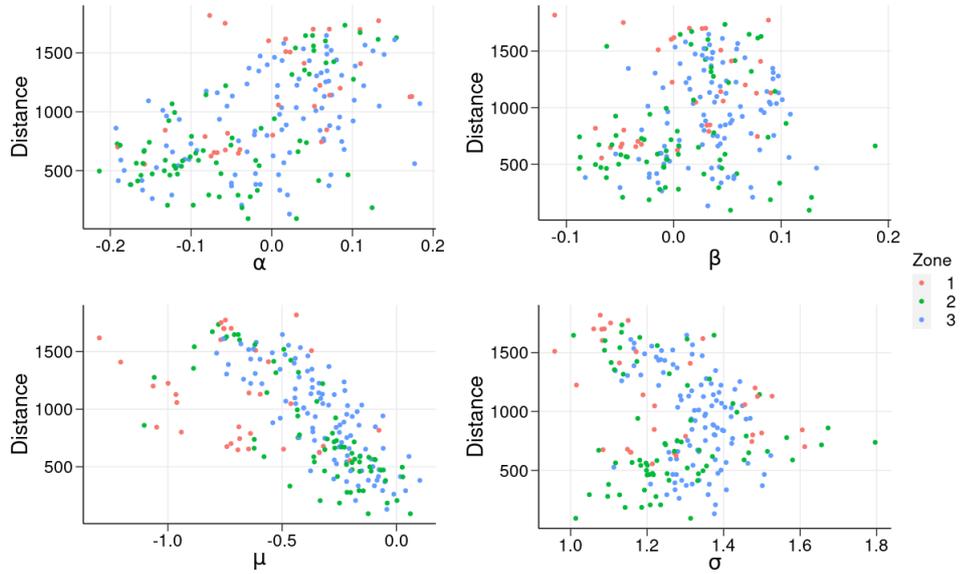
Figure 46: Parameter estimates of the conditional distribution for conditioning cell 5 against the distance between the conditioning cell and other cell.

consistently seen at all conditioning cells, but where they are seen they appear similar in form. This suggests that our extremal modelling approach should consider trends between parameters and distance but also differences in behaviour for different areas of the ice sheet. However, one aspect that this example does not fully display is the difference in dependence at different spatial scales. We now choose smaller scale data samples to look more closely at the dependence in the extremes. Conditioning cell 1 from the previous analysis showed the clearest trends for both α and β parameters, so we choose a square of 100 (10 by 10) cells close to the centre of the ice sheet that are situated approximately 16km apart along horizontal and vertical axes. This allows us to examine whether the trends observed are consistent for multiple different conditioning cells in the same area, and whether there are different trends in the parameters at smaller scales. The four corners of the new sample area - which are chosen as the conditioning cells - can be seen in Figure 47.

The impact of choosing a smaller sample area on dependence can be seen from the estimates for α and β in Figure 48. The ranges of both sets of parameter estimates differ from those seen for the entire ice sheet sample. Estimates of α
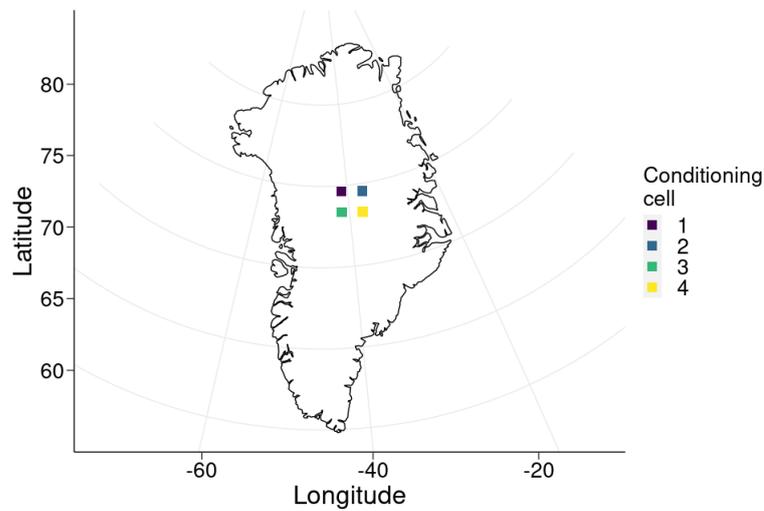
Figure 47: Conditioning cells for the sample of cells in the centre of the ice sheet. The four conditioning cells act as the four corners of the sample area.

are centred around 0.9, whereas there was previously only a single estimate above 0.9, and β has estimates centred around 0.25 rather than around 0.05. Both sets of estimates suggest strong dependence between cells, providing further evidence of the relationship between distance and extremal dependence. Neither parameter has a particularly strong relationship with distance, although this may be due to the distances now being too small for any variability in the surface conditions of the cells.

The estimates of μ show broadly similar trends to those seen across the entire ice sheet. The range of estimates is comparable, with the main difference being fewer estimates below -1, which were typically seen for cells further away from the conditioning cell. The trend with distance is also linear and decreasing as seen previously. However, the estimates of σ show a clearer trend than was seen for the entire ice sheet. With the new sample focusing on cells a maximum of 200km apart, we can see that σ increases consistently and linearly as distance increases. This trend is consistent regardless of conditioning cell, suggesting that a linear component is most appropriate either at small distances or for cells with similar surface conditions.
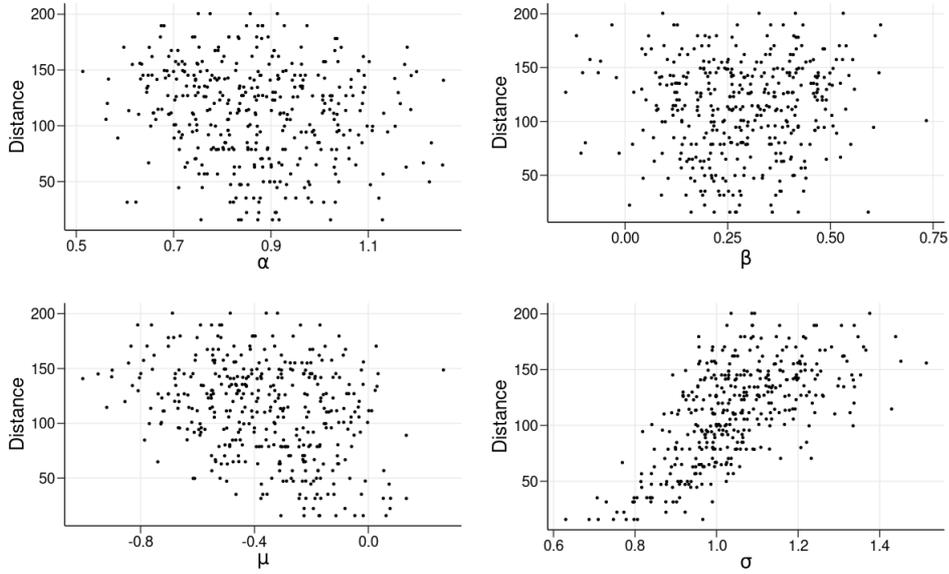
118

Figure 48: Parameter estimates for the extreme conditional distributions fit to the central ice sheet sample for the four conditioning cells shown in Figure 47 plotted against the distance away from the conditioning cell.

5.4.3 Basin fits (east and west)

Having identified trends at small scales with consistent surface conditions, we next move to examine whether these trends are consistent across other areas of the ice sheet. We take two samples of points in two drainage basins located on opposite sides of the ice sheet. This allows us to examine a variety of surface conditions within a single sample, maintain relatively small distances between cells, and assess stationarity of dependence by comparing similar surface characteristics and distances at different locations on the ice sheet. The locations of the two drainage basins and the associated samples can be seen in Figure 49. For the basin samples, we examine the parameter estimates considering model fits that consider each cells as the conditioning cell, rather than selecting only a small number of conditioning cells. This is partially due to the smaller number of cells in each of the basin samples, but also to account for the potential impact that the choice of conditioning cell could have upon our inference.

For the east basin (Figure 50), we see similar trends as previously observed for
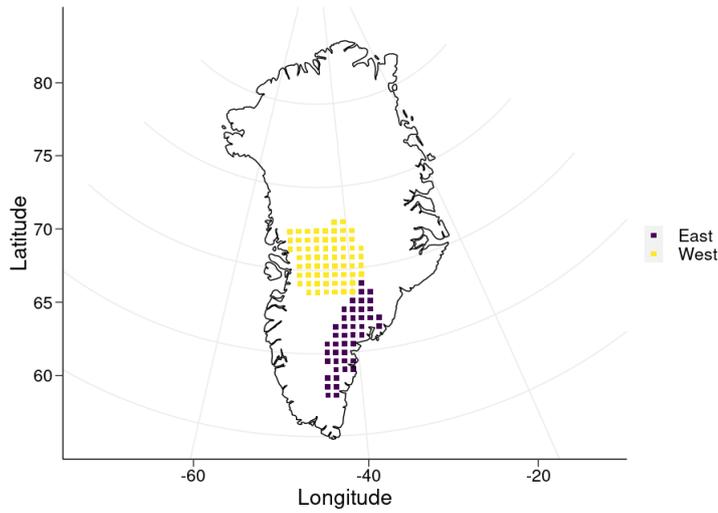
Figure 49: Locations of the cells in the two drainage basin samples.

the estimates of α and β. Cells at small distances have generally higher values of α, although the overall distribution of estimates still has its mode around 0. As distance increases, fewer cells have estimates close to 1 suggesting that once again extremal dependence decreases with distance. The estimates for β show little relationship with distance, but again are centred around 0. Furthermore, there are more negative estimates of β than in previous samples, suggesting different dependence relationships than we have previously seen in other areas. This corroborates previous trends of the east coast ice sheet, as it has displayed different temperature and dependence behaviour in multiple previous models we have examined.

