Supplement to:

Subglacial sedimentary basins focus key vulnerabilities of the Antarctic ice-sheet

**Lu Li1, Alan R.A. Aitken1,2, Mark D. Lindsay1,3,4, Bernd Kulessa5,6**

1 – School of Earth Sciences, The University Of Western Australia, Perth, Western Australia, 6009, Australia

2 – The Australian Centre for Excellence in Antarctic Science, The University of Western Australia, Perth, Western Australia, 6009, Australia

3 – CSIRO Mineral Resources, Perth, Western Australia, 6151, Australia

4 – ARC Industrial Transformation and Training Centre in Data Analytics for Resources and the Environment (DARE)

5 – School of Biosciences, Geography and Physics, Swansea University, Wales, SA2 8PP, UK

6 – School of Geography, Planning, and Spatial Sciences, University of Tasmania, Hobart, Tasmania, 7001, Australia

This document includes:

[1 Sedimentary basin mapping with Random Forest prediction 3](#_Toc88860744)

[1.1 Workflow 3](#_Toc88860745)

[1.2 Training points 4](#_Toc88860746)

[1.2.1 Outcrop 4](#_Toc88860747)

[1.2.2 Seismic observation 5](#_Toc88860748)

[1.2.3 Potential field interpretation 6](#_Toc88860749)

[1.3 Evidence layers 7](#_Toc88860750)

[1.3.1 Bedrock topography 10](#_Toc88860751)

[1.3.2 Gravity field 13](#_Toc88860752)

[1.3.3 Magnetic field 15](#_Toc88860753)

[1.3.4 Cryosphere and Solid-Earth Structure 17](#_Toc88860754)

[1.4 Methods 17](#_Toc88860755)

[1.4.1 Evidence layer processing 17](#_Toc88860756)

[1.4.2 Evidence layer reduction 17](#_Toc88860757)

[1.4.3 Training Point Sampling 18](#_Toc88860758)

[1.4.4 RF Model training 19](#_Toc88860759)

[1.4.5 Accuracy measurement 20](#_Toc88860760)

[1.5 Result 21](#_Toc88860761)

[1.5.1 Evidence selection 21](#_Toc88860762)

[1.5.2 Model turning 24](#_Toc88860763)

[1.5.3 Model accuracy and uncertainties 25](#_Toc88860764)

[1.5.4 Evidence importance 31](#_Toc88860765)

[1.5.5 Sensitivity test 32](#_Toc88860766)

[1.6 Model limitation and robustness 33](#_Toc88860767)

[1.6.1 Uncertainty in training point calibration 33](#_Toc88860768)

[1.6.2 Uncertainty in evidence layers 34](#_Toc88860769)

[1.6.3 Model robustness 34](#_Toc88860770)

[2 Hydro-mechanical model 36](#_Toc88860771)

[2.1 Method 36](#_Toc88860772)

[2.1.1 Geomechanical Modelling 36](#_Toc88860773)

[2.1.2 Ice Sheet Evolution 37](#_Toc88860774)

[2.1.3 Hydrology modelling 38](#_Toc88860775)

[2.2 Model setting and boundary conditions 38](#_Toc88860776)

[2.3 Discussion 40](#_Toc88860777)

[2.3.1 Impact of permeability to basal water flux 41](#_Toc88860778)

[2.3.2 Impact of sedimentary basin thickness to basal water flux 43](#_Toc88860779)

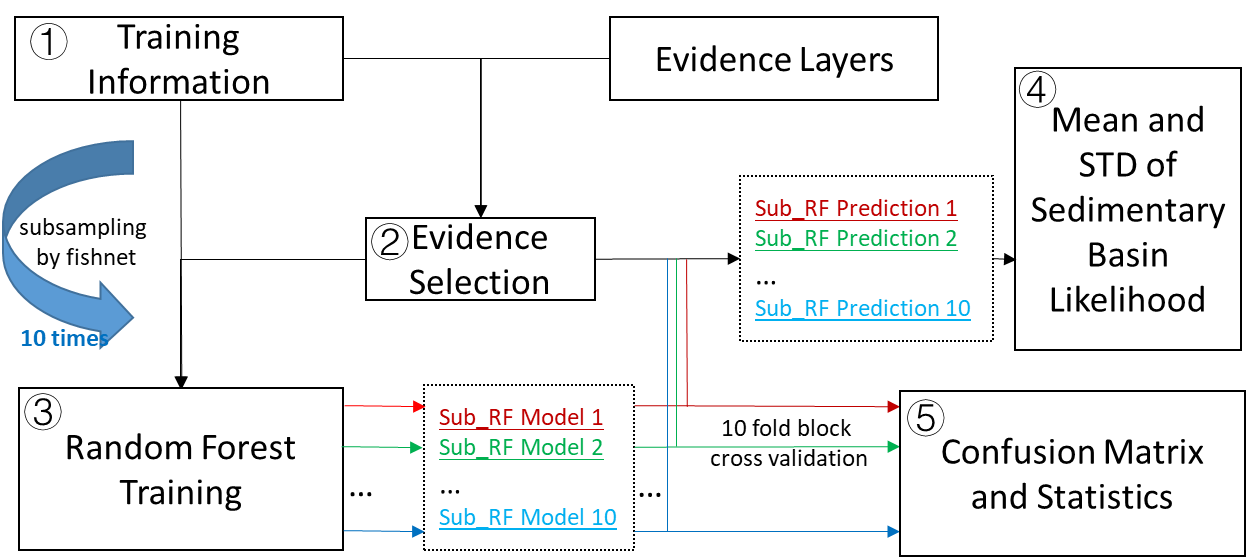
[2.3.3 Impact of basal heat flux due to water transportation 44](#_Toc88860780)

1. Sedimentary basin mapping with Random Forest prediction

We use a supervised machine learning Random Forest (RF) method to build a bedrock type probability model, aiming at applying a consistent data-based rule. The analysis was performed in R 4.0.2 (R Core Team, 2018) in R-Studio 1.3.1073, and is documented in the <https://github.com/LL-Geo/ANT_SEDI>.

## Workflow

The general workflow for building a spatial classification model in RF involves five main steps (Fig. S1): 1) training information generation, 2) evidence layer selection, 3) model training, 4) model predictions, and 5) model evaluation.



**Fig. S1. Workflow in building Random Forest classification in this study.**

In this study, we generate training information based on the current understanding of the sedimentary basin and crystalline basement structure in Antarctica. We choose several geophysical datasets as evidence layers to generate training information. We then select the most important yet uncorrelated evidence layers to build the RF model. Sparse and spatial equally distributed training information are proven to be valid to construct RF model based on prior knowledge47,48. Here, we subsample the whole training information by a reference fishnet, and then train to build an RF model. The entire training information subsampling and RF training process are repeated 10 times in 10 different seeds to reduce bias during the random subsampling process. We use the mean of all 10 RF model responses to represent Antarctica's sedimentary basin likelihood. Finally, we use 10 fold block cross-validation to validate the model.

## Training points

In this study, we calibrate the model based on different observations, including bedrock outcrop, seismic measurement, and geophysical data interpretations. Each type of data has unique uncertainties due to the method of observation.

In general, we define a sedimentary basin as a place with a substantial thickness of sedimentary rock and/or sediment accumulation, without significant reworking and tectonic deformation. In contrast, we label the pre-Jurassic sedimentary outcrop as the basement rock, as these rocks are most likely to have been deformed or metamorphosed during the breakup of Gondwana. A typical example is the Beacon Supergroup in the front of Transantarctic Mountains, and the Cambrian sedimentary rocks forming the basement in the Ellsworth-Whitmore Mountains.

### Outcrop

Rock outcrop is very sparse in the whole continent due to thick ice coverage. We use the results derived from Geomap49 to represent the location and type of rock outcrop. Cenozoic and Mesozoic sedimentary rocks are marked as the modern sedimentary basin found in Victoria Land, and Antarctica Peninsular. Beacon Supergroup sedimentary rock and Late Palaeozoic sedimentary strata are marked as the ancient sedimentary basin. The Beacon Supergroup sedimentary rocks are mainly distributed in Transantarctic Mountains. The late Palaeozoic sedimentary strata is found in Antarctic Peninsula. We label the remaining classes of granite, volcanic, and metamorphosed rock as crystalline basement. We further merge ancient sedimentary basin rocks as basement rock, as these are deformed and metamorphosed. The typical example is the thick Cambrian sedimentary rock in the Ellsworth-Whitmore Mountains, which form the basement for post-Jurassic sediment accumulation50.

The outcropping information may introduce sample noise due to variable geology within a small area. To reduce this uncertainty, we apply 20 times of 3\*3 averaging filter to geological map in a 250 m cell size. This result is an outcrop geology map with a minimum spatial dimension of 10 km. In each cell, the total likelihood of all geology units remains at 1. We then regrid the outcrop geology map into 10 km grid. We keep the geology class with a likelihood larger than 90% for the final analysis.

### Seismic observation

Active seismic reflection and refraction methods are powerful methods to image upper crustal structures beneath ice and ocean. The major active Antarctic seismic observation is the continental margin recorded by Antarctic SDLS (Seismic data library system). In West Antarctica, Trey, et al.51 find a thick sedimentary layer in the Ross Sea area using wide-angle reflection/refraction profile. In the Weddell Sea, seismic results show massive sedimentary deposits up to 13 km thick under Filchner-Ronne Ice shelf52. Combining and interpolating the SDLS data, the GlobSed model has shown the total sedimentary basin thickness along the continent margin up to date53. Here, we select the total sediment thickness > 1 km and within the 20 km of the SDLS data to represent the offshore sedimentary basin. We define basement structure where GlobSed shows sediment thickness < 100 m within the SDLS measurement range. These major basement structures located in the Eastern Amundsen Sea Embayment, Central High in the Ross Sea, and the Gunnerus Ridge in the Riiser-Larsen Sea.

Over the onshore area, the combined seismic reflection and refraction experiment on Siple Coast reveal continuous and thick sedimentary basins are located at the current onset area of ice stream54,55. Seismic imaging also helps to constrain both soft sediment and consolidated sedimentary rocks at the bottom of Lake Vostok56.

Besides active seismic reflection/refraction profiles, the passive seismic data also images the velocity structure (see Table. S1 for the details of the seismic station locations). Receiver function studies reveal subglacial sedimentary layer in West Antarctica Rift System57,58, where its thickness is typical < 1 km. Ramirez, et al.59 report the HOWD seismic station has 2.5 km thick low velocity layer in receiver function, which is interpreted as a sedimentary basin near the Ellsworth Mountains. In East Antarctica, major receiver function study focus on the Wilkes Subglacial Basin. In the North Wilkes Subglacial Basin, WISE seismic station reveal 200 to 400 m sediments near the Eastern Basin60. Further south, Chai, et al. find a low P wave velocity, indicate a low density layer beneath E028 station61. In south Wilkes Subglacial Basin, grid search58 and horizontal-to-vertical spectral ratio method62 show sediment deposited beneath N052 and N060, respectively. In the centre pole area, Anandakrishnan and Winberry58 found a sediment layer in the SPA seismic station by receiver function study. In central DML, Gupta, et al.63 find a 1.5 km sediment beneath the ice sheet. We classify this location as sedimentary basin except where at stations where thickness error is larger than the sediment thickness (ST09, ST13, KOLR, SIPL)57. For the basement location, it is assigned where low velocity sediment is unnecessary in the inversion, where a two layer ice and crystalline crust model could fit the observed receiver well57,58,64.

