A statistical fracture model for Antarctic glaciers

Veronika Emetc¹, Paul Tregoning¹, and Malcolm Sambridge¹

¹Research School of Earth Science, Australian National University, Canberra, Australia

Correspondence to: Veronika Emetc (Veronika.Emetc@anu.edu.au)

Abstract.

Antarctic and Greenland hold more than 99% of all fresh water on Earth and, therefore, can significantly influence global sea level. Predicting future ice sheet mass balance depends upon ice sheet modelling, but it is limited by knowledge of a number of processes, some of which are still poorly understood. One such process is the calving of the ice shelves, where blocks of ice break off from the ice front. However, large scale ice flow models do not include an accurate representation of this process and the most commonly used damage mechanics and fracture mechanics methods have a large number of uncertainties. Here we present an alternative, statistics-based method to model the most probable zones of nucleation of fractures. We test our theory on all main ice shelf regions in Antarctica, including the Antarctic Peninsula. We can model up to 99% of observed fractures, with an average rate of 77% which represents a 50% improvement over previously used damage-based approaches, thus providing the basis for modelling calving of ice shelves. We found that classifying Antarctic ice shelf regions based on the factors that controlled fracture formation led to grouping of ice shelves/glaciers with similar physical characteristics and geometry.

Keywords. Antarctica, glaciers, probability, calving, ice shelves, fracture nucleation, crevasse, logistic regression, bayesian

1 Introduction

In recent years, increased positive temperature anomalies have been observed in Antarctica (Jansen et al., 2007; Vaughan et al., 2003; Johanson and Fu, 2007; Steig et al., 2009). Future climate changes in this area may be even more pronounced (Vaughan et al., 2003), which may cause the state of the Antarctic ice sheet to change significantly. This could lead to a release of fresh water currently stored in the ice sheet and a consequent rise in sea level. West Antarctica alone can contribute up to 4.3 metres to global sea level (Fretwell et al., 2013). Almost 30% of the global population lives in coastal areas (Small and Nicholls, 2003), therefore, the effect of sea level rise (SLR) can significantly impact the economy and society in these regions. However, projections of global sea level rise have a number of uncertainties, with the greatest related to the calving rate of the ice shelves (Jansen et al., 2007). Thus, understanding the factors that control the mass balance of the Antarctic ice sheet is crucial if we want to better understand the future impact of climate change and contribution of Antarctic ice mass loss to global SLR.

The mechanics of ice mass loss of Antarctica is controlled by three processes: surface ablation, basal melting and calving (Paterson, 2000), where the later relates directly to SLR for grounded ice and contributes to buttressing effect on floating ice shelves. Due to the fact that the surface ablation in Antarctica is relatively low, it was previously believed that the Antarctic ice
Increased calving from the major ice shelves between 1998 and 2003 led to a growing concern of the total ice sheet stability (Borstad et al., 2013). It has been shown that the main part of ice mass loss from Antarctica comes from an increased number of iceberg calving events (Mercer, 1978; Jacobs et al., 1992; Katz and Worster, 2010; Gudmundsson, 2013; Borstad et al., 2013) and estimates of the calving effect on the mass balance are as high as 50% (Depoorter et al., 2013; Rignot et al., 2013). Also, studies by Jezek (1984); De Angelis and Skvarca (2003); Dupont and Alley (2005); Goldberg et al. (2009); Katz and Worster (2010); Gudmundsson (2013); Borstad et al. (2013) showed that increased calving can lead to destabilization of ice shelves and thus to a loss of the supporting mechanism they provide on the inland ice in Antarctica. This support can be crucial for the overall stability of the ice sheet (Miles et al., 2013).

However, none of the IPCC projections of the future state of Antarctica considers the calving mechanisms (Stocker, 2014). Research on crevasse propagation started as early as 1955 and calving parametrisation has being under development for the last 20 years. However, a method that can universally describe calving at any ice shelf in Antarctica has not been found. In most of the large scale ice sheet models this process is still either described with a large number of limitations or not implemented at all (e.g. Benn et al. (2007b); Stocker (2014); Bassis (2011)). There are a number of calving parameterisations, but still none of them includes propagation of fractures in both depth and latitude-longitude space. Also, most of the available models are specific to a particular case and set of predictors, or a location, and therefore can not be applied to any chosen ice shelf. Many experiments have tried to find an optimal way of implementing a full calving process to ice sheet models, but none of them has worked for all locations in Antarctica. However, a number of different approaches based on Linear Fracture Mechanics (LEFM), Continuum Damage Mechanics (CDM) or "crevasse-depth" have advanced the understanding of the calving process. The history of development of these methods is presented in Table 1.

ELMER/Ice and PISM models only include calving built on simplified physics (Alley et al., 2008). PISM can perform a calving algorithm based on eigen vector theory introduced by Levermann et al. (2012) (Eq. 1) (Winkelmann et al., 2011), which is only a first order approximation and does not include initiation and propagation of crevasses. They calculate calving rate as:

\[ C = K \det(\dot{\varepsilon}) = K\dot{\varepsilon}_+\dot{\varepsilon}_-, \text{ for } \dot{\varepsilon}_+ > 0, \]  

where \( K \) is a proportionality constant and \( \dot{\varepsilon}_\pm \) are the eigenvalues of the horizontal strain rate tensor.

Another method they use is calculating a calving rate based on the critical ice thickness, which is mainly used to model calving of marine terminated glaciers rather than floating ice shelves (due to different physics governing calving between the grounded ice and floating ice). Also, most of the experiments with ELMER/Ice calving were performed for Greenland glaciers, which have a different calving mechanism to the floating ice shelves in Antarctica (Van der Veen, 2002).

In Table 1 we can see that to date Continuum Damage Mechanics (Kachanov, 1958) and Linear Fracture Mechanics (Van der Veen, 1998a) are the most common methods to model fracture formation. Also, Krug et al. (2014) showed that combining these two methods provides a significant improvement towards modelling of calving.
In order to develop a reliable calving law we need to have a strong basis that would include both physics and observational data. In other words, if we want to model propagation of fractures and the consequent calving it is essential to know the location of these fractures on the ice shelf surface or where they are initiated. In this study, we focus on modelling of crevasses (surface fractures less than 200 metres wide) on the surface of the Antarctica ice sheet. We implement a method based on a probabilistic approach that can be used as an alternative to the damage-based approach mentioned above. Our method is build on the logistic regression algorithm (LRA) and includes finding a relationship between the observed fractures we take from satellite images and various factors such as geometry, mechanical properties, flow regime as well as climate forcing (that we use as predictors). Our best results were achieved combining LRA with Bayesian as well as Jensen-Shannon Divergence theory described in section 3. We apply our method to most of the ice shelf regions in both West and East Antarctica as well as the Antarctic Peninsula. From modelling of 34 ice shelves/glaciers we found that we can match the observations of fractures with a success rate from 18% to 99% with an average rate of 77% (Figure 1a).

2 Background

2.1 Current state of calving computations in ice sheet models

Damage mechanics is used to calculate zones where ice is weakened due to formation of small fractures. However, this method depends on a number of control predictors that have been calculated only for one specific glacier in Greenland (Duddu and Waisman, 2012) and therefore can lead to incorrect results when applied to a randomly selected antarctic ice shelf/glacier. Furthermore, in situ observations in Antarctica are not feasible due to a very large regional extent of the ice shelves as well as very high frequency of fracturing (sometimes every 50 metres). Also, satellite observations alone are difficult to use because of the very high resolution required to see all the small fractures and snow can often cover the fractures in the image. Therefore, the main factors that creates weaknesses in ice shelves/glaciers in Antarctica remain unclear.

The Linear Fracture Mechanics approach is focused on calculating the stress intensity around fractures and their propagation until the stress intensity reaches a certain critical value. Although the "crevasse-depth" criterion based on the shear rate estimation with depth can produce interesting results when applied to marine terminated glaciers in Greenland (Nick et al., 2010), it can not describe calving at the floating ice shelves in Antarctica due to the different mechanics. The Continuum Damage Mechanics approach, based on the method suggested by Kachanov (1958), includes estimation of damaged zones where the ice is softened due to presence of fractures. Damage can be calculated from the inversion of velocities at the ice shelves (Borstad et al., 2012) which is based on minimising the cost function that describes the misfit between the observed and modelled velocities. It is proposed that it can allow to model zones where the ice is weakened and thus we can estimate where the initiation of small fractures takes place. Krug et al. (2014) used damage mechanics to model the slow development of small fractures in ice. Consequently, the estimation of the depth of these initial fractures is performed using the stress intensity calculations, which is based on Linear Fracture Mechanics. It can describes propagation of the formed fractures with depth and a consequent calving that occurs mainly at the ice shelf front.
The ELMER/Ice model (Gagliardini et al., 2013) includes calving combining CDM and LEFM, but this method was performed for Greenland only (Krug et al., 2014) and has not been proven to work for any ice shelf in Antarctica. The Community Ice Sheet Model (CISM) assumes that calving takes place when the water depth reaches a certain value (Price et al., 2014), which is, as mentioned above, suitable only for tidewater glaciers, but not for ice shelves. The Parallel Ice Sheet Model (PISM) have an implemented calving parametrisation, but it is based on simplified physics and includes only along-flow ice spreading (Alley et al., 2008). They suggest two options for calving. The first is based on the assumption that calving takes place when thickness at the ice calving front reaches a critical ice thickness \( H_{cr} \sim 150 – 250\) m. The second applies the Eigen vector principal components proposed by Levermann et al. (2012), which is based on the correlation between the large-scale calving rate and the local ice-flow spreading rate. However, it considers only large-scale behaviour and does not take into account the formation and propagation of crevasses.