The estimates for μ and σ are also consistent with trends from previous sample areas. Estimates of μ decrease with distance from a maximum of around 1 to around -1, with some exceptional cells having estimates as low as -3. Estimates of σ again range from around 0.8 for low distances up to around 1.5 for greater distances. These trends are consistent in their form and can therefore be considered by the model structure for the spatial extremes model. There is minimal difference in the estimates between difference zones though, and estimates from all three zones share the similar ranges of estimates described above.
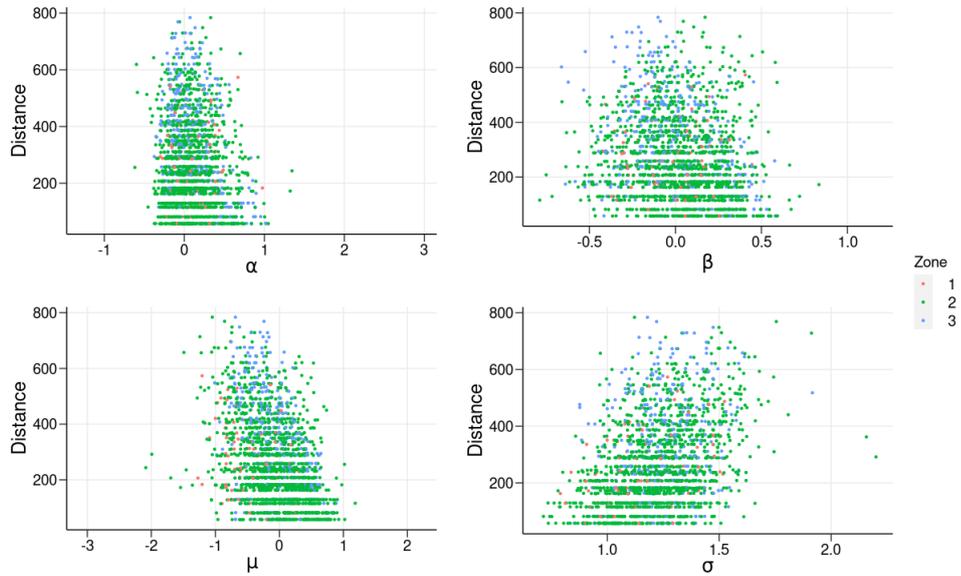
Figure 50: Conditional parameter estimates for the east basin plotted against the distance from the conditioning cell.

The estimates for the dependence parameters in the west basin (Figure 51) also share similarities with previous estimates, as well as distinct differences to the east basin. The estimates of α have a similarly lower range, but in the west basin there are far more estimates that exceed 1 and more estimates close to 1 at larger distances. Most of the higher estimates are from cells in zone 3, however this may simply be due to the west basin having many more cells in zone 3 i.e cells closer to the centre of the ice sheet. The same point could be made for the estimates of β, which match the range of those seen in the east basin, except for in zone 3 which are higher at larger distances. This suggests that although there may be differences between the behaviour in different drainage basins, the trends observed could represent different smaller partitions of the same overall large-scale trends.

The estimates of μ mirror the trends seen from the estimates of α. The overall distribution of estimates reasonably closely resembles the estimates from the east basin, with zone 3 showing different behaviour as it did for α. Rather than more frequent high estimates of μ, we instead see more frequent and greater magnitude low estimates than seen in the east basin, predominantly from cells in zone 3. This
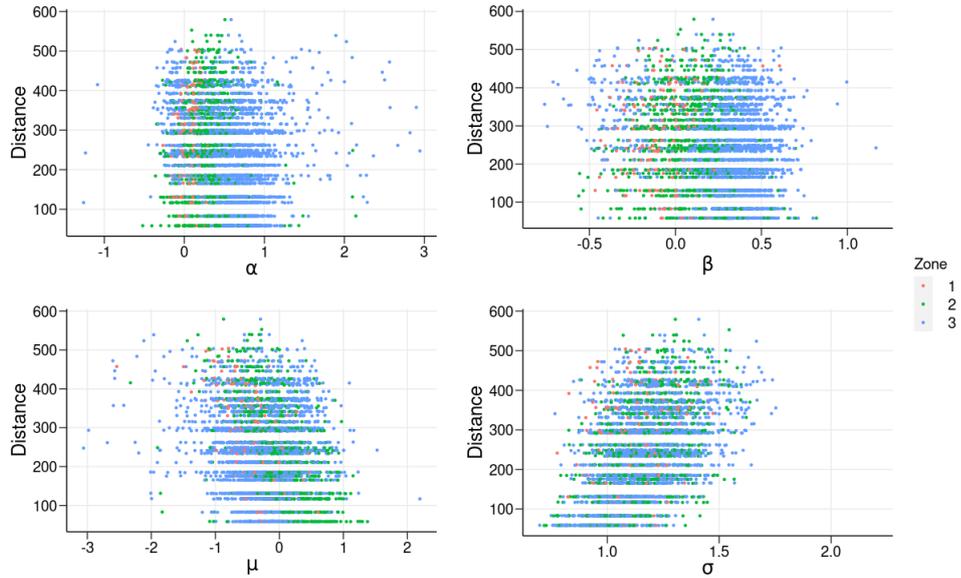
Figure 51: Conditional parameter estimates for the west basin plotted against the distance from the conditioning cell.

isn't unsurprising given the close relationship previously observed between estimates of α and μ (Figure 45). Of the four parameters compared, the estimates for σ match most closely between the different basins. Aside from the difference in the number of cells compared, the two sets of estimates have similar ranges and trends with distance.

Overall, the comparison of drainage basins gives us confidence that the trends we have observed in the conditional distribution parameter estimates are consistent across different parts of the ice sheet. There are clear trends in all parameters, the forms of which appear to be relatively simple to parameterise. The dependence is not limited to a specific spatial scale or area of the ice sheet, despite minor differences in areas of the ice sheet that have previously shown more unique behaviour, i.e the south east coast.

5.4.4 Pairwise fits with Generalised normal distribution

As well as models for the dependence parameters, the conditional spatial model requires a model for the margins of the conditional spatial process. (Wadsworth

and Tawn, 2018) showed that this distribution may change with distance between cells. It is therefore important to check this distribution before moving to the full spatial extremes model.

We fit the conditional distribution to the 206 sites of the entire ice sheet previously used as the first example of the conditional distribution, this time using the generalised normal distribution rather than the normal distribution. This gives us a conditional distribution of:

$$X_j \mid X_c > \mu_c \sim N(\alpha x_c + \mu x_0^\beta, \sigma x_c^\beta, \delta). \tag{5}$$

Using this new conditional distribution, we examine the new parameter δ. From Figure 52, we see that there is considerable variance in the estimates of δ, particularly for higher values. The higher values correspond to distributions closer to a uniform distribution, suggesting that the tails of these distributions are not heavily weighted. However, the highest density of estimates lies between 1 and 2. It is also important to note that the mode is not at 2, which would be the expected modal estimate if the most appropriate distribution was the normal distribution. This further justifies the use of the generalised normal distribution since the optimal distribution is heavier tailed than the normal distribution.

Lastly, we examine the other four parameters of the conditional distribution, since the addition of the extra generalised normal distribution parameter could cause other parameter estimates to adjust. From Figure 53, we see that the trends are a mixture of those observed earlier for the full ice sheet sample and those observed for the samples from the east and west basins. Estimates of α decrease linearly as distance increases, with a decrease in the range of the estimates for larger distances. The estimates of β see a slight decrease as distance increases but at a more gradual rate than α, and remain closer to 0; μ decreases away from 0 and with an increasing range as distance increases, and σ increases up to distances around 1000 km before decreasing slightly for greater distances. This gives us a reasonable
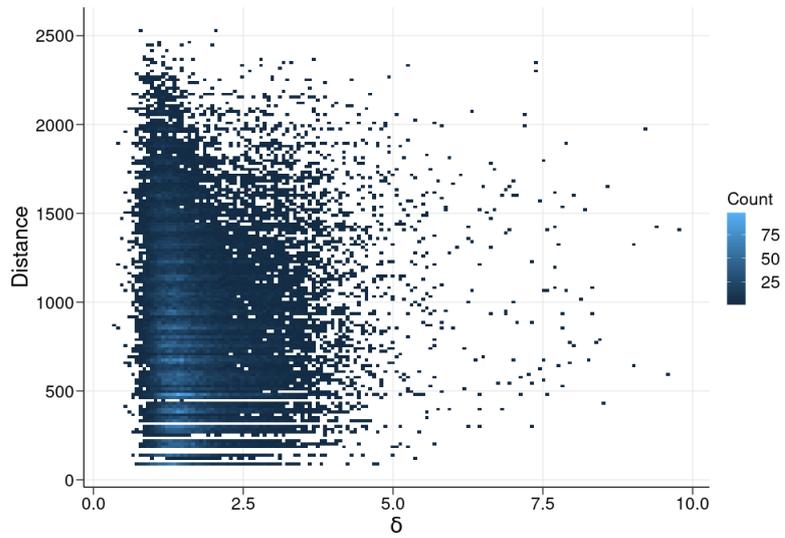
Figure 52: Heatmap of estimates for δ against the distance from the conditioning cell. Values for δ are limited to 10 for clarity, as a small number ($< 0.01\%$) of estimates exceed this.

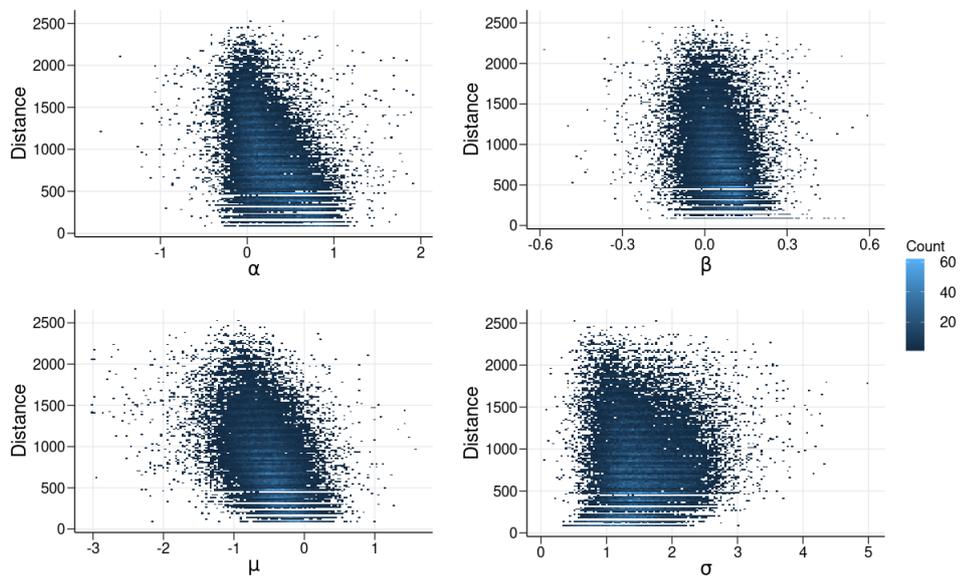basis to determine functional forms of parameters for our spatial extremes model.

Figure 53: Heatmap of the conditional parameter estimates for the Greenland sample using the generalised normal distribution against the distance from the conditioning cell.

5.5 Spatial Conditional extremes fit

While assessing the fit of different model structures, we apply the models to a sample of 73 cells spread across the entire ice sheet (1 in every 200 cells in both latitude and longitude dimensions) for the years 2010 to 2019. This provides us with enough data to capture the dependence structure, while keeping a feasible computational time for model assessment. Using a consistent sample for all models enables fair comparisons of the model structure and parameter estimates.