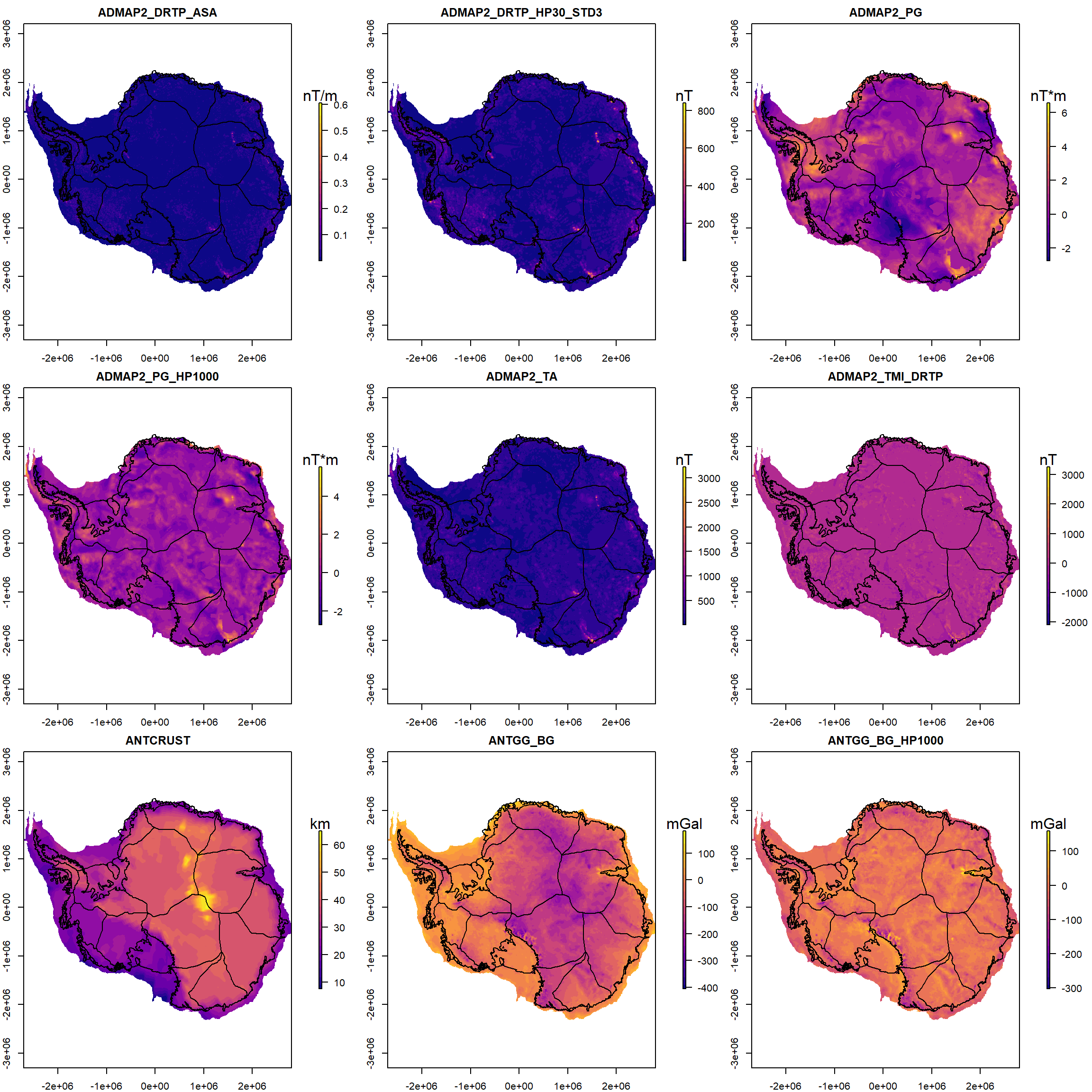
### Potential field interpretation

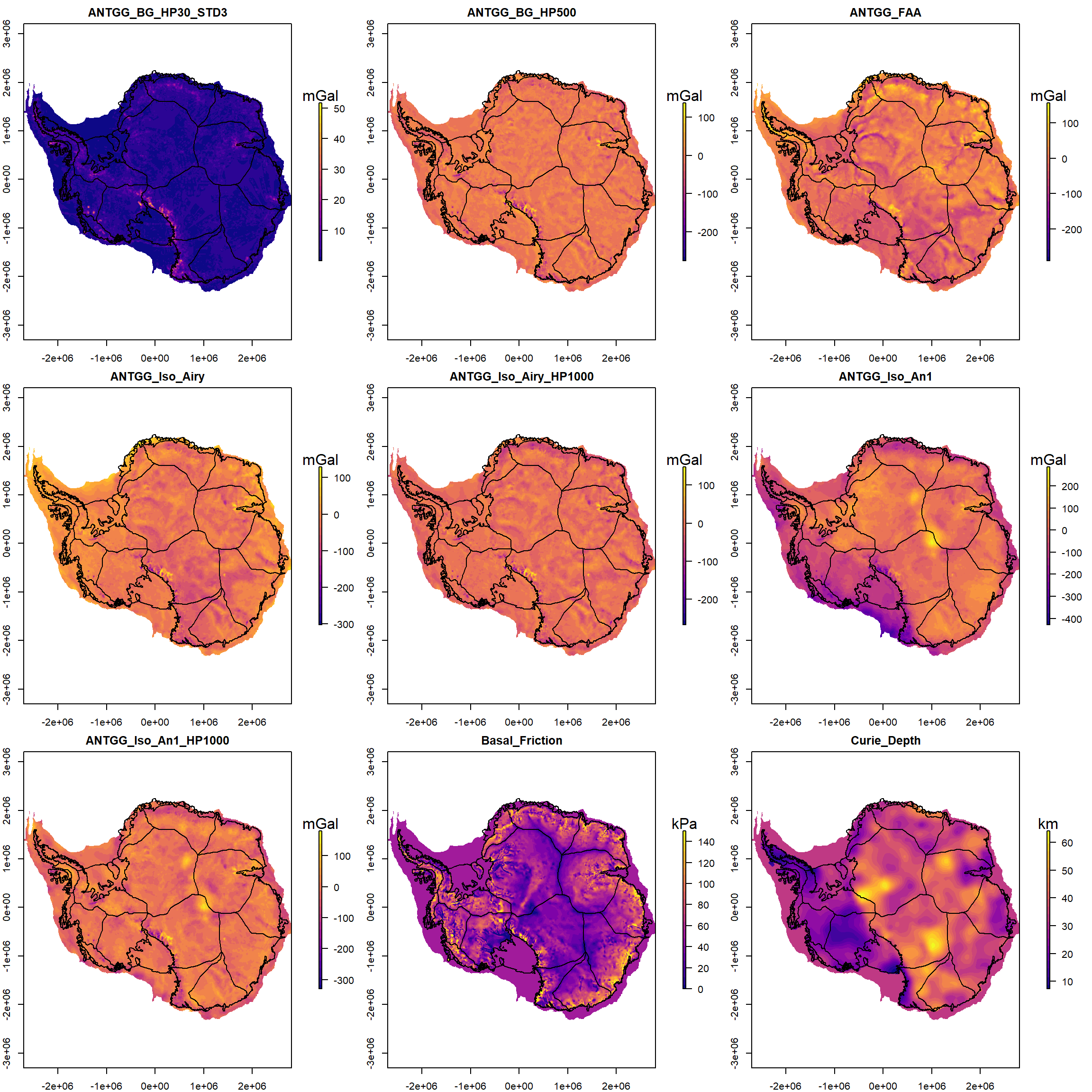
Topography, gravity and magnetic data help interpret subglacial sedimentary basin distribution where seismic imaging is not available. Marine and rift sediments are inferred by gravity and magnetic studies in West Antarctica. Studinger, et al.65 evaluate regional blanketing marine sediments and linear fault-bounded sedimentary basins by isostatic adjusted topography, free air gravity and Bouguer gravity studies over the West Antarctica Ice Sheet. In the Interior Ross Embayment, Bell, et al.66 find three sedimentary basins deeper than 3.6 km according to depth to magnetic source modelling and interpret these basins to have formed during late Cretaceous due to positive gravity anomaly. The Ellsworth–Whitmore Mountains area preserve thick Cambrian sedimentary rock up to 13 km. These sedimentary rocks were reworked and now form the basement for the post-Jurassic sediment50. Subglacial sediments are also found in the interior in East Antarctica Ice Sheet. In Coats Land, Bamber, et al.67 report a 3 km depth subglacial basin in Slessor half-graben using 2D Werner deconvolution depth estimation. In Wilkes Land, deep magnetic basement structure indicates sedimentary rock infill68. Further, the variation of sediment thickness by gravity inversion in the Sabrina Subglacial Basin indicate the dynamic retreat and advance of the East Antarctica Ice Sheet69. In northern Wilkes Subglacial Basin, the high-frequency short-wavelength magnetic signature indicates the Ferrar rock pervasively intrude Beacon sandstone70. The negative magnetic field indicates the fault bounded sedimentary basin preserve in the Central Basin and Western Basin, and Ross Orogen metasediment in Eastern Basin70. The gravity inversion of sediment thickness constrained by magnetic depth to basement suggests thick sediment in Eastern Basin area, with 500-1000 m sediments in the fault bounded Central Basin and Western Basin at northern Wilkes Subglacial Basin area71. Recent, Pensacola Pole Basin is revealed using by newly collected PolarGAP data72. Based on aeromagnetic, aerogravity, satellite imagery and confirmed volcanoes, van Wyk de Vries, et al.73 identify 138 volcanoes in West Antarctica. We make a 10 km buffer to these volcanoes to represent the volcanic basement rock.

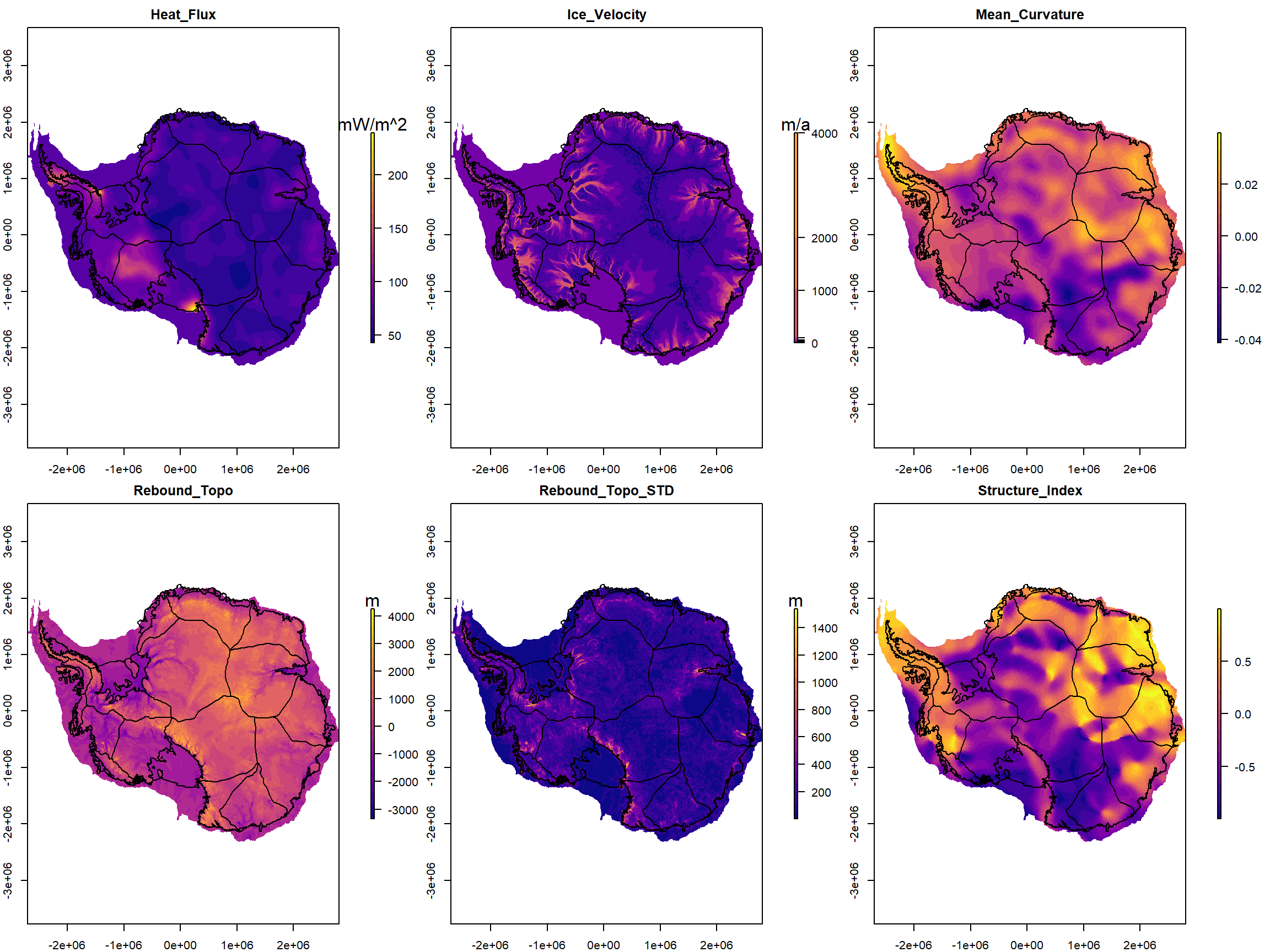
We draw the polygon of the mentioned area to indicate the sedimentary basin by potential field interpretation. For the basement area, we mark the pre-Jurassic sedimentary basin in the Ellsworth-Whitmore Mountains as the basement rock50. In Ross Embayment, we label the shallow magnetic basement (< 2.5 km below sea level) as the volcanic basement66,74. We further generate the inferred basement location with sediment thickness < 300 m in gravity inversion result69,71,75, and several deep magnetic provinces lack of nagetive isostatic gravity response68.

## Evidence layers

Summarising from the sedimentary basin study in the training point generation, the sedimentary rocks are indicated by a smooth bedrock surface, low residual gravity, and deeper depth to magnetic basement. Low free-air gravity signals are interpreted to be low density sedimentary rocks that infill topographic lows. We use bedrock elevation, gravity field, and magnetic field to represent solid-earth properties of sedimentary basins and crystalline basement.







**Fig S2. Evidence layer used in this study.** All evidence layers are regrided into a 10 km grid. TMI (total magnetic intensity), DRTP (Differential reduction to the pole), HP (High-Pass filter), STD, (Standard Deviation), PG (Pseudo-Gravity), FAA (Free-Air gravity Anomaly), BG (Bouguer Gravity), Iso (Isostatic gravity), Rebound\_Topo (isostatic Rebound Topography).

### Bedrock topography

#### Isostatic rebound topography

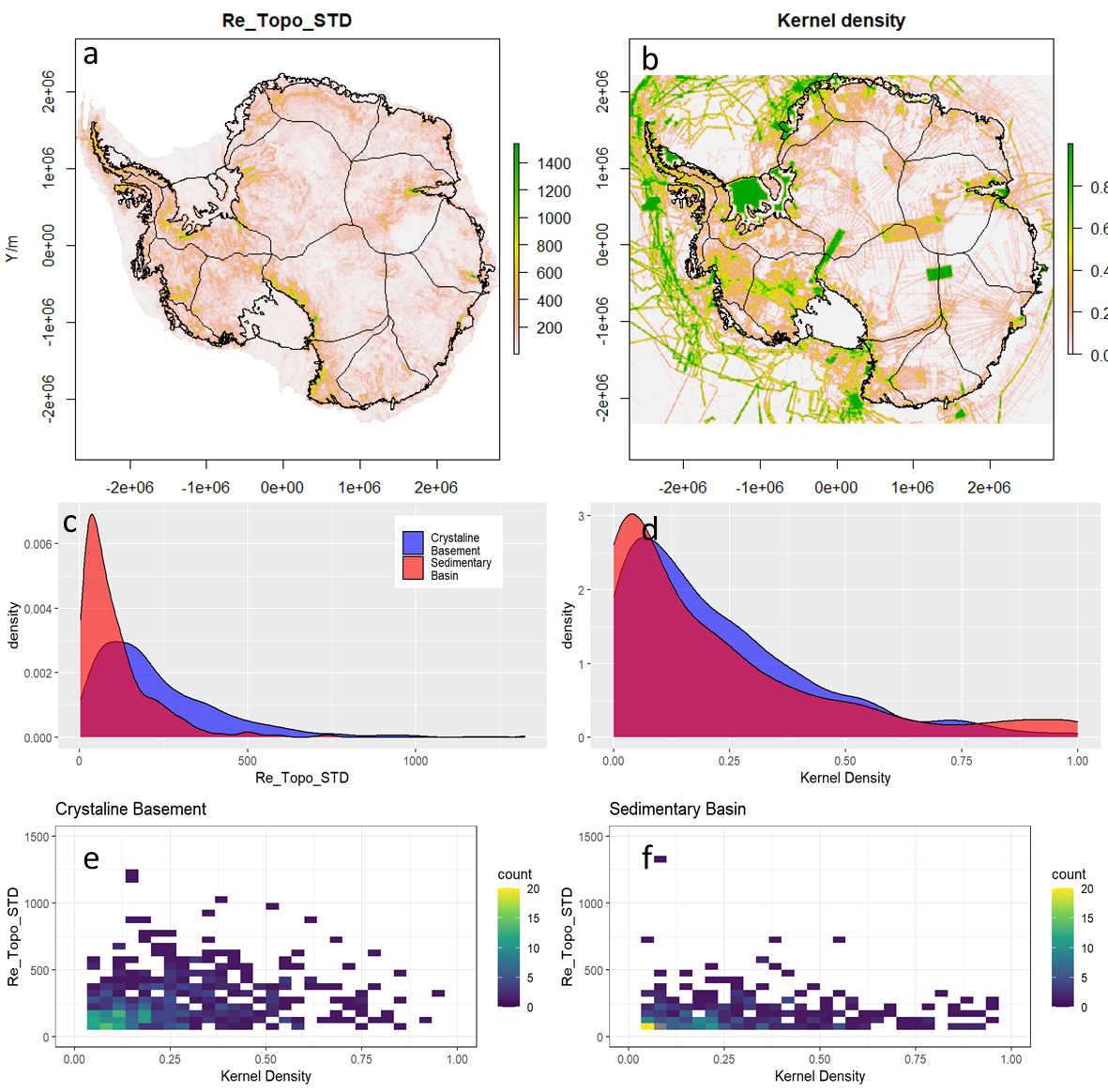
The topography is the primary indicator of geological processes as well as bedrock type. In the sediment erosion, transportation and deposition process, the sedimentary rock in the highland area are likely to be eroded to expose bedrock. The eroded materials are then redistributed in the low topography area to form a sedimentary basin.

By releasing the current ice load, we remove the subsidence caused by ice thickness. We estimate bed elevation before the ice expansion in Antarctica by calculating the isostatic topographic rebound using the bedrock topography and ice thickness from BedMachine Antarctica76. We assume an ice density of 917 kg/m3, sea water density of 1030 kg/m3, and mantle density of 3330 kg/m3 to remove the modern ice sheet load. For areas below sea level, we iteratively calculate the load of water that replaces ice until the total load change is < 2 meters.

#### Isostatic rebound topography roughness

‘Bed roughness’ is the irregular variation of bed surface with its distance77. Sediment deposited in the low topographic areas during ice-free time periods and long term erosion and deposition process would cause the sedimentary basin to have an inherently smooth surface. In contrast, the erosion process of hard basement rock is less efficient to makes it have rough terrain. When the ice sheet expands over the bed, the bed roughness could be preserved or modified by the ice sheet during multiple glacial cycles69.

Here we calculate bed roughness by the standard deviation of isostatic rebound topography in 30 km window size. We note the potential bias of bed roughness information due to the heterogeneous radar data measurement and also the inherent smooth data interpolation in BedMachine Antarctica76. Hence, we test the relationship of kernel density of the radar measurement and bed roughness with the calibrated bedrock type. We calculate the kernel density of data measurement by a 16 km radius circular kernel. We use the area where bedrock error is less than 200 meter to represent the reliable data source in BedMachine data76. The correlation of each dataset indicates the discrimination of bedrock type mainly depends on the bed roughness but not associated with the data density. For sedimentary basin class, the smooth topography distributed through all kernel densities. The crystalline basement class has a rougher bed in both high and low kernel densities region. Further, we calulate the Maximal Information Coefficient (MIC) of bed roughness with kernel densities (detail of MIC see 1.4.2). The MIC result (0.295) indicates a weak correlation between bed roughness and kernel densies. Overall, our that training process is not highly biased by the data heterogeneous distribution (Fig. S3).

****

**Fig S3. Topography roughness versus kernel density of data measurement.** The kernel density is calculated based on 16 km radius circular window, which matches the area for topography roughness estimation (30 km square window). The second row shows the density plot of topography roughness and kernel density of data measurement for both sedimentary basin and crystalline basement classes. The third row shows cross plot between topography roughness and kernel density for crystalline basement and sedimentary basin respectively.

### Gravity field

Gravity measurements respond to the density variations in the subsurface. The low-density sedimentary rock fills in topography low, which makes both free-air gravity and Bouguer gravity to be negative and smooth. Furthermore, isostatic gravity has removed the long wavelength signal in the Bouguer gravity due to the crust and mantle interface. Isostatic gravity anomaly could more directly reflect the upper crust density variation.68

#### Free air gravity

The free-air gravity anomaly response is interpreted to dominant by bedrock topography signal. The low free-air gravity anomalies correspond to topographicy basin and trench, and high free air gravity anomalies corresponds to mountain and dome78. The strong negative free-air gravity anomaly could be expected that low density material fills and preserve in these low topographic areas72.

We merge the continental scale free air gravity dataset AntGG78 with the latest gravity measurement in the South Pole72, centre Marie Byrd Land79, Ross Ice Shelf80, and Recovey Lakes Region81. We also use Global Marine Gravity V2782 to fill the data gap without near surface measurement in offshore area. To merge these datasets, we upward continue these airborne free air gravity measurement to 5 km ellipsoid height. Then all these datasets are merged by blending to reference GOCO06s83 satellite gravity model.

#### Complete Bouguer Gravity and its residual

By removing the topography effect in free-air gravity, the Bouguer gravity reflects long wavelength information caused by crustal thickness variation and short wavelength variation caused by upper crust density variation.

We calculate the gravity effect of surface load using the reference density of 917 kg/m3, 1030 kg/m3, 1000 kg/m3 and 2670 kg/m3 for ice, ocean water, lake water and topography layers respectively. These layers are regridded into geodetic coordinates at 0.05° resolution from BedMachine Antarctica76. The 3D full terrain effect is then calculated using Tesseroids84 within a 3° (333 km) radius around each grid point at 5 km ellipsoid height. The full terrain effect is subtracted to get the complete Bouguer gravity.

The continental scale Bouguer response shows most of WA has positive Bouguer anomaly and EA has negative Bouguer anomaly. The response is dominated by a long wavelength signal interpreted to be caused by thin crust in WA and thick crust in EA. This long wavelength dominated signal hinders the utility from using Bouguer gravity amplitude to distinguish bedrock type. To remove the effect of deep crustal structure signal, we calculate the residual gravity using a Gaussian high-pass filter of 500 and 1000 km wavelength.

#### Complete Bouguer Gravity Roughness

The infilled sedimentary rock will flatten the short wavelength roughness response caused by the basement variation. The Moho signature is smooth. Hence, the roughness of Bouguer gravity is assumed to represent the upper crust density variation, indicating the subglacial geology variation.

Variable data sources and fight line spacing means that the compilation of continent-scale datasets mixes different wavelength information. To match with data density evaluation and bedrock topography roughness analysis, we use a Gaussian high-pass filter in 60 km wavelength to calculate the residual gravity to reflect the near surface density variation. After that, we calculate the standard deviation of the filtered Bouguer gravity in a 30 km window size to compute the roughness of the residual Bouguer gravity anomalies.

