Antarctic Topographic Realizations and Geostatistical Modeling Used to Map Subglacial Lakes

E. J. MacKie1, D. M. Schroeder1,2, J. Caers3, M. R. Siegfried1,4, and C. Scheidt5

1Department of Geophysics, Stanford University, Stanford, CA, USA, 2Department of Electrical Engineering, Stanford University, Stanford, CA, USA, 3Department of Geological Sciences, Stanford University, Stanford, CA, USA, 4Department of Geophysics, Colorado School of Mines, Golden, CO, USA, 5Department of Energy Resources Engineering, Stanford University, Stanford, CA, USA

Abstract  Antarctic subglacial lakes can play an important role in ice sheet dynamics, biology, geology, and oceanography, but it is difficult to definitively constrain their character and locations. Subglacial lake locations are related to factors including heat flux, ice surface slope, ice thickness, and bed topography, though these relationships are not fully quantified. Bed topography is particularly important for determining where water flows and accumulates, but digital elevation models of the ice sheet bed rely on interpolation and are unrealistically smooth, biasing estimates of subglacial lake location and surface area. To address this issue, we use geostatistical methods to simulate realistically rough bed topography. We use our simulated topography to predict subglacial lake distribution across the continent using a binomial logistic regression, which uses physical parameters and known lake locations to calculate the probabilities of lake occurrences. Our results suggest that topography models interpolated without appropriate geostatistics overestimate subglacial lake surface area and that total lake surface area is lower than previously predicted. We find that radar-detected lakes are more likely to occur in the interior of East Antarctica, while altimetry-detected (active) lakes are expected to be found in West Antarctica and near the grounding line. We observe that radar-detected lakes have a high correlation with heat flux and ice thickness, while active lakes are associated with higher ice velocity.

1. Introduction

Subglacial lakes are water bodies between the ice and bedrock that are produced through a combination of the insulation and pressure of the ice sheet, frictional heating, and geothermal heat flux (Pattyn, 2010; Robin et al., 1970; Siegert, 2000). They are an integral part of the Antarctic subglacial hydraulic system and are known to influence ice sheet dynamics by interaction with overlying ice, drainage, and the reduction of basal friction (e.g., Kamb, 1987; Siegfried et al., 2016; Stearns et al., 2008). Constraints on water storage are necessary for estimating the magnitude of past outburst events, predicting sea level rise contribution from meltwater discharge, and modeling ice sheet history (Dowdeswell & Siegert, 1999). The character, size, and spatial distribution of subglacial lakes provide constraints for present and former ice sheet dynamics and the processes that drive ice sheet movement (e.g., Bell, 2008). In addition to being important for ice sheet dynamics, subglacial lakes have biological and geological implications: With no sunlight, no exposure to the atmosphere, and cold temperatures, subglacial lakes are an extreme habitat that could contain unique forms of life (e.g., Christner et al., 2006, 2014; Siegert et al., 2001), and these lakes may contain sediment columns that record up to thousands or millions of years of ice sheet history (Smith et al., 2018).

Subglacial lakes were first identified in radio-echo sounding (RES) data in the 1960s (Robin et al., 1970). In radar images (Figure 1a), lakes are characterized by bright, flat reflections at the ice/water interface (e.g., Carter et al., 2007; Oswald & Robin, 1973; Wright & Siegert, 2012). In some cases, the radar record is ambiguous (Matsuoka, 2011) and requires more refined processing techniques for interpreting the state of the basal interface (Schroeder et al., 2015). Less distinct lakes and basal water features have been identified with radar by their specularity (Carter et al., 2007; Schroeder et al., 2014; Young et al., 2016), scattering properties (Jordan et al., 2017), and reflectivity after correcting for radar attenuation (Chu et al., 2016). Since their initial discovery, RES surveys have led to the discovery of hundreds of lakes beneath the Antarctic ice sheet (e.g., Siegfried & Fricker, 2018; Wright & Siegert, 2012), shown in Figure 1b. The largest of these, Vostok Subglacial Lake, has a surface area of 14,000 km² (Siegert et al., 2005).
Efforts have been made to use radar-detected lakes to constrain the total amount of liquid water beneath the ice sheets in Antarctica. Dowdeswell and Siegert (1999) estimated that subglacial water bodies cover 54,000 km² in Antarctica. This study used about 70 lakes found in 400,000 line kilometers of radar data surveyed between 1967 and 1979 to extrapolate a continent-wide estimate by assuming that the amount of water found along the survey lines is proportional to the rest of the continent. This estimate assumed that the distribution of lakes is uniform throughout the continent and that the fraction of the flight lines containing lakes is representative of the rest of Antarctica. The radar data used in the Dowdeswell and Siegert (1999) study were recorded on 35-mm optical film (Drewry et al., 1982). The low image quality and analog nature of this data set make it difficult to observe features, so it is possible that not all the lakes were found.

Livingstone et al. (2013) modeled lake locations by calculating hydraulic potential and estimated that subglacial lakes make up \( \sim 4\% \) of the ice/bed interface. A minimum area of 5 km² of water surface was used to identify lakes. After masking out cold-bedded regions using a thermodynamic ice sheet model, the fractional lake coverage dropped to 2.7%, or a cumulative area of about 350,000 km², an order of magnitude larger than Dowdeswell and Siegert (1999). This suggests that bed temperature plays an important role in modeling both lake surface area and locations. Goeller et al. (2016) also mapped expected lakes using hydraulic potential with a grid resolution of 5 km, yielding 4.9% lake coverage, and then scaled down their modeled cumulative lake surface area by an order of magnitude based on the number of predicted lakes that were not confirmed by radar observations, yielding an estimated coverage of 0.6%, or 77,000 ± 18,000 km².

The Livingstone et al. (2013) and Goeller et al. (2016) models relied on Bedmap2 bed topography (Fretwell et al., 2013), which was generated by combining all available radar measurements of topography and interpolating between measurements. This interpolation created unrealistically smooth bed topography. Bedmap2 contains regions with bed elevation uncertainty of up to 1,000 m (Fretwell et al., 2013), and interpolation artifacts from the smoothing process likely biased continental-scale models of subglacial lake size distribution toward larger lakes (Goeller et al., 2016; Livingstone et al., 2013). This bias is exacerbated by the 5-km grid resolution used in the Goeller et al. (2016) model (more than half of known lakes are shorter than 5 km Dowdeswell & Siegert, 1999).

Logistical and financial limitations on RES surveys used to measure bed topography impede the acquisition of high-resolution bed measurements for the entire continent. Gaps in data coverage are typically filled by kriging interpolation (e.g., Fretwell et al., 2013; Lythe & Vaughan, 2001), which has a smoothing effect. An alternative approach is to use available data to stochastically simulate topography so that unsurveyed areas
have statistically realistic roughness. This method of "conditional simulation" is commonly used in mining applications and reservoir modeling (e.g., Deutsch & Wang, 1996; Journel, 1974) and can generate multiple simulations that reflect the spatial uncertainty in the model. Conditional simulations of bed topography have been implemented at a catchment scale (Goff et al., 2014), but no realistically rough digital elevation models (DEMs) exist at a continental scale for Antarctica.

Recent studies also suggest that subglacial lakes can exhibit dynamic behavior. Observations of vertical changes of the ice surface with interferometric synthetic aperture radar techniques have led to the discovery of subglacial water bodies with volume changes on time scales as short as weeks, referred to as "active" subglacial lakes (Gray et al., 2005) (Figure 1c). ERS-2 satellite radar altimetry and ICESat laser altimetry data improved the spatial and temporal coverage of active lake observations and uncovered over 100 lakes (Fricker et al., 2007; Smith et al., 2009; Wingham et al., 2006). These lakes have been observed to be hydrologically connected, with water cascading through multiple lakes and ultimately emptying across the grounding line (Fricker & Scambos, 2009). Subglacial lake drainage has been linked to changes in ice velocity (Siegfried et al., 2016; Smith et al., 2017; Stearns et al., 2008), so the subglacial water storage and drainage system may play a role in ice sheet models and sea level rise projections. Active lakes have also been identified beneath the rapidly flowing Thwaites Glacier (Smith et al., 2017). However, subglacial lakes identified through surface deformation are not characterized by a flat, bright bed in RES observations (e.g., Siegert et al., 2014; Wright & Siegert, 2012), suggesting that our physical understanding of these features remains incomplete. Mapping of active subglacial lakes is also limited by the relatively brief time scale over which we have continent-wide surface observations (Siegfried & Fricker, 2018).