We present the main studies of calving in Table 1 and among them the most comprehensive parametrisation of calving by Krug et al. (2014) includes combining damage and fracture mechanics. This method is more complex in comparison to the other approaches and can help to understand calving mechanism in Antarctica much better. Apparently, the propagation of fractures can be modelled only if the fractured zones are known or computed in a good agreement with the observed surface fractures. Therefore, modelling of the formation of the fractured zones is an important basis for the consequent estimation of the fracture depth as well as calving and it has to be described in the ice sheet models in an accurate way.

Damage is an arbitrary variable used to determine failure of ice and nucleation of fractures (usually when the damage predictor reaches a certain value) and can be described as a density function. However, the damage approach might not be able to provide us with reliable results that agree with observations, because the nature of fracturing of the ice sheet might be too complex to be comprehensively described by the methods of damage mechanics. Therefore, damage may not be justified for glacier ice, especially at the ice front (Levermann et al., 2012). The propagation of damage is usually calculated using an advection equation and a source function. However, this method has a number of uncertainties. First of all, the damage is not a physical predictor and modelling its propagation includes many complications when choosing an appropriate source function and when including a healing process. In fact, the source function is the controlling factor in the damage propagation, but the hypothesis of what should be used as the source function has not been well justified. The one that is used in the models can be constructed in either a complex or simplified way, but always includes many assumptions, arbitrary predictors and uncertainties (Pralong and Funk, 2005; Krug et al., 2014). Secondly, the estimation of the main predictors that determine damage is based on experiments that have a large number of uncertainties (Krug et al., 2014). It also does not take into account such important factors as ice fabric, impurities and non-linear behaviour due to presence of crevasses (Borstad et al., 2012). Also, it is sensitive to a number of constants such as initial damage \( D_0 \) (Borstad et al., 2012), especially when applied to a randomly chosen ice shelf/glacier. Thirdly, when calculating damage from an inversion in Ice Sheet System Model (ISSM) (Larour et al., 2012) (described in the next section) it is impossible to do an inversion on both the basal friction and damage since they have similar effects on the ice velocity (Morlighem, 2017). Therefore, the inversion for damage is performed only on floating ice, where there is no friction, and, thus, there is no information about the damage on grounded ice.
2.2 Ice Sheet System Model ISSM. Model setup

First, we used ISSM (Larour et al., 2012) to compute factors such as velocities, stresses, strains, backstresses, the dynamics of the ice sheet in time as well as friction coefficient and viscosity (calculated from inverse modelling) in order to use them in our statistical model as quasi-observations. All our experiments were performed for 34 regions in Antarctica (see Figure 1b), each including both ice shelves and the grounded ice around 100 kilometres from the grounding line (further referred as ice shelf regions or ice shelf/glacier). Second, we calculated damage on floating ice for each region from inversion of velocities (Borstad et al., 2013). This method finds a value of damage from an inversion based on minimising the misfit between observed and modelled velocities.

ISSM is a fully dynamic model that includes both 2-D and 3-D components and is based on the equations of conservation of mass, momentum and energy as well as Glen’s law (shows the relation between shear strain rate and shear stress). There are a number of approximations to solve these equations that are available in ISSM: Shallow Ice Approximation (SIA) is generally used for grounded ice and assumes that the height-to-width ratio is small, Shelfy Stream Approximation (SSA) and Pattyn’s higher order model (HO that can be applied for a 3D case). We modelled glacier dynamics in a two-dimensional case in order to achieve faster computational time while having a relatively high spatial resolution and apply the SSA approximation (assumes that the vertical shear is zero, thus velocities do not depend on depth (MacAyeal, 1989)) as it is more suitable for modelling velocities at the floating ice shelves in a 2-D case.

To perform a simulation ISSM requires forcing data, geometry of the region, boundary conditions as well as friction and rheology coefficient. We used seaRISE air temperature, snow accumulation and geothermal heat flux (Le Brocq et al., 2010) as climate forcing data. Also, we calculated changes of the surface temperature with the lapse rate and imposed it on the ice surface. The data for geometry of the ice shelves and surrounded grounded ice (such as bedrock topography, ice thickness and glacier surface) were taken from Bedmap2 at 1 km spatial resolution bedrock topography. We used them for both: as an input for modelling with ISSM and the predictor factors in our probabilistic method. Basal friction for grounded ice and rheology for floating ice were calculated from inversion of velocities (Khazendar et al., 2007). The horizontal ice velocities for the inverse modelling were taken from InSAR (450-metres resolution) (Rignot et al., 2011b, a) and we applied Dirichlet conditions at the inflow boundary. The ice mask to distinguish between grounded and floating ice is based on hydrostatic equilibrium (mask $\geq$ 0 if ice is present and mask $\leq$ 0 for the ocean). We set boundary conditions as follows: the upper surface is considered stress-free and at the ice-bedrock interface the friction is applied. The mean basal melting rate for grounded and floating ice (Depoorter et al., 2013) were set as 2 and 4 m/yr, respectively, and the thickening rate at the base of the floating ice (due to ice accretion) was set to 2 m/yr (a default setting in ISSM). It is a first approximation, but in this study we do not model the basal effect on horizontal pattern of fracturing at the surface. We ran one simulation for stress balance solution per region (ice shelf/glacier) which allowed us to obtain the factors required as an input in the calculation of the probability of fracturing.

In addition, it is important to have a finer resolution mesh in order to model surface fractures, as the distance between them is normally around 50-100 metres. However, using a mesh of 50-metre or 100-metre resolution creates a significant increase of the computational time in our model. Therefore, for our experiments, we constructed a 200-metre resolution triangular mesh.
for the chosen domains in East and West Antarctica as well as the Antarctic Peninsula. First, all the main predictors were calculated on a 200-metre resolution mesh (to achieve a faster computational speed) and then the results were interpolated to a 100-metre resolution mesh (to use in our fracture model resolved at 100-metres). All further computations and analysis were performed on this finer mesh.

5 Methods

To develop an alternative method for modelling fracture formation in ice shelves/glaciers, we took into account that the damage varies from 0 to 1, in a probabilistic sense with 0 being not fractured and 1 being fully fractured. We can substitute this probability with a likelihood function. Thus, to implement our ice calving into the ice sheet models, we developed a statistical-based parametrisation for fracture formation in ice shelves/glaciers. First, we present the main method (logistic regression) used for predicting formation of fractures. Second, we describe the predictor factors (predictors) we used in our analysis. Finally, two methods used for optimisation of a set of predictor factors are presented (Bayesian based algorithm and Jensen-Shannon Divergence).

3.1 The logistic regression algorithm (LRA)

The function $P$ is called a logistic function. Taking any range of data it produces values from 0 to 1 and thus it can be used as a probability (Hosmer Jr and Lemeshow, 2004). The logistic function allows us to calculate the probability of an event as a function of different predictor factors (see Table 2).

Our set of predictors $x_i$ initially consisted of 24 vector arrays. Some of them are known to influence fracture formation (stresses, strains, changes of the velocity gradient, ice rheology). Others we considered to be potentially important (various geometrical properties, proximity to the ice front and the grounding line, etc). We analysed the data for each predictors and found that there were a linear dependency between first and second components of the principal stress axis. Also, temperature and accumulation were excluded from the list of predictors due to the incompatibility of their spatial resolution with the relatively fine 100-metre mesh we used to model fractures. However, they might be important for the formation and propagation of fractures as warmer temperature can increase the number of fractures, but a better resolution climate data would be needed. Excluding these two factors, we obtained a set of 20 predictors that we describe further in this section.