#### Isostatic residual gravity

Similar to the complete Bouguer gravity, the variation of Moho as well as mantle structure is interpreted to dominate the continental scale gravity datasets. The broad crust-scale signals mask the near surface information, and hinder the ability use gravity to define upper crust information. Here, we assume a local Airy isostatic compensation. We also use the Moho depth (AN1)85 driven by seismic tomography to remove the Moho signal in gravity data.

We convert ice, seawater, and lake water layer into equivalent topography, using density 917 kg/m3, 1030 kg/m3, 1000 kg/m3 respectively. For the Airy isostatic model, we use the compensation depth is 30km with density 2800 kg/m3 for the crust, with 450 kg/m3 density contrast86. After subtracting isostatic response from Bouguer gravity anomaly, we use a 1000 km high pass filter to remove deep mantle structure.

#### Curvature from GOCE

The curvature of GOCE gravity gradient data provide a new perspective of the lithosphere domain as well as the tectonic boundary in the Antarctica continent87. The shape index and minimum curvature from topography and isostatic corrected GOCE gradient show several major tectonic structures in Antarctica. These tectonic settings reveal the different geology under the ice sheet.

### Magnetic field

Sedimentary rocks usually contain little or no magnetisation. The magnetisation of basement rock is usually much higher than sedimentary rocks. The short wavelength signal produced by crystalline bedrock is masked when overlain by sedimentary layers to produced and eventual relatively smooth magnetic signal. Moreover, high frequency signals are very rare in sedimentary basin, except where volcanoes are present.

#### TMI and residual

The total magnetic intensity (TMI) of the PolarGAP72 ,ROSETTA-Ice80 and IceGRAV81 surveys are smoothly stitched to the ADMAP-2 dataset using a fast Fourier Transform suturing method. The new magnetic compilation is reduced to the magnetic pole using the differential reduction to pole (DRTP) method88. We filter the wavelengths > 1000 km to remove the effect of deep magnetic sources.

#### TA

Total Amplitude (TA) of magnetic vector is the amplitude of magnetic vector (three components) data, which has been shown to be a useful tool for qualitative interpretation of magnetic data at large area89. It weakly dependent on the magnetization direction, thus less effected by the remanent magnetizations. We transform the TMI field to TA by assuming geomagnetic field direction as the magnetization direction90.

#### TMI roughness

We calculate the magnetic roughness using the standard deviation of the DRTP-TMI response in 30 km window size after high pass filtering 60 km wavelength. We also use the Analytic Signal Amplitude of DRTP-TMI to highlight the short wavelength information. The roughness coincides with continental margins with exposed basement rock due to glacial erosion69. Extreme high short wavelength variation distributed from Ross Ice Shelf to the Marie Byrd Land is interpreted to be volcanic rock. The smooth magnetic signal is associated with the inner continent and thick ice coverage. The Transantarctic Mountains show the non-magnetic character within meta-sedimentary and meta-volcanic outcrops area. In West Antarctica Rift System, we can clearly see the variation of the rough magnetic anomaly is interrupted by a less magnetised layer, which corresponds to the inferred sediment fill by the Werner filter method66. However, due to the strong magnetised volcanic rocks, the overall area shows rough magnetic respond.

#### Pseudo-gravity and its residual

Pseudo-gravity transform enhances the long-wavelength information of the magnetic source, which can illustrate regional deeper magnetic source variation. Follow the work of Salem, et al.91, we integrate the DRTP-TMI field, and then normalise with the ambient magnetic field. The pseudo-gravity amplify the long-wavelength magnetic information. We calculate the residual of psedo-gravity by a Gaussian high-pass filter of 1000 km to remove the magnetic signal due to very deep crustal architectures.

The residual pseudo-gravity shows several strong negative anomalies in the thick metasedimentary basin proximal to the Ellsworth–Whitmore Mountains as well as the Beacon supergroup in the Transantarctic Mountains and metasedimentary basin in Victoria Land.

### Cryosphere and Solid-Earth Structure

The relationship between earth structure and ice sheet properties to the bedrock type is complicated and remains unclear. Here we use crustal thickness85, heat flux92, ice flow speed93 and basal friction94 as evidence layers, and we aim to test the relationship with these basal responses to the subglacial bed condition, and understand the potential interactions between the cryosphere and solid earth.

## Methods

### Evidence layer processing

While geophysical data collection in Antarctica is ongoing, there are still large data gaps in the current continental scale data compilation and we have endeavoured to filled data gaps with the latest publicly available airborne geophysical measurement.

To maximise the utility of the multidimensional datasets, we impute the evidence layer to generate a complete multi-dimension dataset. We fill the data gaps using the global mean value of each dataset, as the majority of missing values are associate with ice sheet properties at the ocean or beneath the ice shelf, which is not associated with the underlying bedrock properties.

### Evidence layer reduction

Highly correlated evidence layers indicate duplicate and unnecessary information. Inclusion of duplicate information lowers model performance and model interpretability. We use a two-step approach for evidence layer selection95. We estimate the importance of evidence layers using the Boruta method96. We then use the Maximal Information Coefficient (MIC) to measure the independence of evidence layers. The highly correlated, thus unimportant evidence layers, are removed according to the RF out of bag error metrics.

The Boruta algorithm adds shadow attributes by copying and randomising predictor variables to the model domain96. It trains a RF model and computes the Z scores. The Maximum Z score among Shadow Attributes (MZSA) is then used to measure the importance of variables. The original predictor variable is important if its variable importance score is significantly higher than MZSA, while its unimportant with variable importance score is significantly lower than MZSA. The algorithm stops until the importance is calculated for all variables, or reach the maximum RF runs limitation.

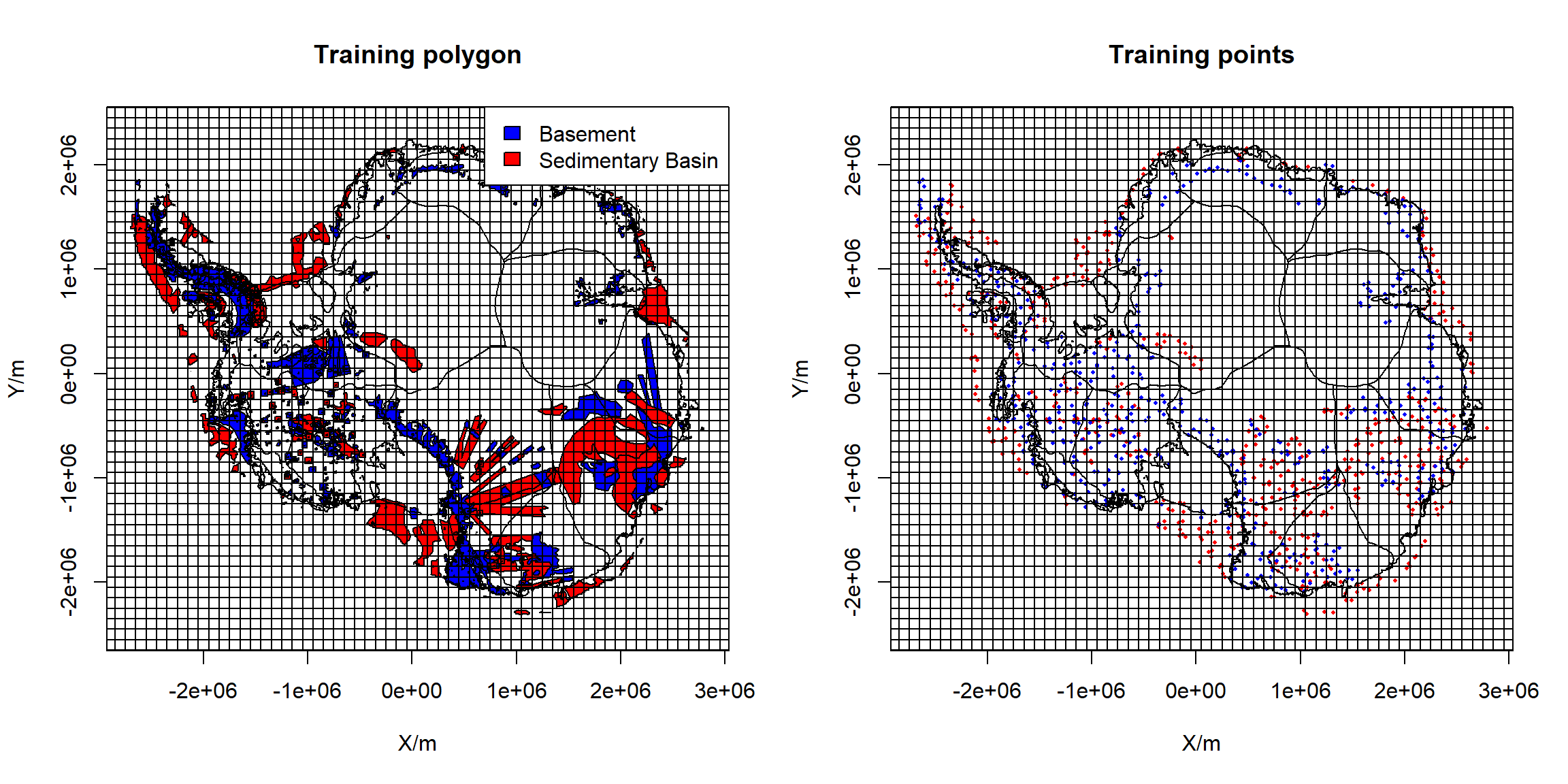
The MIC estimates the highest normalised mutual information of two variables estimated by different bins97. Mutual information is based on whether the relationship between two variables could be captured by scatterplots from different subsets of the data. MIC is zero for two statistically independent variables and close to one for highly correlated variables. We use the R package Minerva to calculate the MIC for all variables98.

In the following step, we test the influence of the model performance with the correlation of evidence layers. We test using evidence with MIC ranging from 0.1 to 1 and assess using the model OOB (out-of-bag) errors in RF. In each test, we remove the correlated evidence layers with lower importance and keep the higher important evidence according to the Boruta evidence importance.

### Training Point Sampling

The performance of the machine learning method depends greatly on the training point set used for model training. For the lithology mapping problem, sparse training data with a spatially balanced class sample is preferred, as it limits the bias associated with prior knowledge and reduces the risk of overfitting47,48. This is an important consideration in this study, as we generate training information from different sources with associated variable uncertainties. The spatially balanced sparse training points could maintain the prior information and reduce the chance of overfitting in the misclassified or complex bedrock type.

Training points are located in a reference grid (fishnet) with 100 km cell size to maintain an even spatial distribution. For each cell in the fishnet, we randomly sample one point in each bedrock type within the overlapping area of the training polygon and the fishnet cell. By doing this, we obtain spatial evenly distributed training information with 613 points for sedimentary basin and 665 points for crystalline basement. We further reduce crystalline basement training points to 613 points by random subsampling. This generates a balanced in the subclass and spatially even distributed training set with near 3.2% of total training area. The sampling process is illustrated in Fig. S4.



**Fig. S4 Illustration of the sampling process using a reference fishnet with 100 km grid**. At left: Training polygon overlie with reference fishnet. At right: Red dots indicate sampled sedimentary basin; blue dots indicate sampled basement rock. We sample maximum 1 point of sedimentary basin and crystaline bed in each fishnet cell.

### RF Model training

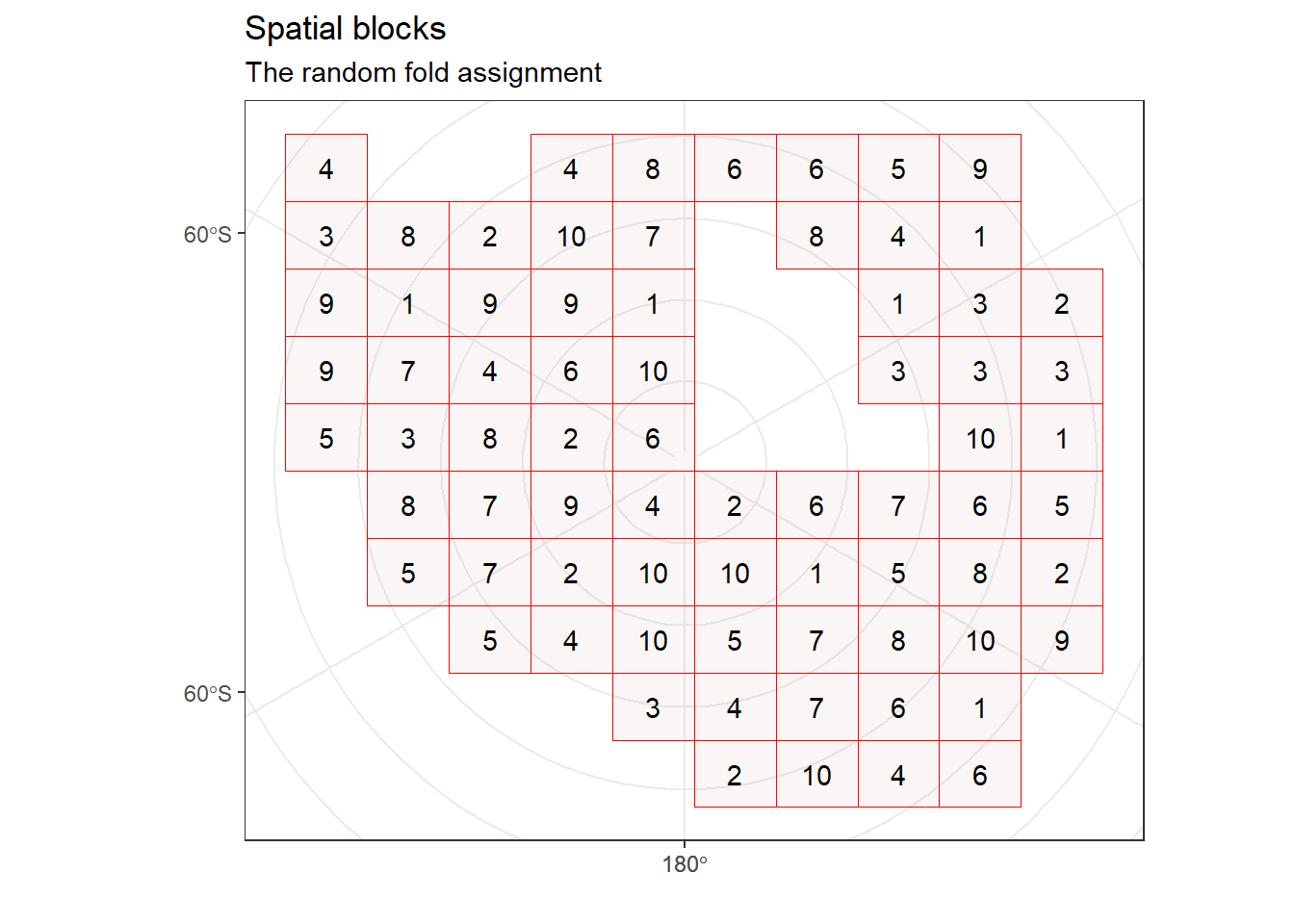
RF has several hyper-parameters to control the training process, including the number of trees (ntree), number of evidence layers to use at any given split (mtry). These hyper-parameters might has a large influence on model accuracy. We use a grid search method to test the influence of these two hyper-paremeters to the RF training accurancy. The RF training accuracy is estimated by 10-fold cross-validation with three times repetition.

After selecting the hyper-parameter, we use a balanced number of sedimentary basin and crystalline bedrock training points for model training (613 points in each class). In order to maintain a stable and unbiased result, we repeat the training point sampling process using the reference fishnet grid 10 times. In each time, we generate a sub-RF model based on the sampled training points at this time. Each sub-RF model result indicates a representation of subglacial bedrock type. We use the mean of 10 sub-RF predictions to determine the final likelihood for classified subglacial sedimentary basin distribution.

RF use the mean decrease in accuracy or node purity to estimate the relative importance of evidence layers. The node purity might be biased when evidence layers vary in their dynamic range99. Here we use the mean decrease in accuracy to measure the importance of evidence layers. The mean decrease in accuracy measures the model performance of the evidence layer when it is assigned randomly, with the rest of the evidence layers unchanged. The evidence layer is more important when model accuracy decreases in its absence.

### Accuracy measurement

Cross-validation is widely used to evaluate the model performance without independent validation data, such as this case given we lack ground truth information. However, cross-validation tend to overestimate the model performance due to geophysical data representing the volume response of rock properties with spatial correlation100. To overcome this problem, we use a 10 fold block cross-validation to evaluate the model performance100. In 10 fold block cross-validation, the model domain is subdivided into 10 equal-sized blocks. In here, we divide the whole model domain into 10 rows with 10 columns blocks. Blocks are then randomly assigned to a ‘fold’ until each fold has a similar amount of training points (Fig. S5).