5.5.1 Full ice sheet fit

Having fit the spatial extremes model, we assess the fit empirically by comparing simulations from the model to the data. For each cell and each day the cell has an exceedance, we calculate the total number of exceedances at all cells. For each exceedance date, given the observed value at the conditioning site, we simulate from the spatial extremes model 10 times and calculate the number of cells that have an exceedance. To account for missing data in the original data set, we subset the simulations to the same set of cells present in each day of data, so that simulated values are only compared against cells that have a valid temperature in the data.

Figure 54 demonstrates the difference in event sizes from the data compared to simulations from the fitted model. For most cells, the model underestimates the size of spatial extreme events compared to the data; while the range of the size of events from the model is around 8 to 16 cells, the range for the data is from around 6 to 29. This is the key difference between the two sets of extreme events - the simulated extreme events from the model are realistic, but not as widely varying as those seen in the data.

We can view the dependence structure by examining the form of the dependence functions a and b over a range of distances (Figure 55). From this, we see that the range of dependence as estimated by the model tails off relatively quickly from high initial values. For a, the initial high estimates quickly decrease to below 0.25 for distances above 250 km. Although the estimates do not reach 0, for the largest
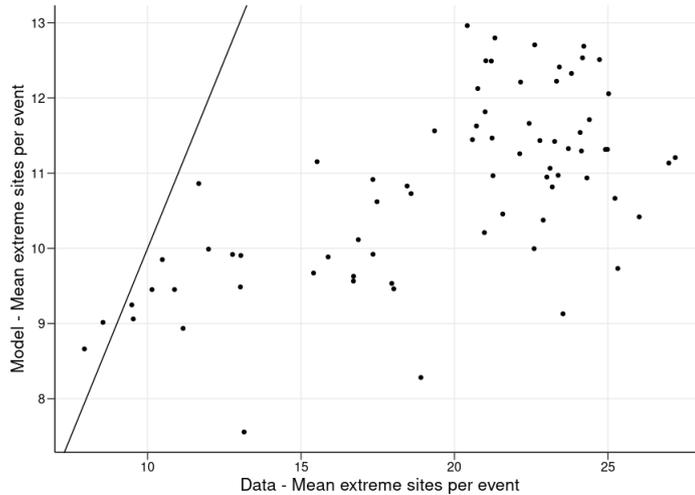
Figure 54: Number of exceedances calculated empirically from the data compared to simulations from the spatial extremes model for the entire ice sheet fit, both conditioning on the same cell.

distances the estimates are very small (0.014) suggesting asymptotic independence at large distances. This is also suggested by the estimates for b, which decay to values very close to 1 for distances greater than 1500 km. Since the estimates of b decrease at a faster rate than a, this suggests that there is some asymptotic dependence at shorter distances. These estimates and trends are somewhat corroborated by the earlier conditional distribution estimates, since there was a decrease in the dependence of the data at greater distances (Figure 53).

The estimates for δ also show distinct trends as the separation distance increases. Figure 56 shows that the estimates decrease from their starting value of around 2 to values around 1.4 for the largest distances. This does not quite reach the parameter estimates that would suggest a Laplace distribution for large distances, but does show clear changes in the conditional distributions as the separation distances changes. The trend is not quite linear, but is closer to linear than the trends from the dependence parameters, thereby showing the complexity of the conditional distributions and their changes with distance.

Using the fitted model, we can simulate extreme temperature events and examine their properties. To do this, we first either sample or choose a conditioning cell then
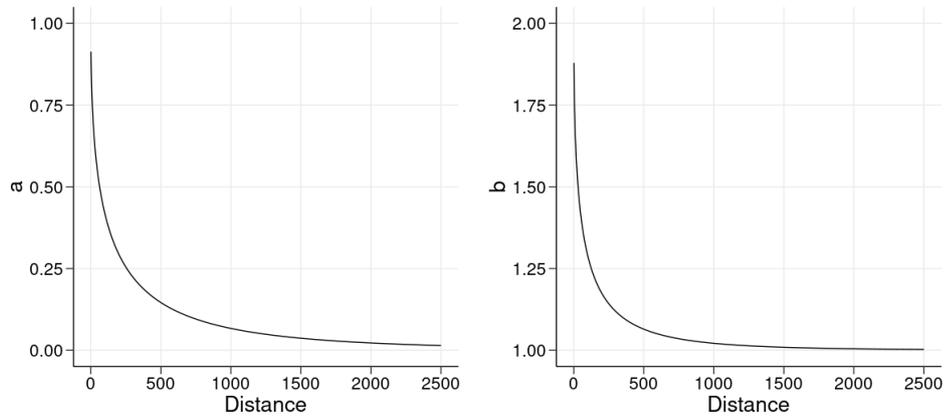
126

Figure 55: Functions for a and b evaluated over a range of distances using the fitted model parameter estimates for the entire ice sheet fit.
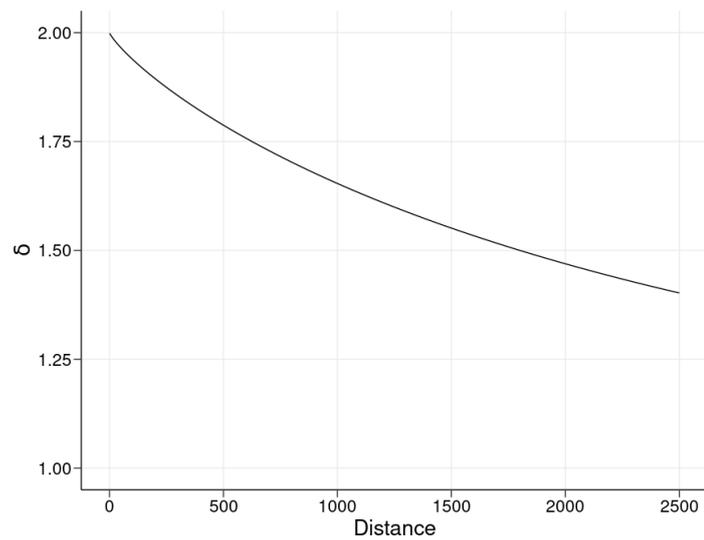

Figure 56: Function for δ evaluated over a range of distances using the fitted model parameter estimates for the entire ice sheet fit.
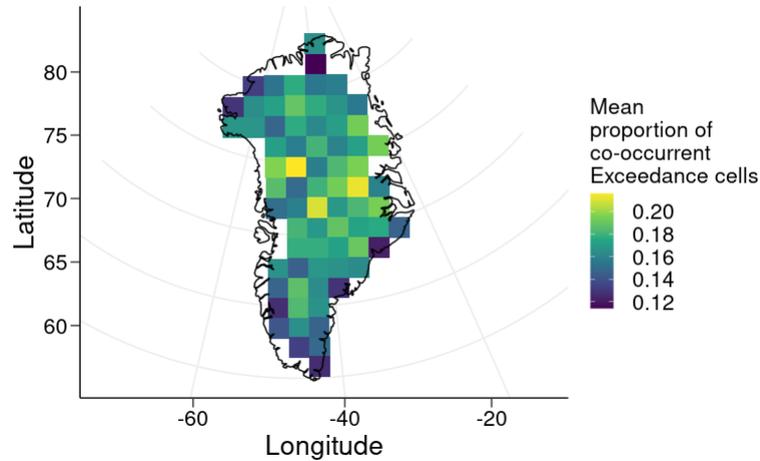
Figure 57: Mean proportion of co-occurrent exceedances from 1000 simulations considering each cell as the conditioning cell for the entire ice sheet model fit.

simulate all other cells conditionally depending on whether our inference is centred on extreme events affecting the entire ice sheet or events that include one specific cell. For the inference below, we simulate 100 extreme events using each cell as the conditioning cell.

We first examine the average size of the extreme events. For each simulation at a given conditioning cell, we calculate the proportion of other cells that also had an exceedance, then take an average over all simulations. We use the proportion of co-occurrent cells rather than the absolute number due to the different numbers of cells in each region that we compare to make comparisons more valid. From Figure 57, we can see that there is an overall trend of larger extreme events occurring more often in the centre and north east of the ice sheet. Smaller events taking place at cells closer to the coast is expected. Firstly, considering the dependence of the model is based on distance, cells on the coast have on average fewer cells close by. Furthermore, the coastal cells have more varied surface conditions than cells closer to the centre and thereby more localised high temperature events are more likely.

From the simulations, we can also consider the properties of the temperature values that are observed. We examine quantiles of the temperatures, in contrast to

128

the melt measures obtained from the mixture model, which use the mixture model's ice and melt components to estimate a probability that each observation is a melt observation (2.4.1). The thresholds used for this analysis are based on quantiles rather than the melt threshold, which results in a wider range of temperatures for the simulated events (Figure 58). Even so, the overall trends reflect those previously seen, with the coasts experiencing higher average temperatures. Central cells have much lower averages, many of which are not very close to the melt thresholds. This confirms that the spatial trends seen in the mean temperatures are also present in the extreme temperatures after conditioning on observing an extreme temperature.
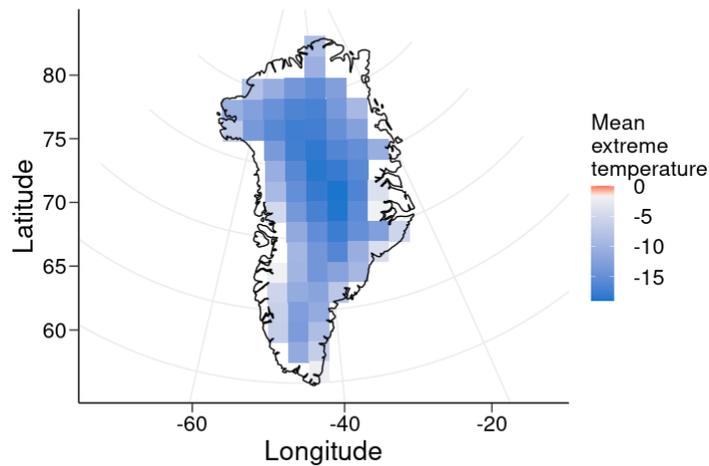

Figure 58: Mean temperatures of each cell averaged across simulations including all cells as conditioning cells for the entire ice sheet model fit.