To apply the logistic regression algorithm we constructed a function $P$ (Eq. 2) that describes the probability of fracture nucleation in a certain node as a function of the predictor factors $x_i$. Each $j$ element in the $P$ array can be calculated as:

$$P_j = \text{Prob}(X = 1|x_i) = \frac{\exp(\beta_0 + \beta_1 \cdot x_{1j} + \beta_2 \cdot x_{2j} + \beta_3 \cdot x_{3j} + \ldots)}{1 + \exp(\beta_0 + \beta_1 \cdot x_{1j} + \beta_2 \cdot x_{2j} + \beta_3 \cdot x_{3j} + \ldots)},$$  \hspace{1cm} (2)
where for $x_{ij}$ is an element in a predictor array ($i$ is the number of the predictor and $j$ is a row number).

The unknown coefficients $\beta_i$ can be found by maximising the likelihood function $L$ (Eq. 3).

$$L(\beta_j) = \prod_{j=1}^{n} P_j^{\delta_j} (1 - P_j)^{1-\delta_j},$$

(3)

where $n$ is the number of observations and $\delta$ is the Kronecker symbol:

$$\delta = \begin{cases} 
1, & \text{when there is a surface fracture visible on a satellite image} \\
0, & \text{otherwise.} 
\end{cases}$$

(4)

All the predictor factors are either taken from InSAR and Bedmap2 or calculated using formulas described below. The calculation of each predictor was performed using methods already implemented in ISSM (e.g. stresses, strains, friction coefficient) or by applying new algorithms (e.g. calculation of curvature, distances to ice front, grounding line, mountains). First, the effective strain rate $\dot{\epsilon}_E$ had to be included in our analysis because it is known that a crevasse initiation is linked to strain rates (Campbell et al., 2013). We calculated this predictor in ISSM as:

$$\dot{\epsilon}_E = \sqrt{\frac{\dot{\epsilon}_{xx}^2 + \dot{\epsilon}_{yy}^2}{2} + \dot{\epsilon}_{xy}^2},$$

(5)

where $\epsilon_{ik} = \frac{1}{2}(\frac{\partial v_i}{\partial x_k} + \frac{\partial v_k}{\partial x_i})$ and $v_i$ are horizontal components of the ice flow velocities (InSAR).

Also, we included principal strains, calculated as:

$$\mu = \frac{\epsilon_{xx} + \epsilon_{yy}}{2} \pm \sqrt{\left(\frac{\epsilon_{xx} - \epsilon_{yy}}{2}\right)^2 + \epsilon_{xy}^2},$$

(6)

Second, principal stress $\lambda$ is important as it has a direct effect on the opening and closing of crevasses. Also, principal axis $s$ shows the direction of the stresses (compressive or tensile):

$$\sigma \cdot s = \lambda s$$

(7)
Thirdly, when ice is flowing over a vertical bend on the surface it might experience fracturing, therefore we included the surface and bedrock slopes calculated from:

\[
\begin{align*}
\frac{dS}{dx} &= \sum_{i=0}^{M} \alpha_i S_i \\
\frac{dS}{dy} &= \sum_{i=0}^{M} \gamma_i S_i \\
\text{Slope} &= \left(\frac{dS}{dx}\right)^2 + \left(\frac{dS}{dy}\right)^2 
\end{align*}
\] (8)

where \( M \) is the number of vertices, \( S \) is the surface, \( \alpha_i = \frac{d\phi_i}{dx} \) and \( \gamma_i = \frac{d\phi_i}{dy} \) are the coefficients of a nodal function defined as \( \phi_i = \sum_{i=1}^{3} \alpha_i \cdot x + \gamma_i \cdot y \) (calculated for each triangular element on the mesh, the constant term is equal to 0).

Fourth, curvature of the glacier channel can play a role in fracture formation as we can observe more fractures around a horizontal bend of the glacier. The calculation of this predictor was performed using:

\[
\alpha = \arccos\left(\frac{v_x(P) \cdot v_x(E) + v_y(P) \cdot v_y(E)}{|v(P)| \cdot |v(E)|}\right),
\] (9)

where \( v(P) \) is the ice velocity at the point of observations and \( v(E) \) is the velocity \( D \) metres away from the point. The distance \( D \) is based on the velocity magnitude \( v(P) \), because if the velocity is high we need to increase \( D \) so that two subsequent points capture the geometry of the bend. Thus, if \( v(P) \) is greater than 2000 m/yr we assign \( D = 3v(P) \), otherwise \( D = 6v(P) \).

By looking at the satellite images we can see that more fractures occur at a certain distance from the ice front as well as the grounding line. We found that the relation between the occurrence of fractures and distance to the ice front as well as the distance to the grounding line is non-linear. For most ice shelves/glaciers we can see more fractures 3-5 km as well as 10-13 km away from the front and a slightly smaller number of fractures closer than 3 km to the front or between 5 and 10 km. Therefore, instead of using \( d_{IF} \) and \( d_{GL} \) (distance to the ice front and the grounding line in km) as predictor variables, we construct dummy variables: \( DM_{IF} \) and \( DM_{GL} \), respectively, which represent two-column arrays in the following form:

\[
DM_{IF} = \begin{cases} 
(1,1), & \text{when } 3 \text{ km} \leq d_{IF} < 5 \text{ km} \\
or & 10 \text{ km} \leq d_{IF} < 13 \text{ km} \\
(1,0), & \text{when } 5 \text{ km} \leq d_{IF} < 10 \text{ km} \\
(0,1), & \text{when } d_{IF} < 3 \text{ km} \\
(0,0), & \text{else}
\end{cases}
\] (10)

\[
DM_{GL} = \begin{cases} 
(1,1), & \text{when } 5 \text{ km} \leq d_{GL} < 15 \text{ km} \\
(1,0), & \text{when } d_{GL} < 5 \text{ km} \\
(0,1), & \text{when } 15 \text{ km} \leq d_{GL} < 20 \text{ km} \\
(0,0), & \text{else}
\end{cases}
\] (11)
Lastly, because all the predictor factors have different units as well as significantly different magnitudes we normalised all variables, using:

\[ x_i^* = \frac{x_i - \mu_i}{\sigma_i} \], where \( \mu_i \) and \( \sigma_i \) are mean and standard deviation of the predictor variables, respectively. (12)

5 3.2 Observing fractures using satellite images

In order to obtain information about the location of fractures on the ice sheet surface we used satellite images taken from Google Earth-Pro, where images of the Antarctic ice sheet were available at different spatial resolutions. However, to be able to see small surface fractures, our choice was limited only to the images with a resolution smaller than 10 metres for the period between 2011 and 2015. Also, we included only regions with a good satellite coverage (at least one hight resolution satellite image) and where it is relatively easy to identify surface fractures.

Many features can be observed on the ice surface and it is important to distinguish the surface fractures from other patterns such as surface troughs due to bottom crevasses or subglacial channels. It was suggested by Luckman et al. (2012) that the features on the images that are wider and have a larger spacing between them are more likely to be troughs linked to bottom crevasses. Also, Alley et al. (2016) proposed a way to distinguish basal channels and fractures on the satellite images. They classified channels into sub-glacially sourced, ocean-sourced and grounding-line-sourced. Most of them either follow the ice flow direction and begin either at the grounding-line or downstream. Therefore, we avoided including such features in the set of chosen surface fractures.

In addition, it is important to choose satellite images of different regions in Antarctica in order to build correct predictions for fracturing \( P_i \) as the diversity in sampling provides a better estimation of the \( \beta \) coefficients (the number of observation points is less important). Thus, by choosing multiple glaciers we can more accurately construct an approximate surface that separates fractured from non-fractured nodes (the plane is determined by \( \beta \) coefficients).

To construct the set of observed fractures we manually selected fractured points as well as non-fractured points on the satellite images. Non-fractured points are more difficult to identify, therefore we mainly take them from areas where high-resolution images are available and they are mostly located in blue ice regions. These are the areas with low snow accumulation or where the snow was removed by the wind. In such areas we can clearly see where the ice is not damaged. We construct a field of fractures by assigning fracture nodes to 1, non-fractured nodes to 0. We select fractured nodes where we can see surface fractures that do not have features cause by present basal fractures or channels.

The resolution is not high enough in all areas to clearly see every fracture and also some of them can be covered in snow. This creates a large uncertainty in cells where we are not able to observe any visible fractures. Sometimes it is impossible to say whether there are no fractures or whether we just cannot see them. Thus, we corrected a field of fractures by using the probability of observed fractures instead of the observations of fractures in order to make the observation field continuous as well as to account for the uncertainties in our ability to observe accurately whether or not areas are fractured. Firstly, we assumed that if there is a fracture in one node then the neighbour nodes are more likely to be fractured as well. In the node
where we could not see a fracture we assigned the observed probability to 1. Within a radius of 500 metres away from the fracture we decreased the probability linearly from 1 to 0.55. If a non-fractured node was found within a high resolution area we assigned the probability of fracturing in this node to 0.05. Within a 500 metres radius of non-fractured nodes we allowed the probability to increase linearly from 0.05 to 0.4. In all other nodes we set a value of 0.5, assuming that, since the information in the satellite image is ambiguous, it is an equal probability of being fractured or non-fractured.