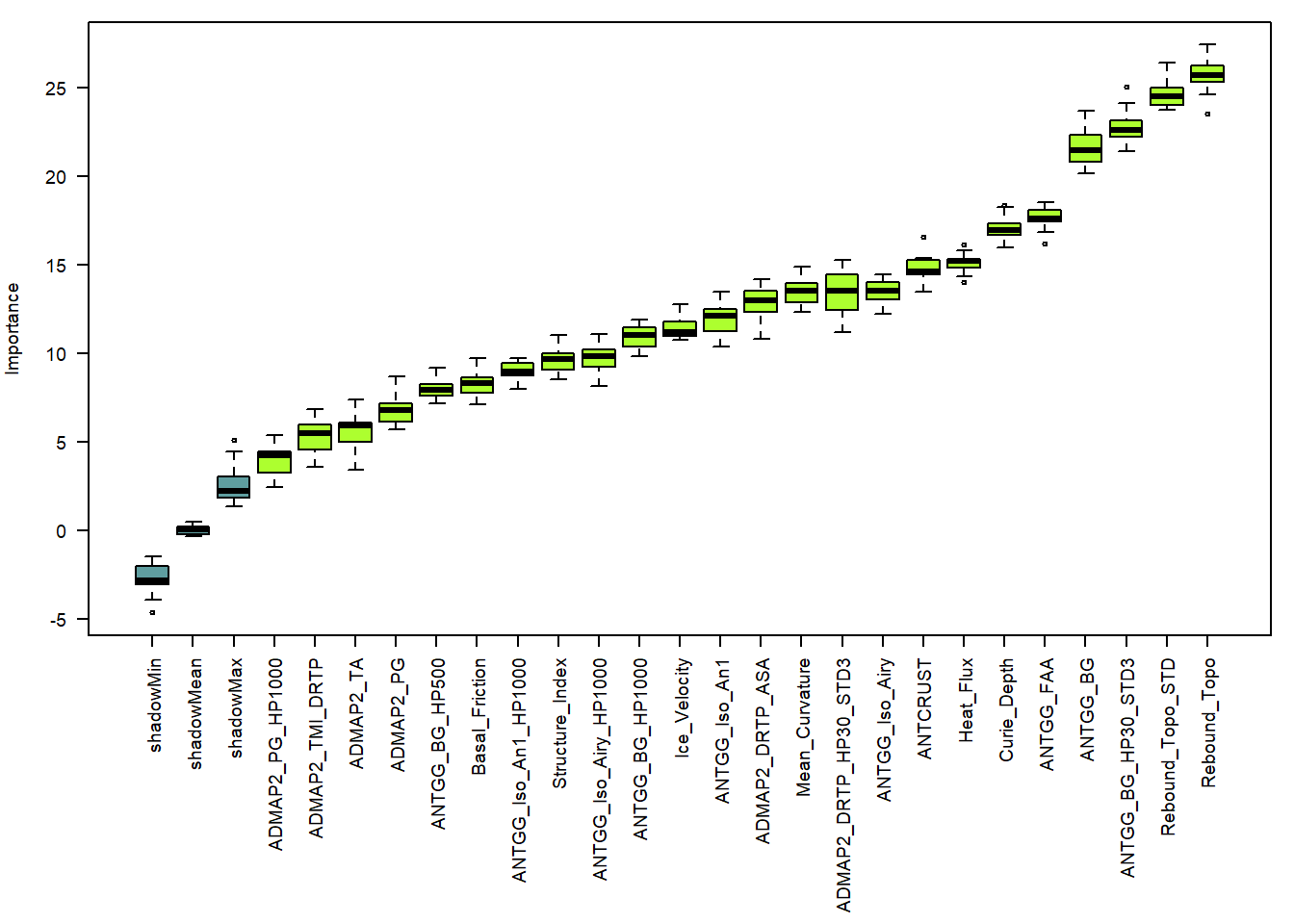
**Fig. S5 Illustration of the block fold in cross-validation process.**

Nine folds are used to train the model, with the remaining fold used for validation. Cross-validation allows the generation of a confusion matrix that represents model accuracy (Table. S2).

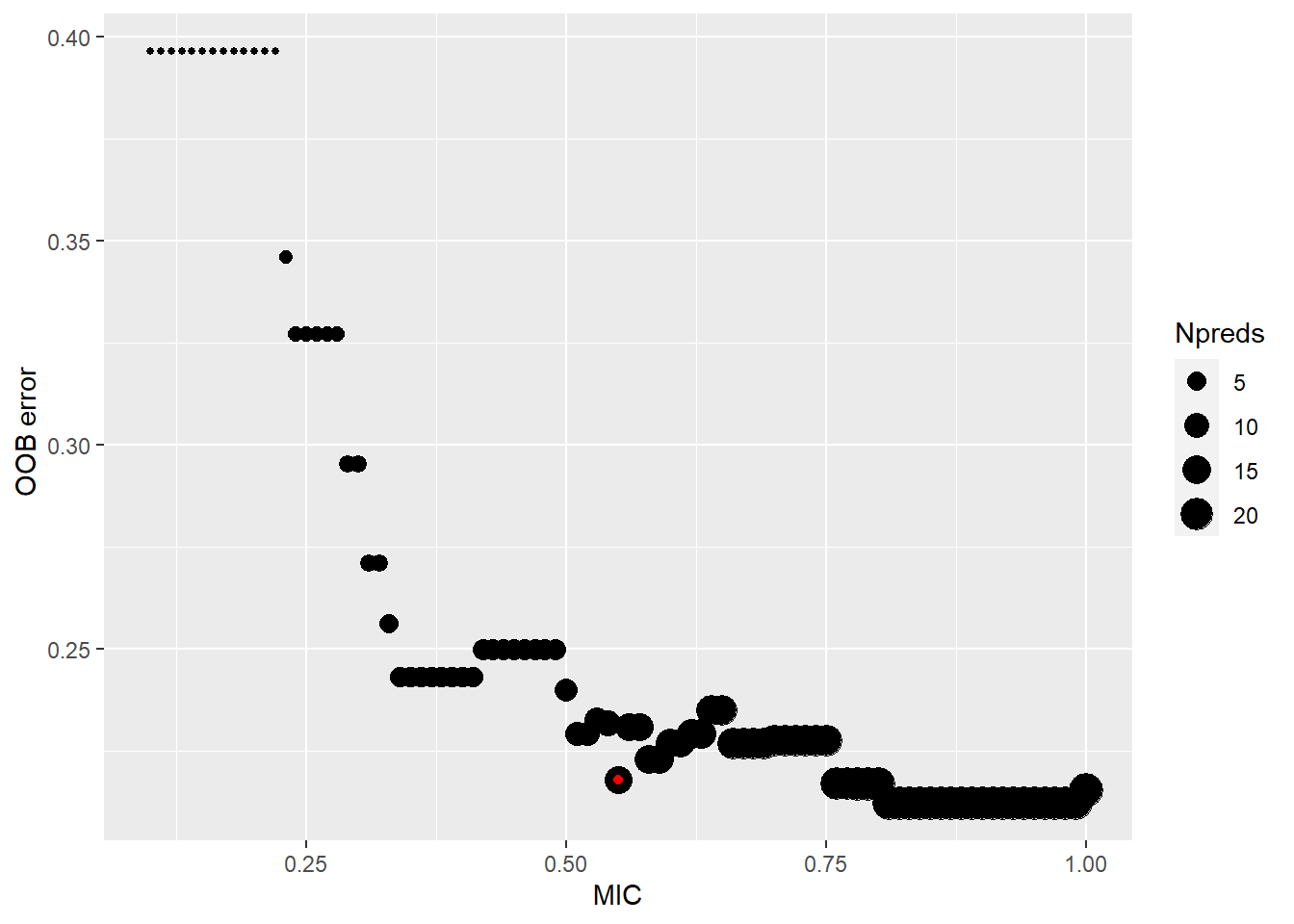
## Result

### Evidence selection

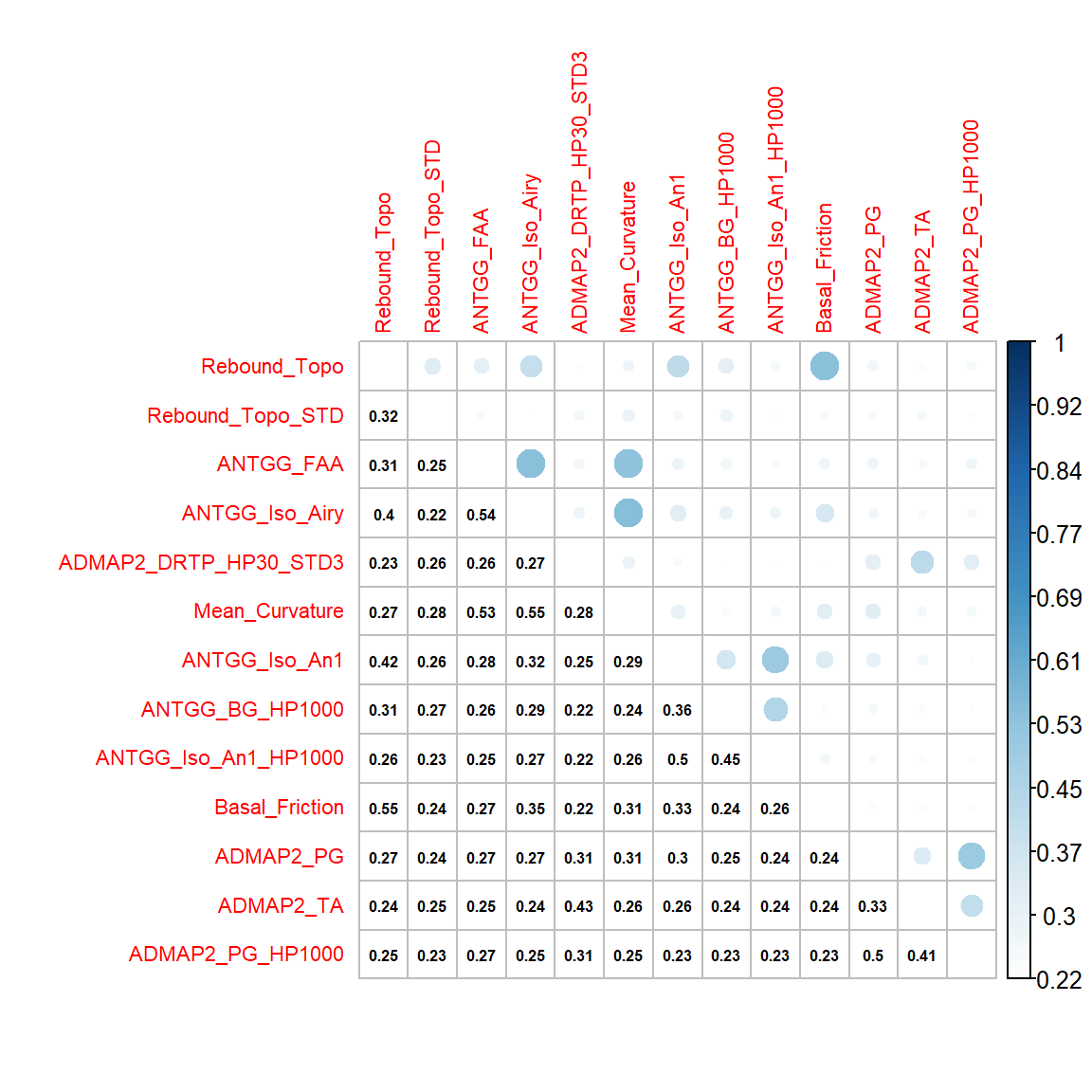
23 evidence layers were all deemed important in the Boruta analysis, with max 500 times importance source runs and a confidence level of 0.5 (Fig. S6).

** Fig. S6 Model importance generated by Boruta variable selection**. All evidence layers show important as they all higher than maxmum shadow variables.

Based on a plot of the RF OOB error estimates over the MIC, a value of 0.55 was selected (Fig. S7). This selection ensured higher model performance while at the same time minimising the number of predictor variables by removing highly correlated and unimportant evidence layers. The final 13 evidence layers were used for the RF model training (Fig.S8). These layers in cluding: isostatic rebound topography, topography roughness (STD of isostatic rebound topography), Free-Air anomaly, Airy isostatic gravity, Mangtic Roughness (STD of DRTP-TMI), mean curvature, AN1 isostatic gravity, Bouguer gravity residual (HP1000), Airy isostatic gravity residual (HP1000), Basal friction, Pseudo gravity, Total Amplitude of magnetic vector, Pseudo gravity residual (HP1000).

****

**Fig. S7 Influence of MIC on the out-of-bag error of RF model.** The red dot (MIC=0.55) shows the cut off value of MIC with low OOB error and number of evidence layers.

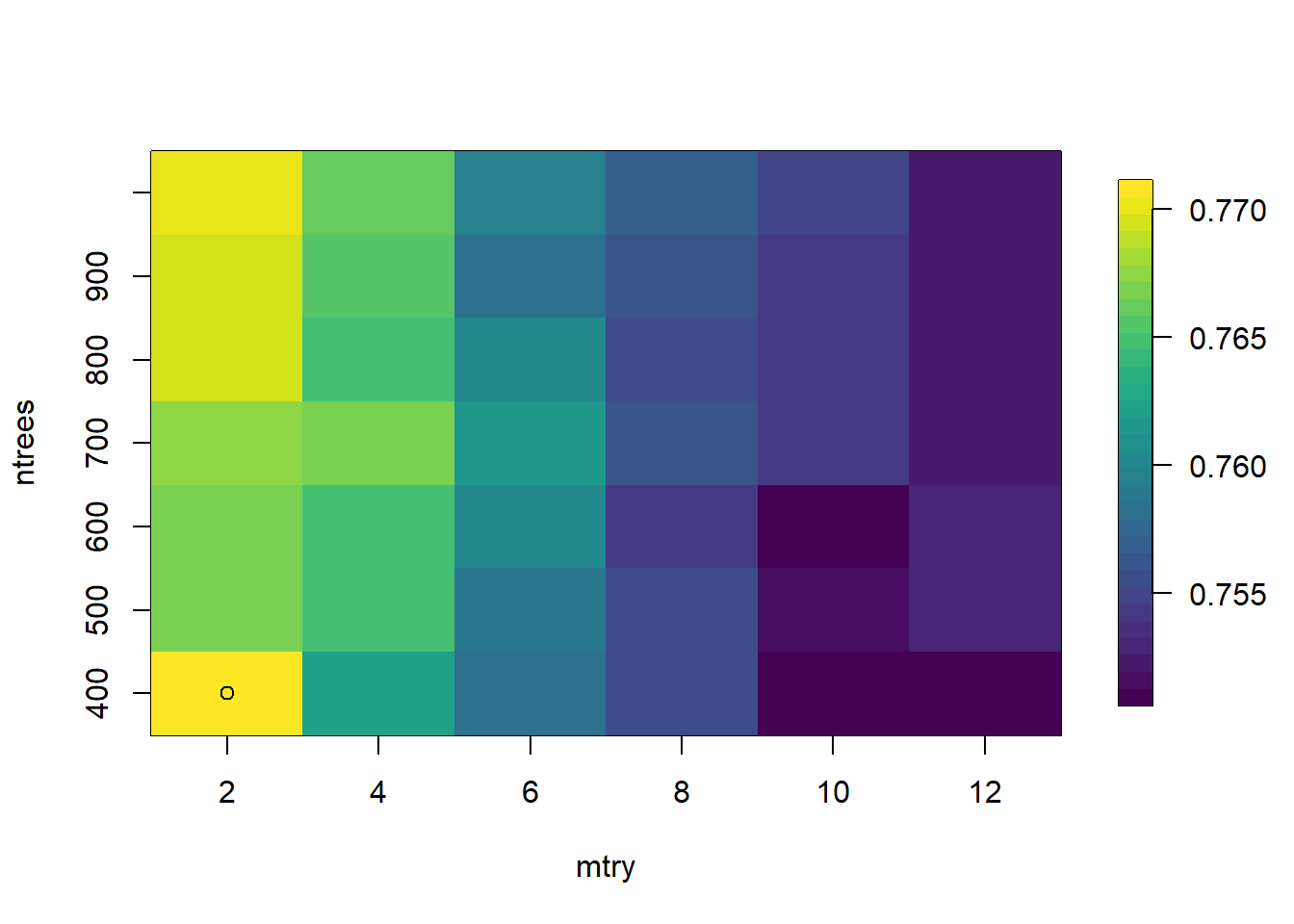


**Fig. S8 Correlation plot (MIC) for all selected evidence layers.** The maximum correlated layer is basal friction with the isostatic rebound topography data.

### Model turning

We use a grid search method to check the RF out-of-bag error with the influence mtry and ntrees. The mtry change from 2 to 12 with 2 intervals. The ntrees change from 400 to 1000 with 100 intervals.

Based on the CV accuracy, the model performance is stable by changing hyperparameters (Fig. S9). The maximum difference is less than 0.15. We find the maximum CV accuracy is achieved with mtry = 2, and ntrees = 400. We will use this value to build the RF model.



**Fig. S9 CV accuracy with various mtry and ntrees for RF model turning.** The maximum accruacy is achived with mtry=2 and ntrees=400.

### Model accuracy and uncertainties

In here, we would like to communicate the model accuracy and uncertainties in three ways. At first, we compare the RF prediction with the prior training information. This indicates the discrepancy between all prior information with the RF model result. Then, we check the spatial distribution of the consistent and inconsistent prediction during the block cross-validation process. This information could reflect how RF model performs with an area without training information. In the end, we show the internal variation of 10 sub-RF model output, which indicates the uncertainty of the RF model during the training point subsampling and training process.

For a cell to cell comparison with initial all training information (Fig. S10), 2421 cells (6.33% of the total) are inconsistently classified as sedimentary basin, 1547 pixels (4.04%) are inconsistently classified as crystalline bedrock. The remaining 34287 pixels (89.63%) is consistent during the RF training process.

The 10 fold block cross-validation resulted by summarising all 10 sub\_RF model has an overall model accuracy of 75.11%, with 95% confidence limits of 74.33% and 75.87% (Table S.2). We further calculate the 10 fold block cross-validation by combining all training points to train and validate the model. This result shows the model has an overall accuracy of 75.87%, with 95% confidence limits of 75.11% and 76.63% (Table. S3).