In this paper, we generate conditional simulations of Antarctic bed topography to map subglacial lakes and investigate differences between active and radar-detected lakes. To overcome limitations from smoothing in Bedmap2, we simulate 100 DEMs with realistic roughness and demonstrate the influence of topographic roughness on subglacial water coverage. Multiple simulations create an ensemble of topographic realizations that reflect spatial uncertainty.

To predict lake locations and thereby further improve our surface area estimate, we implement a binomial logistic regression. This method uses statistical relationships between known subglacial lakes and bed elevation, heat flux (Martos et al., 2017; Maule et al., 2005; Shapiro & Ritzwoller, 2004), ice velocity (Rignot et al., 2017), strain (Alley et al., 2018), the curvature (i.e., second derivative) of hydraulic potential, ice surface slope, ice thickness, flow accumulation models (Schwanghart & Scherler, 2014), and the distance along the ice flow path from the grounding line to calculate the relative probability of finding a subglacial lake at a given location. This process improves the accuracy of lake surface area estimates by incorporating other conditions into determining lake locations rather than only using bed geometry. The regression analysis is done separately for active and radar-detected lakes in order to map their respective locations and suggest regions that may contain undiscovered lakes. The correlation coefficients between lake probability and each physical parameter allow us to determine the relative importance of certain conditions for lake occurrence, providing insights into the complexities of the Antarctic subglacial hydrological environment.

2. Methods

We conducted a geostatistical analysis of subglacial conditions in order to generate statistically realistic maps of bed topography and hydrology and to examine trends and properties of subglacial lakes. First, bed topography was simulated at long and short wavelengths. Longer-wavelength roughness was added to regions of Bedmap2 with bed uncertainty of 1,000 m. This was done through multiple-point direct sampling (Mariethoz et al., 2010), a geostatistical method that fills in gaps in bed topography by using other regions as training data (section 2.1.2). Small-scale roughness was added to the large-scale roughness results by simulating small-wavelength topography through a Fourier analysis of radar bed topography data, similar to the method used in Hu and Tonder (1992) (section 2.1.3). The simulated DEMs were used to calculate ice thickness, water routing models (Schwanghart & Scherler, 2014), bed slope, and curvature of hydraulic potential. Curvature was used to find areas of hydraulic concavity where water would be likely to collect. An eigenvalue decomposition was used to orthogonalize or reduce dependency between the data sets; this is needed to satisfy assumptions in logistic regression and avoid overfitting the model (Mandel, 1982). A binomial logistic regression analysis was used to statistically model subglacial lake occurrence. This regression outputs maps
Figure 2. Flow diagram describing the construction of our simulated bed topography, simulated subglacial lake locations, and evaluation of resulting lake occurrence.

of the probabilities of lake occurrence across Antarctica (section 2.2.4). We applied a threshold to the probability maps in order to classify the continent into lake versus no-lake areas. To better understand which factors influence lake occurrence, the correlation coefficients between the physical predictors and the lake occurrence probability maps were used to study sensitivity. The general workflow, illustrated in Figure 2, is described in detail in the sections that follow.

2.1. DEM Simulation
We created multiple DEM realizations at 1-km resolution by simulating topographic roughness at short and long wavelengths and adding it to Bedmap2. We simulated large-scale topographic roughness through multiple-point direct sampling (Mariethoz et al., 2010) in regions of Bedmap2 with uncertainty greater than or equal to 1,000 m (section 2.1.2). For short spatial wavelengths, we added small-scale (high-frequency) roughness to the entire continent, similar to the approach used in Goff et al. (2014) where short-wavelength roughness is “draped” over a smoother surface. Small-scale roughness was based on the difference between Bedmap2 and radar measurements of bed topography (CReSIS, 2018). The smoothest and roughest extremes of this residual roughness were selected to generate a random surface using methods from Hu and Tonder (1992) (section 2.1.3). Fifty simulations were made with each roughness and added to the direct sampling results, totaling to 100 synthetic DEMs.

2.1.1. Topography Data
We used Bedmap2 topography (Fretwell et al., 2013) as the basis for our topographic simulations. Bedmap2 was created using traditional interpolation methods (e.g., cosine filtering and kriging). Topography was interpolated between radar flight lines with bed elevation picks at 1-km resolution. Eighty-three percent of the Bedmap2 cells are within 20 km of a data point (Fretwell et al., 2013), but some regions are as far as hundreds of kilometers from a data point. The unsurveyed regions were isotropically interpolated, creating unrealistically smooth bed topography. Bed elevation uncertainty can reach up to 1,000 m (Fretwell et al., 2013). Even cells containing radar-derived bed data were smoothed for internal consistency, reducing short-scale roughness.

2.1.2. Multiple-Point Direct Sampling Simulation (Large-Scale Roughness)
Multiple-point statistics (MPS) simulations were used to simulate topography in any region of Bedmap2 where bed uncertainty is 1,000 m or greater, such as the sample region in Figure 3a. MPS methods (e.g., Meerschman et al., 2013; Straubhaar et al., 2011) extract and store patterns and spatial dependencies between points in a training image in order to simulate information elsewhere. The multiple-point direct sampling method, developed by Mariethoz et al. (2010), implements MPS through direct sampling of the training image and simulates the grid cell of interest without having to store statistical information. MPS, in
Figure 3. (a) Sample of Bedmap2 (Fretwell et al., 2013) topography in East Antarctica. (b) Bedmap2 topography uncertainty. (c) Multiple-point direct sampling applied to regions with uncertainty of 1,000 m (with the most obvious differences shown in the dashed oval). (d) Small-scale roughness simulated from MCoRDS data (CReSIS, 2018) was added to the Bedmap2-MPS topography grid, resulting in realistic topography, using the rougher roughness.

particular direct sampling, has been used in reservoir modeling to fill gaps by learning from the data where no gap exists (e.g., Caers et al., 2001; Strebelle, 2002). The DeeSse commercial software was used to implement direct sampling in our simulations. We defined Bedmap2 grid cells with uncertainty of 1,000 m or more as unknown (Figure 3b), and all other areas as part of the training image. We assumed that the surveyed regions of Antarctica are representative of the areas with gaps. Fifty realizations were generated so that many possible perturbations of topography could be used to estimate uncertainty in the subglacial lake maps due to long-wavelength topographic uncertainty. The added long-wavelength topography is shown in Figure 3c. Full methods and implementation are described in Mariethoz et al. (2010), Meerschman et al. (2013), and Straubhaar et al. (2011).

2.1.3. Fourier Analysis of Topography Data (Small-Scale Roughness)

We added short-wavelength roughness to the direct sampling results (Figure 3d) through a Fourier analysis of radar bed topography from the Multichannel Coherent Radar Depth Sounder (MCoRDS) operated by the Center for Remote Sensing of Ice Sheets (CReSIS, 2018). We analyzed over 70 flight lines from 2010 and 2011 to view a wide sample of roughness behavior, shown in Figure 4a. We calculated the difference between MCoRDS topography data and Bedmap2 for the flights. Following the roughness parameterization described in Taylor et al. (2004) and Bingham and Siegert (2009), the residual roughness of each flight line was characterized by its roughness energy, or the integral of the power spectral density. The roughest and smoothest extremes of the residual topography were used as parameters for the simulations in order to encapsulate the roughness extremes in our model. The roughest residual topography was found in the Transantarctic Mountains, and the smoothest came from Thwaites Glacier. To create synthetic roughness, the power spectral density was converted to amplitude, and circular symmetry was imposed to represent amplitude in two dimensions. A random phase shift between 0 and 2π was applied to each element, and then the inverse Fourier transform was computed to produce a 2-D surface (e.g., Hu & Tonder, 1992). We generated 50
Figure 4. (a) MCoRDS radar bed measurements (CReSIS, 2018) used to determine small-scale added roughness. The roughness energy of the difference between the measured bed and Bedmap2 are plotted. (b) MCoRDS radar topography compared to Bedmap2 and our simulated topography. The plotted segments show the bed segments with the roughest and smoothest residual roughness (Bedmap2 topography subtracted from measured topography). The residual roughness from the “smooth” and “rough” segments were used to simulate small-scale roughness, which was added to the long-wavelength simulations to form the “smooth” and “rough” DEMs. The yellow lines in (b) show one realization of simulated topography at the sample locations.

roughness simulations using the smoothest residuals, referred to as “smooth topography” throughout this paper, and 50 with the roughest residuals, referred to as “rough topography” (Figure 4b). This roughness was added to the MPS results, giving a total of 100 DEMs that include statistically realistic roughness. We assessed the quality of simulated topography by comparing the frequency content of MCoRDS bed topography and our simulated topography along flight lines.