3.3 Optimisation problem

3.3.1 Random walk

To implement our method we used a 100-metre resolution models for 34 ice shelf regions (including floating ice and surrounded grounded ice). We started with an apriori model of fracturing and then improved it based on three methods: random walk, Bayesian and Jensen-Shannon Divergence algorithm. The main difficulty in constructing the probability function for each glacier was to identify a set of the predictor factors we need to include in LRA. As the logistic regression formula was based on the correlation coefficients between the observed fracture field and a certain set of 20 predictors in our model, we needed to determine which predictors we need to include in the formula. For each glacier we performed a 100000-step run with random sets of predictors for each step (number of predictors and the selection of predictors are selected at random every step). We defined the best-fitting model to be a result with a success of identifying fractures more than 70% and the error of over-estimation smaller than 15% (however, if a good-fit model was not found after 2000 steps we looked for a model with a 65% success and 20% error). Once a good fit was found, we saved it and continued running the model with different sets of predictor factors for the remaining number of steps to see if a better model can be found. This also provided a mean set of factors needed for a good-fit model.

3.3.2 Bayesian

To test the behaviour of the models with more precision and to choose an optimal set of factors from the full set we also performed a non-linear Bayesian inversion. This process has the advantage of allowing us to take into account uncertainties in the observed data.

We assume that the prior PDF is uniformly distributed between 1 and 20 (U[1,20], because the maximum number of predictor factors in a set is equal to 20). As a prior model we take a calculated probability at every time step. Each step included two criteria: if a new likelihood is greater than the prior likelihood or it is greater than a certain percentage (taken at random at each step) of the old likelihood we accept the model. However, the main difficulty arose when we tried to find the expression for a likelihood function for our model. We tested a number of different commonly used expressions, such as:

\[
\begin{align*}
L_i &= \sum \log(f_i) \\
fi &= (1 - pi)^{1 - di} + pi^{di},
\end{align*}
\]  

(13)
\[
\begin{align*}
L_i &= \sum \log(f_i) \\
f_i &= (1 - p_i) \cdot (1 - d_i) + p_i \cdot d_i,
\end{align*}
\]

(14)

where \(d_i\) and \(p_i\) are observed and modelled fractures on a glacier, respectively.

However, all of them produced very large likelihoods that increased dramatically with a percentage change in probability density function, reaching an order of \(10^5\). Thus, it was important to choose a representative likelihood function. We constructed the likelihood function assuming that the measure \(R\) of the total agreement between two models (the sum of all probabilities) follows a Gaussian distribution with a mean \(E\) (Eq. 15) and a standard deviation as a square root of variance \(\sigma\) described in Equation 18.

We calculated the expected values for both the data and a chosen model as:

\[
E(f_i^{\text{pred}}) = \sum_{i=1}^{N} f_i
\]

(15)

\[
E(f_i^{\text{obs}}) = \sum_{i=1}^{N} d_i^2 + (1 - d_i)^2
\]

(16)

\[
E(f_i^{\text{best}}) = \sum_{i=1}^{N} p^* \cdot d_i + (1 - p^*) \cdot (1 - d_i),
\]

(17)

where \(p^*\) is the best-fit probability and the probability \(f_i\) that two predictions agree in a cell \(i\) is calculated using Equation 14.

Second, we found the variance as a difference between the two expected values (Eq. 18).

\[
\sigma = |E(f_i^{\text{obs}}) - E(f_i^{\text{best}})|
\]

(18)

Our idea was to calculate the likelihood \(L_i\) as an exponential function of the misfit between data and the model assuming that either data (observed fractures) or the analysed model have an error (Eq. 19). Where misfit \(\phi_d(m)\) can be calculated as the square of a ratio between the expectation for the data minus the expectation for the model and the variance of the data.

\[
L_i = e^{-\frac{1}{2} \phi_d(m)}, \text{ where } \phi_d(m) = \frac{(E(f_i^{\text{obs}}) - E(f_i^{\text{pred}}))^2}{\sigma^2}.
\]

(19)

In addition, the area of each ice shelf region is an order of \(10^8\ km^2\) which leads to a very large sum of all modelled probabilities between 0 and 0.5 and therefore an extremely large likelihood (it is important to notice that this values should not be
confused with 0 and 1 values set for observed fractures only). In order to achieve a more realistic magnitude of the likelihood function we needed to re-calculate the estimated probabilities by scaling them between 0.55 and 1. To do this we assigned everything below 0.55 to non-fractures (zero values) and scaled the remaining values to the range 0 to 1.

For a prior model and prior scores, we took the best-fitting model from the random walk search described above. First, we perform a Bayesian analysis for 500 steps and after we narrowed down the selection and we accepted only those models that have likelihoods greater than 90% of the best likelihood.

3.4 Glaciers classification and Jensen-Shannon Divergence (JSD)

We started with a construction of a binary array for each glacier, where the number of rows represent the number of good-fitting models for a glacier and the number of columns represent the 20 predictor factors. Next, we found the average occurrence of each predictor:

\[ A_i = \frac{\sum_{j=1}^{N} k_j}{N}, \]  

(20)

where \( i \in [1, 20] \) is the predictor index, \( N \) is the number of good-fitting models and \( k_j = 1 \) when the predictor is included in the good-fit model \( j \) and 0 otherwise.

This allowed us to find how often a certain predictor was included in good-fit models. If a predictor was selected more than 50% of the time then it was assigned as a potential for a best-fit mode, otherwise it is set to 0. Thus, we obtained a 34x20 array (34 glaciers vs. 20 predictors) that consisted of 1 when the predictor can be included in the best-fit model and 0 otherwise.

Next, we classified the glaciers in 4 groups. There were a large number of possible combinations to select such groups. Therefore, we constructed a test that can run through every possible combination and calculate the percentage of a similarity between glaciers in a group (Eq. 21).

\[ S = \frac{M}{21} \cdot 100, \]  

(21)

where \( M \) is the number of matches between sets of predictors for two glaciers and \( S \) is a group number.

Thus, we placed all 34 glaciers in 4 different groups with Group 1 having glaciers that can be more easily combined and Group 4 being a more narrow group of specific glaciers that can not be placed in any of the other three groups.

In order to validate of our method we applied the additional Jensen-Shannon Divergence method (JSD) to identify the similarity between the best probability for each glacier and a probability calculated by placing the glacier in a certain group.
JSD formula is widely used in statistics to measure a divergence of one probability distribution from another. The Kullback-Leibler divergence is defined as:

\[
JSD(P_1||P_2) = \frac{1}{2}D(P_1||M) + \frac{1}{2}D(P_2||M),
\]

(22)

where \(D(P_1||M) = \sum(P_1 \log(P_1/M))\), \(D(P_2||M) = \sum(P_2 \log(P_2/M))\) are conditional PDFs, \(M = \frac{P_1 + P_2}{2}\) and \(P_1\), \(P_2\) are the new (probability when assigning a glacier to a new group) and the old probabilities (the best-fit model), respectively. It is important to note that both probabilities \(P_1\) and \(P_2\) have to be normalised before applying Equation 22.

Thus, JSD can be used as a tool to measure the distance between two distributions and can provide a value that we can use to assign a particular glacier to one of the groups.

4 Results

We applied the LRA method combined with the random walk method to 34 ice shelf regions that include both ice shelves and surrounding grounded ice (the corresponding names and the location can be found in Table 5 and Fig. 1b, respectively) and found a best-fitting model for 32 of them. The fracturing of the other two ice shelves/glaciers cannot be described using the predictors we have, producing too large or too small probabilities. In total, for each ice shelf/glacier the random walk analysis gave us a number of possible sets of predictors that can produce a good-fitting model. We combined all of these possible sets for each glacier to see which predictors are always present in the good-fitting model and which ones are never included. The results of the random walk and the Bayesian showed a good agreement. Most of the essential predictors for each particular glacier selected in Bayesian were also chosen when performing the random walk. In most cases, the Bayesian analysis showed equal importance of most of the predictors although effective strain rate and velocity had a slightly higher rate. There were no universal set of factors that could be used to model all ice shelf regions. However, subsets of glaciers had some similarities in terms of the predictors that had to be included in order to achieve a good-fitting model.

