



1 Geochemical zones and environmental gradients for soils from the 2 Central Transantarctic Mountains, Antarctica

3 Melisa A. Diaz^{1,2*}, Christopher B. Gardner^{1,2}, Susan A. Welch^{1,2}, W. Andrew Jackson³, Byron J.
4 Adams⁴, Diana H. Wall⁵, Ian D. Hogg^{6,7}, Noah Fierer⁸, W. Berry Lyons^{1,2}

5 ¹School of Earth Sciences, The Ohio State University, Columbus, OH, USA

6 ²Byrd Polar and Climate Research Center, The Ohio State University, Columbus, OH, USA

7 ³Department of Civil, Environmental, & Construction Engineering, Texas Tech University, Lubbock, TX, USA

8 ⁴Department of Biology, Evolutionary Ecology Laboratories, and Monte L. Bean Museum, Brigham Young University,
9 Provo, UT, USA

10 ⁵Department of Biology and School of Global Environmental Sustainability, Colorado State University, Fort Collins, CO,
11 USA

12 ⁶Canadian High Arctic Research Station, Polar Knowledge Canada, Cambridge Bay, Nunavut, Canada

13 ⁷School of Science, University of Waikato, Hamilton, New Zealand

14 ⁸Department of Ecology and Evolutionary Biology and Cooperative Institute for Research in Environmental Science,
15 University of Colorado Boulder, Boulder, CO, USA

16
17 *Correspondence to:* Melisa A. Diaz (diaz.237@osu.edu)

18
19 **Abstract.** Previous studies have established links between biodiversity and soil geochemistry in the McMurdo Dry Valleys,
20 Antarctica, where environmental gradients are important determinants of soil biodiversity. However, these gradients are not
21 well established in the Central Transantarctic Mountains, which are thought to represent some of the least hospitable
22 Antarctic soils. We analyzed 220 samples from 11 ice-free areas along the Shackleton Glacier (~85 °S), a major outlet
23 glacier of the East Antarctic Ice Sheet. We established three zones of distinct geochemical gradients near the head of the
24 glacier (upper), central (middle), and at the mouth (lower). The upper zone had the highest water-soluble salt concentrations
25 with total salt concentrations exceeding 80,000 $\mu\text{g g}^{-1}$, while the lower zone had the lowest water-soluble N:P ratios,
26 suggesting that, in addition to other parameters (such as proximity to water/ice), the lower zone likely represents the most
27 favorable ecological habitats. Given the strong dependence of geochemistry with geographic parameters, we established
28 multiple linear regression and random forest models to predict soil geochemical trends given latitude, longitude, elevation,
29 distance from the coast, distance from the glacier, and soil moisture (variables which can be inferred from remote
30 measurements). Confidence in our model predictions was moderately high, with R^2 values for total water-soluble salts,
31 water-soluble N:P, ClO_4^- , and ClO_3^- of 0.51, 0.42, 0.40, and 0.28, respectively. These modeling results can be used to predict
32 geochemical gradients and estimate salt concentrations for other Transantarctic Mountain soils, information that can
33 ultimately be used to better predict distributions of soil biota in this remote region.



34 1. Introduction

35 From an ecological standpoint, the least biologically diverse terrestrial systems are those found in extreme physical
36 and chemical environments. The abundance and diversity of life in soils is dependent on a number of environmental
37 parameters, including temperature, precipitation, organic matter content, and nutrient availability (Wall et al., 2012). Hot
38 deserts are typically viewed as one of the least biologically diverse environments. However, cold deserts can often be even
39 less diverse (Freckman and Virginia, 1998). Soils in Antarctica typically serve as end-members for low habitat suitability
40 due to their high salt concentrations, low organic carbon, low soil moisture, and low mean annual temperatures (Courtright et
41 al., 2001).

42 In the McMurdo Dry Valleys (MDV), organic matter and salt concentrations influence soil communities, where
43 soils with higher amounts of organic carbon, lower water-soluble N:P ratios, and lower total water-soluble salt
44 concentrations generally harbor the greatest biomass and biodiversity (Barrett et al., 2006; Bottos et al., 2020; Caruso et al.,
45 2019; Magalhães et al., 2012). These Antarctic ecosystems are relatively simple and are the only known soil systems where
46 nematodes and microarthropods (Collembola, Acari) are at the top of the food chain (Freckman and Virginia, 1998; Hogg
47 and Wall, 2012). Studies of soils in the MDV and Transantarctic Mountains (TAM) have been key to understanding
48 ecosystem structure and function in extreme terrestrial environments (e.g. Caruso et al., 2019; Collins et al., 2019; Freckman
49 and Virginia, 1998).