2.2. Subglacial Lake Simulation

We employed a logistic regression to estimate subglacial lake locations. This method used statistical relationships between physical predictors and known lake locations to estimate the probability of having a lake at any given location in Antarctica. Our predictors included heat flux, ice velocity, surface slope, bed elevation, ice thickness, bed slope, strain rate, distance along ice flow path from the grounding line, flow accumulation, and hydraulic potential curvature. For each DEM, we calculated geometric parameters (ice thickness, bed slope, flow accumulation, and hydraulic potential curvature) from the simulated bed topography; other parameters (heat flux, ice velocity, surface slope, strain, and distance along flow path) were kept unchanged between simulations. We used this set of parameters in a regression to fit a model that calculates the
Table 1
Simulation Scenarios

| Scenario | Topography         | Lake type        |
|----------|--------------------|------------------|
| 1        | Bedmap2            | Radar-detected   |
| 2        | Bedmap2            | Active           |
| 3        | 50 smooth simulations | Radar detected  |
| 4        | 50 smooth simulations | Active         |
| 5        | 50 rough simulations | Radar detected  |
| 6        | 50 rough simulations | Active         |

Note. A logistic regression analysis was performed for both active and RES lakes with each topographic simulation and Bedmap2.

probability of lake occurrence across Antarctica. We ran simulations with each topography data set and lake type combination (Table 1). We also modeled active and RES lakes with Bedmap2 topography for comparison.

2.2.1. Subglacial Lake Data

Known lake locations of RES-identified lakes were used to train the regression model. The largest database of subglacial lake positions comes from the Wright and Siegert (2012) inventory. This inventory contains both active and radar-identified lakes. We removed the active lakes from the inventory, leaving 246 RES lakes. We also included 105 RES lakes found in the Gamburtsev Mountain Range, identified by Wolovick et al. (2013). We ran separate models with active lakes or lakes that are inferred by rapid changes in surface elevation. One hundred thirty-one lakes came from the Siegfried and Fricker (2018) database, which collated all known active subglacial lakes that have an outline. We included four additional lakes from the Adventure Subglacial Trench (Wingham et al., 2006) and four lakes located at the onset of Recovery Glacier (Bell et al., 2007). The upper Recovery lakes have been detected by radar with mixed success (Langley et al., 2011), but because of their proposed dynamic nature (Langley et al., 2011), we included them in the active lake data set.

Our logistic regression is informed by predictors at known lake locations, but it also requires information at known “non-lake” locations or places where radar surveying has shown that there are no lakes. The non-lake locations for RES lakes were identified using a map of radar coverage used in Bedmap2 (Fretwell et al., 2013). The locations with a known lake, or within a radius of the observed lake length, were removed so that the remainder of the flight lines could be considered “known” non-lakes. We assumed that all (non-active) subglacial lakes beneath radar flight lines were detected by radar and correctly identified. The non-lake locations for active lakes were identified using ICESat coverage (Zwally et al., 2014). Any point surveyed by ICESat that did not contain an active lake was considered a non-lake. Because there is a significantly greater number of non-lake data points than known lake points with which to train the regression model, the estimated probabilities of lake occurrence will be very low and should not be interpreted in an absolute sense (see further discussion in section 4.2).

2.2.2. Other Data Sets

We included geothermal heat flux to provide information on which areas have sufficient geothermal heat for the bed to be at the pressure melting point. We included the Shapiro and Ritzwoller (2004), Maule et al. (2005), and Martos et al. (2017) heat flux data sets. The Maule et al. (2005) and Martos et al. (2017) data sets were determined using magnetic susceptibility, where satellite magnetic data were used to estimate the Curie depth or depths at which rocks reach the Curie temperature and become nonmagnetic. These were used in a thermal model to estimate heat flux at the bed surface (Maule et al., 2005; Martos et al., 2017). The Shapiro and Ritzwoller (2004) heat flux distribution employed a seismic model to extrapolate heat flux measurements to tectonically and structurally similar regions that do not have measurements.

There is not an established connection between ice velocity and subglacial lakes, but we included ice velocity as a parameter so that we can explore a potential relationship. We used interferometric synthetic aperture radar-derived ice velocity data from the National Aeronautics and Space Administration’s Making Earth System Data Records for Use in Research Environments (MEaSUREs) program (Rignot et al., 2017).

We included distance along ice flow path from the grounding line to investigate a possible relationship between lakes and position along flow pathways. This distance was determined by calculating the flowline
at each grid cell and summing over the distance. The flowlines were calculated using the Antarctic Mapping Tools “flowline” function, which uses MEaSUREs ice velocity to compute the streamline from a given location (Greene et al., 2017).

We included strain rate in our logistic regression to explore a possible relationship with lake occurrence, particularly active lakes. We used the along flow, across flow, and shear strain calculated from satellite-derived ice velocity (Alley et al., 2018).

Flat ice surfaces have been used to predict the locations of subglacial lakes (e.g., Ridley et al., 1993). We calculated maximum surface slopes from Bedmap2 surface elevation data (Fretwell et al., 2013), and bed slope from our population of synthetic DEMs.

The synthetic DEMs were used to calculate the curvature of hydraulic potential and water routing models for inclusion in the logistic regression. Hydraulic potential gradient is often used to predict water movement (e.g., Goeller et al., 2016; Livingstone et al., 2013; Pattyn, 2010), but because it is a function of bed slope and surface slope, which are already used in our model, hydraulic potential gradient would be redundant. Instead, we used hydraulic potential curvature to find areas of concavity where water would be likely to collect. Hydraulic potential was evaluated as follows:

$$\phi = \rho_w gh + \rho_i gH,$$  

where $$\rho_w$$ is the density of water ($$1,000 \text{ kg/m}^3$$), $$\rho_i$$ is the density of ice ($$917 \text{ kg/m}^3$$), $$g$$ is gravitational acceleration, $$h$$ is bed elevation, and $$H$$ is ice thickness (Shreve, 1972). We calculated hydraulic potential and then took the second derivative.

We also estimated subglacial water flow paths using a water routing model to show the effect of upstream catchment area on lake location. We used the FLOWobj function from TopoToolbox (Schwanghart & Scherler, 2014) to delineate subglacial water drainage pathways based on subglacial hydropotential gradients and calculated flow accumulation. We created models using both “carve” and “fill” algorithms, which route water in different ways. The “fill” option routes water through the center of flat areas, while the “carve” option finds the deepest points in flat areas and adjusts the flow path to reach the deep spots.

### 2.2.3. Eigenvalue Decomposition

The regression input parameters were not independent of each other so they were decorrelated to avoid over-fitting the regression model. For example, there is presumably some correlation between the heat flux data sets. To avoid attributing too much weight to trends that exist within multiple variables, we conducted an eigenvalue decomposition or orthogonal linear transformation to decorrelate the variables (Joliffe & Morgan, 1992; Mandel, 1982). We normalized each data set, then calculated the covariance matrix, and solved for the eigenvectors, or principal components. This results in 15 principal components that are an orthogonal representation of the data.

### 2.2.4. Binomial Logistic Regression

We used a binomial logistic regression to model a relationship between known lake locations and the principal components (i.e., orthogonalized parameter inputs). A binomial logistic regression model uses a logistic function to relate predictor variables to a binary dependent variable, in this case the location of subglacial lakes. It is commonly used to predict landslide hazards (e.g., Ohlmacher & Davis, 2003). We used a logistic regression to fit coefficients between the parameters and known lakes and then used the coefficients to calculate the log odds of lake occurrence, which were converted to probability:

$$\log \left( \frac{P(\text{lake})}{P(\text{no lake})} \right) = C_1X_1 + C_2X_2 + C_3X_3 + \ldots + C_{15}X_{15}$$  

where the $$X$$ values are the principal components and the $$C$$ values are the coefficients.