As with previous models, we can use the marginal model to convert the simulated temperatures to melt estimates to examine the melt implications of the simulated extreme temperature events. We now compare on the probability of melt given that an extreme temperature event has taken place. Figure 59 shows that the melt trends are as expected. Of particular interest, even after conditioning on an extreme temperature event occurring, the centre of the ice sheet still has a very low probability of experiencing melt. This suggests that melt events that affect most or

129

nearly all of the ice sheet are even more unlikely than previously thought. Other cells - mostly on the coasts - show a much higher probability of melt temperatures in extreme events, particularly on the west and east coasts.
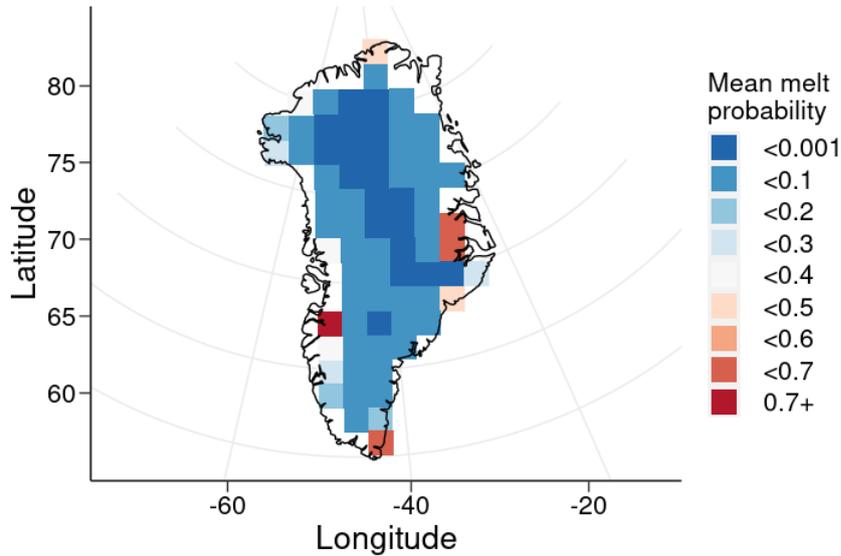

Figure 59: Mean melt probability for each cell from spatial extreme event simulations for the entire ice sheet model fit.

5.5.2 Drainage basin fits

We next examine if the full spatial events occur at different scales and in different regions. To examine this, we split the ice sheet into drainage basins and fit the model locally to each drainage basin. Each basin has between 15 and 36 cells, and the basin at the southern tip of the ice sheet is ommitted due to its small size (6 cells). The specific samples and drainage basin labels can be seen in Figure 60.

We follow the same analysis sequence as for the entire ice sheet, starting with the parameter estimates and their implications for spatial extreme dependence. From Figure 61, we can see there is considerable variation in the dependence parameters between the different basins. Basins 5 and 6 have noticeably lower dependence from their estimates of a than the other basins. Both basins have very low estimates of a above distances of around 500 km, whereas the other basins have higher estimates
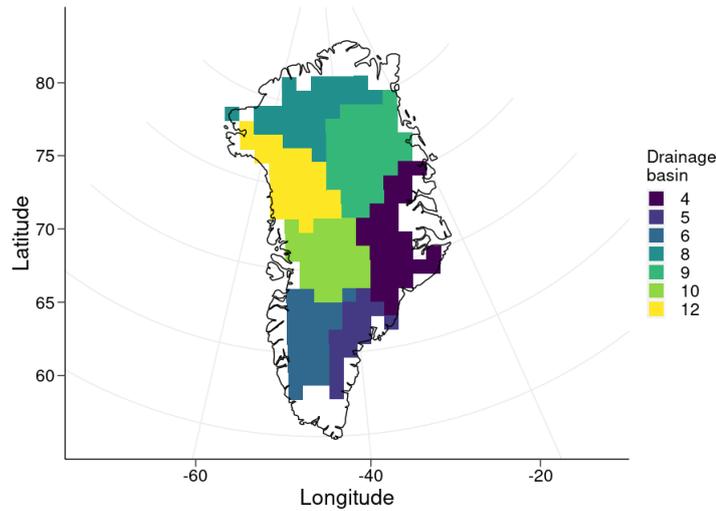
Figure 60: Map of the drainage basins on the Greenland ice sheet, and the selected sample at the chosen resolution.

for much larger distances. This trend can be extended to include drainage basin 10, since these three basins are the most southern basins and they all have the lowest three sets of estimates for a and b. If we consider this as a trend with latitude, we see that the basin at the highest average latitude - basin 8 - has the highest values of a along with basin 12, which has lower a estimates but the highest estimates for b over almost all distances. Given the size of the basins, these trends may be a result of latitude or could be due to the specific characteristics of the ice sheet in each basin. In either case, it is clear that there are different dependence properties of extreme temperatures in different areas of the ice sheet, suggesting that the extremal dependence observed may differ at different scales of the ice sheet.

The difference is even more evident in the estimates for δ (Figure 62). However, the trends differ from those seen in the dependence parameters. The basins form two groups; the first group consists of basins 6, 8 and 9, in the very north (8 and 9) and south west (6) of the ice sheet. Here, the estimates for δ do not decrease to levels close to 0, instead tailing off to around 1.5 for the largest distances on the ice sheet. These estimates suggest that the distributions for cells in these basins do not get close to normal even at large distances, potentially indicating that the dependence
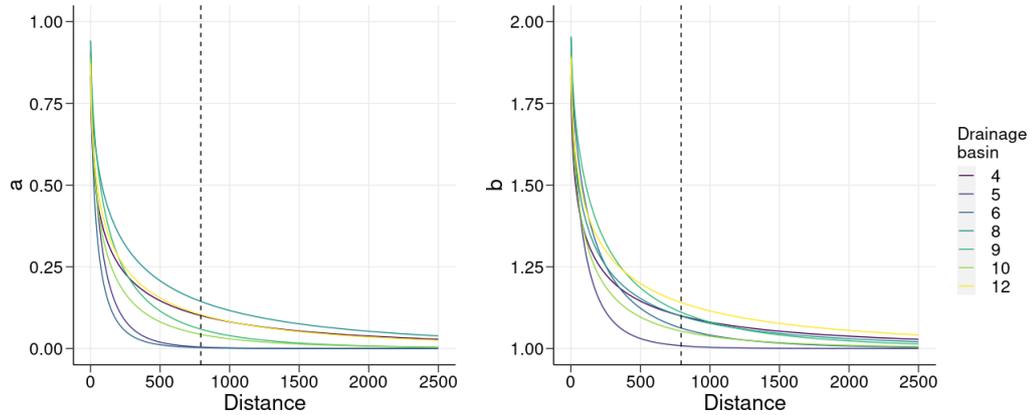
Figure 61: Functions for a and b evaluated over a range of distances using the fitted model parameter estimates for the drainage basin fits. Dotted line represents the mean maximum distance observed from all drainage basins.

here does not reach asymptotic independence. The second group consists of basins 4, 5, 10, and 12, situated on the west (12 and 10) and east (4 and 5) of the ice sheet. The estimates for these basins either tail off at 0 or get very close to 0, suggesting that there are distances and cells within these basins that effectively reach asymptotic independence.

This difference does not appear to be due to the shapes of the basins, since all basins have cells a relatively large distance apart that are considered in the model fitting process, so the difference is not due to untrustworthy extrapolation. It also does not appear to be due to the form of the function used to model δ, since basin 5 also has a convex functional shape yet has estimates that are much closer to 0.

Figure 63 shows the event sizes simulated by the basin models. The differing dependence structures result in a wide range of event sizes, with more regional variance than seen for the entire ice sheet fit. South areas of the ice sheet generally have smaller extreme events than previously seen as the two basins on the southern peninsula of ice sheet (basins 5 and 6) have the lowest mean proportion of co-occurrent exceedance cells. The main area that differs to the entire ice sheet fit is the north west. Here, the basin model produces events that cover more cells than would have been predicted from the full model, suggesting that the local behaviour
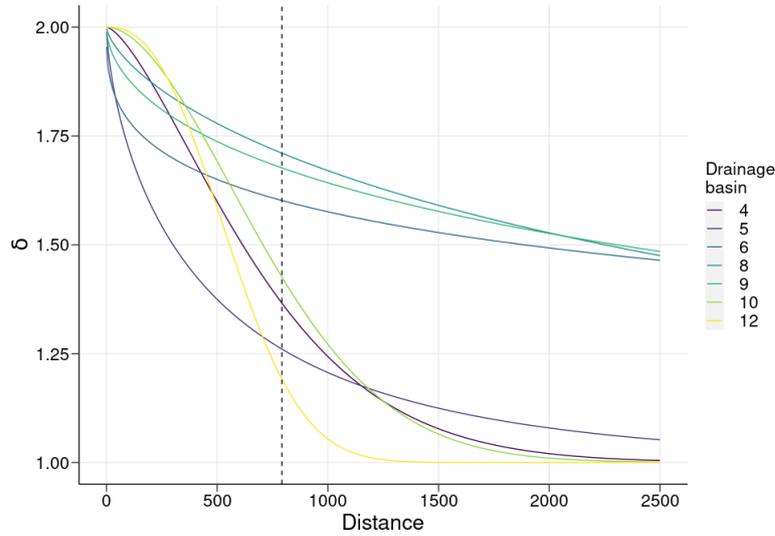
Figure 62: Function for δ evaluated over a range of distances using the fitted model parameter estimates for each drainage basin.

in this region is distinct from the overall trends seen across the ice sheet. The north east also has relatively large events, while the east has generally smaller events than the west of the ice sheet.

The mean temperatures of the extreme events (Figure 64) reflect the same broad trends as for the entire ice sheet fit. This may in part be due to the common marginal models used in both model fits. The estimated melt probabilities also resemble the previously seen trends, although with some different trends which are potentially due to the higher resolution. Similar coastal areas show the highest melt probabilities. The trends in the centre of the ice sheet have only minor differences, which can be attributed to sampling variability rather than to changing characteristics.