50 Biological processes in Antarctic soils are largely dependent on the availability, duration, and proximity of soils to
51 liquid water (Barrett et al., 2006). Due to the seasonality in freezing and thawing events, liquid water acts as a pulse to the
52 ecosystem, providing water for organisms, but also wetting surface soils and dissolving soluble salts (Webster-Brown et al.,
53 2010; Zeglin et al., 2009). Experiments of salt thresholds on Antarctic nematodes found that no individuals survived in
54 highly saline soils ($\sim 2,600$ mg L⁻¹ TDS) (Nkem et al., 2006). Concentrations of soluble salts exist at these concentrations or
55 higher for high elevation and inland locations in the TAM (Bockheim, 2008; Lyons et al., 2016). Additionally, studies on
56 TAM soils have found that increased salt concentrations lead to a decrease in soil biodiversity in older soils compared to
57 younger soils (Magalhães et al., 2012). Yet, despite these inhospitable conditions (e.g. high salt concentrations and glacial
58 advance and retreat), some organisms are postulated to have found suitable refugia in TAM soils and persisted in isolation
59 for millions of years and through glacial cycles (Beet et al., 2016; Stevens et al., 2006; Stevens and Hogg, 2003).

60 It is generally accepted that habitat suitability for invertebrate species in Antarctic soils is driven by a combination
61 of geochemical, geographic, and geomorphic variables (Bottos et al., 2020; Courtright et al., 2001; Freckman and Virginia,
62 1998; Magalhães et al., 2012). Geographic variables, such as elevation, can be measured with advanced mapping tools and
63 satellite imagery; however, surface exposure ages, soil geochemistry and nutrient content require extensive logistical support
64 and resource allocation for sample collection and analysis. More efficient estimation tools are needed to aid in our ability to
65 understand and predict habitat suitability for invertebrates throughout the TAM.



66 With this study, we determined and evaluated geochemical patterns and gradients of water-soluble ions in soils
67 collected from 11 ice-free areas along the Shackleton Glacier, Central Transantarctic Mountains (CTAM). Particular
68 attention was given to total water-soluble salt concentrations, N:P ratios, and ClO_4^- and ClO_3^- concentrations, based on their
69 influence on biodiversity, as determined in previous studies (e.g. Ball et al., 2018; Barrett et al., 2006b; Courtright et al.,
70 2001; Dragone et al., 2020; Nkem et al., 2006). The geochemical data were compared to geographic parameters to
71 understand how the physical environment influences the observed geochemical variability. Our results show that water-
72 soluble ion concentrations and distributions are driven largely by soil geography and surface exposure age. Finally, we
73 implemented statistical and machine learning techniques to interpolate and predict the soil geochemistry across the region
74 using geographic variables. Our multiple linear regression and random forest models show that latitude, longitude, elevation,
75 distance from the coast, distance from the glacier, and soil moisture (all variables currently or soon to be remotely
76 measurable using maps and satellites) are moderately effective at estimating spatial patterns in TAM soil geochemistry, with
77 R^2 values as high as 0.87. These data will be particularly useful for ecologists seeking to understand refugia and habitat
78 suitability in Antarctica and similarly harsh, desert environments.

79 2. Study sites

80 The Shackleton Glacier (~84.5 to 86.4°S; ~130 km long and ~10 km wide) is a S-N trending outlet glacier of the
81 East Antarctic Ice Sheet (EAIS) located to the west of the Beardmore Glacier and flows through the Queen Maud Mountains
82 (CTAM) into the Ross Sea (Fig. 1). The elevations of exposed soils range from ~150 m.a.s.l. to >3,500 m.a.s.l. from the
83 coast towards the Polar Plateau. Long-term climate data are not yet available, but the Shackleton Glacier region is a polar
84 desert regime, similar to the Beardmore Glacier region, with average annual temperatures well below freezing and little
85 precipitation (LaPrade, 1984).

86 During the Last Glacial Maximum (LGM) and glacial periods throughout the Pleistocene, the size and thickness of
87 the EAIS has been suggested to be greater than current levels (Golledge et al., 2013; Nakada and Lambeck, 1988; Talarico et
88 al., 2012; Wilson et al., 2018). Outlet glaciers, such as the Shackleton Glacier, may have had the greatest increases in extent,
89 especially towards the glacier terminus (Golledge et al., 2012; Golledge and Levy, 2011). The behavior of local alpine and
90 tributary glaciers is not well-constrained, but these glaciers are also believed to have advanced and retreated over the last two
91 million years (Diaz et al., 2020a; Jackson et al., 2018). As a result, currently exposed soils were overlain and reworked by
92 fluctuations of the Shackleton Glacier and other tributary and alpine glaciers in the region. Exposure ages range from the
93 early Holocene to the Miocene, and generally increase with distance from the coast and distance from the glacier (Balter et
94 al., 2020; Diaz et al., 2020a).