### 2.3. Lake Statistics and Surface Area Calculation

#### 2.3.1. Expected Area and Size Distribution

The calculated lake location probabilities from our logistic regression are relative, not absolute, so the expected cumulative lake surface area cannot be calculated by taking the sum of the probabilities of the grid cells multiplied by their areas. Instead, we set a threshold or cutoff point where all probabilities above a certain number were considered a lake, and all of the grid cells with lower probabilities were considered a non-lake. We used data from the Antarctic Gamburtsev Province Project (AGAP) (Bell et al., 2011; Ferraccioli et al., 2011) over the Gamburtsev Mountains to constrain lake coverage for RES lakes. We used
Figure 5. We used the densely surveyed Gamburtsev mountain range to set a cutoff threshold for determining lakes. (a) The number of known lakes and known non-lake grid cells are calculated at each test threshold to find the optimal area of coverage. (b) Lake probability map generated by one of the rough bed topographies. (c) Example of AGAP (Bell et al., 2011; Ferraccioli et al., 2011) RES survey lines and expected lake coverage after setting an optimal threshold.

AGAP because it is densely surveyed and there are a large number of lakes with which to tune this threshold. We assumed that all lakes in the area were detected. For each map, we analyzed the percent of known lakes and non-lakes generated as a function of the percent of the area covered by lakes. The intersection between these curves is the optimal threshold. We chose our search radius such that the cumulative length of lakes calculated along radar flight lines was within an order of magnitude of known lake lengths in this area. In the Gamburtsev Mountains, there are ~80 km of observed lakes along the flight lines, so we chose our radius so that our optimal threshold generates lakes that intersect roughly 80 km of flight lines (Figure 5).

This was used to pick a cutoff point that was applied to the rest of the continent to develop a binary map of lakes (section 3.3.1), which was used to calculate total surface area. We made histograms of lake length to show the distribution of lake sizes (section 3.3.2). Lake lengths were determined by finding the the area of each expected lake in the binary map, assuming it had a circular shape and calculating the diameter.

2.3.2. Correlation Coefficients
The correlation coefficients between each physical parameter and the lake probability maps shows the sensitivity of lake occurrence to each variable. For each simulation, we calculated a correlation coefficient for each physical parameter. This result suggests which variables are most important for predicting subglacial lake occurrence.

2.3.3. Model Accuracy Evaluation
To demonstrate the effectiveness of the lake location predictions, we computed the percent of known lakes that were matched by the simulations at a given probability cutoff threshold. We evaluated lake accuracy at different cutoff thresholds. Matches were determined using both a 5-km search radius, as used in Goeller et al. (2016), and a 20-km search radius. The 20-km search radius was used to account for topographic variability in the simulations that may have shifted lake positions. The results were averaged across all the smooth simulations and all the rough simulations.

3. Results
3.1. Topographic Models
We generated 100 synthetic DEMs, the first Antarctic DEMs with statistically realistic topography. Fifty were simulated from smoother radar bed picks (Figures 6a–6c), and 50 were simulated from rougher observed bed topography (Figures 6da and 6e). The standard deviation of bed elevation between our set of simulations, shown in Figure 6g and 6h, ranges from 100 to 700 m.

The power spectral density plots in Figure 7 show the wavelength dependent power of the Center for Remote Sensing of Ice Sheets radar bed measurements (Figure 3), Bedmap2, and the simulated topography. The mean of the power spectrum for each smooth (Figure 7a) and rough (Figure 7b) simulation is plotted. The power spectral densities for the simulated DEMs are more similar to the radar topography than Bedmap2 at all wavelengths. For the smooth topography, the spectral behavior of the simulated
Figure 6. Example synthetic DEMs produced by our geostatistical simulations. (a) Continent-wide example of a smooth bed realizations. Example simulated DEMs from a smaller region in East Antarctica using our (b, c) smooth realizations, (d, e) rough realizations, and (f) Bedmap2. (g) Standard deviation of elevation for smooth realizations and (h) rough realizations.
Figure 7. Power spectral density comparisons of simulations with Bedmap2 topography (blue), simulated data (yellow), and the RES bed picks (red) for the segment shown in Figure 4b. The simulated curves show the averaged power spectral density of all the (a) smooth simulations and (b) rough simulations.

Topography resembles that of the radar bed measurements but has a slightly lower power at 1-km wavelengths. The rough topographic simulations closely match the spectral behavior of the bed measurements but have a higher power at subkilometer wavelengths. The rough topography has a higher corner wavelength than the smooth topography.

3.2. Subglacial Lake Simulations
We generated a subglacial lake simulation for each of the scenarios in Table 1: 1 with Bedmap2 and RES lakes, 1 with Bedmap2 and active lakes, 50 with smooth simulated topography and RES lakes, 50 with smooth simulated topography and active lakes, 50 with rough simulated topography and RES lakes, and 50 with rough simulated topography and active lakes. Here we describe the results of the RES lake and active lake simulations:

3.2.1. RES Lake Simulations
The probability maps for RES lakes (Figure 8) showed relatively high probability ($P > 0.003$) areas clustered in the interior of East Antarctica, especially at Dome A, Dome C, and Dome F. Vostok Subglacial Lake emerged with high probability. Dome F contained a large probable lake that persisted across different roughnesses and simulations. There were no probable lakes in West Antarctica or near the coast. The regional trends remained the same across simulations with different roughness, though the small-scale behavior varied (Figures 8f–8h). Known lakes in the interior of East Antarctica aligned well with probable lakes, but known lakes in West Antarctica or the South Pole region did not have high probabilities.

3.2.2. Active Lake Simulations
Relatively high probabilities of active subglacial lake occurrence ($P > 0.0004$) were widespread in West Antarctica and some sectors of East Antarctica (Figure 9). High probabilities occurred near known active subglacial lakes at Rutford, Institute, Foundation, and Recovery ice streams. Evans Glacier had high probabilities though no active lakes have been detected there. Pine Island and Thwaites glaciers showed high probabilities of active lake occurrence, as did Byrd Glacier. The regional trends were consistent across different roughnesses.

3.3. The Effect of Roughness on Surface Area and Lake Size for RES Lakes

3.3.1. Surface Area
The probability threshold values for determining what constitutes a lake were 0.0029, $\sim 0.0035$, and $\sim 0.0033$ for the Bedmap2, smooth, and rough simulations, respectively. These were used to generate maps of radar lake locations (Figure 10). The subglacial lakes are clustered in the center of East Antarctica around Vostok, Dome C, and Dome A and extend upward toward Dome F. No lakes were generated near the coast or in West Antarctica.

The added roughness reduced the expected total surface area of subglacial lakes. Using the Bedmap2 topography 0.60% or 77,000 km$^2$ of the continent had a subglacial lake (Figure 10d). For the smooth topographic simulations 0.41 ± 0.12% of subglacial Antarctica had a lake or about 52,000 ± 15,000 km$^2$ (Figure 10e). And for the rough simulations, subglacial lake surface area fell to 0.32 ± 0.18% or 41,000 ± 23,000 km$^2$ (Figure 10f).
3.3.2. Size Distribution

The Bedmap2 lake model generates 6,078 lakes with a mean size of 12.6 km². The smooth simulations generate more than double as many lakes and have a mean area of 3.3 km². The rough simulations have a mean lake total area about half the size of that in Bedmap2 but have a similar number of lakes. About 30% of known lakes have an observed length of 1 km or less (Figure 11a), while in each simulation, roughly 80% of the lakes have a length of 1 km (Figures 11b–11d). The smooth, rough, and Bedmap2 simulations have similar lake length distributions.

Figure 8. (a) Probability map of radar-detected lakes. (b) Dome F contains an abundance of probable lakes and possibly a large lake. Dome A topography for (c) Bedmap2, (d) a smooth simulation, and (e) a rough simulation, along with their corresponding lake simulations (f–h).
Figure 9. (a) Continent-wide active lake probability generated using a smooth topographic simulation, with (b) Recovery Glacier and (c) West Antarctic ice streams and outlet glaciers.

3.4. Correlation With Physical Parameters
The correlation coefficients between the probability maps and the input parameters, shown in Figure 12, reflect the importance of each variable for predicting lake occurrence. Both types of lake have correlation coefficients of about 0.2 or greater for ice thickness and heat flux, though ice thickness is more strongly correlated with RES lakes. Distance along flow path from the grounding line is more strongly correlated with RES lakes than active lakes. Ice velocity has a correlation coefficient of \( \sim 0.35 \) with active lakes but is independent of radar lakes. The water routing models are independent of RES lakes but have a positive correlation with active lakes. The “fill” water routing algorithm has a stronger correlation than the “carving” algorithm. The three heat flux data sets correlate similarly well with both active and radar lakes, though the Martos et al. (2017) data set performs slightly better for both lake types. Both varieties of lake are largely independent of strain rate, bed slope, surface slope, and hydraulic curvature.