Thus, we assigned the 34 glaciers into 4 groups not allowing the deviation from the best-fit models to exceed 5%. After that we tested if it was the optimal choice by estimating the deviation from the best solution using the Jensen-Shannon Divergence algorithm. We assigned each glacier to a particular group based on its minimum value of the deviation from the best-fitting model in JSD analysis. By doing this we slightly modified the members of each group that we had previously created. For example, Glacier 27 belonged to Group 1 previously and it fit well with only a slight change of the best-fit score. However the JSD showed that if we move this glacier to Group 2 the deviation from the best-fit decreases from 0.01 to 0.003. Also, the JSD algorithm measures the total distance to the best-fit probability, therefore we had to take into account the fact that it can decrease the overestimation error but at the same time significantly decrease the success rate (Glaciers 10,13, 15, 11, 30, 32).

Thus, since these six glaciers are of a similar type to the glaciers in Group 1 and their JSD was similar for Group1 and Group 2 (e.g JSD=0.02 in Group 2 and JSD=0.0205 in Group 1) they were assigned to Group 1 to avoid a decrease in the success rate of identifying fractures.
Finally, to obtain a higher number of identified fractures we assigned each glacier to a particular group and the set of factors for each group are presented in Table 3 and Table 4, respectively. The characteristics of each group are described below.

4.1 Group 1

This was the largest group of glaciers and the best-fit model includes all the predictors except for the bed and surface slopes. The analysis of the estimated coefficients in LRA showed that predictors with the highest weight in our model for this group of glaciers were: effective strain rate, proximity to mountains and the surface elevation gradient (see Table 6). We present some of our results in Figs. 2d and 3b. The main pattern of surface fractures is well represented in our model for Vanderford IS (see Fig. 2d) even though high resolution images were not available for this glacier. There were still a small number of areas where there are no observed surface fractures but the model showed as high as 80% chance of fracturing. On the other hand, Drygalski ice shelf has a larger number of high resolution areas and, as a possible result, less over-estimation of fracturing (see Fig. 3b). We can see that the "definitely non-fractured nodes" (selected in blue ice areas) are successfully represented in our model. For this glacier none of the observed non-fractured nodes was assigned to have a high probability of fracturing. In these nodes the modelled probability is as low as 0.1. Also, we can see that in the regions with a large number of observed fractures the probability is as high as 0.9 and it’s slightly less in the areas with a lower number of observed fractures (between 0.6 and 0.8).

We compared our results to the damage-based approach based on damage inversion and the results for the Cook ice shelf are shown in Fig. 7b. Our probability-based method can capture the fracture pattern much better than the damage-based approach. The damage approach shows almost zero chance of fracturing in many areas where the frequency of the observed fractures is very high. Also, the over-estimation of fractures on the eastern side of the glaciers is more significant than in the probability-based approach. In some areas where we do not see fractures, our method assigns the probability to non-zero values, but they are mostly of a lower magnitude, not exceeding 50% chance of fracturing (e.g Figs. 4a and 3b).

4.2 Group 2

The second group of glaciers has the best-fit when the bed slope and one of the principal stress axes components are excluded. Effective strain rate and surface slopes were found to be the most important predictors in the model for this group (see Table 6).

Our model almost precisely represents the observed fracture pattern for Larsen D ice shelf (Fig. 2b). In most of the regions with observed fractures the modelled probability reaches 0.9, except for the small fracture pattern on the left hand side of Fig. 2b, where there are no high resolution images available. In all cases the model represents the non-fractured nodes with high precision. We can see that the field of high probability of fracturing does not include the observed non-fractured nodes. Similar situations are observed for Edward VII IS and Rayner Thyner IS (Fig. 2c and 2a, respectively). The modelled probability captures most of the fractured as well as non-fractured nodes with an exception of a few very small regions that are outside of the high resolution image areas.
Comparing our results with the damage method for Shirase IS (Fig. 3e) we found that damage did not capture any observed fractures in areas where the velocities are relatively low. All the high resolution areas on the right hand side are well represented in the LRA model, whereas the damage-based approach incorrectly assigns 0 values in areas where fractures are observed. Also, in Fig. 5b for Rennik IS we see that the damage-based method completely ignores the observed pattern of fractures. In comparison, the probabilistic method captures most of the surface fractures with a very low over-estimation rate. Very interesting results were found for Mariner IS (see Fig. 6d) showing a very good agreement between our model and the observations. The observations show that there are no fractures in the centre of the ice streams. We can see that the nodes observed to be non-fractured and located within the high-resolution regions are also showed to have less chance of fracturing when using LRA method. This probabilistic approach can capture the observed fractures, whereas the damage-based model shows that the ice is damaged wherever the flow velocities are high.

4.3 Group 3

These glaciers were very sensitive to the choice of predictor factors and the JSD process could not assign them to either of the two groups mentioned above. Overall this group includes 4 ice shelf regions: Totten IS, Nivl IS, Dibble IS and Holmes IS. Interestingly, for Totten only a certain set of factors can produce a good fit to observed fractures. For this glacier, including the back stresses in the model produces a slightly smaller number of modelled fractures. Potentially, in the model for this group we could include the proximity to the ice front since it produces slightly better results for 3 of the 4 glaciers. However, it lowers the success rate for Nivl IS significantly. Thus, in order to achieve a set that would give a good-fitting model for all of the glacier we exclude back stress and the ice front proximity from the list of predictor factors for this group. Overall, unlike the other two groups above, the most important predictors for fracturing in this group are strain change, ice thickness and velocity (see Table 6).

When using the damage-based model for Nivl and Abbot IS (see Fig. 4b and 4d) the ice is completely un-fractured in the regions with a large number of observed fractures that can be captured using LRA (Fig. 4a and 4c). For Dibble IS (see Fig. 5d) the damage method produces no damage in the observed fractured areas and high damage where we do not see fractures, whereas the LRA method captures the fracture pattern with just a slight over-estimation of fractures at the front of the ice shelf. Finally, we present the result for the Totten IS (see Fig. 7d). The images we used for this glacier were very hard to interpret due to the presence of many features on the ice surface as well as the low resolution of the imagery. Our model captures all the fractures we could observe on the images, but the over-estimation is relatively hight at the front. However, it is still able to represent the overall fracture pattern better than the damage-based model.

4.4 Group 4

Overall, this group includes Larsen C, Amery, George IV and Borchgrevnik IS. The most important predictor factors in this group are effective strain rate, proximity to mountains and surface change (see Table 6). For all of the glaciers in this group we found that including the ice front and grounding line proximity distorts the model. It significantly increases the error due to over-estimation of fractures. For Borchgrevnik IS it also leads to a drop of the success rate of fracture prediction. Also,
Larsen C and Amery ice shelves can be grouped together but cannot be included in any of the groups mentioned above. For these glaciers only a small number of predictors need to be included in the model. We used the effective strain rate, and the strain change, principal axis stresses and principal values, curvature, friction coefficient and surface elevation change. Bayesian analysis also confirmed the sensitivity of the Amery fracture model to this set of predictor factors.

4.5 Summary

To estimate how well our probability model and the damage model identify observed fractures we calculated the percentage of success and error for each ice shelf/glacier model. First, we found the number of cases when there is a modelled fracture in the vicinity of an observed fracture (within 100-metres radius). Then, we divide this number by the total number of observed fractures and find the percentage of success as:

$$P_s = \frac{S}{N} \cdot 100,$$

where $S$ is the number of successfully identified fractures and $N$ is the total number of observed fractures.

To find the percentage of failure we calculated how many times there is a modelled fracture when there are no observed fractures within a 100-metres radius. We divide this number by the total number of non-fractured nodes and find the failure (error) as:

$$P_f = \frac{F}{NF} \cdot 100,$$

where $F$ is the number of overestimated fractures and $NF$ is the total number of observed non-fractures.

The results for damage on a 100-metre mesh show that the method is able to identify only 34% of all fractures on average with a maximum of 64% of fractures with an overall over-estimation error of 14%. For all four groups we can see that damage does not capture regions of fractured nodes and even did not capture the main pattern, showing no risk of fracturing where there are fractures observed on satellite images and high chance of fracturing in regions where there are no fractures (see Fig. 5d, 5b, 7b). On the other hand, our method is able to identify up to 99% of the exact location of fractures with the average 77%, with an average over-estimation error of 23%, which is only slightly higher than the damage error. We can also see that where we could not achieve a high score using LRA the damage-based method did not produce a high success score either (see Fig. 1a). Overall, there were no cases when damage method would correctly identify more fractures than LRA method (see Fig. 1a).

5 Glacier Characteristics

The formation of fractures is a very complex process that cannot be effectively described by only applying damage mechanics. We showed that our statistical approach can be a very strong tool for identifying the location of fractures on the surface of ice
shelves/glaciers. We observe the strongest dependence of fracturing on principal axis components, strain change, curvature, friction coefficient, proximity to mountains as well as ice rheology. In most of the cases we can see that damage is highly correlated to the regions where the flow velocities are high. However, as we show using our method the process of fracture initiation is more complex and can be governed by a very large range of predictors apart from just ice velocity and stress rates.