95 The soils contain a range of water-soluble salts derived primarily from atmospheric deposition and chemical
96 weathering (Claridge and Campbell, 1968; Diaz et al., 2020b). The major salts are typically nitrate and sulfate salts,
97 especially at higher elevations and further inland from the coast of the Ross Sea (Diaz et al., 2020b). The solubilities of the



98 salts vary, but nitrate salts are highly soluble and their occurrence at high elevation and inland locations suggests that those
99 soils have maintained persistent arid conditions.

100 3. Methods

101 3.1. Sample collection and preparation

102 During the 2017-2018 austral summer, 220 surface soil samples (~top 5 cm) were collected from 11 distinct ice-free
103 areas (Roberts Massif, Schroeder Hill, Mt. Augustana, Bennett Platform, Mt. Heekin, Thanksgiving Valley, Taylor Nunatak,
104 Mt. Franke, Mt. Wasko, Nilsen Peak, and Mt. Speed) along the Shackleton Glacier, including a subset of 27 samples
105 previously analyzed for S, N, and O isotopes in nitrate and sulfate (Diaz et al., 2020b). At each area, we collected samples in
106 transects (ranging from ~200 m to ~2,000 m in length) to maximize the geochemical variability. Our transects were also
107 designed to capture the LGM transition, with some soils exposed throughout the LGM and others exposed following glacier
108 retreat. GPS coordinates and elevations were recorded with each sample and later used to estimate the distance from coast
109 and distance from the glacier (defined as linear distance from the nearest glacier – Shackleton, tributary, or alpine). Once
110 collected, the samples were stored and shipped frozen (-20 °C) to The Ohio State University.

111 Prior to geochemical analysis, the samples were dried at 50 °C for at least 72 hours with the loss in mass attributed
112 to soil moisture content. The dried soils were leached at a 1:5 soil to DI water ratio, and the leachate was filtered through 0.4
113 µm Nucleopore membrane filters (Diaz et al., 2018, 2020b; Nkem et al., 2006). Due to the low sediment to water ratio, this
114 leaching technique only dissolves the more water-soluble salts (Toner et al., 2013). These include salts with ClO_4^- , NO_3^- , Cl^- ,
115 SO_4^{2-} , ClO_3^- , and $\text{CO}_3^{2-} + \text{HCO}_3^-$. Process blanks were generated and analyzed to account for any contamination from the
116 leaching process.

117 3.2. Analytical analysis of water-soluble anions, cations, and nutrients

118 The analytical techniques used here are similar to those reported by Diaz et al. (2020b). In brief, the analytes
119 included anions (F^- , Cl^- , Br^- , and SO_4^{2-}) which were measured on a Dionex ICS-2100 ion chromatograph, cations (K^+ , Na^+ ,
120 Ca^{2+} , Mg^{2+} , and Sr^{2+}) which were measured on a PerkinElmer Optima 8300 Inductively Coupled Plasma-Optical Emission
121 Spectrometer (ICP-OES), and nutrients ($\text{NO}_3^- + \text{NO}_2^-$, PO_4^{3-} , H_4SiO_4 , and NH_3) which were measured on a Skalar San++
122 Automated Wet Chemistry Analyzer at The Ohio State University. Perchlorate (ClO_4^-) and chlorate (ClO_3^-) were measured
123 using an ion chromatograph-tandem mass spectrometry technique (IC-MS/MS) at Texas Tech University (Jackson et al.,
124 2012, 2015). All analytes are reported as listed. Total water-soluble salt concentration was calculated as the sum of all
125 measured cations and anions. The precision of replicated check standards and samples was typically better than 10% for all
126 major anions, cations and nutrients, and better than 20% for perchlorate and chlorate. Accuracy was typically better than 5%
127 for all major anions, cations, and nutrients, as determined by the NIST 1643e external reference standard and the 2015 USGS
128 interlaboratory calibration standard (M-216), and better than 10% for perchlorate and chlorate, as determined by spike



129 recoveries. Precision and accuracy for individual analytes are located in Table S1. Detection limits for the analytes have been
130 previous reported (Diaz et al., 2018; Jackson et al., 2012).

131 3.3. Data interpolation and machine learning

132 Inverse distance weighted (IDW) interpolations were performed for Bennett Platform, Thanksgiving Valley, and
133 Roberts Massif using the Geostatistical Analyst tool in ArcMap 10.3. Since IDW is a deterministic method where unknown
134 values are predicted based on proximity to known values, we chose those three sites as they had the most defined transects
135 and relatively higher sample density. The interpolation parameters were constant with a power of 4, maximum neighbors of
136 15, minimum neighbors of 5, and 4 sectors, and a variable search radius. These parameters were chosen such that they
137 optimize for the lowest mean absolute error.