5.5.3   Central area fit

Lastly, we fit the model to a smaller sample of cells in the centre of the ice sheet. The sample consists of 81 cells approximately 40 km apart in latitude and longitude in a square grid, close to the centre of the ice sheet. We apply the model here to examine events at a much smaller scale than in previous applications, and to examine the model's behaviour in an area that may have more spatial stationarity compared to
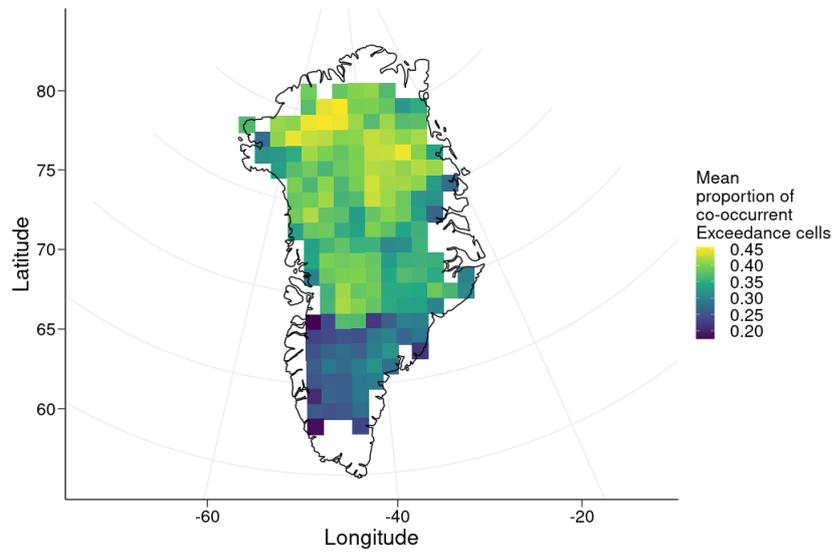
133

Figure 63: Mean proportion of co-occurrent exceedances from 1000 simulations considering each cell as the conditioning cell for the drainage basin fits.
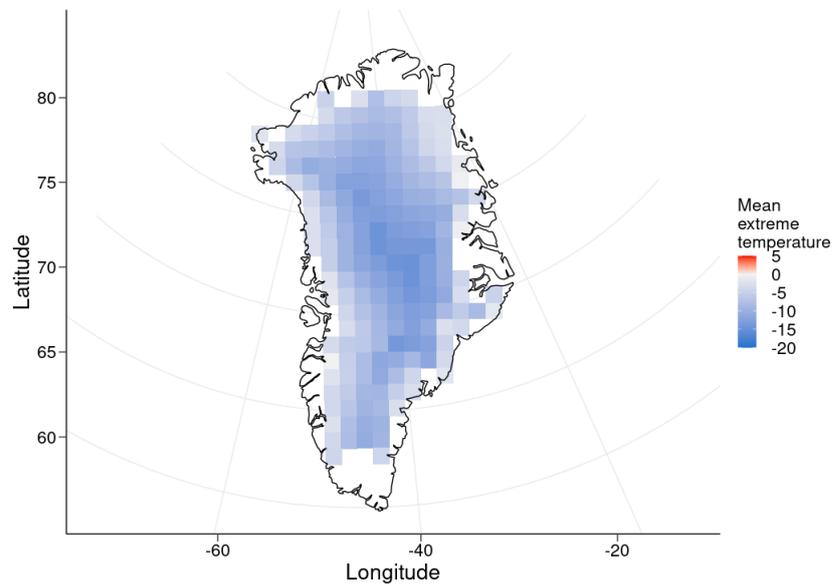

Figure 64: Mean temperatures of each cell averaged across simulations including all cells as conditioning cells for the drainage basin fits.

areas closer to the coasts.

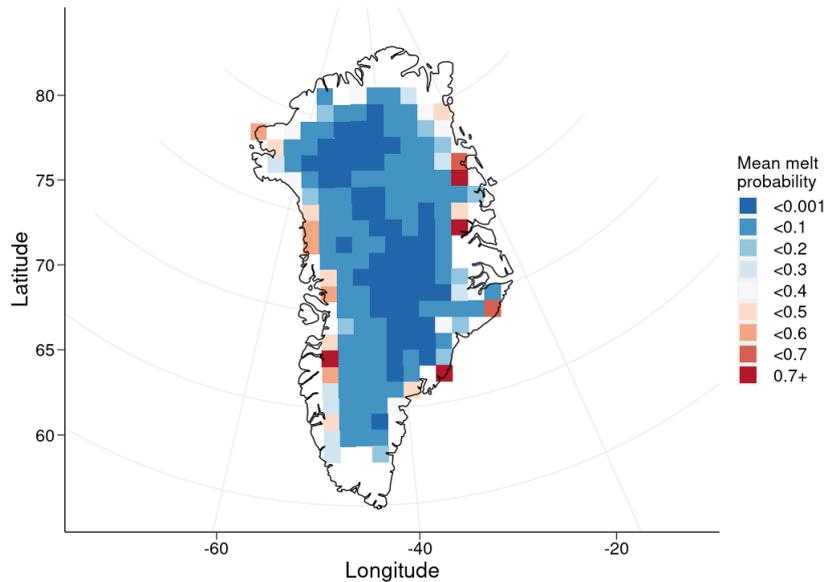From the simulated events (Figure 66) we see that this smaller scale model

Figure 65: Mean melt probability for each cell from spatial extreme event simulations for the drainage basin fits.

predicts event sizes much closer to those seen in the data than was the case for the entire ice sheet fit (Figure 57), and in fact the model slightly overestimates the size of the extreme events. The absolute difference in the size of the events is larger than for the entire ice sheet model, however the events for the central area are also much larger in general. As a proportion of the number of cells in the events, the event sizes are more realistic for the central area than for either of the other applications.

The estimates for the dependence parameters in Figure 67 have the same overall form as the estimates for the drainage basins and entire ice sheet, but differ distinctly in how quickly the dependence decays. For the entire range of distances observed within the central sample, the estimates are above 0.3 for a and above 1.3 for b. This shows that within the area, all cells have at least some dependence with all other cells. This dependence does not even decay to 0 for the largest distances seen across the ice sheet. However, this does require extrapolating the function to distances not observed in the data. The estimates for a and b closely resemble each other in terms of shape, largely due to the estimate of β being close to 1.

The estimates for δ decrease almost linearly (Figure 68). Extrapolating to the
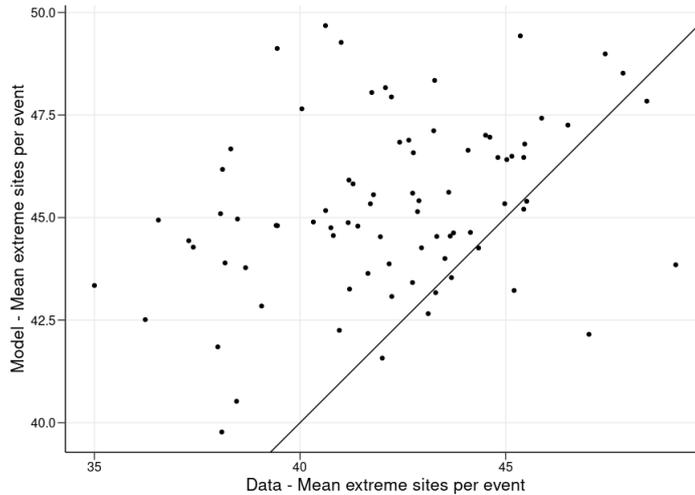
Figure 66: Number of exceedances calculated empirically from the data compared to simulations from the spatial extremes model for the central area fit, both conditioning on the same cell for the central area fit.

full range of the data, the estimates most closely resemble the δ estimates observed in the drainage basins around the centre latitudes of the ice sheet, which makes sense considering its location. For the central area, the distributions stay close to two and thereby normal distributions. The lowest value observed within the sample is around 1.6, clearly not close to one as would be expected if the data was independent across the region. This adds further weight to the inference that all of the cells in the central area are linked regardless of distance between them, and suggests that the trends observed in δ in the drainage basin fits are not particularly influenced by the exclusion of coastal cells.

From Figure 69, we see that the mean size of the simulated events show little spatial trend. Although not as informative of trends in the extreme properties of the area, this demonstrates that the region selected appears to have spatial stationarity in its extreme events and that this property can be identified by the model. Furthermore, the importance of scale can be seen, since such a clear trend was not observed by the other applications.

Similarly, the mean temperatures are also quite consistent over the central area (Figure 70). Rather than the variation in the mean temperatures, the main point
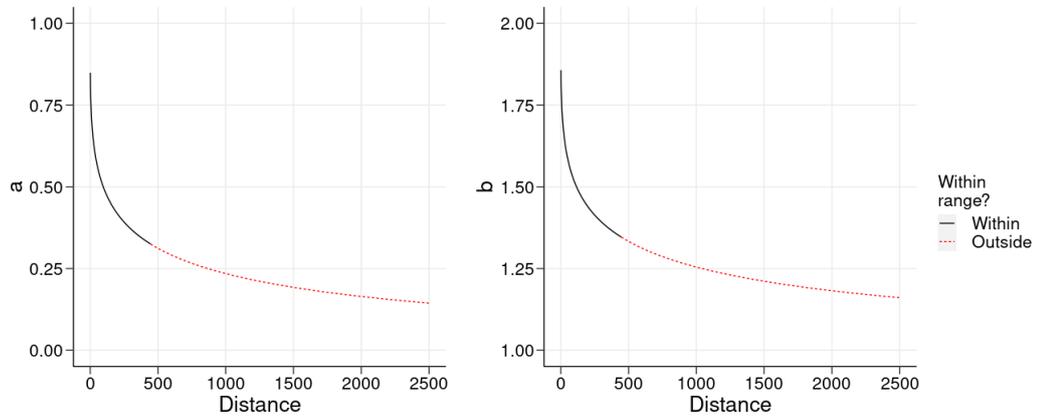
136

Figure 67: Functions for a and b evaluated over a range of distances using the fitted model parameter estimates for the central area fit.
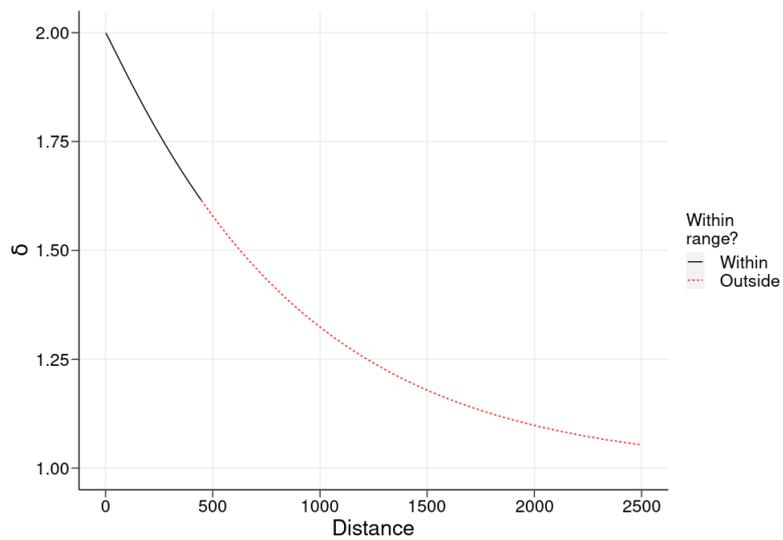

Figure 68: Function for δ evaluated over a range of distances using the fitted model parameter estimates for the central area fit.
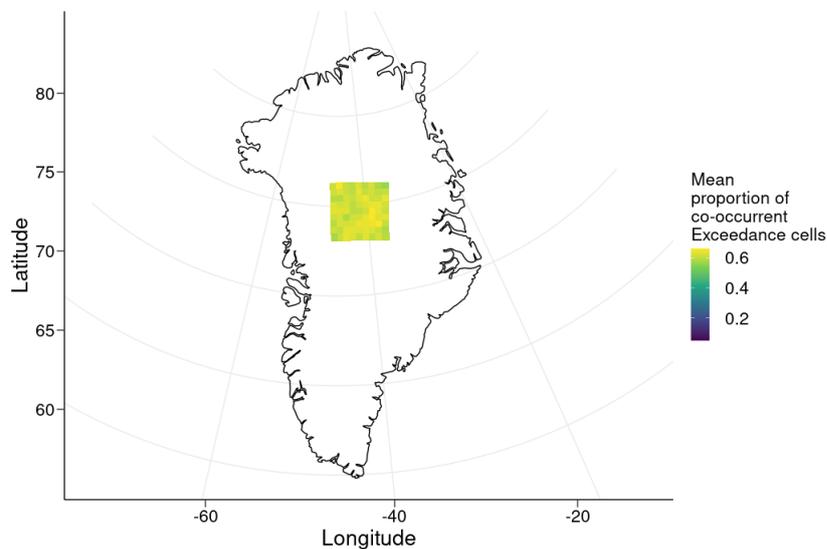
Figure 69: Mean proportion of co-occurrent exceedances from 1000 simulations considering each cell as the conditioning cell for the central area fit.

of interest is that the temperatures are consistently below 0° C, so that even when modelling only the extreme temperatures in the centre of the ice sheet, they are still not that high in absolute terms. Although the model is applicable only on non-cloudy days, this does highlight the rarity of temperature events with temperatures close to melt levels that affect the entire ice sheet. Even an average extreme temperature event in this central area does not see temperatures close to the melt thresholds.