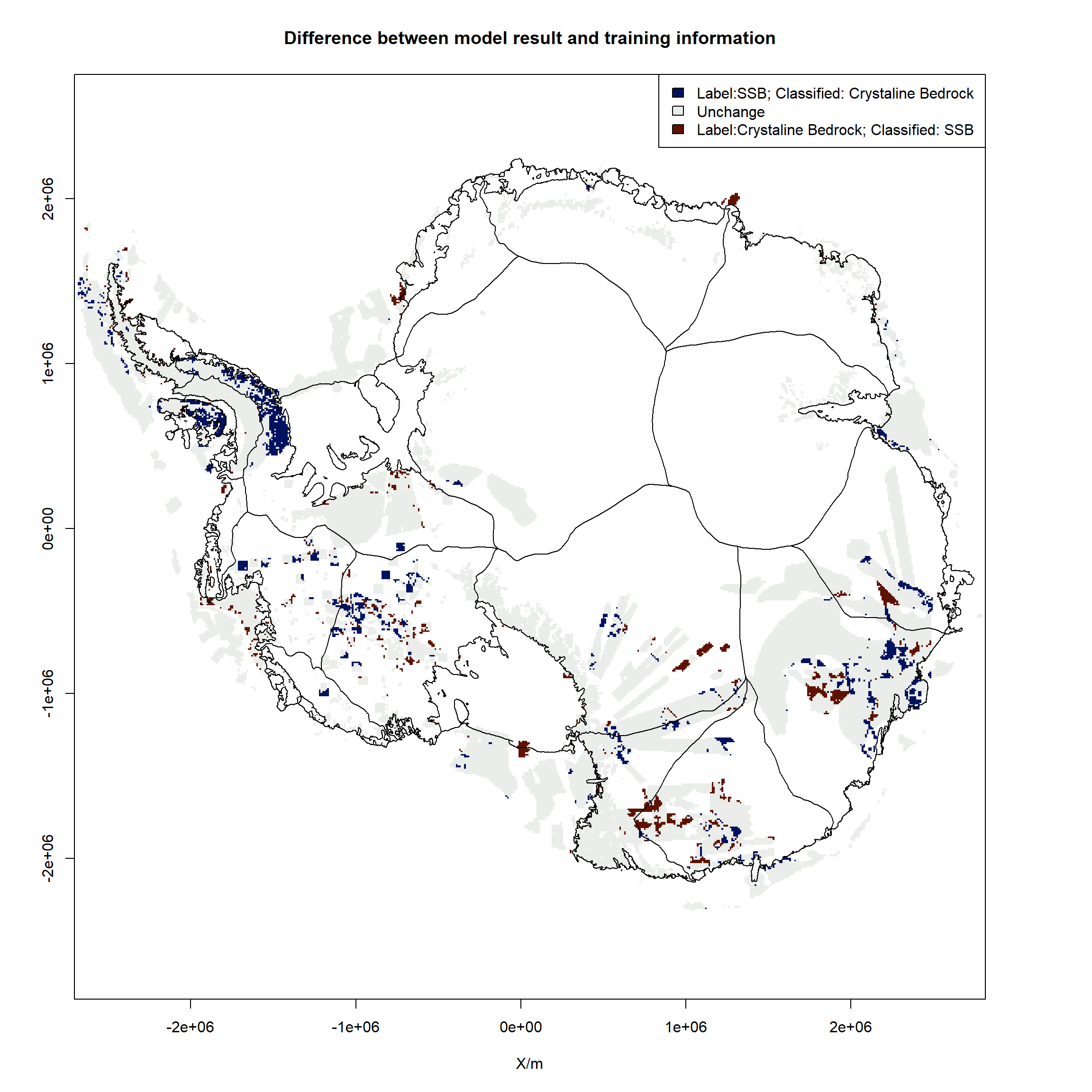
Although the model performance during block cross-validation does not directly translate to the unknown area without training information, it indicates the utility and weakness of the RF model when applied to areas without training information. The incorrectly classified training points during block cross-validation are shown in Fig. S11. We note the majority of training points inconsistent with the prediction are located at the boundaries between sedimentary basin and crystalline bed. Most other training points are relatively consistently correct during cross-validation process.

The standard deviation of 10 sub-RF models are shown to represent the variation of subglacial sedimentary basin likelihood (Fig. S12). In general, we find a large likelihood variation at the geology boundaries. The highest likelihood variation is located at the Antarctica Peninsula.

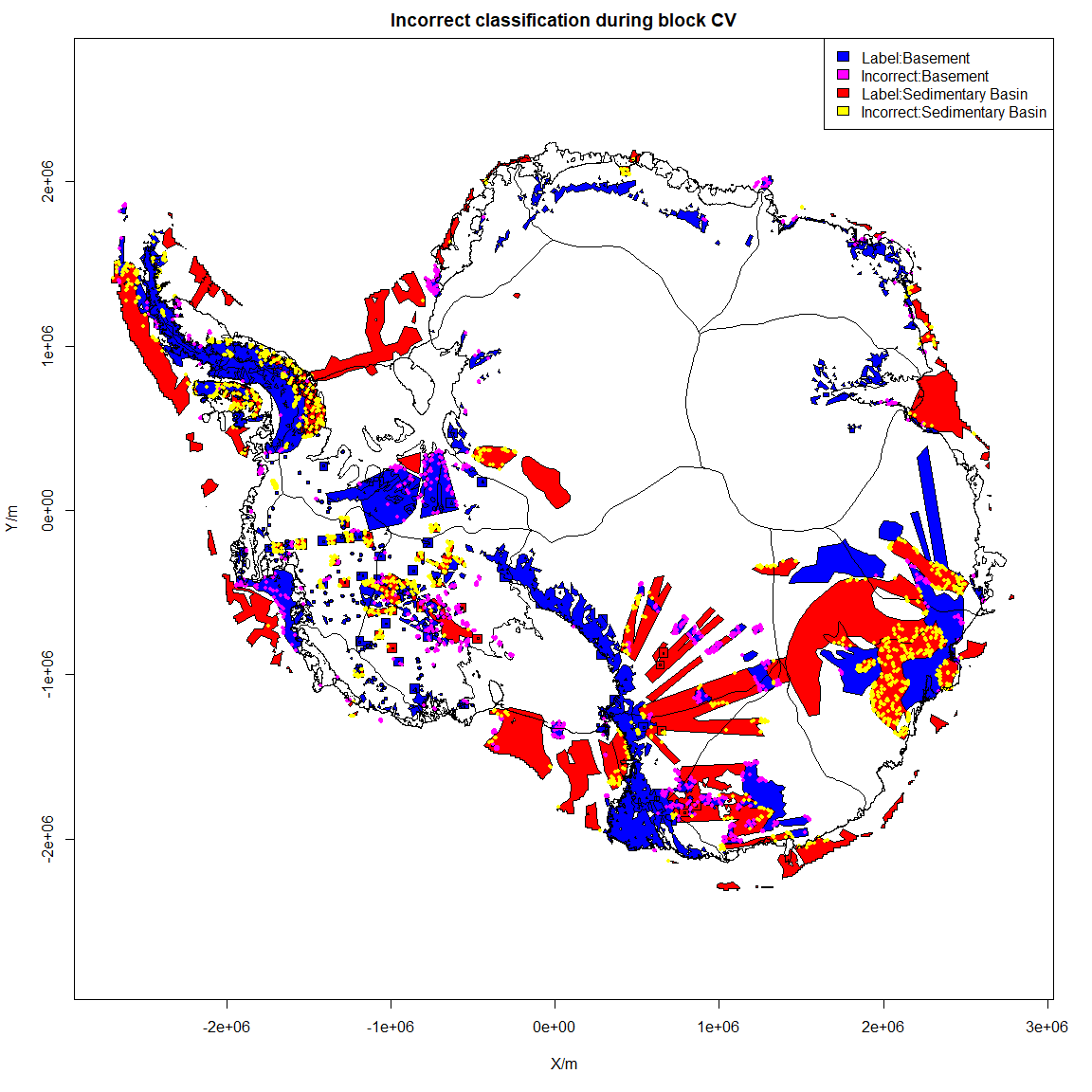
Fig. S12d shows the frequency when a location is predicted as sedimentary basin (likelihood > 0.5) in each RF model. Although the sedimentary basin likelihood value varies in the sub\_RF models during the training point sampling process, the classified sedimentary basin in each RF model are consistent as shown. This pattern suggests the broader sedimentary distribution is consistent, but with a small variable at the margin. The overall classification result is robust during the subsampling and training process.

We further show the correlation of the consistent and inconsistent training points with the STD of all sub\_RF models (Fig. S13). For the mean RF prediction result compared with the initial training information (all information in the training polygon), the consistent prediction is located at the low likelihood variation area, while the inconsistent prediction has a relatively higher likelihood variation (Fig. S13d). This effect is further amplified during the block cross-validation process. We see a large difference for the distribution of consistent and inconsistent predictions. The median of a consistent classification has a STD of 0.65, while the inconsistent classification has a STD of 0.14 (Fig. S13b).

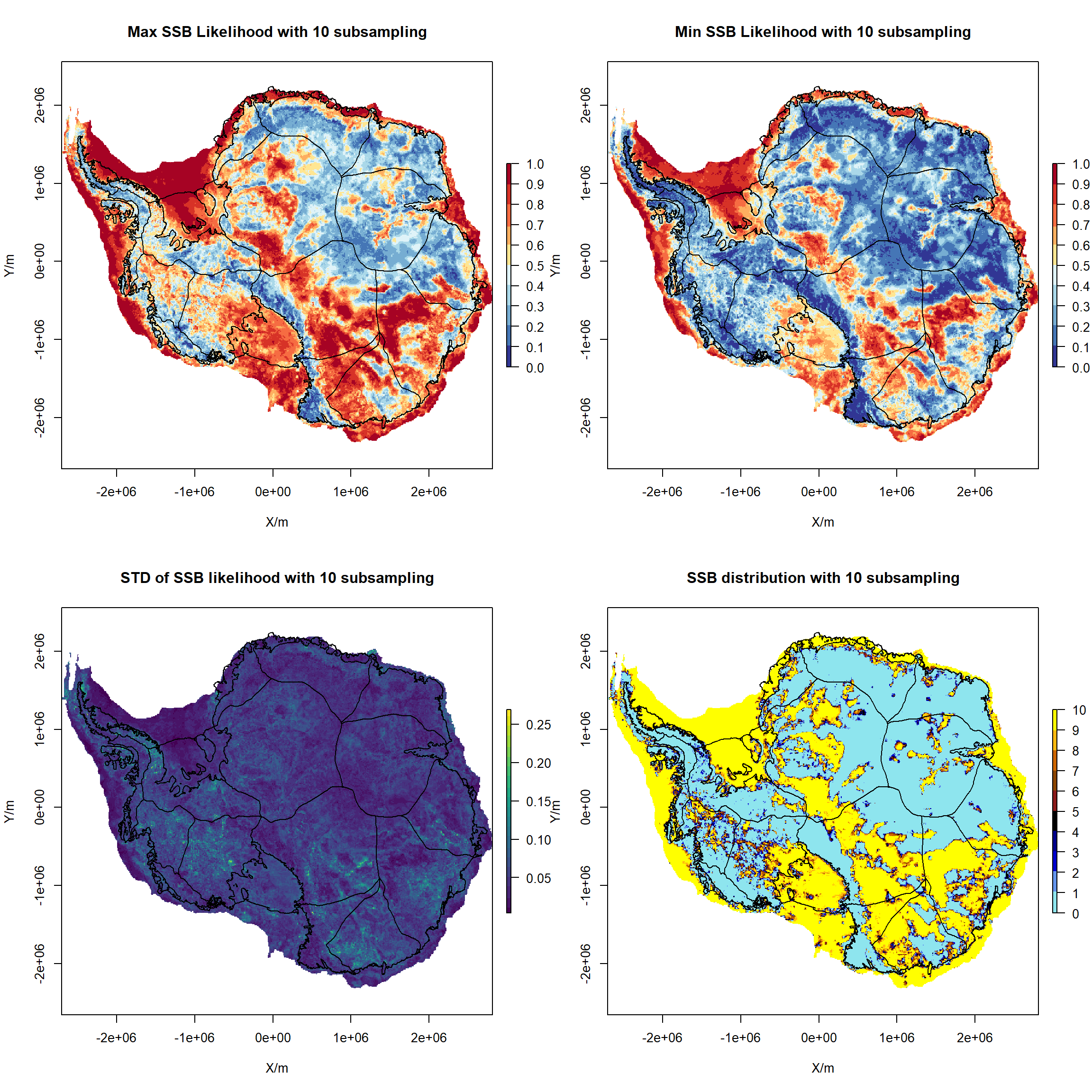
These incorrectly classified points during block cross-validation could be attributed to the inherent complex geology in its setting, and/or the geology boundary, and/or potential incorrect training information. In Antarctica Peninsula, the outcrop sedimentary rock is misclassified as crystalline bed, as it is extensively intruded by volcanic rocks101. The misclassified sedimentary basin in West Antarctic Rift System associated with the positive gravity response and surrounded by volcanic rock, supports a complex signal in its geophysical respond66. In Wilkes Land, the narrow Knox subglacial basin and strong reverse magnetic signature in Law Dome all indicate the unique geology setting. The Sabrina Subglacial Basin shows serval misclassification at its internal area. The overlying ice sheet has been proposed to retreat and advance serval times69,102. The past glacier activity are likely to thin the sedimentary basin and change its geophysical response in this area.



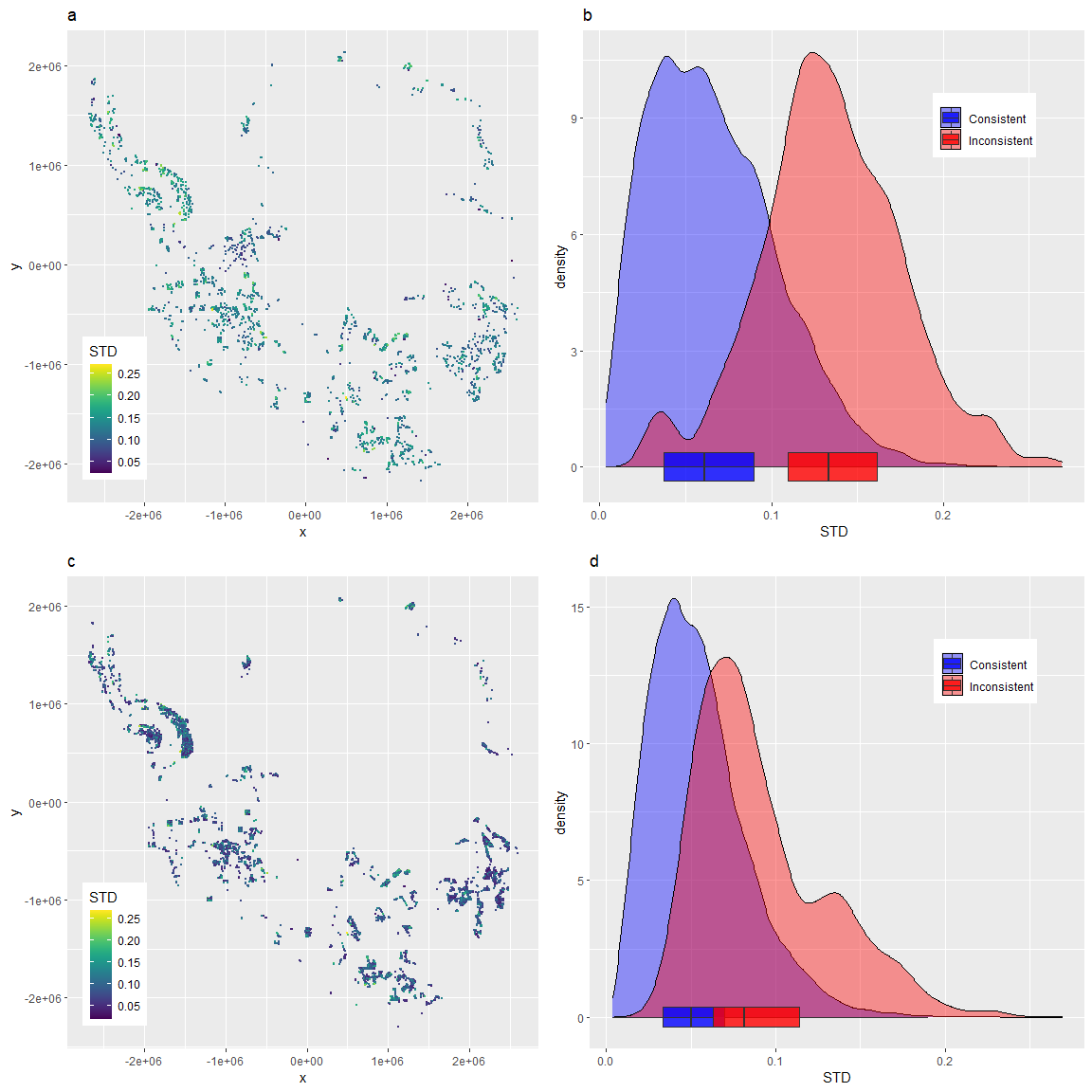
**Fig. S10 Comparison with the initial training information as consistent (white) and inconsistent (red and blue).** Red represents locations labelled as crystalline bedrock but classified as subglacial sedimentary basin (SSB). Blue represents locations labelled as subglacial sedimentary basin (SSB) but classified as crystalline bedrock.



**Fig. S11 Incorrect classification by 10 fold block cross-validation.** Yellow dot shows labelled sedimentary basin is classified as basement rock during cross-validation process; magenta shows labelled basement rock is classified as sedimentary basin.