138 Multiple linear regressions were generated for all geochemical analytes, except H_4SiO_4 (total of 15 dependent
139 variables), with latitude, longitude, elevation, distance from the coast, distance from the glacier, and soil moisture as
140 independent variables using built-in functions in R 3.6.3 (R Core Team, 2020). Random forest regression models were
141 similarly generated using the randomForest library. The random forest model is a machine learning algorithm that utilizes
142 supervised learning algorithms to predict values given input predictor variables (Breiman, 2001). Multiple decision trees are
143 run in parallel with a randomized subset of predictor variables, and the aggregate result of each tree is used to generate a
144 predicted outcome. Since each tree is generated using a random sample and random predictor variables, the random forest
145 model is effective at minimizing overfitting and handling outliers (Breiman, 2001).

146 Machine learning algorithms are widely used in variety of disciplines from finance (Patel et al., 2015) to ecology
147 (Davidson et al., 2009; Peters et al., 2007; Prasad et al., 2006), for both data prediction (regression) and classification.
148 Recently, these techniques have been used for Earth Science applications, including geologic mapping (Heung et al., 2014;
149 Kirkwood et al., 2016), air quality monitoring (Stafoggia et al., 2019), and water contaminant tracing (Tesoriero et al., 2017).
150 We developed a novel application of machine learning to predict concentrations and gradients of water-soluble salts in
151 Antarctic soils, given set geographic parameters, similar to the approaches developed for stock market and real estate
152 predictions (Antipov and Pokryshevskaya, 2012; Patel et al., 2015).

153 For our random forest models, any sparse missing values in Table S2 were estimated by averaging the geochemistry
154 of the samples collected immediately before and after in the same transect. Missing values due to concentrations below the
155 detection limit were input as 0. The new imputed dataset was split into a training set representing 86% of the data ($n = 189$,
156 Table S3) and a testing set representing the remaining 14% ($n = 31$, Table S4). The training dataset was used to generate the
157 random forest models for each analyte. Each of the models were run with 2000 decision trees ($n_{\text{tree}} = 2000$) to minimize the
158 mean square errors. The number of random variables used for each node split in the decision trees was set to the
159 recommended regression default of $\text{variables}/3$ to optimize the model randomness, which in our case was 2 ($m_{\text{try}} = 2$),



160 following parameters described previously (Breiman, 2001). The scripts developed for both the multiple linear regression
161 and random forest models are included in the supplementary materials.

162 4. Results

163 4.1. Geochemistry of upper, middle, and lower zones

164 The maximum, minimum, mean, standard deviation and coefficient of variation are reported in Table 1 for the
165 measured geographic and geochemical data. Concentrations of water-soluble ions span up to five orders of magnitude and
166 are variable across the region. Elevation, distance from the coast, distance from the glacier, and soil moisture are also
167 variable and span up to three orders of magnitude. The highest elevation samples ($> 2,000$ m.a.s.l.) were collected from
168 Schroeder Hill and the greatest soil moisture content is from Mt. Wasko at 12.3%, with a mean of 2.1% for all samples.

169 Shackleton Glacier region surface soils can be separated into three zones based on their water-soluble geochemistry:
170 an upper zone near the Polar Plateau, a middle zone near the center of the glacier, and a lower zone where the glacier flows
171 into the Ross Sea (Figs. 1; 2). The upper zone samples are characterized by the highest total water-soluble salt
172 concentrations, with the highest values greater than $80,000 \mu\text{g g}^{-1}$ at Schroeder Hill, while the lower zone samples have the
173 lowest total salt concentrations, with the lowest values near $10 \mu\text{g g}^{-1}$ at Mt. Wasko (Fig. 2a-c). The middle zone has
174 intermediate values. Water-soluble N:P molar ratios generally follow a similar trend (Fig. 2d-f). The lowest N:P ratios are in
175 the lower zone soils, while the middle and upper zones have more variable values. Concentrations of ClO_4^- and ClO_3^- follow
176 similar trends as the total salts, with less distinction between middle and upper zones, though most concentrations in the
177 lower zone are below the detection limit (Fig. 2g-l; Table S2).

178 Observed trends between the zones appear to be driven, at least partially, by geography. Regressions of total water-
179 soluble salt concentration, water-soluble N:P ratio, and ClO_3^- concentration with elevation, distance from the coast, and
180 distance from the glacier are all positive (Fig. 2). The strongest relationships are between total salts and elevation, and ClO_3^-
181 and distance from the coast, with R^2 values of 0.26 and 0.24, respectively, and p-values < 0.001 (Fig. 2a;2k). The weakest
182 relationships are between ClO_4^- and distance from the coast and distance from the glacier, with R^2 values of 0.01 (Fig. 2h;
183 2i). Distance from the glacier varies widely between individual zones with frequent overlaps, but there appears to be a
184 moderate relationship with N:P ratio and total salts (Fig. 2c; 2f). Overall, total salt concentration has the strongest
185 relationship with geography and ClO_4^- has the weakest relationships.