The rarity of melt from these extreme events is confirmed by the mean melt probabilities as estimated from the model simulations (Figure 71). The results reflect the same trends observed for this region from the previous two models. There is a low chance of observing melt at some cells, alongside a negligible probability for a large proportion of cells. There appears to be a minor trend between the melt probabilities and elevation, although considering the magnitude of the probabilities this could also be due to other sources of variation in the region. Even from extreme temperatures events as defined by the model, the centre of the ice sheet has a less than 0.1 probability of observing melt even at the most likely cell to observe melt.
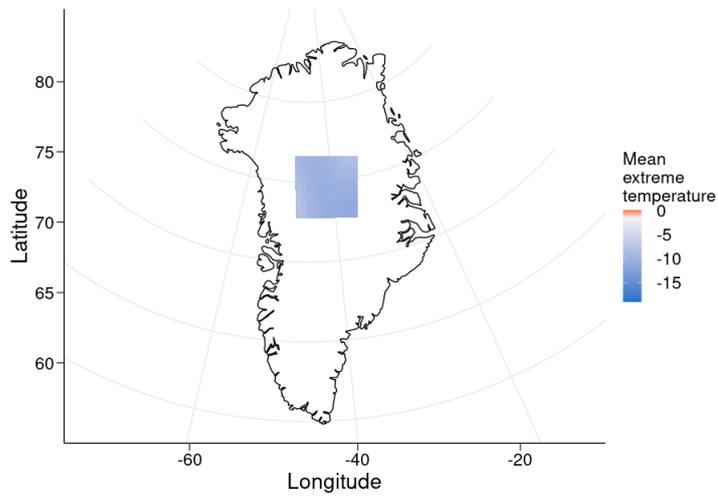
Figure 70: Mean temperatures of each cell averaged across simulations including all cells as conditioning cells for the central area fit.
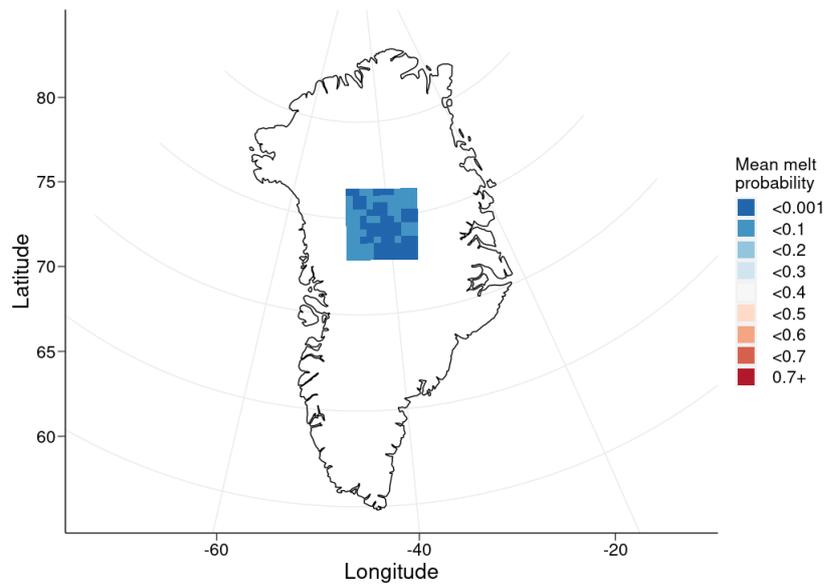
Figure 71: Mean melt probability for each cell from spatial extreme event simulations for the central area fit.

5.6 Discussion

The spatial conditional extremes model captures the form of extreme temperature events and enables them to be simulated given an exceedance at a particular cell. The model builds upon our understanding of dependence from Section 4, focusing

on only extreme temperatures rather than all temperatures. The modelling work in Section 4 gives some indication of what we might expect from the extremal dependence structure, but it is important to look at it in isolation rather than to assume that the same dependence is present in all aspects of the data. This enables us to examine the characteristics of extreme temperature events in more detail than can be achieved from finding isolated examples of extreme temperature events in the data. The model structure and fitted parameters provide insight into the scale and shape of extremal dependence in the data. Furthermore, the simulations from the model can be used to examine the magnitude of temperatures typically seen in extreme temperature events.

The model furthers our ability to examine temperatures by allowing us to simulate extreme temperature events that may cover many cells. Whereas the Gaussian process models simulated temperature events, we can choose a particular site - or any given site - to have an extreme temperature and examine the behaviour of the rest of the sample area based on this condition. Because the model only captures extreme temperature events, the simulations from the model are all instances of the specific phenomena - extreme temperature events - that we are interested in. We therefore require fewer simulations from the model than we would from the Gaussian process models to look at extreme events, and can be more specific with what our simulations show in terms of sites that have an exceedance. We can also have more confidence that the dependence functions and covariance structure found from the model are representative of the extreme temperatures.

The dependence structure of the model applications demonstrate one of the biggest challenges of working with the MODIS data set; the scale of the data and of the ice sheet. The ice sheet has reasonably smoothly varying conditions across its surface, but even the few main variables that influence these result in a complex overlay of trends at different scales and in different directions. This complexity makes it difficult for a single model application to accurately capture every trend seen in the data. Instead, model fits applied to more specific scales or areas may

provide more nuance as to the trends observed than a full ice sheet wide model is capable of.

Although different scales of application can be used for the model, a possible extension that could aim to capture some of the ice sheet's surface conditions is spatial non-stationarity. Including a non-stationary kernel could allow the model to capture more complex trends in the ice sheet in a single model fit, rather than through multiple model fits of smaller areas of the ice sheet. This would further add to the model's already complex structure, but could provide useful insight into some of the trends observed from the stationary kernel fits. This may be of particular interest when comparing coastal and central cells, given their differences in characteristics from previous models.

6   Discussion

In this thesis we have used a statistical modelling approach to develop a detailed understanding of the trends and properties of melt and surface temperature on the Greenland ice sheet. Starting with simple summary statistics of a complex, and previously unexplored, spatio-temporal data set, we have developed progressively advanced models to develop our scientific understanding in specific and useful ways. This includes an understanding of single site behaviour, spatial trends in single site behaviour across the ice sheet, spatial dependence, and finally spatial dependence of extreme temperatures. A core theme running through the thesis is that each part of the analysis uses contextual scientific knowledge of the ice sheet and builds on insights from the previous analyses to build representative models that answer specific questions regarding ISTs.

In Section 2, we define a Gaussian mixture model that is capable of modelling ISTs at locations anywhere on the ice sheet using a consistent model structure. The choice of modelling ISTs using a single distribution allows a comparable description of the IST distribution to be found in what is otherwise highly cyclical and

spatially-varying data. A limitation of the approach as it stands is the assumption of stationarity in time. Time series modelling techniques could be applied to find a different form for the marginal model that includes seasonality, long term climate change-related trends or inter-year variability, which would require a similar consideration of the upper IST temperature limit as made by the mixture model. This would provide a different framework for future models to build on, but could also be used to more specifically describe temperatures at a particular time of year rather than for any time of year. We tested some initial ideas based around Fourier regression techniques for this type of modelling, but found more use and better fitting models using the mixture model. Instead of building an alternative model structure, time could be built into the current mixture model framework by including regression type terms to capture long term trends.

Furthermore, the definition of melt defined in our model could be extended via validation from an external data source. In the absence of information about the state of the surface from MODIS, we infer that any temperature above $-1.65°$ C could correspond to a melt temperature. Our definition uses the physical constraints on water and ice, however, the melt probabilities could be checked using data, such as that obtained from passive microwave readings, that identifies whether temperatures refer to ice or water (melt) temperatures. By linking a passive microwave data set with MODIS data, each temperature observation could be linked to a passive microwave observation to examine which temperatures relate to melt and which to ice. We tested this idea using a logistic regression framework, whereby temperatures were regressed using the passive microwave data (Mote, 2014). Unfortunately, the link between the data sets was not consistent enough to give a reliable estimate of melt probability that could be applied at all cells. This may have been due to the differing spatial resolutions of the data sets, in addition to the difference in time of observation.

Having defined our marginal mixture models, we examined the spatial trends in the mixture model fits. This gave us insight into the spatial trends of ISTs,

142

the covariates most likely to help explain spatial trends, and, using a regression model, we were able to predict the mixture model parameters for locations without observations. Despite the relatively simple methods, there was sufficient information in the spatial variables to predict the mixture model parameters with a good degree of accuracy. We could extend the regression approach to use machine learning methods which allow us to identify non-linear response-covariate relationships. Considering the scale of MODIS data, if more powerful computational resources were available then the work could be expanded include to more mixture models in the fitting process, potentially making machine learning tools viable. We tested neural network models to predict the parameters given a set of input spatial variables, however the fit of the models was poor due to insufficient training data. More powerful computational resources could allow for more mixture models to be fitted, giving more training points for a neural network and allowing faster training of the model. If this were the case, considerations could be made towards the most useful sample of the data for a machine learning model. Closer to the coasts or in areas of rapid elevation change, samples could provide more information than in more uniform areas of the ice sheet such as closer to the centre. Further covariates could be added to the model too, such as aspect of the ice sheet, however due to availability of data at the time of writing, the variables used were the only variables considered.