3.5. Lake Simulation Accuracy
We assessed the accuracy of our lake location simulations by calculating the percentage of known lakes that were identified as a lake in our simulation for a range of probability thresholds (Figure 13). For RES lake simulations, the top 10% highest probability cells account for about 70% of known lakes. For active lakes, about 90% of known lakes were captured in the top 10% of the simulations with rough topography and about 70% for the smooth topography. Simulated active lake location accuracy increased with roughness.
Figure 10. Expected coverage of radar-detected lakes in Antarctica using (a) Bedmap2 topography, (b) smooth topography, and (c) rough topography. Sample area in East Antarctica shows the detailed difference between subglacial lake locations for (d) Bedmap2, (e) smooth, and (f) rough topography. Uncertainty range shown for surface area and lake number are one standard deviation.

4. Discussion

4.1. Bed Topography

To the best of our knowledge, our topographic simulations are the first full-continent DEMs of Antarctica with simulated roughness and are the first glaciological application of multiple-point direct sampling. The direct sampling and Fourier analysis allow us to overcome the spatial limitations of RES surveys. It also helps ensure that inferred regional variations in subglacial hydrology are not impacted by kriging artifacts nor variations in the density of bed elevation measurements. While radar coverage across Antarctica has expanded in the past decades and mass conservation techniques can be used to provide topographic constraints (e.g., Morlighem et al., 2011), geostatistical simulation is the only way to realistically represent bed roughness. Furthermore, mass conservation techniques rely on ice surface velocity and make physical assumptions about the bed (e.g., Morlighem et al., 2011). Our DEMs were generated independently of surface observations, making them an unbiased parameter.

In this study, the synthetic topography enabled the creation of realistic lake size distributions, but it has the potential to improve other models as well. Topographic roughness has implications for water storage (e.g., Cael et al., 2017) and should be taken into account when estimating subglacial water storage and the magnitude of drainage events. Topographic simulations could be used to route subglacial water, constrain ice volume estimates, or allow us to accurately model sliding behavior. The size of topographic features in part controls basal sliding and deformation behavior (e.g., Weertman, 1964), so realistically representing
Figure 11. RES lake length histograms for (a) known lakes, (b) expected lakes from simulation using Bedmap2, (c) expected lakes from simulations using smooth topography, and (d) expected lakes from simulations using rough topography. Lake lengths are shown on a logarithmic scale.

roughness could improve ice sheet models. The ensemble of multiple topographic realizations provides a basis for uncertainty quantification in ice sheet modeling.

4.2. Subglacial Lake Probabilities and Surface Area

Based on the characteristics of known lake and non-lake locations, we developed a probabilistic metric for subglacial lake location and integrated surface area. Because significantly more non-lake data points than known lake locations were used in the regression analysis, even the highest lake probability values were on the order of $10^{-3}$. The active lake regression analysis uses more non-lake data than in the RES lake analysis, so probabilities were about an order of magnitude lower than radar-detected lake probabilities. Therefore, the probabilities of active subglacial lake occurrence should be interpreted as relative, not absolute. The probabilities within each map are internally consistent but cannot be compared to probabilities for a different lake or topography type. Therefore, these are not the true probabilities of lake occurrence.

We calculated surface area estimates for RES lakes by using AGAP data to set a probability threshold for determining which probabilities constitute a lake (Table 2) but were unable to do so for active lakes. This is because we did not have reliable non-lake data set for active lakes with which to find an optimal surface area. Our surface area estimates show that cumulative RES lake surface area decreases with increasing bed roughness and that calculations made using Bedmap2 topography can overestimate surface area by as much as a factor of 4. Our cumulative subglacial lake estimate was between 0.14% and 0.53%, or roughly 18,000 to 68,000 km$^2$. This is comparable to the Dowdeswell and Siegert (1999) and Goeller et al. (2016) estimates, which were 54,000 km$^2$ and $77,000 \pm 18,000$ km$^2$, respectively, though neither study generated a map of lake locations. The Dowdeswell and Siegert (1999) and Goeller et al. (2016) estimates are likely within range of
Figure 12. Correlation coefficients between the input parameters and lake probability maps. Results shown for smooth topography realizations.

Figure 13. The percent of known subglacial lakes that are matched by the simulations as a function of increasing lake surface area for (a) RES-detected and (b) active lakes. From left to right, the probability cutoff threshold is decreased so that a greater percent of the continent is covered by lake. At each point, the percentage of known lake matches is evaluated. Lake matches are computed for both a 5- and 20-km search radius. For the smooth and rough results, the mean recovery is shown across all simulations. For example, approximately 80% of known RES lakes are in the 20% highest probability pixels.

Ours because they use the ratio of lakes intercepted by airborne radar flight lines as a constraint on surface area. Our estimate is an order of magnitude smaller than Livingstone et al. (2013), which estimated 2.7%, or 350,000 km². This difference is likely because we used rough topography and set a probability cutoff point (section 2.3.1) for determining surface area.

Our expected radar-detected lake number and size distribution differs from previous studies and known lake length measurements. Only a third of known lakes have an observed length of 1 km or less, while our simulations show that roughly 80% of lakes are expected to be this size. This difference is likely explained by...
Table 2
Cumulative Subglacial Lake Surface Area and Lake Number Across Different Studies

| Study                                      | Surface area     | Expected number of lakes |
|--------------------------------------------|------------------|--------------------------|
| Dowdeswell and Siegert (1999)              | 54,000 km²       |                          |
| Livingstone et al. (2013)                  | 2.7%, 350,000 km²| 9,360                    |
| Goeller et al. (2016)                      | 0.6% ± 18,000 km²| 921 ± 300                |
| Scenario 1 (Bedmap2 topography)            | 0.6%, 77,000 km² | 6,078                    |
| Scenario 3 (smooth topography)             | 0.41 ± 0.12%     | 15,119 ± 9,637           |
| Scenario 5 (rough topography)              | 0.32 ± 0.18%, 23,000 km² | 5,034 ± 2,500 |

The sampling bias in known lakes as smaller lakes are harder to detect. Livingstone et al. (2013) only counted lakes larger than 5 km², so smaller lakes are not included. Goeller et al. (2016) used a grid resolution of 5 km, although more than 90% of our expected lakes and 70% of the known lake lengths used in our study are smaller than 5 km. Each of our expected lake maps had a median lake area of 1 km², but the mean sizes varied with topographic roughness. Our lake simulation results using Bedmap2 yielded lakes with a mean surface area of 12.6 km², while our smooth topography simulation had a mean lake size of 3.3 km². Although the Bedmap2 topography generated larger lakes and a greater cumulative surface area, the smooth topography had a greater number of lakes. The smooth topographic simulations yielded more than twice as many expected lakes as the Bedmap2 topography and more than 15 times as many lakes as the Goeller et al. (2016) model.

The rough topography yields fewer, larger lakes than the smooth topography. This may be due to the difference in corner wavelengths (Figure 7). Because the rough topography has a higher corner wavelength than the smooth topography, it will have a larger dominant wavelength. Therefore, the rough topography has both higher relief and wider basins, which produces larger lakes.

These results suggest that most undiscovered lakes would have to have an area less than 5 km² in order to match our expected size distributions and number of lakes. Our simulations can not generate lakes with an area smaller than 1 km², though there are likely many. This would not substantially influence cumulative lake surface area, but it could affect the distribution of lake sizes.

4.3. Lake Simulations and Locations

Our lake location probability maps confirmed the locations of many known lakes and also showed potential locations of undiscovered lakes. For RES lakes, Dome A, Dome F, and Princess Elizabeth Land have promising regions that lack dense radar surveys. For active lakes, Evans Glacier, Pine Island Glacier, Bailey Ice Stream, and the lower portion of Byrd Glacier all had high probabilities, though they have no known active lakes. While satellite altimetry does not have the spatial constraints of ice-penetrating radar and can theoretically image the entire Antarctic surface, active lakes may remain undetected because the ice surface is too rough, the lake was too small relative to the ice thickness, or the lake was not active during the period of observation. Our probability maps can assist in locating active lakes where altimetry methods might not be suited for detection.

The high correlation coefficients with heat flux suggest that heat flux plays an important role in modulating lake locations. This was also found to be the case in the Livingstone et al. (2013) study, which used the Pattyn (2010) model of warm and cold-bedded regions. However, known lake locations were used in the Pattyn study, which introduces circular logic in the Livingstone et al. (2013) model. Our method of incorporating heat flux data through regression allowed us to account for the basal thermal regime without bias.