186 Ternary diagrams highlight the specific geochemical gradients within and between the zones. The anion ternary
187 diagram only includes SO_4^{2-} , NO_3^- , and Cl^- , which are the major water-soluble salts in the region (Claridge and Campbell,
188 1968; Diaz et al., 2020b). Though carbonate and bicarbonate salts have been identified in both lacustrine sediments and soils
189 in Antarctica, previously measured concentrations in the Shackleton Glacier region were low, ranging from 0.07 to 2.5%,
190 and bicarbonate salts were not identified in the highest elevation and furthest inland soils (Claridge and Campbell, 1968;
191 Diaz et al., 2020b; Lyons et al., 2016). The most abundant anion for the upper zone is SO_4^{2-} , which is greater than 99% of the



192 total anion budget in some Schroeder Hill and Roberts Massif samples, though other locations are dominated by NO_3^- (Fig.
193 3). The anions are more evenly distributed in the middle zone, though the majority of samples are most abundant in NO_3^- and
194 Cl^- . The lower zone has much lower SO_4^{2-} fractions than the upper zone and the dominant anion is generally Cl^- . The cation
195 distribution is very similar for all three zones (Fig. 3). $\text{Na}^+ + \text{K}^+$ is the most abundant cation pair representing over 90% of
196 the total cations for many upper and middle zone samples, while Ca^{2+} is the second most abundant. In general, Mg^{2+} is the
197 least abundant cation across all sampling locations.

198 4.2. Statistical geochemical variability

199 A principal component analysis (PCA) was performed in R (using factoextra (Kassambara and Mundt, 2017) and
200 built in software libraries) to determine which geochemical variables most strongly differ across the samples. For the PCA,
201 the first two principal components account for over 50% of the total dataset variability at 44.2% and 11.6%, respectively.
202 The different zones are correlated with different principal components (Fig. 4). The samples from the middle zone are
203 positively correlated with PC1 and PC2. In the biplot, they plot in the upper right quadrant with high concentrations of Cl^- ,
204 NO_3^- , water soluble N:P ratio, and Ca^{2+} , with a minor influence from soil moisture and H_4SiO_4 . The upper zone samples
205 generally plot along PC1 and are most associated with Sr^{2+} , SO_4^{2-} , Mg^{2+} , Na^+ , K^+ , F^- , ClO_4^- , and ClO_3^- . The samples from the
206 lower, more coastal zone are negatively correlated with PC1 and are distinguished by their higher PO_4^{3-} concentrations. Most
207 samples from all locations plot within the 95% confidence interval ellipses. However, there are two strong outliers from
208 Schroeder Hill and Mt. Heekin.

209 Similar to the PCA, we performed a simple Spearman's rank correlation for the entire dataset to visualize the
210 statistical dependence between all variables. Since a goal of this study is to relate water-soluble ion concentrations to
211 geography, we focused on latitude, longitude, distance from the coast, distance from the glacier, and soil moisture. The
212 strongest correlation coefficients are between Cl^- and latitude, elevation, and distance from the coast, and Sr^{2+} and soil
213 moisture (Fig. 5). Most other correlations are moderate to weak, though there appear to be notably stronger correlations
214 between ClO_3^- and latitude and distance from coast, Ca^{2+} and longitude, elevation, and distance from coast, NO_3^- and
215 latitude, and SO_4^{2-} with distance from glacier. Longitude, elevation, and distance from coast have the greatest number of
216 strong and moderate correlations with the geochemistry data. Outside of the geographic parameters, Na^+ is highly correlated
217 with total water-soluble salts, likely representative of the high $\text{Na}^+ + \text{K}^+$ percentages (Fig. 3), and Sr^{2+} is highly correlated
218 with K^+ , likely reflecting a common ion source.

219 4.3. Spatial interpolation and machine learning model performance

220 The total salt concentrations of individual samples at Bennett Platform produce the most defined interpolation
221 gradient from the glacier front to further inland compared to Roberts Massif and Thanksgiving Valley (Fig. 6). Bennett
222 Platform also has the smoothest salt concentration contours suggesting that the interpolation model is the strongest and most
223 robust at this location. The second strongest interpolation is Thanksgiving Valley. Contrary to the measurements at Bennett



224 Platform, Thanksgiving Valley has the highest salt concentrations in the center of the valley, with lower concentrations to
225 both the east and west. The lowest concentration contours are closest to the glacier for both Bennett Platform and
226 Thanksgiving Valley, which is likely related to glacial history since the soils near the glacier are relatively younger than
227 those further inland based on meteoric ^{10}Be data (Diaz et al., 2020a). The interpolation from Roberts Massif does not have a
228 distinguishable spatial trend.