Our first view of the spatial dependence structure of the data came from our Gaussian process models in Section 4. These models captured the spatial dependence of the data a multivariate normal distribution with covariance based on Euclidean distance as well as other considered measures of distance. Of these, difference in distance to the coast proved to be the most useful for the model fitting. This makes sense as the ocean has a substantial impact on the weather and therefore temperatures, so cells at similar distances from the coast are more likely to experience similar temperatures. This could be said in some capacity for the other spatial variables too since there is overlap between them, but the modelling shows

that in practice distance to the coast is most informative specifically for modelling dependence. As with the regression work in Section 3, other variables could have been included in the work, and thereby other distance metrics would have been calculated. The main extension that could have been applied to this model is to add non-stationarity to the Gaussian processes. Although some of the non-stationarity was accounted for from the split of the ice sheet into two drainage basin regions, having a fully non-stationary kernel could result in an improved fit for the model. For our work, the stationary kernels gave a useful, interpretable view of dependence in the data set.

The final Section 5 examined the spatial extremal dependence in the data set, building from the Gaussian process work to answer a more specific question about the nature of extreme temperatures. The spatial conditional extremes model can be used to characterise the dependence of extremes at different areas of the ice sheet by estimating parameters in localised fits, and then use simulations to describe and predict extreme temperature events across the ice sheet. As with the Gaussian processes, this model could also be extended to consider non-stationarity, particularly considering the different trends observed in extremal dependence in different areas of the ice sheet. Previous methods based around alternative definitions of distance could be applied to the spatial conditional extremes model, giving it more flexibility to capture extreme temperature events. Fits of the model could also be applied to other subsets of the ice sheet, potentially defined by their glaciological zone rather than by drainage basins as is done in our analysis.

Having established the spatial characteristics of extreme temperature events, adding time to the spatial extremes model in different ways could be useful for a broader understanding of extreme events over time. The current model fits describe extreme temperature events in a given time period, but could be extended to show how these events change over time. This could either be done by fitting the model to subsets of time and examining how the parameters and simulations change between time periods, or by including time in the model structure itself. The second of

these ideas would consider when the extreme events in the data occur, in order to generate events in a particular year or other time period. This could use the year of the event as a covariate to fit an overall trend to, rather than having to separate out the data into distinct time periods to observe temporal trends. This would enable a description of extreme temperature events over time, to examine if events are becoming larger or changing form in certain areas of the ice sheet.

The approach taken in this thesis of gradual development of applied statistical modelling greatly helped to shape the analysis around the research questions of greatest importance. Each model used has been motivated by the data and the specific goal of each modelling step, without limits to a particular category of statistical models. What began as an intended application of extreme value methods became an analysis using a much broader range of modelling techniques. Rather than attempt to apply the intended models regardless of the appropriateness to the context at hand, we adapted our entire modelling approach to fit the data that we observed. This included building up our analysis and models progressively rather than immediately applying more complex models. The benefits of this are that once we apply the more complex spatial models, we already have insight into the expected relevant spatial variables and a deep understanding of the marginal behaviour to fully translate any spatial trends back to individual cells.

This work was only possible due to the combined input from environmental experts and statisticians working in tandem. Statistical models alone without contextual information would have been unable to fully capture the temperature process observed. The soft upper limit of ISTs, nature of the missing data, and expected characteristics of the ice sheet could only be captured by the models thanks to domain specific knowledge. Conversely, without the statistical modelling framework, even with domain knowledge we would be limited to simple summary statistics. It is the combination of domain knowledge and statistical modelling that allows us to accurately model complex environmental processes such as ISTs. This creates analyses that provide useful scientific inference while simultaneously

developing statistical methodology in the presence of new challenges posed by the data.

1 Appendix: EM algorithm

1.1 Truncated normal distribution

Let $X \sim TN(\mu, \sigma^2, a, b)$ where μ is the mean, σ is the standard deviation, and a (b) is the lower (upper) truncation point. Furthermore, let $\alpha = \frac{a-\mu}{\sigma}$ and $\beta = \frac{b-\mu}{\sigma}$. Then X has probability density function:

$$f_{TN}(x) = \frac{f_N(\frac{x-\mu}{\sigma})}{\sigma(F_N(\beta) - F_N(\alpha))},$$

where f_N and F_N are the probability density function and the cumulative distribution function of a standard normal distribution respectively.

1.2 Algorithm

Let $(\mu_k, \sigma_k, \alpha_k, \beta_k)$ denote the parameters for the kth truncated normal distribution. To initialise the algorithm, randomly sample without replacement three values of $x \in X$ and set them as μ_k for $k = 1, 2, 3$. We set $\mu_4 = 0$ to ensure that one of the model components starts in the region of melt temperatures. Let σ_k be the sample variance and the component weights $\phi_k = 1/4$ for $k = 1, 2, 3, 4$. For simplicity we refer to the truncated normal probability density function and cumulative distribution function as $f(x)$ and $F(x)$ respectively. The EM algorithm consists of iterating between two stages, the expectation and maximisation steps, until convergence is obtained. For the expectation step, we set:

$$\hat{\gamma}_{ik} = \frac{\hat{\phi}_k f\left(x_i \mid \hat{\mu}_k, \hat{\sigma}_k\right)}{\sum_{j=1}^{4} \hat{\phi}_j f\left(x_i \mid \hat{\mu}_j, \hat{\sigma}_j\right)}$$

where $\hat{\gamma}_{ik}$ is the estimated probability that observation i belongs to model component k.

For the maximisation step, let:

$$\hat{\phi}_k = \sum_{i=1}^{N} \frac{\hat{\gamma}_{ik}}{N}$$

$$\hat{\mu}_k = \frac{\sum_{i=1}^{N} \hat{\gamma}_{ik} x_i}{\sum_{i=1}^{N} \hat{\gamma}_{ik}} + \hat{\sigma}_k \left(\frac{f(\alpha_k) - f(\beta_k)}{F(\beta_k) - F(\alpha_k)} \right)$$

$$\hat{\sigma}_k^2 = \frac{\sum_{i=1}^{N} \hat{\gamma}_{ik} (x_i - \hat{\mu}_k)^2}{\sum_{i=1}^{N} \hat{\gamma}_{ik}} \left[1 + \frac{\alpha_k f(\alpha_k) - \beta_k f(\beta_k)}{F_N(\beta) - F_N(\alpha)} - \left(\frac{f(\alpha_k) - f(\beta_k)}{F_N(\beta) - F_N(\alpha)} \right)^2 \right].$$

We iterate between these two steps until the parameters converge to the final estimates (800 iterations was sufficient in this case). The algorithm is considered to have converged if the difference between parameters in each iteration is sufficiently small. We found a difference of 10^{-5} between iterations to be sufficient indication of convergence for all parameters.

Bibliography

Acero, F., García, J. A., Gallego, M. C., Parey, S. and Dacunha-Castelle, D. (2014), 'Trends in summer extreme temperatures over the Iberian Peninsula using nonurban station data', *Journal of Geophysical Research: Atmospheres* **119**(1), 39–53.

Benson, C. S. (1960), Stratigraphic studies in the snow and firn of the Greenland ice sheet, phd, California Institute of Technology.
URL: *https://resolver.caltech.edu/CaltechETD:etd-03232006-104828*

Burke, E. J., Perry, R. H. and Brown, S. J. (2010), 'An extreme value analysis of uk drought and projections of change in the future', *Journal of Hydrology* **388**(1-2), 131–143.

Cazenave, A. and Nerem, R. S. (2004), 'Present-day sea level change: Observations and causes', *Reviews of Geophysics* **42**(3). _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2003RG000139.
URL: *https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2003RG000139*

Cheng, L., AghaKouchak, A., Gilleland, E. and Katz, R. W. (2014), 'Non-stationary extreme value analysis in a changing climate', *Climatic change* **127**(2), 353–369.

Cooley, D. (2009), 'Extreme value analysis and the study of climate change', *Climatic change* **97**(1), 77–83.

Cristóbal, J., Ninyerola, M. and Pons, X. (2008), 'Modeling air temperature through a combination of remote sensing and GIS data',

Journal of Geophysical Research: Atmospheres **113**(D13). _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2007JD009318.

URL: *https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007JD009318*

Davison, A. C. and Smith, R. L. (1990*a*), 'Models for exceedances over high thresholds', *Journal of the Royal Statistical Society: Series B (Methodological)* **52**(3), 393–425.

Davison, A. C. and Smith, R. L. (1990*b*), 'Models for Exceedances over High Thresholds', *Journal of the Royal Statistical Society. Series B (Methodological)* **52**(3), 393–442.

URL: *https://www.jstor.org/stable/2345667*

Eastoe, E. F. and Tawn, J. A. (2009), 'Modelling non-stationary extremes with application to surface level ozone', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **58**(1), 25–45.

Fausto, R. S., van As, D., Mankoff, K. D., Vandecrux, B., Citterio, M., Ahlstrøm, A. P., Andersen, S. B., Colgan, W., Karlsson, N. B., Kjeldsen, K. K., Korsgaard, N. J., Larsen, S. H., Nielsen, S., Pedersen, A. , Shields, C. L., Solgaard, A. M. and Box, J. E. (2021), 'Programme for Monitoring of the Greenland Ice Sheet (PROMICE) automatic weather station data', *Earth System Science Data* **13**(8), 3819–3845. Publisher: Copernicus GmbH.

URL: *https://essd.copernicus.org/articles/13/3819/2021/*

Fawcett, L. and Walshaw, D. (2006), 'A hierarchical model for extreme wind speeds', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **55**(5), 631–646.

Fettweis, X., Tedesco, M., van den Broeke, M. and Ettema, J. (2011), 'Melting trends over the Greenland ice sheet (1958–2009) from spaceborne microwave data and regional climate models', *The Cryosphere* **5**(2), 359–375. Publisher: Copernicus

GmbH.
URL: *https://tc.copernicus.org/articles/5/359/2011/*

Friederichs, P. (2010), 'Statistical downscaling of extreme precipitation events using extreme value theory', *Extremes* **13**(2), 109–132.

Gnedenko, B. (1943), 'Sur la distribution limite du terme maximum d'une serie aleatoire', *Annals of mathematics* pp. 423–453.

Gouldsborough, L., Hossain, R., Eastoe, E. and Young, P. (2021), A Temperature Dependent Extreme Value Analysis of UK Surface Ozone, 1980 – 2019. Submitted.