5.1 Group 1

The ice shelves/glaciers in this group has a number of characteristics that distinguish them from other ice shelves/glaciers. Most of them are relatively wide with a large floating area. The floating part is not restricted by any channel walls and the width of the shelf is similar to its length. All glaciers in this groups are relatively static, with less curvature or significant surface elevation changes. However, there were two exceptions: Abbot IS and Drygalski IS have slightly different characteristics. First, Abbot is a wide glacier that carries most of the properties of Group 1. However it has a large number of glaciers that restrict its outflow towards the ocean and, therefore, it has similarities with the glaciers from Group 4. This observation is in good agreement with the JSD results that showed that Abbot can be assigned also to Group 4 as the change in JSD distance in this case would be very small. Second, the JSD results showed that Drygalski IS could be as well placed in Group 2 or Group 3. We can see that this ice shelf has some characteristics similar to Group 2 (large number of mountains) and Group 3 (a very long floating tongue).

Therefore, we suggest that some glaciers have mixed features of Group 1-Group 2 (such as Vanderford) or Group 1-Group 3 (Ekstroöm, Tracy-Tremenchus, Rennik), however they still have more characteristics of Group 1 and produce a better-fitting results when assigned to this group.

5.2 Group 2

This group includes relatively smaller number of ice shelves/glaciers. All of the ice shelves/glaciers have a large amount of mountains, smaller ice thickness as well as many small narrow channels and fast ice streams. They are mostly located at the Antarctic Peninsula or near the Trans-Antarctic Mountains. They all have many small channels and fast ice streams. Also, all of the ice streams are relatively steep which may explain why it is necessary to include surface slopes in order to achieve a good-fitting probability.

5.3 Group 3

Group 3 glaciers were found to have many similar features. Most of the ice shelf regions have one relatively long glacier that flows inside an embayment, for most of them the length is much higher than the width. Also, they all have a very low glacier channel curvature. The velocities of these ice shelves/glaciers are relatively high which can explain why the changes in strain rate and velocity are the most important predictors for this group.

Interestingly, although the average back stress at Totten is one of the highest out of all 34 ice shelf regions, including it in the model does not significantly change the probability of a fracture. Thus, apparently, even having large magnitudes these predictors make just minor contributions to the constructed probability and the other predictors govern the fracture formation
for the Totten IS. The effective strain rate is also one of the highest for Totten, but we found that it is not this predictor that mostly contributes to fracture formation, rather it is the effective strain rate gradient. Thus, sudden changes in the flow regime of the glacier would more likely promote an increase of the number of fractures.

5.4 Group 4

The JSD analysis has shown that Borchgrevink IS could also be assigned to Group 1, but it produced slightly better results being placed in Group 4. On the other hand, Amery and Larsen C need to be strictly assigned to Group 4 only. George IV IS and Amery IS have similar characteristics as they are both narrow and long (in fact much longer than any other glaciers of this type in Antarctica) and are located inside an embayment. Furthermore, George IV IS and Larsen C IS are both located next to each other on the Antarctic Peninsula. Although Larsen C IS is not inside an embayment, it is a significantly long and narrow ice shelf stretching around the coast. Borchgrevink IS also has similar features to the Amery and George IV ice shelves as it of a similar shape and is located inside a narrow channel. However, it does not have exactly the same characteristics as the other ice shelf regions in this group as it is much shorter, which could be why JSD showed that it can be placed in Group 1.

On the Amery IS (see Fig. 3c) most of the fracturing occurs upstream at the grounding line as well as at the walls of the glacier channel. There were a number of fractures (right hand side of the Fig. 3c, close the the ice front of the glacier) that could not be represented by our models. However, the uncertainty of the fracture observations in this area is high due to the difficulty distinguishing them from the surface features caused by basal crevasses.

Furthermore, we suggest that adding the proximity to the grounding line and the ice front did not produce a good fit for the Borchgrevink IS because of the specific shape of this region. The distance between the front and the grounding line is very small relatively to other glaciers in our analysis.

5.5 Discussion

We found that, in general, the most important predictor factors to model fractures for all analysed glaciers are effective strain rate, strain rate gradient, surface elevation gradient and proximity to mountains, in agreement with the theory of possible mechanism of fracture formation (Colgan et al., 2016). Previous analysis based on damage accounts for stresses, thickness and viscosity, but ignores such predictors as proximity to mountains and grounding line as well as the curvature of a channel, which might be crucial for modelling of the fracture formation in Antarctica. Our results can be used in Antarctic logistics in order to determine snow covered crevasses as they can be a potential hazard for navigation in Antarctica. Many researchers use ground penetrating radar to find hidden crevasses, but it is a real time assessment method that requires both financial and human resources and, therefore, can not cover all the areas in Antarctica.

To understand the effect of $\beta$ coefficients on the probability function we assign one of the $\beta$ coefficients to be negative, whereas we set all the other coefficients to 0. Then, it is obvious from Eq. 2 that the probability of fracturing in this case will be smaller than in a case when this $\beta$ coefficient is positive (in the situation when the predictor has positive values). Similar algorithm can be applied to analyse coefficients of negative predictors.
We do not claim that all the predictors that were chosen in the final set for each group represent the exact fracture mechanism for each glacier. We can see that, for some glaciers, sets containing different predictors can lead to results close to the best-fitting model. However, for some cases such as Amery and Totten ice shelves the number of a good-fitting models is very limited. For example, including the friction coefficient and proximity to the ice front in the analysis we can achieve a better fit to the observations. Therefore, we conclude that some factors have a very strong effect on fracture formation process and some still have an effect but it is minor for some glaciers.

The main uncertainty of our method comes from the over-estimation of the number of fractures. We do not know how many fractures are not visible on satellite images, therefore it is uncertain what over-estimation rate we should allow in our model. A possible solution to this could be to supplement satellite images with radar and seismic measurements (Navarro et al., 2005; Delaney and Arcone, 2005) or to acquire higher resolution satellite images that can help to identify location of fractures even if they are hidden under the snow surface. However, for our method it would require a large set of observational data. In our paper we assume that 20-25% rate of over-estimation should be reasonable. We can see that most of our results show over-estimation of fracture formation in the areas around observed fractures. This shows as well that our over-estimation might be reasonable as fractures are more likely to be formed if there are other fractures already present nearby. Our assumption is based on the fact that the ice regime conditions are similar within a 500 metres radius, not implying any direct influence of the old fractures to the new fractures (Colgan et al., 2016). Also, the area of high probability of a fracture is larger than the number of observed fractures mainly due to the fact that it is impossible to select all of the fractures on the satellite images manually. The selected observed fractures capture the main areas of fracturing assuming that surrounded nodes are likely to be fractured as well.

The over-estimation of the fracturing in the case of Vanderford (see Fig. 2d) can be seen mainly at the front of the ice shelf. There is a very high chance of developing surface fractures in that area and therefore it might be that the fractures are just not visible on the satellite images. In fact that region has a relatively higher snow accumulation rate reaching 1 m/yr. After a closer inspection of the satellite image areas where we see the over-estimation error we can recognise presence of surface fractures but, due to a very large number of fractures around Antarctica, the manual selection of all the fractured data points is a time intensive process. Sampling with a higher spatial density would require an automated algorithm. However, it is important to note that the low spatial density sampling does not influence the results of the fracture probability (as mentioned previously only diversity in sampling is important), it affects only the observations data set that we use to estimate the success of our model.

The damage-based approach produces areas of high damage downstream from the observed fractures (see Figs. 6b, 5d). If we assume that there are some fractures that are invisible on the images it is still unclear why the modelled ice is not damaged upstream where observed fractures are present. If we could see damaged ice upstream from the observed fractures we could assume that the ice was damaged and transported downstream where we can see fractures. However, in most of the results based on damage approach we could not capture this type of behaviour, whereas in the probability-based model we do not observe this type of error and in most of the cases overestimation of fractures occurs in a vicinity or upstream from observed fractures (e.g. Fig. 5a). In Figure 3a we show modelled probability of fracturing for Holmes IS. We can see that the over-estimation rate...
is quite small and it all can be seen upstream from the observed fractures, which might be due to fractures that were formed further upstream being transported downstream where they are invisible on the satellite images.

In addition, we looked at various properties of two ice shelves/glaciers for which we could not find a good approximation using any of the 20 predictors (part of the Ronne IS and Larsen B IS). We found that some estimated predictors had extreme values. For example, we found that rheology of Larsen B is the lowest out of all glaciers in our analysis. Also, the ice thickness and proximity to the grounding line are the smallest. Furthermore, it has the highest strain rate change as well as the highest surface elevation change. On the other hand, Ronne IS has the lowest elevation change. Also, one of the principal stresses components is the lowest. We do not have enough samples to cover values that are non-typical for the majority of glaciers, which might be the reason we could not find a good-fit model for these two glaciers. The analysis of these two ice shelves/glaciers using Bayesian analysis produced similar results, confirming that including any of the 20 predictors do not produce a good likelihood in those regions.