229 The multiple linear regression and random forest models vary in their strength for the individual analytes. The
230 highest R^2 value from the linear regression is 0.55 for Sr^{2+} , while total water-soluble salts, water-soluble N:P ratio, ClO_4^- ,
231 and ClO_3^- have values of 0.37, 0.37, 0.10, and 0.33, respectively (Table 2). The lowest R^2 value is for Cl^- at 0.05. The p-
232 values for nearly all analytes are $\ll 0.001$, with Cl^- having the only value above 0.05. The highest out-of-the-bag explained
233 variance values from the random forest models are for K^+ and Sr^{2+} at 62% for both analytes. Values for NO_3^- , PO_4^{3-} , ClO_4^- ,
234 and N:P ratio are negative. The explained variance for total salts is 45% and the variance for ClO_3^- is 43%. We also
235 evaluated the most important and least important variables in the random forest models based on node purity. The most
236 important variable for the majority of analytes is elevation, while distance from the glacier is most important for N:P ratio
237 and latitude for ClO_3^- (Table 2). The least important variable is distance from the coast for every analyte, except ClO_3^- and
238 NH_3 , for which distance from the glacier is least important.

239 5. Discussion

240 5.1. Implications for ecological habitat suitability

241 By establishing geochemical zones for the Shackleton Glacier region, we can better understand the relationship
242 between geochemistry and geography, and ultimately biogeography. As stated in the introduction, we focused particularly on
243 total water-soluble salt concentrations, water-soluble N:P ratios, and ClO_4^- and ClO_3^- concentrations.

244 5.1.1. Elevation and moisture controls on total water-soluble salt gradients

245 The elevational trends of total salt concentrations at the Shackleton Glacier are similar to those previously described
246 in the TAM, where soils from higher elevation sites typically have higher salt concentrations (Bottos et al., 2020; Lyons et
247 al., 2016; Magalhães et al., 2012). Our results are also consistent with those from Scarrow et al. (2014), who found that salt
248 concentrations typically decreased with distance from the glacier. Our total water-soluble salt interpolation maps highlight
249 the spatial variability in Shackleton Glacier region soils (Fig. 6). The most spatially variable location is Robert Massif, which
250 does not appear to follow local elevational, latitudinal, and/or distance inland gradients. This heterogeneity is not necessarily
251 due to currently active soil leaching, as the soil moisture values are not drastically different between the samples (Table S2).
252 Though the variability in cation concentrations is likely due to weathering of tills, scree, and bedrock (Claridge and
253 Campbell, 1968), recent work on the isotopic composition of water-soluble nitrate and sulfate, the major anions in the upper
254 zone, suggests a common, atmospheric source (Diaz et al., 2020b).



255 We argue that the heterogeneity in the total salt concentrations at Roberts Massif (Figs. 2; 6) is probably related to
256 different and complex wetting history, where seasonal snow patch melt may pool in local depressions, transporting water-
257 soluble salts from slightly higher elevations and/or from saline wet-patches (Levy et al., 2012). This is demonstrated on a
258 larger scale at Thanksgiving Valley, a glacially carved valley, where the higher concentrations of salts in the center of the
259 valley are likely due to the transport of salts from nearby higher elevation slopes during melting events. This is further
260 evidenced by the presence of two small, closed-basin ponds in the center of the valley, which likely formed from glacial melt
261 and may have been larger in size in the recent past (Diaz et al., 2019). Similarly, streams and meltwater tracks in the MDV
262 leach soils and carry salts into closed basin, brackish to hyper-saline lakes, where salts are cryoconcentrated over time
263 (Lyons et al., 1998). Our results suggest that elevation and wetting history are important contributors to total salt gradients in
264 the Shackleton Glacier region, as they influence the accumulation of salts and subsequent leaching from soils.

265 5.1.2. Influence of glacial history on water-soluble N:P ratios

266 Stoichiometric dependencies have been identified for Antarctic terrestrial organisms, where nutrient concentrations,
267 in addition to soil aridity, limit ecosystem development (Nkem et al., 2006). Since nitrate is primarily derived from
268 atmospheric deposition and phosphorus is primarily liberated from minerals by chemical weathering in the CTAM, many
269 inland and higher elevation soils have accumulated high concentrations of NO_3^- , resulting in stoichiometric imbalance with
270 soluble PO_4^{3-} (Ball et al., 2018; Barrett et al., 2007; Diaz et al., 2020b; Lyons et al., 2016; Nkem et al., 2006). As in the
271 MDV, younger and coastal soils at lower elevations in the Shackleton Glacier region have the lowest water-soluble N:P
272 ratios, driven by relatively low concentrations of NO_3^- and high concentrations of PO_4^{3-} due to an increase in moisture
273 content and chemical weathering (Heindel et al., 2017) (Fig. 2; 4). It is not surprising that life was conspicuous in these soils,
274 with thick lichen growth on several rocks and the presence of both Collembola and mites at Mt. Speed and Mt. Wasko (Fig.
275 S1). However, despite overall elevational and latitudinal gradients, some inland locations in the middle and upper zones have
276 water-soluble N:P ratios near those from the lower zone (Fig. 2).