Gumbel, E. J. (1958), *Statistics of Extremes*, Columbia University Press.
URL: *https://doi.org/10.7312/gumb92958*

Hall, D. K., Cullather, R. I., DiGirolamo, N. E., Comiso, J. C., Medley, B. C. and Nowicki, S. M. (2018), 'A Multilayer Surface Temperature, Surface Albedo, and Water Vapor Product of Greenland from MODIS', *Remote Sensing* **10**(4), 555. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.
URL: *https://www.mdpi.com/2072-4292/10/4/555*

Hall, D., Key, J., Casey, K., Riggs, G. and Cavalieri, D. (2004), 'Sea ice surface temperature product from modis', *IEEE Transactions on Geoscience and Remote Sensing* **42**(5), 1076–1087.

Hanna, E., Fettweis, X., Mernild, S. H., Cappelen, J., Ribergaard, M. H., Shuman, C. A., Steffen, K., Wood, L. and Mote, T. L. (2014), 'Atmospheric and oceanic climate forcing of the exceptional Greenland ice sheet surface melt in summer 2012', *International Journal of Climatology* **34**(4), 1022–1037.
URL: *https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.3743*

Hanna, E., Navarro, F. J., Pattyn, F., Domingues, C. M., Fettweis, X., Ivins, E. R., Nicholls, R. J., Ritz, C., Smith, B., Tulaczyk, S., Whitehouse, P. L. and Zwally,

H. J. (2013), 'Ice-sheet mass balance and climate change', *Nature* **498**(7452), 51–59. Number: 7452 Publisher: Nature Publishing Group.
URL: *https://www.nature.com/articles/nature12238*

Heffernan, J. E. and Tawn, J. A. (2004), 'A conditional approach for multivariate extreme values (with discussion)', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66**(3), 497–546. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9868.2004.02050.x.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2004.02050.x*

Heinonen, M., Mannerström, H., Rousu, J., Kaski, S. and Lähdesmäki, H. (2016), Non-Stationary Gaussian Process Regression with Hamiltonian Monte Carlo, *in* 'Proceedings of the 19th International Conference on Artificial Intelligence and Statistics', PMLR, pp. 732–740. ISSN: 1938-7228.
URL: *https://proceedings.mlr.press/v51/heinonen16.html*

Hooker, J., Duveiller, G. and Cescatti, A. (2018), 'A global dataset of air temperature derived from satellite remote sensing and weather stations', *Scientific Data* **5**(1), 180246. Number: 1 Publisher: Nature Publishing Group.
URL: *https://www.nature.com/articles/sdata2018246*

IPCC (2013), Summary for Policymakers, *in* T. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P. Midgley, eds, 'Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change', Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1–30. Section: SPM Type: Book Section.
URL: *www.climatechange2013.org*

Jonathan, P. and Ewans, K. (2013), 'Statistical modelling of extreme ocean environments for marine design: a review', *Ocean Engineering* **62**, 91–109.

Katz, R. W. (1999), 'Extreme value theory for precipitation: sensitivity analysis for climate change', *Advances in water resources* **23**(2), 133–139.

Katz, R. W., Parlange, M. B. and Naveau, P. (2002), 'Statistics of extremes in hydrology', *Advances in water resources* **25**(8-12), 1287–1304.

Keef, C., Tawn, J. and Svensson, C. (2009), 'Spatial risk assessment for extreme river flows', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **58**(5), 601–618. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9876.2009.00672.x.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9876.2009.00672.x*

Keribin, C. (2000), 'Consistent estimation of the order of mixture models', *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)* **62**(1), 49–66.
URL: *http://www.jstor.org/stable/25051289*

Koenig, L. S. and Hall, D. K. (2010), 'Comparison of satellite, thermochron and air temperatures at Summit, Greenland, during the winter of 2008/09', *Journal of Glaciology* **56**(198), 735–741.
URL: *https://www.cambridge.org/core/journals/journal-of-glaciology/article/comparison-of-satellite-thermochron-and-air-temperatures-at-summit-greenland-during-the-winter-of-200809/63FA4BACF7C88FCA057BA4BC2C4EEFD7*

Kulp, S. A. and Strauss, B. H. (2019), 'New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding', *Nature Communications* **10**(1), 4844. Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Climate-change impacts;Projection and prediction Subject_term_id: climate-change-impacts;projection-and-prediction.
URL: *https://www.nature.com/articles/s41467-019-12808-z*

Mernild, S. H., Mote, T. L. and Liston, G. E. (2011), 'Greenland ice sheet surface melt extent and trends: 1960–2010', *Journal of Glaciology* **57**(204), 621–628. Publisher: Cambridge University Press.

URL: *https://www.cambridge.org/core/journals/journal-of-glaciology/article/greenland-ice-sheet-surface-melt-extent-and-trends-19602010/92981710E4A25E39E1C0156EC993ED33*

Mote, T. L. (2007), 'Greenland surface melt trends 1973–2007: Evidence of a large increase in 2007', *Geophysical Research Letters* **34**(22). _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2007GL031976.

URL: *https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007GL031976*

Mote, T. L. (2014), 'Measures greenland surface melt daily 25km ease-grid 2.0, version 1'.

URL: *https://nsidc.org/data/NSIDC-0533/versions/1*

Nghiem, S. V., Hall, D. K., Mote, T. L., Tedesco, M., Albert, M. R., Keegan, K., Shuman, C. A., DiGirolamo, N. E. and Neumann, G. (2012), 'The extreme melt across the Greenland ice sheet in 2012', *Geophysical Research Letters* **39**(20).

URL: *https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2012GL053611*

Nicholls, R. J. and Cazenave, A. (2010), 'Sea-Level Rise and Its Impact on Coastal Zones', *Science* **328**(5985), 1517–1520. Publisher: American Association for the Advancement of Science.

URL: *https://www.science.org/doi/full/10.1126/science.1185782*

Pandita, P., Tsilifis, P., Awalgaonkar, N. M., Bilionis, I. and Panchal, J. (2021), 'Surrogate-based sequential Bayesian experimental design using non-stationary Gaussian Processes', *Computer Methods in Applied Mechanics and Engineering* **385**, 114007.

URL: *https://www.sciencedirect.com/science/article/pii/S0045782521003388*

Pickands, J. (1975), 'Statistical inference using extreme order statistics', *the Annals of Statistics* pp. 119–131.

Rahmstorf, S., Foster, G. and Cahill, N. (2017), 'Global temperature evolution: recent trends and some pitfalls', **12**(5), 054001. Publisher: IOP Publishing.
URL: *https://doi.org/10.1088/1748-9326/aa6825*

Reeh, N. (1991), 'Parameterization of Melt Rate and Surface Temperature in the Greenland Ice Sheet'. ISSN: 0032-2490 Issue: 3 Number: 3 Pages: 113-128 Place: Bremerhaven Publisher: Alfred Wegener Institute for Polar and Marine Research & German Society of Polar Research Volume: 59.
URL: *https://epic.awi.de/id/eprint/28262/*

Reich, B., Cooley, D., Foley, K., Napelenok, S. and Shaby, B. (2013), 'Extreme value analysis for evaluating ozone control strategies', *The annals of applied statistics* **7**(2), 739.

Rignot, E., Mouginot, J., Scheuchl, B., van den Broeke, M., van Wessem, M. J. and Morlighem, M. (2019), 'Four decades of Antarctic Ice Sheet mass balance from 1979–2017', *Proceedings of the National Academy of Sciences* **116**(4), 1095–1103. Publisher: Proceedings of the National Academy of Sciences.
URL: *https://www.pnas.org/doi/10.1073/pnas.1812883116*

Rignot, E., Velicogna, I., van den Broeke, M. R., Monaghan, A. and Lenaerts, J. T. M. (2011), 'Acceleration of the contribution of the Greenland and Antarctic ice sheets to sea level rise', *Geophysical Research Letters* **38**(5). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2011GL046583.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1029/2011GL046583*

Rogers, N. C., Wild, J. A., Eastoe, E. F., Gjerloev, J. W. and Thomson, A. W. (2020), 'A global climatological model of extreme geomagnetic field fluctuations', *Journal of Space Weather and Space Climate* **10**, 5.

Smith, D. M., Cusack, S., Colman, A. W., Folland, C. K., Harris, G. R. and Murphy, J. M. (2007), 'Improved Surface Temperature Prediction for the Coming Decade from a Global Climate Model', *Science* **317**(5839), 796–799. Publisher: American Association for the Advancement of Science Section: Report.
URL: *https://science.sciencemag.org/content/317/5839/796*

Sterl, A., Severijns, C., Dijkstra, H., Hazeleger, W., van Oldenborgh, G. J., van den Broeke, M., Burgers, G., van den Hurk, B., van Leeuwen, P. J. and van Velthoven, P. (2008), 'When can we expect extremely high surface temperatures?', *Geophysical Research Letters* **35**(14).

Tedesco, M., Fettweis, X., Mote, T., Wahr, J., Alexander, P., Box, J. E. and Wouters, B. (2013), 'Evidence and analysis of 2012 Greenland records from spaceborne observations, a regional climate model and reanalysis data', **7**(2), 615–630.
URL: *https://doi.org/10.7916/D8J38SGV*

Thomson, A. W., Dawson, E. B. and Reay, S. J. (2011), 'Quantifying extreme behavior in geomagnetic activity', *Space Weather* **9**(10).

Towe, R., Eastoe, E., Tawn, J. and Jonathan, P. (2017), 'Statistical downscaling for future extreme wave heights in the North Sea', *The Annals of Applied Statistics* **11**(4), 2375–2403.

van den Broeke, M. R., Enderlin, E. M., Howat, I. M., Kuipers Munneke, P., Noël, B. P. Y., van de Berg, W. J., van Meijgaard, E. and Wouters, B. (2016), 'On the recent contribution of the Greenland ice sheet to sea level change', *The Cryosphere* **10**(5), 1933–1946. Publisher: Copernicus GmbH.
URL: *https://tc.copernicus.org/articles/10/1933/2016/*

Vermeer, M. and Rahmstorf, S. (2009), 'Global sea level linked to global temperature', *Proceedings of the National Academy of Sciences* **106**(51), 21527–

21532.

URL: *https://www.pnas.org/content/106/51/21527*

Von Mises, R. (1936), 'La distribution de la plus grande de n valuers', *Rev. math. Union interbalcanique* **1**, 141–160.

Wadsworth, J. L. and Tawn, J. A. (2018), 'Spatial conditional extremes', p. 23.

Winter, H. C., Tawn, J. A. and Brown, S. J. (2016), 'Modelling the effect of the El Nino-Southern Oscillation on extreme spatial temperature events over Australia', *The Annals of Applied Statistics* **10**(4), 2075–2101.

Zhengming, W. and Dozier, J. (1989), 'Land-surface temperature measurement from space: physical principles and inverse modeling', *IEEE Transactions on Geoscience and Remote Sensing* **27**(3), 268–278. Conference Name: IEEE Transactions on Geoscience and Remote Sensing.