Not all known lake locations were matched by our model. Some of this error is likely due to topographic uncertainty. Because the topography, ice thickness, and subglacial water flow paths change with each realization, lake locations will vary. Some known RES lakes were located in very low probability areas, such as West Antarctica and Victoria Land (Figure 8a). This is why the accuracy curve in Figure 13a did not converge to 100% until almost the entire continent was made to be a lake. While some of the known RES lakes had low RES lake probabilities, they coincided with high probability active lake regions, such as at Mercer Ice Stream. This suggests that these lakes may have more in common with active lakes.
that some of the known lakes in these locations are false positives. Lake identification typically relies heavi-
ly on human interpretation and rarely undergoes a stringent verification and classification process (Carter
et al., 2007).

The active lake simulations with rough topography have a higher accuracy than those with smooth topog-
raphy (Figure 13b). This is likely an effect of the water routing model, which correlates strongly with active
lake location. The ice surface slope has 10 times the effect on hydraulic potential that bed slope does (Shreve,
1972), so a steep bed slope is needed to create an area with high flow accumulation. Therefore, the steep bed
slopes resulting from rougher topography are more conducive to creating areas with high flow accumulation
and local storage.

The lake maps shed light on the complex network of subglacial hydrology and are helpful for a variety of
purposes. They could be used to inform future field campaigns to study lakes, classify hydrological regions
of Antarctica, estimate subglacial water storage, or predict frozen and thawed regions. They could also be
used to parameterize sliding behavior in ice sheet models, constrain the magnitude of outburst events, or
investigate drainage networks. Some of the highest probability locations for active lakes in Antarctica were
Pine Island and Thwaites glaciers, which have experienced significant mass loss (Rignot et al., 2019). These
maps could provide insights on the processes that govern the movement and retreat of these glaciers.

4.4. Active Versus Radar-Detected Lakes

The maps and correlation coefficients showed similarities and differences between active and radar lakes.
The advantage of this approach is that we can use all the information available in our model without an a
priori assumption about importance. Both active and RES lake probabilities correlated well with ice thick-
ness and heat flux. The higher importance of ice thickness for RES lakes suggests that they form through
pressure melting in regions with thick ice and remain in that location. Radar lakes have a higher correlation
with distance along flow path from the grounding line than active lakes, indicating that radar lakes tend to
exist further upstream. This is also confirmed by the probability maps, which show radar lakes clustered in
the center of East Antarctica and active lakes grouped closer to the coast. Active lakes were correlated with
the water routing models while RES lakes were not. This suggests that active lakes are governed by water
routing, while RES lake locations depend on ice thickness. The strong positive correlation of active lakes
with ice velocity also suggests that active lake meltwater is generated through frictional heating or that these
lakes cause increased ice velocity.

Our results suggest that active and RES lakes are fundamentally different types of hydrological features.
RES lakes are located near the divides with small hydraulic catchments, while active lakes are closer to the
grounding line and accumulate water from large upstream catchments and frictional melting. Active lakes
were predicted beneath ice streams, while RES lakes were not. Active lakes appeared in West Antarctica,
while RES lakes were predominantly in East Antarctica, suggesting that West Antarctica and East Antarctica
have extremely different hydrological environments.

The active lake locations and correlation coefficients also help explain why active lakes generally cannot be
detected with ice-penetrating radar. Active lakes were linked with high ice velocity, ice streams, and water
flowing in from upstream. This suggests that active lakes may be too dynamic to form the flat ice/water
interface required for a clean radar reflection (e.g., Siegert et al., 2014).

Certain bed-dependent variables did not correlate well with either lake type because they change dramati-
cally with each topographic simulation. Therefore, bed slope and hydraulic potential curvature were nearly
independent of lake probability, though in reality they may not be. The different topography with each sim-
ulation introduced noise into these parameters that overwhelmed the original signal. The importance of
these variables was likely underestimated by our modeling study.

4.5. Assumptions and Uncertainty

While our DEMs offer realistic representations of bed topography, they are simplified, particularly in that
simulated roughness does not vary with region or direction. For example, the same roughness statistics
are used to simulate small-scale topography on the ocean floor and the Transantarctic Mountains, but the
seafloor is likely much smoother due to sedimentation. Bed topography is not constrained by ice veloc-
ity or mass conservation. The simulations do not recreate morphological features or drainage pathways.
The power spectral density plots (Figure 7) show that the simulated topography behaves more similarly
in the frequency domain to radar topography than Bedmap2 does, but there is still some difference between
the simulated and actual topography. This difference could be attributed to artifacts that are generated by adding simulated small-scale roughness to long-wavelength topography. Depending on the constructive or destructive interference of the small-wavelength topography with long-wavelength topography, the resulting topography could have somewhat different spectral content from the measured topography. Future work on topographic simulation methods is needed to account for regional variations in roughness, possibly through using local topographic data to inform the roughness characteristics of nearby topography. Bed topography has also been linked to ice surface topography (e.g., Bindschadler & Choi, 2007); this relationship could be used to constrain topographic predictions.

Although the ice-bed interface of Antarctica exceeds $12 \times 10^6$ km$^2$, only 351 RES lakes and 139 active lakes are used in the model. This is an exceedingly small amount of data with which to train a continent on through regression analysis. We are unable to use available lake length data because we do not know if the given coordinate for each lake is the center point. We also assume that all radar flight lines without lakes are non-lakes, though dim or smaller lakes could easily go undetected by radar or are not identified when processing radargrams. Future hydrological models would benefit from a definitive non-lake database. It would also be useful to have a range of coordinates for each observed lake rather than one coordinate and a length. Similarly, our use of ICESat coverage for determining non-lake locations does not completely eliminate observational bias. Active lake detection relies on repeat measurements before and after drainage, so the lack of a detected active lake does not necessarily mean that one is not there. We assume that this does not substantially alter our predictions.

The cutoff thresholds for determining which probabilities qualify as a lake were set using the AGAP radar data in the Gamburtsev Mountains (Bell et al., 2011; Ferraccioli et al., 2011). This data set was chosen for its dense coverage and large quantity of lakes, but it covers only a small area. The AGAP surveys are the best training data available, but it is by no means representative of the entire continent. While our model places lakes in the locations with the highest probabilities and does not put any lakes in low probability areas, a small proportion of the low probability regions likely will still have lakes in reality. These maps should be used to complement, not replace other studies.

4.6. The Use of Geostatistics

This project synthesized the available physical data of Antarctica and simulated bed topography to calculate subglacial lake surface area in a way that accounted for uncertainty in bed roughness and physical drivers. The use of geostatistics in this model had several advantages: (1) We directly incorporated the roughness of topographic data in bed topography simulations, (2) multiple realizations of topography and lake simulations were created in order to encapsulate a range of possible bed conditions, (3) multiple data sets were integrated to predict lake occurrence while minimizing physical assumptions, (4) it showed the relationships between physical predictors and lake occurrence, and (5) by applying the same model to active and RES lakes, we could objectively investigate how they differ. This allowed us to mitigate typical issues associated with data limitations and physical uncertainty that frequently hamper physical models.

5. Conclusion

We have developed a protocol for simulating subglacial topography and generated 100 new DEMs with statistically realistic roughness. These DEMs enabled the development of 200 probability maps of radar and active lake probabilities and 100 binary maps of radar lakes. All data products are available online (http://www.usap-dc.org/view/dataset/601213). Radar lakes are expected to cover 0.32–0.41% of Antarctica with 5,000 to 15,000 lakes. Our statistical simulation of lake occurrence results in a lower total surface area and higher number of lakes than past estimates and demonstrates that lake surface area and number depends on bed roughness. Therefore, topographic roughness exerts a control on the hydrological conditions of Antarctica. Our simulations are a realistically rough alternative to DEMs made by kriging and can be used as a boundary condition in other models. This has implications for basal sliding, water routing, and sea level rise contributions from meltwater.

Our results suggest that active and RES lakes are distinct hydrological features that occur in different locations and are linked with different physical conditions. Active lakes are more likely to be found beneath West Antarctic ice streams and outlet glaciers, including a very high likelihood at Thwaites Glacier and Pine Island Glacier, while radar lakes occur in interior East Antarctica near the divides. No active lakes have...
been detected at Evans Glacier or the lower portion of Byrd Glacier, but our results suggest that these locations have a high likelihood of having active lakes. Ice thickness, heat flux, and distance inland along flow path from the grounding line were the dominant predictors for RES lakes. Active lakes have a strong positive correlation with ice velocity, heat flux, and flow accumulation models. This suggests that RES lakes are stable feature that are far upstream of the grounding line and accumulate water through pressure melting, while active lakes are further downstream, collect water from upstream sources, and are part of a dynamic hydraulic system closer to the grounding line.