277 The interpolation model from Bennett Platform shows that some locations near the glacier have lower total water-
278 soluble salt concentrations (Fig. 6), similar to soils surveyed in the MDV (Bockheim, 2002). However, the samples near the
279 glacier at Bennett Platform not only have lower total salt concentrations, they also have lower N:P ratios than samples
280 collected further inland. This is also the case for the middle zone locations (Fig. 2f). We argue this is due to differences in
281 glacial history between the locations. Our previous work showed that soils near the glacier are younger than soils further
282 inland in the Shackleton Glacier region (Diaz et al., 2020a). These soils are shielded from nitrate accumulation during glacial
283 periods, and the recently exposed rocks likely serve as fresh mineral weathering material for PO_4^{3-} mobilization (Heindel et
284 al., 2017). Recently exposed and relatively nutrient rich soils might be important refugia for invertebrates. Previous
285 hypotheses have suggested that organisms may have persisted at higher elevations during glacial periods (Bennett et al.,
286 2016; Stevens and Hogg, 2003). However, abiotic gradients in the Beardmore Glacier region suggest that higher elevation
287 soils have salt concentrations that would classify them as unsuitable habitats (Lyons et al., 2016). If few organisms survived



288 glaciations, the near-glacier, relatively P-rich soils may be important in helping communities recover and restructure post-
289 glaciation.

290 5.1.3. High and variable ClO_4^- and ClO_3^- concentrations

291 Our ClO_4^- and ClO_3^- concentrations include the highest measured in Antarctica to date and are comparable to
292 concentrations from the Atacama and Mojave Deserts (Jackson et al., 2015). Though not a strong correlation, the highest
293 elevation samples (upper zone) have the highest ClO_4^- and ClO_3^- concentrations (Fig. 2g; 2j). Similar to NO_3^- , ClO_4^- and
294 ClO_3^- are derived from atmospheric deposition and because of their solubilities, their accumulations are related to wetting
295 and glacial histories (Jackson et al., 2016, 2015). Therefore, soils which have been exposed for long periods of time and have
296 not experienced snow or ice melt, such as those from Schroeder Hill and Roberts Massif, are able to accumulate high
297 concentrations of ClO_4^- and ClO_3^- . Interestingly, our ClO_4^- concentrations are lower (maximum of $\sim 1.9 \text{ g L}^{-1}$) than the
298 highest recorded tolerance (1.1 M ($\sim 130 \text{ g L}^{-1}$) NaClO_4) for the extremotolerant bacteria *Planococcus halocryophilus*, yet a
299 recent study shows no detectable biomass for Schroeder Hill samples (Dragone et al., 2020). (Per)chlorates are strong
300 oxidizers and are well established as toxic, thus the concentrations of ClO_4^- and ClO_3^- might be additional, crucial indicators
301 of habitat suitability. However, the concentrations are highly heterogenous across our sampled locations (Fig. 2k-l), and
302 unlike ClO_3^- , neither the multiple linear regression nor random forest models were able to adequately capture the variability
303 in ClO_4^- concentrations (Table 2).

304 5.2. Machine learning as a tool to predict soil geochemical trends

305 We sought to evaluate our multiple linear regression and random forest models using a testing dataset from the
306 Shackleton Glacier region ($n = 31$) and a second dataset from the Darwin Mountains ($\sim 80^\circ\text{S}$) ($n = 10$) (Magalhães et al.,
307 2012). Few published/available TAM dataset include sample GPS coordinates, soil moisture, and water-soluble ion
308 geochemistry. As stated in Section 3.3, the Shackleton Glacier region test data were not included in the random forest model
309 generation so we could evaluate our models with an independent dataset. For the Darwin dataset, distance from the glacier,
310 distance from the coast, and elevation were determined using the Reference Elevation Model of Antarctica (REMA), while
311 location, soil moisture and geochemistry were retrieved from the literature (Howat et al., 2019; Magalhães et al., 2012). We
312 evaluated all 15 analytes from the original models with the Shackleton dataset and due to a lack of data, only evaluated 7
313 analytes from the Darwin soils (Figure 7).

314 Both the multiple linear regression and random forest model outputs are moderately well-correlated for the
315 Shackleton dataset, as determined by Pearson correlations between the measured and predicted values (Fig. 7a; Table 3). The
316 random forest models outperform the linear regression models for nearly every analyte, with the notable exceptions of F^- ,
317 Na^+ , and NO_3^- , and nearly all p-values are < 0.001 . Mg^{2+} is the most accurately predicted, with R^2 values of 0.79 and 0.52 for
318 the random forest and linear regression models, respectively (Fig. 7a). In terms of our analytes of interest regarding habitat
319 suitability, total salts have the strongest correlation in the random forest model ($R^2 = 0.51$), followed by water-soluble N:P



320 ratio ($R^2 = 0.42$), ClO_4^- ($R^2 = 0.40$), and ClO_3^- ($R^2 = 0.28$). N:P ratio in particular performs significantly better than the linear
321 regression model ($R^2 = 0.05$). Mean absolute error (MAE) and root mean squared error (RMSE) values indicate that the
322 random forest models also have a smaller error compared to the multiple linear regression models (Table 4). MAE values are
323 lower than RMSE values for both models, indicating the strong presence of outliers in the testing dataset. This is
324 unsurprising as the standard deviation and coefficient of variation values for the entire dataset are relatively large for all
325 analytes.