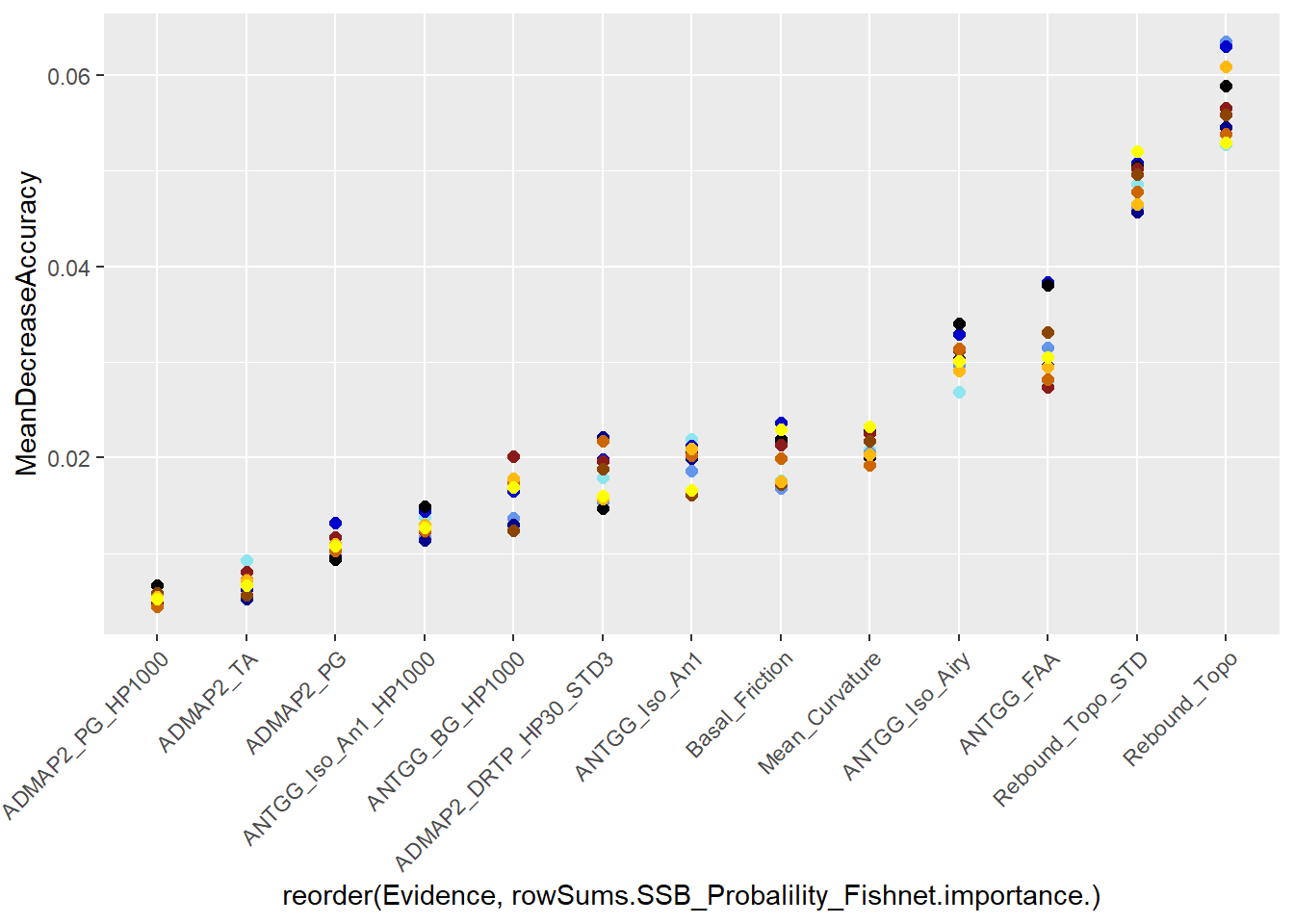
**Fig. S12 Maximum, Minimum, and STD of subglacial sedimentary basin (SSB) likelihood and classified sedimentary basin in the 10 times subsampling and training process.**



**Fig. S13. The correlation of consistent and inconsistent classcifation with the STD of all sub\_RF model.** a, the STD of SSB likelihood of inconsistent training points during the 10-fold block cross-validation. b, density plot of the STD of SSB likelihood of consistent and inconsistent of training points during block CV. c, the STD of SSB likelihood of inconsistent intial training information. d, density plot of the STD of SSB likelihood of consistent and inconsistent of intial training information.

### Evidence importance

The RF evidence importance for all sub-RF 10 models is shown in Fig. S14. The most important information is related to the topography, its roughness. Then the following evidence layers are related to the Solid Earth structure (Free-Air gravity, Airy isostatic gravity, gravity mean curvature). The texture information of magnetic data is ranked as a relative higher import evidence, as it has a strong power in mapping volcanic rocks.

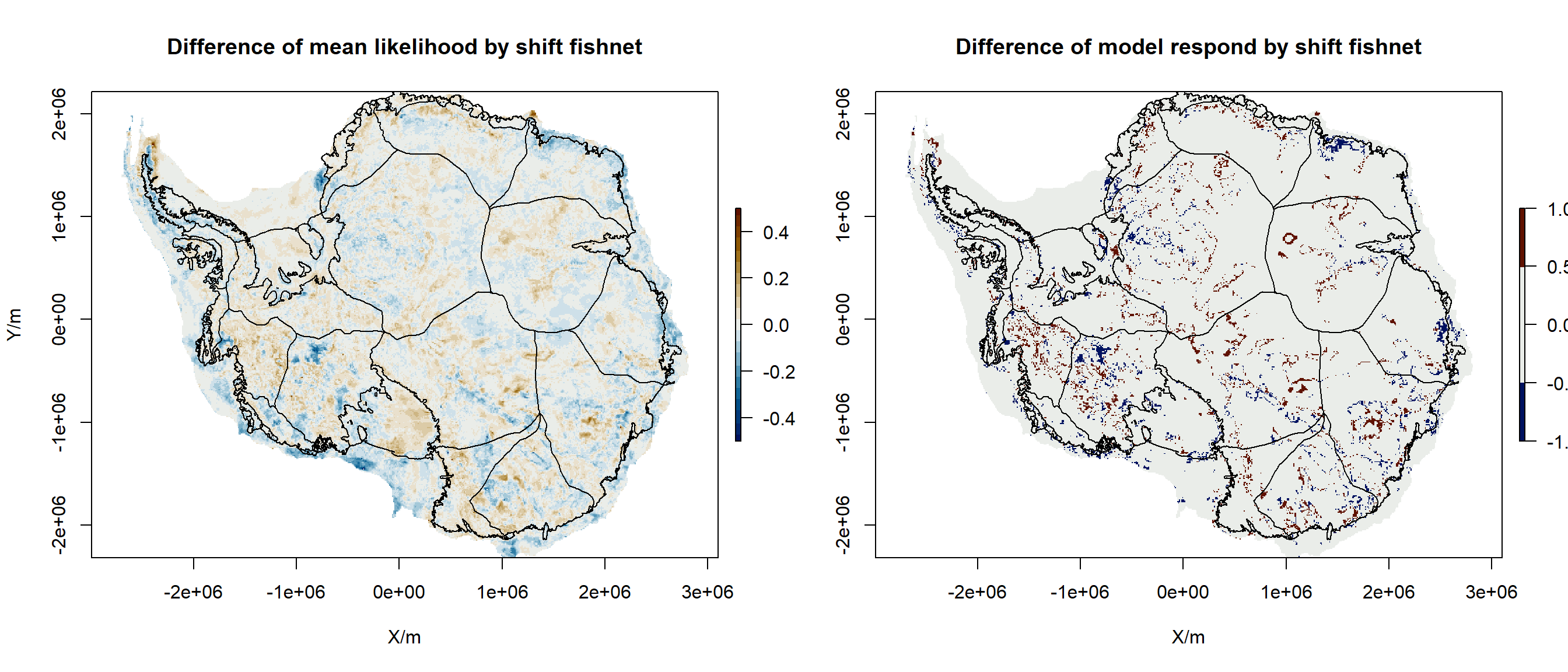


**Fig. S14 Mean decrease accuracy of evidence layer ranked by RF.** The colored dot shows the mean decrease accuracy variation in the 10 sub-RF models. We rank the evidence importantce based on the mean value of mean decrease accuracy of all 10 sub-RF models.

### Sensitivity test

We further test the model change due to the change of reference sampling fishnet by shifting the whole fishnet in the top and right half-cell size (50 km). We repeat the same training point sampling and model classification 10 times and test the model performance after the change of the reference training point.

The overall RF classification performance difference is very small (Fig. S15). We see the majority difference of sedimentary basin likelihood is less than 0.1, with the largest change located near the grounding line of Ross Ice Shelf, and inland area at East Antarctica. We note, the lack of airborne geophysical measurement at this sector, which support an inherent uncertainty in this region. The other large change is located in the inner Sabrina Subglacial Basin. The natural setting of this area is mixed sedimentary basin with basement zone due to prior glacial retreat and advance69.



**Fig. S15 The likelihood variation and classfication difference of RF model by using differnet fishnet.** The positive likelihood variation indicates a reduction of sedimentary basin likelihood using a shifting fishnet approach. The negative likelihood variation indicates an increase of sedimentary basin likelihood using a shift fishnet. The blue area means the bedrock is classified as basement by orginal fishnet, but as sedimentary basin in shifted fishnet. Red area means the bedrock is classified as sedimentary basin by original fishnet, but as the basement in shifted fishnet.

## Model limitation and robustness

This study utilises a predefined subglacial geology map to reclassify subglacial bedrock types beneath Antarctic ice-sheet. The result has defined sedimentary basins in Antarctica to an accuracy of 75.11%. However, this approach has suffered some drawbacks due to missing knowledge (inaccessible of Antarctic ice-sheet bed) and the limitation of the current generation of Antarctica datasets.

### Uncertainty in training point calibration

The training points are primarily generated by the integration of interpreted subglacial geology. The inaccessible Antarctica ice sheet makes the training point selection difficult and uncertain. A typical example is the misclassified information in the North Wilkes Subglacial Basin, where crystalline basement training points are classified as sedimentary basin. The interpretation of aeromagnetic data indicates thick metasediment preservation in the East Basin in WSB, and modern sedimentary basins are mainly preserved in the magnetic low area, the rest positive anomaly region showing basement structure70. Inversion of gravity data suggests a broader distribution of sedimentary basin with variable sedimentary basin thickness. This misclassified area shows 100 – 300 meters sedimentary basin thickness by gravity inversion. These thin sedimentary layers are close to the uncertainty range of the current continental-scale gravity dataset (10 km cell size). This uncertain information does have a contribution to the model uncertainty. However, this uncertainty is unavoidable with the majority of subglacial geological interpretations performed using airborne geophysics data.

### Uncertainty in evidence layers

The model importance shows the topography, topography roughness and Free-ari gravity ranked top-3 among all evidence layers to classify sedimentary basin and crystalline basement. The inherent smoothing by data interpolation in continental-scale geophysical datasets could lead to excessively producing a smooth texture, which artificially inflates sedimentary basin distribution and underestimates basement rock distribution. It’s especially in the central East Antarctica, where large line spacing geophysical measurements smooth the geology signal. As the note in newly collect airborne gravity data in Dronning Maud Land, that the West Ragnhild Trough anomaly is much sharper and deeper compare with the ANTGG dataset103. However, the latest bedrock topography model (BedMachine), and updated AntGG gravity model has greatly improved the data coverage and reduced uncertainties in the continental scale dataset. Especially at the fast-flow region, the mass conservation approach has resolved the geometry of bedrock in detail.

### Model robustness

Although the model result is limited by the uncertainty in both training point and evidence layers, our result is still robust in the most upper stream area in Antarctica. These areas have a consistent RF model prediction with a relative low STD. Meanwhile, since the 1990s, airborne geophysical measurement has started to focus the onset of ice streams74. The overall data quality over the fast-changing ice stream is high compared with the interior continental region. These rich data measurements improve the confidence of bed classification at these regions. Meanwhile, we loosely constrain the subglacial geology during the data classification by subsampling the interpreted bedrock type distribution using a reference fishnet. The 10 times subsampling and training process gives a good balance of model variability and accuracy. The result shows a large area with a consistent RF classification result. We also test the influence of reference fishnet on the model performance by shifting its location. The model results with different fishnets are mainly consistent. The major difference is located at several geology boundaries.

Overall, we implement a robust workflow in the subglacial sedimentary basin distribution mapping. The model could be easily updated when new knowledge (training points) and new measurement (updated evidence layer) become available.

1. Hydro-mechanical model

CVFEM\_Rift2D is a muti-physical model including geomechanical modelling, hydrologic modelling, solute transportation and heat transformation104. The geomechanical modelling calculates the deformation and stress change during the ice sheet advance and retreat. The mean normal stress change act as a source loading term in the hydrologic modelling. In CVFEM\_Rift2D, the geomechanical and hydrologic modelling are partially coupled by modifying the source loading term using interpolation to drive the groundwater flow.

## Method

### Geomechanical Modelling

The lithosphere deformation and due to ice sheet weight is solved by a 2-D plane strain elasticity equation:

where are the material coefficients, including Young’s modulus and Poisson’s ratio . It is defined by:

After solving the lithosphere deformation, the horizontal , vertical , out of plane normal stress , and total normal stress can be calculated as follows:

### Ice Sheet Evolution

The ice sheet evolution is represented by a parabolic polynomial equation, the ice thickness at the location and time is shown as:

where and show maximum ice sheet thickness and expansion at time .

For the normal glacial retreat scenario, we simulate a 30 ka glacial cycle with linear expansion and retreat. From 0 ka to 19 ka, ice sheet grows from 0 to 3 km thickness to 1300 km lateral extent. The ice sheet thickness remains 3 km during glacial maximum for another 1 ka. After glacial maximum, the ice sheet retreat to 0 km in the next 10 ka.

The net estimation of Antarctica glacial retreat indicates that fast-flowing ice sheet (>800 m/a) retreat 110 meters with per meter of ice thinning105. In here, we keep the ice sheet expansion phase unchanged, but using different ice sheet retreat rate to test the influence of groundling retreat rates on mean basal water flux. We calculate the ice sheet thinning rate as the mean ice sheet thickness change within 64 km to the grounding line. We show that our moderate retreat scenario has a comparable groundling-line retreat and ice thinning rate to Thwaites Glacier. Fast glacier retreat rate has been modelled with glacier retreat over 500 km at 300 years at Thwaites Glacier (1,677 m/a)106.

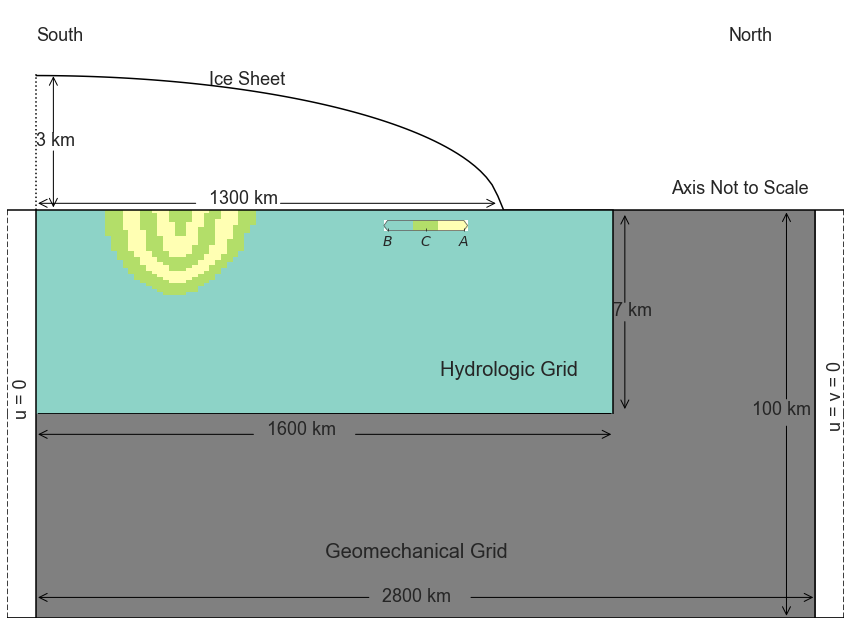
### Hydrology modelling

The subglacial hydraulic head change is solved by the Darcy flow:

where and represent the permeability of basal unit in the horizontal and vertical directions, is the hydraulic head to be solved, is time, is the gravitational acceleration, is fluid viscosity, is the reference fluid density, is the fluid density, is the relative density of fluid density to reference fluid density, is the total normal stress due to ice sheet loading. The Skempton’s coefficient control the percentage of ice loading support by the pore fluid. Measurement of sandstone, shale, and carbonate has a Skempton’s coefficient range from 0.58 - 1, 0.98 - 0.99, and 0.73 - 1 respectively107. We use in that all loading due to ice sheet is supported by the subsurface pore fluid. is the specific strorage. Details of model parameters are listed at Table. S4.

## Model setting and boundary conditions

In CVFEM\_Rift2D, the geomechanical model is coupled with the hydrologic model based on interpolation. Here, we use a coarser model to solve the lithosphere deformation and stress change associate with ice sheet loading and unloading. A fined hydrology grid is implemented to solve groundwater transportation (Fig. S16). In here, we assume that any upward flow water at the surface will transport through ice sheet bed or subglacial till layer.



**Fig. S16 Hydro-mechanical model setting.** The ice sheet over the top of sedimentary basin and crystalline bedrock (B) has a maximum extension of 1300 km and a maximum thickness of 3 km. Ice sheet flows from South to North. The sedimentary basin includes aquifer unit A and confine unit B. The sedimentary basin extends 400 km horizontally with a maximum depth of 3 km. We solve the groundwater flow and heat transportation at the top 7 km in the refined hydrologic grid. The geomechanical model is conducted in a coarser grid with a maximum extension 100 km.

The geomechanical model extends 2800 km in horizontal and 100 km in vertical. There are 281 nodal columns and 11 nodal rows with 10 km interval. We solve the geomechanical deformation with a time interval of 100 years.