Limitations in data coverage are a hurdle to characterizing the properties and systems of subglacial environments. Our use of geostatistical methods allowed us to realistically represent subglacial conditions and investigate hydrological complexities. The protocols established in this paper could enable further investigations of subglacial conditions.

Acknowledgments
This research was supported by NSF Award 1745137. M. R. S. was supported in part by the George Thompson Postdoctoral Fellowship and NASA Grant NNX17AB03G. We acknowledge the use of data and/or data products from CReSIS generated with support from the University of Kansas, NASA Operation IceBridge Grant NNX16AH54G, NSF Grant ACI-1443054, Lilly Endowment Incorporated, and Indiana METACy Initiative. We thank Karen Alley for providing us with strain rate data. We appreciate the constructive reviews by John Goff and two anonymous reviewers. We would also like to thank the United States Antarctic Program for hosting our data in their repository. See http://www.usap-dc.org/view/dataset/601213 for data products generated by this study.

References
Alley, K. E., Scambos, T. A., Anderson, R. S., Rajaram, H., Pope, A., & Haran, T. M. (2018). Continent-wide estimates of Antarctic strain rates from Landsat 8-derived velocity grids. Journal of Glaciology, 64(244), 321–332.
Bell, R. E. (2008). The role of subglacial water in ice-sheet mass balance. Nature Geoscience, 1(5), 297.
Bell, R. E., Ferraccioli, F., Creyts, T. T., Braaten, D., Coral, N., Das, I., et al. (2011). Widespread persistent thickening of the East Antarctic Ice Sheet by freezing from the base. Science, 331(6024), 1592–1595.
Bell, R. E., Studdinger, M., Shuman, C. A., Fahnestock, M. A., & Joughin, I. (2007). Large subglacial lakes in East Antarctica at the onset of fast-flowing ice streams. Nature, 448(7130), 904.
Bindschadler, R., & Choi, H. (2007). Increased water storage at ice-stream onsets: A critical mechanism?. Journal of Glaciology, 53(181), 163–171.
Bingham, R. G., & Siegert, M. J. (2009). Quantifying subglacial bed roughness in Antarctica: Implications for ice-sheet dynamics and history. Quaternary Science Reviews, 28(3-4), 223–236.
CReSIS (2018). McORDS data, Lawrence, Kansas, USA. Digital Media. http://data.crest.isu.edu/
Cael, B. B., Heathcote, A. J., & Seekell, D. A. (2017). The volume and mean depth of Earth’s lakes. Geophysical Research Letters, 44, 209–218. https://doi.org/10.1002/2016GL071378
Caers, J., Avseth, P., & Mukerji, T. (2001). Geostatistical integration of rock physics, seismic amplitudes, and geologic models in North Sea turbidite systems. The Leading Edge, 20(3), 308–312.
Carter, S. P., Blankenship, D. D., Peters, M. E., Young, D. A., Holt, J. W., & Morse, D. L. (2007). Radar-based subglacial lake classification in Antarctica. Geochemistry, Geophysics, Geosystems, 8, Q03016. https://doi.org/10.1029/2006GC001408
Christner, B. C., Priscu, J. C., Achberger, A. M., Barbante, C., Carter, S. P., Christianson, K., et al. the WISSARD Science Team (2014). A microbial ecosystem beneath the West Antarctic ice sheet. Nature, 512(7514), 310–313.
Christner, B. C., Royston-Bishop, G., Foreman, C. M., Arnold, B. R., Tranter, M., Welch, K. A., et al. (2006). Limnological conditions in subglacial Lake Vostok, Antarctica. Limnology and Oceanography, 51(6), 2485–2501.
Chu, W., Schroeder, D. M., Seroussi, H., Creyts, T. T., Palmer, S. J., & Bell, R. E. (2016). Extensive winter subglacial water storage beneath the Greenland Ice Sheet. Geophysical Research Letters, 43, 12,484–12,492. https://doi.org/10.1002/2016GL071538
Deutsch, C. V., & Wang, L. (1996). Hierarchical object-based stochastic modeling of fluvial reservoirs. Mathematical Geology, 28(7), 857–880.
Dowdeswell, J. A., & Siegert, M. J. (1999). The dimensions and topographic setting of Antarctic subglacial lakes and implications for large-scale water storage beneath continental ice sheets. Geological Society of America Bulletin, 111(2), 254–263.
Drewry, D. J., Jordan, S. R., & Jankowski, E. (1982). Measured properties of the Antarctic ice sheet: Surface configuration, ice thickness, volume and bedrock characteristics. Annals of Glaciology, 3, 83–91.
Ferraccioli, F., Finn, C. A., Jordan, T. A., Bell, R. E., Anderson, L. M., & Damaske, D. (2011). East Antarctic rifting triggers uplift of the Gamburtsev Mountains. Nature, 479(7373), 388.
Fretwell, P., Pritchard, H. D., Vaughan, D. G., Bamber, J. L., Barrand, N. E., Bell, R., et al. (2013). Bedmap2: Improved ice bed, surface and thickness datasets for Antarctica. The Cryosphere, 7(1), 375–393.
Fricker, H. A., & Scambos, T. (2009). Connected subglacial lake activity on lower Mercer and Whillans ice streams, West Antarctica, 2003-2008. Journal of Glaciology, 55(190), 303–315.
Fricker, H. A., Scambos, T., Bindschadler, R., & Padman, L. (2007). An active subglacial water system in west Antarctica mapped from space. Science, 315(5818), 1544–1548.
Goeller, S., Steinhaege, D., Thoma, M., & Grosfeld, K. (2016). Assessing the subglacial lake coverage of Antarctica. Annals of Glaciology, 57(72), 109–117.
Goff, J. A., Powell, E. M., Young, D. A., & Blankenship, D. D. (2014). Instruments and methods conditional simulation of Thwaites Glacier (Antarctica) bed topography for flow models: Incorporating inhomogeneous statistics and channelized morphology. Journal of Glaciology, 60, 635–646.
Gray, L., Joughin, I., Tulaczyk, S., Spikes, V. B., Bindschadler, R., & Jeffes, K. (2005). Evidence for subglacial water transport in the West Antarctic Ice Sheet through three-dimensional satellite radar interferometry. Geophysical Research Letters, 32, L03501. https://doi.org/10.1029/2004GL021387
Greene, C. A., Gowther, D. E., & Blankenship, D. D. (2017). Antarctic mapping tools for MATLAB. Computers & Geosciences, 104, 151–157.
Hu, Y. Z., & Tonder, K. (1992). Simulation of 3-D random rough surface by 2-D digital filter and Fourier analysis. International Journal of Machine Tools and Manufacture, 32(1-2), 83–90.
Joliffe, I. T., & Morgan, B. J. T. (1992). Principal component analysis and exploratory factor analysis. Statistical methods in medical research, 1(1), 69–95.
Jordan, T. M., Cooper, M. A., Schroeder, D. M., Williams, C. N., Paden, J. D., Siegert, M. J., & Bamber, J. L. (2017). Self-affine subglacial roughness: Consequences for radar scattering and basal water discrimination in northern Greenland. The Cryosphere, 11(3), 1247.
Journel, A. G. (1974). Geostatistics for conditional simulation of ore bodies. Economic Geology, 69(5), 673–687.
Kamb, B. (1987). Glacier surge mechanism based on linked cavity configuration of the basal water conduit system. Journal of Geophysical Research, 92(B9), 9083–9100.
MACKIE ET AL. 21 of 22

Journal of Geophysical Research: Earth Surface
10.1029/2019JF005420

Langley, K., Kohler, J., Matsuoka, K., Sinisalo, A., Scambos, T., Neumann, T., et al. (2011). Recovery Lakes, East Antarctica: Radar assessment of sub-glacial water extent. Geophysical Research Letters, 38, L05501. https://doi.org/10.1029/2010GL046904

Livingstone, S., Clark, C., Woodward, J., & Kinglake, J. (2013). Potential subglacial lake locations and meltwater drainage pathways beneath the Antarctic and Greenland ice sheets. Cryosphere, 7(6), 1721–1740.

Lythe, M. B., & Vaughan, D. G. (2001). BEDMAP: A new ice thickness and subglacial topographic model of Antarctica. Journal of Geophysical Research, 106(B6), 11,335–11,351.