6 Conclusions

Most of the previous research on fracture formation has been focused on the application of damage mechanics. However, we have shown that it cannot fully reproduce the nature of fracturing at various ice shelf regions and the location of fracture formation has a large error. In this study we looked at this problem from a different perspective by constructing a probabilistic model based on the observations of surface fractures.

We found that Logistic Regression Analysis combined with other statistical methods can significantly improve the prediction of fracture formation for the Antarctic ice shelves/glaciers and can lead to the identification of up to 99% of observed surface crevasses with an average of 77% for all ice shelf regions. It has a number of uncertainties and leads to some over-estimation of the number of fractures in comparison to the observations, but the rate is not significantly higher than the over-estimation error when using the damage-based method.

We classified the Antarctic ice shelf regions into 4 groups. Each group has a set of factors that can be used in the Logistic Regression Analysis to estimate the probability of fracturing in a certain location. Also, we found that the ice shelves/glaciers in each group have similar characteristics. There were ice shelves/glaciers of specific shapes and having specific regimes that are more difficult to describe applying the general set of factors suggested in this study. However, overall our method can work as a tool that can be efficiently used in the analysis of fracture formation for most of the ice shelves/glaciers in Antarctica.

Our model is easy to implement and it can be effectively used as a basis and the first step in implementing a calving parametrisation in ice sheet models. Being able to more accurately predict zones of formation of small fractures provides the basis for modelling of fracture propagation at the locations where we determine fractures using LRA. This can then lead to an implementation of a calving parametrisation. Thus, this statistics-based method can help to expand our current knowledge of the crevasses as well as improve mapping of potential hazards.
7 Tables

Table 1. Development of calving parameterisations

<table>
<thead>
<tr>
<th>Year</th>
<th>Reference</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>1955</td>
<td>Crevasse penetration depth using tensile stress and overburden pressure</td>
<td>Nye (1955)</td>
</tr>
<tr>
<td>1973</td>
<td>Crevasse penetration depth of a single crevasse</td>
<td>Weertman (1973)</td>
</tr>
<tr>
<td>1976</td>
<td>Crevasse penetration depth estimation using LEFM</td>
<td>Smith (1976)</td>
</tr>
<tr>
<td>1993</td>
<td>Strain related fracture formation</td>
<td>Vaughan (1993)</td>
</tr>
<tr>
<td>1997</td>
<td>Sea level dependent calving</td>
<td>Motyka (1997)</td>
</tr>
<tr>
<td>1998</td>
<td>Linear Elastic Fracture Mechanics</td>
<td>Van der Veen (1998a, b)</td>
</tr>
<tr>
<td>2003</td>
<td>Damage mechanics for a single crevasse</td>
<td>Pralong et al. (2003)</td>
</tr>
<tr>
<td>2005</td>
<td>Damage mechanics for a single crevasse</td>
<td>Pralong and Funk (2005)</td>
</tr>
<tr>
<td>2007</td>
<td>Crevasse depth</td>
<td>Benn et al. (2007a, b)</td>
</tr>
<tr>
<td>2010</td>
<td>Crevasse depth</td>
<td>Nick et al. (2010)</td>
</tr>
<tr>
<td>2012</td>
<td>Damage mechanics applied to a crevasse field</td>
<td>Borstad et al. (2012)</td>
</tr>
<tr>
<td>2012</td>
<td>Kinetic 1st order calving</td>
<td>Levermann et al. (2012)</td>
</tr>
<tr>
<td>2012</td>
<td>CDM</td>
<td>Duddu and Waisman (2012)</td>
</tr>
<tr>
<td>2013</td>
<td>CDM</td>
<td>Duddu and Waisman (2013)</td>
</tr>
<tr>
<td>2013</td>
<td>Discrete element models</td>
<td>Bassis and Jacobs (2013)</td>
</tr>
<tr>
<td>2013</td>
<td>Particle-based simulation</td>
<td>Astrom et al. (2013)</td>
</tr>
<tr>
<td>2013</td>
<td>Crevasse depth criterion</td>
<td>Nick et al. (2013)</td>
</tr>
<tr>
<td>2014</td>
<td>Crevasse depth criterion</td>
<td>Cook et al. (2014)</td>
</tr>
<tr>
<td>2014</td>
<td>CDM</td>
<td>Albrecht and Levermann (2014)</td>
</tr>
<tr>
<td>2014</td>
<td>Combining CDM and LEFM</td>
<td>Krug et al. (2014)</td>
</tr>
<tr>
<td>Type</td>
<td>Predictor</td>
<td>Description</td>
</tr>
<tr>
<td>------------------</td>
<td>----------------------------</td>
<td>----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Geometry</td>
<td>Ice thickness</td>
<td>Bedmap2 data for Antarctica at 1 km spatial resolution</td>
</tr>
<tr>
<td></td>
<td>Maximum bed slope</td>
<td>Bedrock and ice surface slopes are calculated using a nodal function (see Eq. 8)</td>
</tr>
<tr>
<td></td>
<td>Maximum surface slope</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Surface slope gradient</td>
<td>Maximum slope change in a 400-600 metres vicinity</td>
</tr>
<tr>
<td></td>
<td>Distance to the ice front</td>
<td>DM_{IP}, calculated using Eq. 10</td>
</tr>
<tr>
<td></td>
<td>Distance to the GL</td>
<td>DM_{GL}, calculated using Eq. 11</td>
</tr>
<tr>
<td></td>
<td>Distance to the mountains</td>
<td>Proximity to the mountains (if present)</td>
</tr>
<tr>
<td></td>
<td>Curvature</td>
<td>Curvature of the glacier channel ( \alpha ), calculated in each node based on the direction and rate of the flow velocities (see Eq. 9)</td>
</tr>
<tr>
<td>Flow parameters</td>
<td>Backstress</td>
<td>Buttressing effect on ice streams calculated in ISSM from inversion</td>
</tr>
<tr>
<td></td>
<td>Velocity</td>
<td>InSAR ice flow velocity</td>
</tr>
<tr>
<td></td>
<td>Rheology predictor (viscosity)</td>
<td>( B ), Glen’s flow predictor, calculated from inversion of velocities (only for floating ice)</td>
</tr>
<tr>
<td></td>
<td>Friction coefficient</td>
<td>calculated from inversion of velocities</td>
</tr>
<tr>
<td></td>
<td>Effective Strain rate</td>
<td>The effective strain rate is calculated using Eq. 5 with observed velocities as an input</td>
</tr>
<tr>
<td></td>
<td>Principal stress (1 and 2)</td>
<td>Eigenvalues ( \lambda ) (normal stresses) in Eq. 7</td>
</tr>
<tr>
<td></td>
<td>Principal axes (two components)</td>
<td>Eigenvectors ( s_i ) in Eq. 7</td>
</tr>
<tr>
<td></td>
<td>Principal strain rate (1 and 2)</td>
<td>Eigenvalues ( \mu ) (see Eq. 6)</td>
</tr>
<tr>
<td></td>
<td>Strain rate gradient</td>
<td>Maximum strain rate change in a 400-600 metres vicinity</td>
</tr>
</tbody>
</table>
### Table 3. Formed groups of ice shelf regions

<table>
<thead>
<tr>
<th>Group1</th>
<th>Group2</th>
<th>Group3</th>
<th>Group4</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>3</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>6</td>
<td>17</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>7</td>
<td>22</td>
<td>5</td>
</tr>
<tr>
<td>12</td>
<td>8</td>
<td>24</td>
<td>16</td>
</tr>
<tr>
<td>13</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Glaciers/ice shelves for which we could not find a good-fitting probability are marked with red.