In the geomechanical model, we set the north end as a non-displacement boundary condition for horizontal and vertical directions. For the south end boundary, we keep the no horizontal displacement boundary condition unchanged, but release the constraint to allow vertical displacement. We assign the weight of ice sheet as the top stress boundary of the model domain, and solve beam bending equation to the bottom displacement boundary condition.

We assume a flat topography surface in the hydrologic model domain. The sedimentary basin is 400 km long sit on the top of the crystalline basement with a maximum depth of 3 km. The lithology in the sedimentary basin includes two basal aquifer units confined by the three aquitard units. We limit the hydrologic model domain at the top 7 km, as the free liquid phase water connects to the land surface only exist in the upper portion of the crust layer. The fine size hydrology grid has horizontal interval 16 km and vertical interval 0.1 km. It gives hydrologic model domain with 101 nodal columns and 71 nodal rows. We use a 10 year time interval to solve the groundwater flow.

For the hydrological model domain, no flux boundaries are used to the bottom and sides boundaries. We specify the 90% of ice sheet thickness as the top hydraulic boundary conditions, where ice sheet is at floating conditions:

For the heat transformation, the initial temperature at the ice sheet sole is set to 0°. We assume insulated side boundary in the hydrology domain side boundaries, and the basal heat flux condition is assigned at 60 mW m-2.

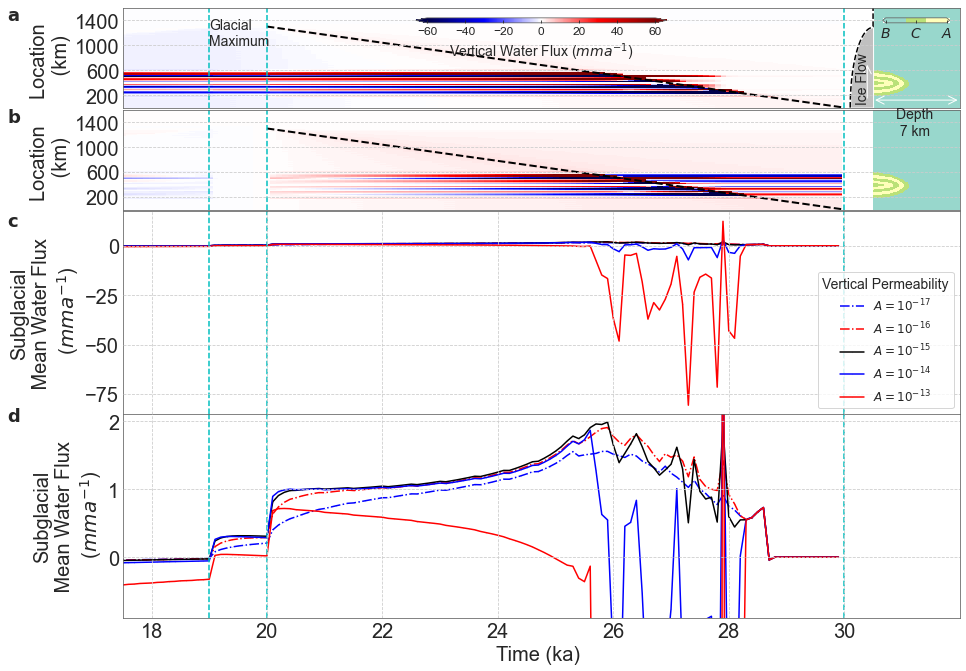
Permafrost generally act as a barrier to the groundwater flow104,108. In our simulation, we ignore the impact of permafrost on groundwater transportation, as a sedimentary basin located at the upper stream with thick ice. A wet base glacier assumption is valid in these thick ice cover area.

## Discussion

### Impact of permeability to basal water flux

Our simulation delineates complex groundwater flow patterns and responds of external loading change controlled by the permeability of the basal aquifer unit (Fig S17). For a normal (vertical permeabilityz =10-15 m2) to low permeability case (z <10-16 m2), the expansion of ice sheet recharge the basal water into sedimentary basin. In turn, the retreat of ice sheet releases the loading to cause major water discharge into the basal water system.

A higher permeable basal aquifer (z ≥10-14) facility groundwater transportation. Hence, it quickly adjusts the ice loading change through groundwater transportation (Supplement Video 2). In this case, water flows into the sedimentary basin (negative vertical water flux in Fig. 17a) at the upper stream, and discharge out at the downstream region (positive vertical water flux in Fig. 17a) during ice sheet expansion. The retreat of the ice sheet causes a reduction of recharging water rate at the upstream of the sedimentary basin, and an enhancement of water discharge at the downstream side (Fig. 17b). This leads to an enhancement of water flux at the basal water system. Once the ice sheet margin retreat over the top of the sedimentary basin, we see the dominant effect of enhanced water recharge at upstream. The major water is transported into the groundwater system and discharged into the ice-free land.

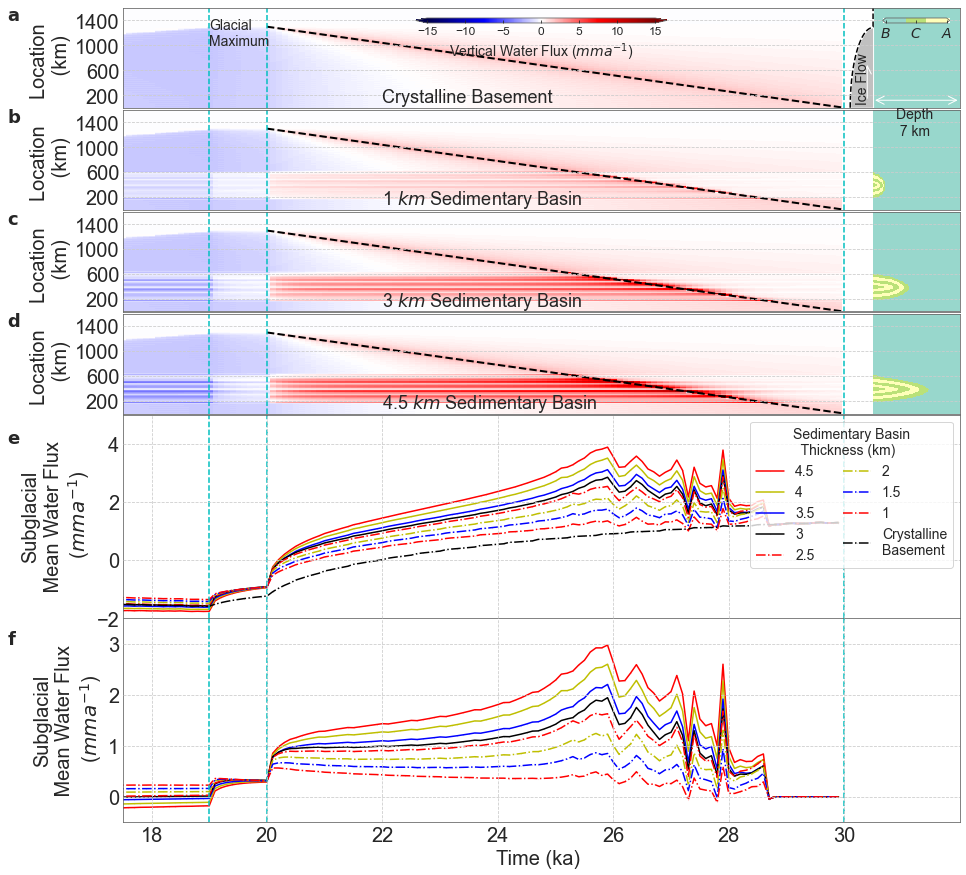
****

**Fig. S17: Impact of permeability to basal water flux. a,** Surface vertical water flux with z = 10-13 m2. The upper stream portion of sedimentary basin shows water recharge, and lower portion of sedimentary basin shows water discharge. **b,** Relative surface water flux change compares with the glacial maximum. At the initial phase of ice sheet retreat, the positive water flux change indicates the reduction of water recharge and enhancement of water discharge. The following stage shows enhancement of both water discharge and recharge. **c**, Mean subglacial water flux change with the permeability of aquifer unit. **d**, A zoomed-in version of mean subglacial water flux change.

Although our simulation indicates groundwater flow pattern varies with the permeability of basal aquifer units, we observe a similar pattern of total water flux at the basal system at the first stage of ice sheet retreat. Before the ice sheet margin retreat over the sedimentary basin, the subglacial water flux is enhanced with a more permeable basal unit. The mean subglacial vertical water flux changed from 1.5 mm a-1 to 2 mm a-1, with the vertical permeability change from 10-17 m2 to 10-13 m2. The following phase shows different subglacial water flux controlled by permeability. A lower to normal permeable support water discharge with a positive subglacial total water flux. The high permeable aquifer shows a complex subglacial total water flux pattern, with major water transform through the groundwater system. This phenomenon is supported by the extreme low erosion rate at limestone109, where major water transport into the Karst limestone system. However, this case is unlikely everywhere in Antarctica, due to the large compaction effect of ice sheet. This could also refer to the case that freshwater discharge on the continental shelf110.

### Impact of sedimentary basin thickness to basal water flux

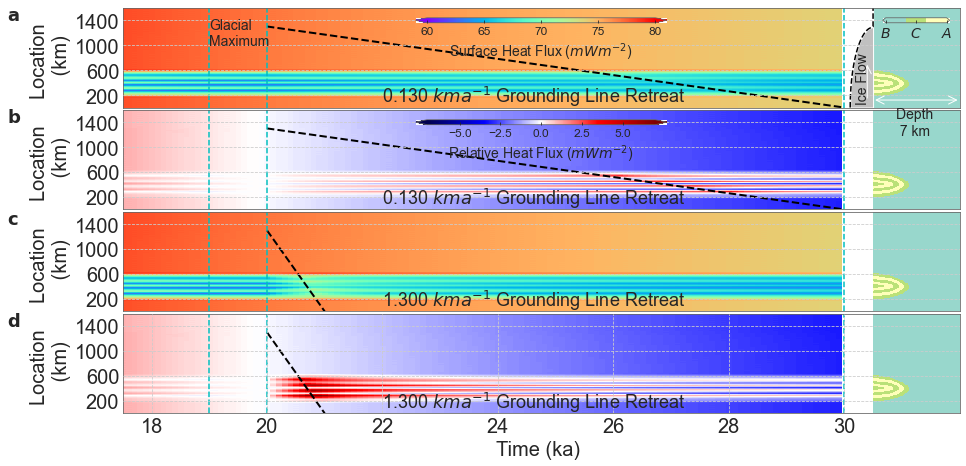
The sedimentary basin thickness also influence the amplitude of basal water flux (Fig. S18). We find an enhancement of basal water flux with a thicker of sedimentary basin. For a 1 km thick sedimentary basin, subglacial water flux is enhanced by 0.4 mm a-1. A 4.5 km thick sedimentary basin could enhance the mean basal water flux up to 3 mm a-1. The sedimentary basin thickness modelled by potential field data could be up to 8 km in Wilkes Subglacial Basin and Aurora Subglacial Basin69,71. These widely distributed thick sedimentary basins show that potential groundwater influence the ice dynamics in East Antarctica111.



**Fig. S18: Impact of sedimentary basin thickness on basal water flux. a-d,** Surface vertical water flux change with sedimentary basin thickness. **a**, crystalline basement. **b**, 1 km sedimentary basin. **c**, 3 km sedimentary basin. **d**, 4.5 km sedimentary basin. **e**, Mean subglacial water flux change with the thickness of sedimentary basin. We see an enhancement of basal water flux with the thickness of sedimentary basin. **f**, The relative mean subglacial water flux compare with the crystalline basement only case.

### Impact of basal heat flux due to water transportation

We further test the impact of groundwater transportation on basal heat flux (Fig. S19). The ice sheet expansion drives basal water recharge into the sedimentary basin. We observe decrease of basal heat flux compare with the crystalline bed covered region. During ice sheet retreat, the surface heat flux is enhanced associate with the water discharge. For a 130 m a-1 retreat case, the surface heat flux is enhanced by 1 mW m-2 compare with the glacial maximum. A faster collapse mode could enhance the basal heat flux by 5 mW m-2.



**Fig. S19: Impact of ice sheet retreat to basal heat flux. a,** Surface heat flux with a 130 m a-1 ice sheet retreat rate. **b**, relative surface heat flux compare with the glacial maximum with a 130 m a-1 ice sheet retreat rate. **c**, Surface heat flux with a 1,300 m a-1 ice sheet retreat rate. **d**, relative surface heat flux compare with the glacial maximum with a 1,300 m a-1 ice sheet retreat rate.

**Supplement reference**

47 Cracknell, M. J. & Reading, A. M. Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. *Comput. Geosci.* **63**, 22-33 (2014).

48 Kuhn, S., Cracknell, M. J. & Reading, A. M. Lithological mapping in the Central African Copper Belt using Random Forests and clustering: Strategies for optimised results. *Ore Geol. Rev.* **112**, 103015 (2019).

49 Cox S.C., B., S. L. & team, T. G. (ed GNS Science) (GNS Science, Lower Hutt, New Zealand., 2019).

50 Jordan, T. A. *et al.* Inland extent of the Weddell Sea Rift imaged by new aerogeophysical data. *Tectonophysics* **585**, 137-160, doi:10.1016/j.tecto.2012.09.010 (2013).

51 Trey, H. *et al.* Transect across the West Antarctic rift system in the Ross Sea, Antarctica. *Tectonophysics* **301**, 61-74, doi:10.1016/s0040-1951(98)00155-3 (1999).

52 Hübscher, C., Jokat, W. & Miller, H. Structure and origin of southern Weddell Sea crust: results and implications. *Geol. Soc. Spec. Pub.* **108**, 201-211 (1996).

53 Straume, E. O. *et al.* GlobSed: Updated Total Sediment Thickness in the World's Oceans. *Geochem. Geophys. Geosyst.* **20**, 1756– 1772, doi:10.1029/2018gc008115 (2019).

54 Anandakrishnan, S., Blankenship, D., Alley, R. & Stoffa, P. Influence of subglacial geology on the position of a West Antarctic ice stream from seismic observations. *Nature* **394**, 62-65 (1998).

55 Peters, L. E. *et al.* Subglacial sediments as a control on the onset and location of two Siple Coast ice streams, West Antarctica. *J. Geophys. Res. Solid Earth* **111** (2006).

56 Filina, I. Y. *et al.* New 3D bathymetry and sediment distribution in Lake Vostok: Implication for pre-glacial origin and numerical modeling of the internal processes within the lake. *Earth Planet. Sci. Lett.* **276**, 106-114, doi:10.1016/j.epsl.2008.09.012 (2008).

57 Chaput, J. *et al.* The crustal thickness of West Antarctica. *J. Geophys. Res. Solid Earth* **119**, 378-395 (2014).

58 Anandakrishnan, S. & Winberry, J. P. Antarctic subglacial sedimentary layer thickness from receiver function analysis. *Global Planet. Change* **42**, 167-176, doi:10.1016/j.gloplacha.2003.10.005 (2004).

59 Ramirez, C. *et al.* Crustal structure of the Transantarctic Mountains, Ellsworth Mountains and Marie Byrd Land, Antarctica: constraints on shear wave velocities, Poisson's ratios and Moho depths. *Geophys. J. Int.* **211**, 1328-1340, doi:10.1093/gji/ggx333 (2017).

60 Piana Agostinetti, N., Roselli, P., Cattaneo, M. & Amato, A. Moho-depth and subglacial sedimentary layer thickness in the Wilkes Basin from receiver function analysis. *IASPEI-General Assembly* (2005).

61 Chai, C., Ammon, C. J., Anandakrishnan, S., Ramirez, C. & Nyblade, A. Estimating subglacial structure using P-wave receiver functions. *Geophys. J. Int.* **209**, 1064-1079, doi:10.1093/gji/ggx075 (2017).

62 Yan, P. *et al.* Antarctic ice sheet thickness estimation using the horizontal-to-vertical spectral ratio method with single-station seismic ambient noise. *The Cryosphere* **12**, 795-810, doi:10.5194/tc-12-795-2018 (2018).

63 Gupta, S., Kanna, N. & Akilan, A. Volcanic passive continental margin beneath Maitri station in central DML, East Antarctica: constraints from crustal shear velocity through receiver function modelling. *Polar Res.* **36**, 1332947 (2017).

64 Dunham, C. *et al.* A joint inversion of receiver function and Rayleigh wave phase velocity dispersion data to estimate crustal structure in West Antarctica. *Geophys. J. Int.* **223**, 1644-1657 (2020).

65 Studinger, M. *et al.* Subglacial sediments: A regional geological template for ice flow in West Antarctica. **28**, 3493-3496, doi:10.1029/2000gl011788 (2001).

66 Bell, R. E., Studinger, M., Karner, G., Finn, C. A. & Blankenship, D. D. in *Antarctica* 117-121 (Springer, 2006).

67 Bamber, J. L. *et al.* East Antarctic ice stream tributary underlain by major sedimentary basin. *Geology* **34**, 33, doi:10.1130/g22160.1 (2006).

68 Aitken, A. R. *et al.* The subglacial geology of Wilkes Land, East Antarctica. *Geophys. Res. Lett.* **41**, 2390-2400, doi:10.1002/2014GL059405 (2014).

69 Aitken, A. R. *et al.* Repeated large-scale retreat and advance of Totten Glacier indicated by inland bed erosion. *Nature* **533**, 385-389, doi:10.1038/nature17447 (2016).

70 Ferraccioli, F., Armadillo, E., Jordan, T., Bozzo, E. & Corr, H. Aeromagnetic exploration over the East Antarctic Ice Sheet: A new view of the Wilkes Subglacial Basin. *Tectonophysics* **478**, 62-77, doi:10.1016/j.tecto.2009.03.013 (2009).