Mandel, J. (1982). Use of the singular value decomposition in regression analysis. The American Statistician, 36(1), 15–24.

Mariethoz, G., Renard, P., & Straubhaar, J. (2010). The direct sampling method to perform multiple-point geostatistical simulations. Water Resources Research, 46, W11536. https://doi.org/10.1029/2008WR007621

Martos, Y. M., Catalán, M., Jordan, T. A., Golyansky, A., Golyansky, D., Eagles, G., & Vaughan, D. G. (2017). Heat flux distribution of Antarctica unveiled. Geophysical Research Letters, 44, 11–417. https://doi.org/10.1002/2017GL075609

Matsuoka, K. (2011). Pitfalls in radar diagnosis of ice-sheet bed conditions: Lessons from englacial attenuation measurements. Geophysical Research Letters, 38, L05505. https://doi.org/10.1029/2010GL046205

Maue, C. F., Purucker, M. E., Olsen, N., & Mosegaard, K. (2005). Heat flux anomalies in Antarctica revealed by satellite magnetic data. Science, 309(5733), 464–467.

Meerschmann, E., Pirot, G., Mariethoz, G., Straubhaar, J., Van Meirvenne, M., & Renard, P. (2013). A practical guide to performing multiple-point statistical simulations with the direct sampling algorithm. Computers & Geosciences, 52, 307–324.

Morlighem, M., Rignot, E., Seroussi, H., Larour, E., Ben Dhia, H., & Aubry, D. (2011). A mass conservation approach for mapping glacier ice thickness. Geophysical Research Letters, 38, L19503. https://doi.org/10.1029/2011GL049690

Oblmacher, G. C., & Davis, J. C. (2003). Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA. Engineering geology, 69(3-4), 331–343.

Oswald, G. K. A., & Robin, G. D. Q. (1973). Lakes beneath the Antarctic ice sheet. Nature, 245(5423), 251.

Pattyn, F. (2010). Antarctic subglacial conditions inferred from a hybrid ice sheet/ice stream model. Earth and Planetary Science Letters, 293(3-4), 451–461.

Ridley, J. K., Cudlip, W., & Laxon, S. W. (1993). Identification of subglacial lakes using ERS-1 radar altimeter. Journal of Glaciology, 39(133), 625–634.

Rignot, E., Mouginot, J., & Scheuchl, B. (2017). MEnSUREs InSAR-based Antarctica ice velocity map, version 2. Boulder, CO: NASA DAAC at the National Snow and Ice Data Center.

Rignot, E., Mouginot, J., Scheuchl, B., van den Broeke, M., van Wessem, M. J., & Morlighem, M. (2019). Four decades of Antarctic Ice Sheet mass balance from 1979–2017. Proceedings of the National Academy of Sciences, 116(4), 1095–1103.

Robinson, G. D. Q., Swithenbank, C. W. M., & Smith, B. M. E. (1970). Radio echo exploration of the Antarctic ice sheet. International Association of Scientific Hydrology Publication, 86, 97–115.

Schröder, D. M., Blankenship, D. D., Raney, R. K., & Grima, C. (2015). Estimating subglacial water geometry using radar bed echo specularity: Application to Thwaites Glacier, West Antarctica. IEEE Geoscience and Remote Sensing Letters, 12(3), 443–447.

Schröder, D. M., Blankenship, D. D., Young, D. A., Witus, A. E., & Anderson, J. B. (2014). Airborne radar sounding evidence for deformable sediments and outcropping bedrock beneath Thwaites Glacier, West Antarctica. Geophysical Research Letters, 41, 7200–7208. https://doi.org/10.1002/2014GL061645

Schwanghart, W., & Scherler, D. (2014). TopoToolbox 2—MATLAB-based software for topographic analysis and modeling in Earth surface sciences. Earth Surface Dynamics, 2(1), 1–7.

Shapira, N. M., & Ritzwoller, M. H. (2004). Inferring surface heat flux distributions guided by a global seismic model: Particular application to Antarctica. Earth and Planetary Science Letters, 228(1-2), 213–224.

Shreve, R. L. (1972). Movement of water in glaciers. Journal of Glaciology, 11(62), 205–214.

Siegel, M. J. (2000). Antarctic subglacial lakes. Earth-Science Reviews, 50(1-2), 29–50.

Siegel, M. J., Carter, S., Tabacco, I., Popov, S., & Blankenship, D. D. (2005). A revised inventory of Antarctic subglacial lakes. Antarctic Science, 17(3), 453–460.

Siegel, M. J., Ellis-Evans, J. C., Tranter, M., Mayer, C., Petit, J. R., Salamatin, A., & Priscu, J. C. (2001). Physical, chemical and biological processes in Lake Vostok and other Antarctic subglacial lakes. Nature, 414(6864), 603.

Siegel, M. J., Ross, N., Corr, H. F. J., Smith, B., Jordan, T. A., Bingham, R. G., et al. (2014). Boundary conditions of an active West Antarctic subglacial lake: Implications for storage of water beneath the ice sheet. The Cryosphere, 8, 15–24.

Siegfried, M. R., & Fricker, H. A. (2018). Thirteen years of subglacial lake activity in Antarctica from multi-mission satellite altimetry. Annals of Glaciology, 59(76p1), 42–55.

Siegfried, M. R., Fricker, H. A., Carter, S. P., & Tulaczyk, S. (2016). Episodic ice velocity fluctuations triggered by a subglacial flood in West Antarctica. Geophysical Research Letters, 43, 2640–2648. https://doi.org/10.1002/2016GL067758

Smith, B. E., Fricker, H. A., Joughin, I. L., & Tulaczyk, S. (2009). An inventory of active subglacial lakes in Antarctica detected by ICESat (2003–2008). Journal of Glaciology, 55(192), 573–595.

Smith, B. E., Gourmelen, N., Huth, A., & Joughin, I. (2017). Connected subglacial lake drainage beneath Thwaites Glacier, West Antarctica. The Cryosphere, 11(1), 451–467.

Smith, A. M., Woodward, J., Ross, N., Bentley, M. J., Hodgson, D. A., Siegel, M. J., & King, E. C. (2018). Evidence for the long-term sedimentary environment in an Antarctic subglacial lake. Earth and Planetary Science Letters, 504, 139–151.

Stearns, L. A., Smith, B. E., & Hamilton, G. S. (2008). Increased flow speed on a large East Antarctic outlet glacier caused by subglacial floods. Nature Geoscience, 1(12), 827.

Straubhaar, J., Renard, P., Mariethoz, G., Froidevaux, R., & Besson, O. (2011). An improved parallel multiple-point algorithm using a list approach. Mathematical Geosciences, 43(3), 305–328.

Strebelle, S. (2002). Conditional simulation of complex geological structures using multiple-point statistics. Mathematical geology, 34(1), 1–21.

Studinger, M., Bell, R. E., & Tikku, A. A. (2004). Estimating the depth and shape of subglacial Lake Vostok’s water cavity from aerogravity data. Geophysical Research Letters, 31, L12401. https://doi.org/10.1029/2004GL019801

Taylor, J., Siegel, M. J., Payne, A. J., & Hubbard, B. (2004). Regional-scale bed roughness beneath ice masses: Measurement and analysis. Computers & Geosciences, 30(8), 899–908.

Weertman, J. (1964). The theory of glacier sliding. Journal of Glaciology, 5(39), 287–303.

Wingham, D. J., Siegel, M. J., Shepherd, A., & Muir, A. S. (2006). Rapid discharge connects Antarctic subglacial lakes. Nature, 440(7087), 1033.
Wolovick, M. J., Bell, R. E., Creyts, T. T., & Frearson, N. (2013). Identification and control of subglacial water networks under Dome A, Antarctica. *Journal of Geophysical Research: Earth Surface, 118*, 140–154. https://doi.org/10.1029/2012JF002555

Wright, A., & Siegert, M. (2012). A fourth inventory of Antarctic subglacial lakes. *Antarctic Science, 24*(6), 659–664.

Young, D. A., Schroeder, D. M., Blankenship, D. D., Kempf, S. D., & Quartini, E. (2016). The distribution of basal water between Antarctic subglacial lakes from radar sounding. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374*(2059), 20140297.

Zwally, H. J., Schutz, R., Hancock, D., & Dimarzio, J. (2014). GLAS/ICESat L1 Antarctic and Greenland Ice Sheet altimetry data (HDF5). Version 34. Boulder, Colorado USA: NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.5067/ICESAT/GLAS/DATA209