### Table 4. Predictors for each formed group

<table>
<thead>
<tr>
<th>Back stress</th>
<th>Effective strain rate</th>
<th>Principal1 strain rate</th>
<th>Principal2 strain rate</th>
<th>PAxis1 stress</th>
<th>PAxis2 stress</th>
<th>PValue1 stress</th>
<th>PValue2 stress</th>
<th>Surface slope</th>
<th>Bed slope</th>
<th>Strain change</th>
<th>Curvature</th>
<th>Friction coefficient</th>
<th>Rheology B</th>
<th>Thickness</th>
<th>at the ice front</th>
<th>at the grounding line</th>
<th>near mountains</th>
<th>surface change</th>
<th>velocity</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Tick-mark stands for an addition of a predictor to the model.
Table 5. A list of analysed ice shelf regions

<table>
<thead>
<tr>
<th>Glacier</th>
<th>Group</th>
<th>Corresponding IS name</th>
<th>Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>George IV</td>
<td>Palmer land, AP</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>Larsen C</td>
<td>Fallieres Coast, AP</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>Larsen D</td>
<td>Black Coast, AP</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>Orville Coast side of the Ronne IS</td>
<td>WA</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>Amery</td>
<td>EA</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>Edward VII</td>
<td>Mawson Coast, EA</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>Rayner Thyner</td>
<td>EA</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>Shirase</td>
<td>Prince Harald Coast, EA</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>Stancomb-Brunt</td>
<td>Caird Coast, EA</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>Riiser-Larsen</td>
<td>Princess Martha Coast, EA</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>Ekstrohm</td>
<td>EA</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>Abbot</td>
<td>Eights Coast, WA</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>Baudoin</td>
<td>Princess Ragnhild Coast, EA</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>Nivl</td>
<td>Princess Astrid Coast, EA</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>Borchgrevnik and Lazarev</td>
<td>Princess Astrid Coast, EA</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>Borchgrevnik</td>
<td>Princess Raghild Coast, EA</td>
</tr>
<tr>
<td>17</td>
<td>3</td>
<td>Dibble IS</td>
<td>Clarie Coast, EA</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>Mertz IS</td>
<td>EA</td>
</tr>
<tr>
<td>19</td>
<td>2</td>
<td>Rennik</td>
<td>Pennell Coast, EA</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>Cook</td>
<td>George V Coast, EA</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>Ninnis</td>
<td>George V Coast, EA</td>
</tr>
<tr>
<td>22</td>
<td>3</td>
<td>Holmes</td>
<td>Banzare Coast, EA</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>Moscow University</td>
<td>Sabrina Coast, EA</td>
</tr>
<tr>
<td>24</td>
<td>3</td>
<td>Totten IS</td>
<td>EA</td>
</tr>
<tr>
<td>25</td>
<td>2</td>
<td>Vanderford IS</td>
<td>EA</td>
</tr>
<tr>
<td>26</td>
<td>1</td>
<td>West IS</td>
<td>Queen Mary Coast, EA</td>
</tr>
<tr>
<td>27</td>
<td>2</td>
<td>Larsen C</td>
<td>Oscar II Coast, AP</td>
</tr>
<tr>
<td>28</td>
<td>2</td>
<td>Larsen B</td>
<td>Nordenskjold Coast, AP</td>
</tr>
<tr>
<td>29</td>
<td>2</td>
<td>Larsen A</td>
<td>Davis Coast, AP</td>
</tr>
<tr>
<td>30</td>
<td>3</td>
<td>Tracy-Tremenchus</td>
<td>Knox Coast, EA</td>
</tr>
<tr>
<td>31</td>
<td>1</td>
<td>Drygalski</td>
<td>Scott Coast, EA</td>
</tr>
<tr>
<td>32</td>
<td>3</td>
<td>Rennik</td>
<td>Lazarev Mountains, Oates Coast, EA</td>
</tr>
<tr>
<td>33</td>
<td>2</td>
<td>Mariner</td>
<td>Borchgrevnik Coast EA</td>
</tr>
<tr>
<td>34</td>
<td>1</td>
<td>Filchner</td>
<td>Coast Land, WA</td>
</tr>
</tbody>
</table>

AP - Antarctic Peninsula, EA - East Antarctica, WA - West Antarctica, IS - ice shelf
Table 6. $\beta$ coefficient for equation 2.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Group1</th>
<th>Group2</th>
<th>Group3</th>
<th>Group4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free coefficient</td>
<td>-1.1739</td>
<td>-0.8674</td>
<td>-0.5932</td>
<td>-1.2868</td>
</tr>
<tr>
<td>Back stress</td>
<td>0.0643</td>
<td>0.0114</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Effective strain rate</td>
<td>-1.5745</td>
<td>-1.6427</td>
<td>-</td>
<td>-1.9979</td>
</tr>
<tr>
<td>Principal 1 strain rate</td>
<td>-0.4520</td>
<td>-0.5270</td>
<td>-0.1069</td>
<td>-</td>
</tr>
<tr>
<td>Principal 2 strain rate</td>
<td>-0.3543</td>
<td>-0.3670</td>
<td>0.0589</td>
<td>-</td>
</tr>
<tr>
<td>PAxis11 stress</td>
<td>-0.0140</td>
<td>-</td>
<td>0.2004</td>
<td>0.1616</td>
</tr>
<tr>
<td>PAxis21 stress</td>
<td>0.0349</td>
<td>0.0276</td>
<td>0.1634</td>
<td>0.1495</td>
</tr>
<tr>
<td>PValue1 stress</td>
<td>0.0210</td>
<td>0.0044</td>
<td>-0.0170</td>
<td>-</td>
</tr>
<tr>
<td>PValue2 stress</td>
<td>0.1318</td>
<td>0.1252</td>
<td>-</td>
<td>0.1075</td>
</tr>
<tr>
<td>Surface slope</td>
<td>-</td>
<td>-1.6080</td>
<td>-</td>
<td>-0.1333</td>
</tr>
<tr>
<td>Bed slope</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Strain change</td>
<td>-0.1243</td>
<td>-0.1411</td>
<td>-0.4509</td>
<td>-0.1029</td>
</tr>
<tr>
<td>Curvature</td>
<td>-0.0436</td>
<td>-0.0131</td>
<td>-0.0870</td>
<td>0.0154</td>
</tr>
<tr>
<td>Friction coefficient</td>
<td>-0.0840</td>
<td>-</td>
<td>-0.1949</td>
<td>0.0388</td>
</tr>
<tr>
<td>Rheology B</td>
<td>0.0985</td>
<td>0.1106</td>
<td>0.1160</td>
<td>-</td>
</tr>
<tr>
<td>Ice thickness</td>
<td>0.2018</td>
<td>0.1459</td>
<td>0.4666</td>
<td>-</td>
</tr>
<tr>
<td>Proximity to the ice front</td>
<td>-0.2112</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Proximity to the grounding line</td>
<td>-0.1791</td>
<td>-0.4151</td>
<td>-0.2237</td>
<td>-</td>
</tr>
<tr>
<td>Proximity to mountains</td>
<td>1.0169</td>
<td>-</td>
<td>0.2977</td>
<td>1.1554</td>
</tr>
<tr>
<td>Surface change</td>
<td>-1.2133</td>
<td>-</td>
<td>-</td>
<td>-1.1823</td>
</tr>
<tr>
<td>Velocity</td>
<td>-0.4416</td>
<td>-0.5026</td>
<td>-0.7624</td>
<td>-</td>
</tr>
</tbody>
</table>

Numbers that have the largest magnitude are marked with red.
Figure 1. Success and error percentages for Logistic Regression Algorithm vs. Damage method, applied to 34 Antarctic ice shelf regions (a). The location of each ice shelf/glacier is shown in panel b.
Figure 2. Modelled probability of a fracture for Group 2: Rayner Thyner IS (a), Larsen D IS (b), Edward VII IS (c) and Group 1: Vanderford IS (d). Observed surface fractures are shown in black and observed non-fractured ice is marked with orange circles. Red polygons represent regions where high resolution images were available.
Figure 3. Modelled probability for Holmes IS (Group 3) (a), Drygalski IS (Group 1) (b), Amery IS (Group 4) (c) and Shirase IS (Group 2) (d) and damage at Shirase IS (Fig. e). Labels the same as Fig. 2.
Figure 4. Modelled probability of a fracture vs. modelled damage for Nivl IS (Group 3) (a, b) and Abbot IS (Group 1) (c, d). Labels are the same as in Fig. 2.
Figure 5. Modelled probability of a fracture vs. modelled damage for Rennik IS (Group 2) (a, b) and Dibble IS (Group 3) (c, d). Labels are the same as in Fig. 2.
Figure 6. Modeled probability of a fracture vs. modeled damage for Moscow University IS (Group 1) (a, b) and Mariner IS (Group 2) (c, d). Labels are the same as in Fig. 2.
Figure 7. Modelled probability of a fracture vs. modelled damage for Cook IS (Group 1) (a, b) and Totten IS (Group 3) (c, d). Labels are the same as in Fig. 2.
Author contributions. Veronika Emetc designed the study, developed the methodology, collected the data, performed the analysis, and wrote the manuscript. Paul Tregoning helped with models and revising the manuscript critically. Malcolm Sambridge developed the source function for the Bayesian analysis.

Acknowledgements. We would like to thank Mathieu Morlighem for his continuous guidance in ice sheet modelling and his help with understanding ISSM. We also would like to thank Teresa Neeman for her help with Logistic Regression Algorithm. We are also grateful to Anthony Purcell for his always helpful advises. Veronika Emetc received scholarship from ARC Discovery DP140103679.
References


Stockert, T.: Climate change 2013: the physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change, Cambridge University Press, 2014.