71 Frederick, B. C. *et al.* Distribution of subglacial sediments across the Wilkes Subglacial Basin, East Antarctica. *J. Geophys. Res. Earth Surface* **121**, 790-813, doi:10.1002/2015jf003760 (2016).

72 Paxman, G. J. *et al.* Subglacial Geology and Geomorphology of the Pensacola‐Pole Basin, East Antarctica. *Geochemistry Geophysics Geosystems* **20**, doi:10.1029/2018GC008126 (2019).

73 de Vries, M. v. W., Bingham, R. G. & Hein, A. S. A new volcanic province: an inventory of subglacial volcanoes in West Antarctica. *Geol. Soc. Spec. Pub.* **461**, 231-248 (2018).

74 Bell, R. E. *et al.* Influence of subglacial geology on the onset of a West Antarctic ice stream from aerogeophysical observations. *Nature* **394**, 58, doi:10.1038/27883 (1998).

75 Maritati, A. *et al.* The tectonic development and erosion of the Knox Subglacial Sedimentary Basin, East Antarctica. *Geophys. Res. Lett.* **43**, 10,728-710,737, doi:10.1002/2016gl071063 (2016).

76 Morlighem, M. *et al.* Deep glacial troughs and stabilizing ridges unveiled beneath the margins of the Antarctic ice sheet. *Nat. Geosci.* **13**, 132-137, doi:10.1038/s41561-019-0510-8 (2020).

77 Shepard, M. K. *et al.* The roughness of natural terrain: A planetary and remote sensing perspective. *Journal of Geophysical Research: Planets* **106**, 32777-32795 (2001).

78 Scheinert, M. *et al.* New Antarctic Gravity Anomaly Grid for Enhanced Geodetic and Geophysical Studies in Antarctica. *Geophys. Res. Lett.* **43**, 600-610, doi:10.1002/2015GL067439 (2016).

79 Young, D. A., Richter, T. G., Powell, E. M., Quartini, E. & Blankenship, D. D. (U.S. Antarctic Program (USAP) Data Center, 2017).

80 Tinto, K. J. *et al.* Ross Ice Shelf response to climate driven by the tectonic imprint on seafloor bathymetry. *Nat. Geosci.*, doi:10.1038/s41561-019-0370-2 (2019).

81 Forsberg, R. *et al.* Exploring the Recovery Lakes region and interior Dronning Maud Land, East Antarctica, with airborne gravity, magnetic and radar measurements. *Geol. Soc. Spec. Pub.* **461**, 23-34 (2018).

82 Sandwell, D. T., Muller, R. D., Smith, W. H. F., Garcia, E. & Francis, R. New global marine gravity model from CryoSat-2 and Jason-1 reveals buried tectonic structure. **346**, 65-67, doi:10.1126/science.1258213 (2014).

83 Kvas, A. *et al.* (ed GFZ Data Services) (2019).

84 Uieda, L., Barbosa, V. C. F. & Braitenberg, C. Tesseroids: Forward-modeling gravitational fields in spherical coordinates. **81**, F41-F48, doi:10.1190/geo2015-0204.1 (2016).

85 An, M. *et al.* S-velocity model and inferred Moho topography beneath the Antarctic Plate from Rayleigh waves. *J. Geophys. Res. Solid Earth* **120**, 359-383, doi:10.1002/2014jb011332 (2015).

86 Pappa, F., Ebbing, J. & Ferraccioli, F. Moho Depths of Antarctica: Comparison of Seismic, Gravity, and Isostatic Results. *Geochem. Geophys. Geosyst.*, doi:10.1029/2018gc008111 (2019).

87 Ebbing, J. *et al.* Earth tectonics as seen by GOCE - Enhanced satellite gravity gradient imaging. *Sci Rep* **8**, 16356, doi:10.1038/s41598-018-34733-9 (2018).

88 Arkani-Hamed, J. Differential reduction to the pole: Revisited. *Geophysics* **72**, L13-L20, doi:10.1190/1.2399370 (2006).

89 Melo, F. F., Gonzalez, S. P., Barbosa, V. C. & Oliveira Jr, V. C. Amplitude of the magnetic anomaly vector in the interpretation of total-field anomaly at low magnetic latitudes. *J. Appl. Geophys.* **190**, 104339 (2021).

90 Guo, L., Meng, X. & Zhang, G. Three-dimensional correlation imaging for total amplitude magnetic anomaly and normalized source strength in the presence of strong remanent magnetization. *J. Appl. Geophys.* **111**, 121-128 (2014).

91 Salem, A. *et al.* Mapping the depth to magnetic basement using inversion of pseudogravity: Application to the Bishop model and the Stord Basin, northern North Sea. *Interpretation* **2**, T69-T78, doi:10.1190/int-2013-0105.1 (2014).

92 Martos, Y. M. *et al.* Heat Flux Distribution of Antarctica Unveiled. *Geophys. Res. Lett.* **44**, 11,417-411,426, doi:10.1002/2017gl075609 (2017).

93 Mouginot, J., Rignot, E. & Scheuchl, B. Continent‐wide, interferometric SAR phase, mapping of Antarctic ice velocity. *Geophys. Res. Lett.* **46**, 9710-9718 (2019).

94 Morlighem, M., Seroussi, H., Larour, E. & Rignot, E. Inversion of basal friction in Antarctica using exact and incomplete adjoints of a higher-order model. **118**, 1746-1753, doi:10.1002/jgrf.20125 (2013).

95 Diesing, M. Deep-sea sediments of the global ocean. *Earth System Science Data* **12**, 3367-3381 (2020).

96 Kursa, M. B. & Rudnicki, W. R. Feature selection with the Boruta package. *J Stat Softw* **36**, 1-13 (2010).

97 Reshef, D. N. *et al.* Detecting novel associations in large data sets. *Science* **334**, 1518-1524 (2011).

98 Albanese, D. *et al.* Minerva and minepy: a C engine for the MINE suite and its R, Python and MATLAB wrappers. *Bioinformatics* **29**, 407-408 (2013).

99 Strobl, C., Boulesteix, A.-L., Zeileis, A. & Hothorn, T. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC bioinformatics* **8**, 25 (2007).

100 Valavi, R., Elith, J., Lahoz‐Monfort, J. J. & Guillera‐Arroita, G. block CV: An r package for generating spatially or environmentally separated folds for k‐fold cross‐validation of species distribution models. *Methods in Ecology and Evolution* **10**, 225-232 (2019).

101 Burton-Johnson, A. & Riley, T. Autochthonous v. accreted terrane development of continental margins: a revised in situ tectonic history of the Antarctic Peninsula. *J. Geol. Soc.* **172**, 822-835 (2015).

102 Young, D. A. *et al.* A dynamic early East Antarctic Ice Sheet suggested by ice-covered fjord landscapes. *Nature* **474**, 72-75 (2011).

103 Eagles, G. *et al.* Erosion at extended continental margins: Insights from new aerogeophysical data in eastern Dronning Maud Land. *Gondwana Res.* **63**, 105-116, doi:10.1016/j.gr.2018.05.011 (2018).

104 Zhang, Y. *et al.* Hydromechanical impacts of Pleistocene glaciations on pore fluid pressure evolution, rock failure, and brine migration within sedimentary basins and the crystalline basement. *Water Resour. Res.* **54**, 7577-7602 (2018).

105 Konrad, H. *et al.* Net retreat of Antarctic glacier grounding lines. *Nat. Geosci.* **11**, 258-262 (2018).

106 DeConto, R. M. *et al.* The Paris Climate Agreement and future sea-level rise from Antarctica. *Nature* **593**, 83-89 (2021).

107 Green, D. H. & Wang, H. F. Fluid pressure response to undrained compression in saturated sedimentary rock. *Geophysics* **51**, 948-956 (1986).

108 Lemieux, J. M., Sudicky, E., Peltier, W. & Tarasov, L. Dynamics of groundwater recharge and seepage over the Canadian landscape during the Wisconsinian glaciation. *J. Geophys. Res. Earth Surface* **113** (2008).

109 Steinemann, O. *et al.* Quantifying glacial erosion on a limestone bed and the relevance for landscape development in the Alps. *Earth Surf. Processes Landforms* **45**, 1401-1417 (2020).

110 DeFoor, W. *et al.* Ice sheet–derived submarine groundwater discharge on Greenland's continental shelf. *Water Resour. Res.* **47** (2011).

111 Gooch, B. T., Young, D. A. & Blankenship, D. D. Potential groundwater and heterogeneous heat source contributions to ice sheet dynamics in critical submarine basins of E ast A ntarctica. *Geochem. Geophys. Geosyst.* **17**, 395-409 (2016).

112 Ramirez, C. *et al.* Crustal structure of the Transantarctic Mountains, Ellsworth Mountains and Marie Byrd Land, Antarctica: constraints on shear wave velocities, Poisson's ratios and Moho depths. *Geophys. J. Int.* **211**, 1328-1340 (2017).

**Supplement table**

Table. S1 The sedimentary basin and basement distribution from seismic study.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| FID | Name | Latitude | Longitude | SSB\_thickness | Source | Label |
| 0 | ATOL | -71.3897 | -68.8701 | 0 | Dunham et al., 202064 | Basement |
| 1 | BREN | -72.675 | -63.021 | 0 |  | Basement |
| 2 | BYRD | -80.017 | -119.474 | 0 |  | Basement |
| 3 | FOWL | -76.8928 | -79.3011 | 0 |  | Basement |
| 4 | HOWD | -77.5286 | -86.7694 | 0 |  | Basement |
| 5 | MA01 | -76.9397 | -97.5596 | 0 |  | Basement |
| 6 | MA02 | -77.4394 | -97.5595 | 0.1 |  | SSB |
| 7 | MA03 | -77.9401 | -97.5585 | 0 |  | Basement |
| 8 | MA04 | -78.4221 | -97.5879 | 0.1 |  | SSB |
| 9 | MA05 | -78.9395 | -97.5573 | 0 |  | Basement |
| 10 | MA06 | -79.4397 | -97.5587 | 0.4 |  | SSB |
| 11 | MA07 | -78.2496 | -93.4991 | 0.1 |  | SSB |
| 12 | MA08 | -77.4014 | -103.002 | 0.8 |  | SSB |
| 13 | MA09 | -79.8993 | -104.998 | 0 |  | Basement |
| 14 | MA10 | -78.5999 | -109 | 0 |  | Basement |
| 15 | MECK | -75.2808 | -72.185 | 0 |  | Basement |
| 16 | PIG1 | -73.9782 | -97.575 | 0 |  | Basement |
| 17 | PIG2 | -74.4557 | -97.683 | 0.1 |  | SSB |
| 18 | PIG3 | -75.0841 | -97.4744 | 0 |  | Basement |
| 19 | PIG4 | -75.7599 | -97.583 | 0.2 |  | SSB |
| 20 | ROTH | -67.52 | -68.1488 | 0 |  | Basement |
| 21 | STEW | -84.1863 | -86.2349 | 0 |  | Basement |
| 22 | THUR | -72.5301 | -97.5606 | 0 |  | Basement |
| 23 | UNGL | -79.7746 | -82.524 | 0 |  | Basement |
| 24 | UPTW | -77.5781 | -109.037 | 0.3 |  | SSB |
| 25 | WAIS | -79.4181 | -111.778 | 0 |  | Basement |
| 26 | WELC | -70.7318 | -63.8274 | 0 |  | Basement |
| 27 | WHIT | -82.6823 | -104.387 | 0 |  | Basement |
| 28 | WILS | -80.0396 | -80.5587 | 0 |  | Basement |
| 29 | ST01 | -83.2279 | -98.7419 | 0.2 | Chaput et al.,201457 | SSB |
| 30 | ST02 | -82.069 | -109.124 | 0.27 |  | SSB |
| 31 | ST03 | -81.4065 | -113.15 | 0 |  | Basement |
| 32 | ST04 | -80.715 | -116.578 | 0.35 |  | SSB |
| 33 | ST06 | -79.3316 | -121.82 | 0.23 |  | SSB |
| 34 | ST07 | -78.6387 | -123.795 | 0 |  | Basement |
| 35 | ST08 | -77.9576 | -125.531 | 0.15 |  | SSB |
| 36 | ST10 | -75.8143 | -129.749 | 0.3 |  | SSB |
| 37 | ST12 | -76.897 | -123.816 | 0 |  | Basement |
| 38 | ST14 | -77.8378 | -134.08 | 0 |  | Basement |
| 39 | CLRK | -77.3231 | -141.849 | 0 |  | Basement |
| 40 | WNDY | -82.3695 | -119.413 | 0 |  | Basement |
| 41 | FALL | -85.3066 | -143.628 | 0 |  | Basement |
| 42 | SILY | -77.1332 | -125.966 | 0 |  | Basement |
| 43 | DNTW | -76.4571 | -107.78 | 0 |  | Basement |
| 44 | WHIT | -82.6823 | -104.387 | 0 |  | Basement |
| 45 | MPAT | -78.0297 | -155.022 | 0 |  | Basement |
| 46 | MECK | -75.2807 | -72.1849 | 0 |  | Basement |
| 47 | HOWD | -77.5285 | -86.7694 | 0 |  | Basement |
| 48 | WILS | -80.0396 | -80.5587 | 0 |  | Basement |
| 49 | DUFK | -82.8619 | -53.2007 | 0 |  | Basement |
| 50 | PECA | -85.6124 | -68.5527 | 0 |  | Basement |
| 51 | LONW | -81.3466 | 152.735 | 0 |  | Basement |
| 52 | MILR | -83.3063 | 156.2517 | 0 |  | Basement |
| 53 | SURP | -84.7199 | -171.202 | 0 |  | Basement |
| 54 | DEVL | -81.4757 | 161.9745 | 0 |  | Basement |
| 55 | FISH | -78.9276 | 162.5652 | 0 |  | Basement |
| 56 | WAIS | -79.4181 | -111.778 | 0 |  | Basement |
| 57 | BYRD | -80.0168 | -119.473 | 0 |  | Basement |
| 58 | THUR | -72.5301 | -97.5606 | 0 |  | Basement |
| 59 | UPTW | -77.5797 | -109.04 | 0.3 |  | SSB |
| 60 | ISDE | -80 | -135 | 0 | Anandakrishnan & Winberry, 200458 | Basement |
| 61 | MBL | -78.0994 | -130.229 | 0.26 |  | SSB |
| 62 | MTM | -79.5044 | -100.021 | 0.55 |  | SSB |
| 63 | SDM | -81.6247 | -148.85 | 0.3 |  | SSB |
| 64 | OND | -80.7539 | -125.738 | 0.25 |  | SSB |
| 65 | STC2 | -82.3625 | -136.41 | 0 |  | Basement |
| 66 | STC6 | -82.4514 | -136.428 | 0.18 |  | SSB |
| 67 | N052 | -79.5511 | 145.7592 | 0.3 |  | SSB |
| 68 | JNCT | -76.9367 | 157.9019 | 0 |  | Basement |
| 69 | SPA | -90 | 0 | 0.2 |  | SSB |
| 70 | 1 | -71.6 | 159.5 | 0 | Agostinetti et al., 200560 | Basement |
| 71 | 3 | -71.7 | 156.8 | 0.25 |  | SSB |
| 72 | 4 | -71.8 | 154.6 | 0.2 |  | SSB |
| 73 | 5 | -71.9 | 152.5 | 0 |  | Basement |
| 74 | Onset | -82.278 | -121.555 | 0.4 | Anandakrishnan et al., 199854 | SSB |
| 75 | N060 | -80 | 142.6 | 1 | Yan et al., 201862 | SSB |
| 76 | HOWD | -77.5286 | -86.7693 | 1 | Ramirez et al., 2017112 | SSB |
| 77 | E028 | -76.3075 | 154.0394 | 1 | Chai et al., 201761 | SSB |
| 78 | Maitri | -70.76 | 11.73 | 1.5 | Gupta et al., 201763 | SSB |

Table. S2 Confusion matrix of 10 sub\_RF model. Observed (reference) classes are in columns; predicted classes are in row. Note, 6 SSB points are excluded during the spatial blocks assignment process.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Basement | SSB | Total | Error |
| Basement | 4681 | 1601 | 6282 | 0.255 |
| SSB | 1449 | 4523 | 5972 | 0.243 |
| Total | 6130 | 6124 |  | |
| Error | 0.236 | 0.261 |

Table. S3 Confusion matrix use all training points. Observed (reference) classes are in columns; predicted classes are in row.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Basement | SSB | Total | Error |
| Basement | 4885 | 1713 | 6598 | 0.260 |
| SSB | 1245 | 4417 | 5662 | 0.220 |
| Total | 6130 | 6130 |  | |
| Error | 0.203 | 0.279 |

Table. S4 Model Setting in Hydro-mechanical model

|  |  |
| --- | --- |
| **Property name** | **Value** |
| Young's modulus of lithosphere (*E*) | 56 GPa |
| Poisson ratio (*v*) | 0.25 |
| Elastic thickness of lithosphere (*L*) | 100 km |
| Flexural rigidity of lithosphere | 6.22 × 1024 Pa\*m3 |
| Diffusivity of asthenosphere | 50 km2/a |
| Mantle density (*ρm*) | 3,380 kg/m3 |
| Ice density (*ρice*) | 900 kg/m3 |
| Gravitational constant (*g*) | 9.812 m/s2 |
| Specific Storage (Ss) | 10−6 m−1 |
| Bedrock permeability (kx = kz; m2) | 10−19m2 |
| Skempton constant (B) | 1 |
| Aquifer permeability (kx, kz; m2) | 10−14, 10−15 |
| Confining unit permeability (kx, kz; m2) | 10−16, 10−17 |