326 Similar to the model performance in the Shackleton Glacier region, the water-soluble ion predictions for the Darwin
327 Glacier region are more strongly correlated with measured values in the random forest models compared to the multiple
328 linear regressions (Fig. 7b). In fact, the linear regression models fail for the Darwin samples and all concentration outputs are
329 negative, which is likely due to overfitting during model generation. MAE and RSME values for both models are much
330 higher than those for the Shackleton dataset (Table 4). On the other hand, the random forest models perform particularly well
331 for some analytes. Though a small sample size, the R^2 values for Mg^{2+} and K^+ are 0.87, with p-values $\ll 0.001$. Total salts is
332 moderately correlated ($R^2 = 0.44$) and N:P ratio has an R^2 value of 0.01, indicating poor model performance. It is unclear
333 why Mg^{2+} and K^+ are the most accurately predicted, though we suspect that this is due to weathering trends of local lithology
334 across the TAM, since chemical weathering is probably the major source of these ions.

335 It should be noted that the R^2 values simply measure the strength of the correlations between the measured and
336 predicted values. We performed slope tests by fitting bivariate lines using the standardized major axis (SMA) to further
337 understand the relationship between the two values using the smatr library in R (Warton et al., 2012). For this test, we
338 specifically evaluated the null hypothesis (H_0) where slope = 1, which would indicate whether an ideal, direct 1:1
339 relationship exists between the measured and predicted values. Test statistic values (t) were used to measure the sample
340 correlation between the residuals and fitted values (Warton et al., 2012). Test statistic values near 1 indicate that we reject
341 the null hypothesis. In other words, higher absolute test statistic values indicate a slope other than 1. Of the 15 analytes in the
342 Shackleton dataset, 7 analytes have slopes near 1 for the multiple linear regression models and 6 for the random forest, as
343 indicated by test statistic values less than 0.5. For the Darwin, only one analyte, NO_3^- , has a test statistic value less than 0.5
344 (Fig. 7; Table 3).

345 These data indicate that while some analytes have high correlations between measured and predicted values, the
346 models perform best with the Shackleton Glacier region soils. Additionally, though the relationship may not be 1:1, the
347 random forest models are effective at predicting the measured geochemical gradients. For example, similar to our data, the
348 Darwin Glacier samples generally have greater water-soluble N:P ratios and total water-soluble salt concentrations further
349 from the glacier and at higher elevations (Magalhães et al., 2012), a trend that is reflected by our model results despite offset
350 values. Additionally, corrections for the offset of the model from a slope = 1 (i.e. multiplying the model output value by the
351 regression slope) can be made to better estimate specific concentrations, though the difference between modeled and



352 measured values can still be up to 2x greater. Our sample size for building the multiple linear regression and random forest
353 models is small. We anticipate that, as more data are collected throughout the CTAM, these data can be added to the model
354 training dataset, expanding our prediction capabilities and increasing model reliability.

355 6. Conclusions

356 The soil ecosystems found in the Transantarctic Mountains are among the least diverse on Earth and their structure
357 is influenced by environmental factors. We characterized environmental and geochemical gradients in the Shackleton Glacier
358 region, which aid in our understanding of the abiotic properties in soils governing biodiversity and biogeography. The 220
359 samples we analyzed represent a wide range of soil environments: those with different elevation, latitude, longitude, glacial
360 history, and geochemistry. We determined three soil zones: an upper zone near the head of the glacier which is characterized
361 by high total water-soluble salt concentrations, high water-soluble N:P ratios, and high ClO_4^- and ClO_3^- concentrations, a
362 lower zone with low total salt concentrations and higher PO_4^{3-} concentrations, and a middle zone with intermediate values.
363 The zones help elucidate the geographic influences on soil geochemistry. In addition, our total water-soluble salt
364 interpolations at Roberts Massif, Bennett Platform, and Thanksgiving Valley reflect the local small-scale variability of salt
365 concentrations and possible influences from soil age and wetting history.

366 Similar to previous studies, our results suggest that high elevation and inland soils, such as those from the upper
367 zone, were likely unsuitable candidates for refugia during the Last Glacial Maximum. However, glacial advance and retreat
368 and climate shifts may leach soils, lowering otherwise toxic total water-soluble salt concentrations and N:P ratios. These
369 more recently exposed soils may be particularly important in maintaining and reviving contemporary and past biological
370 communities.

371 Five geographic variables (latitude, longitude, elevation, distance from the coast, and distance from the glacier) and
372 soil moisture were correlated with soil geochemistry. We used these variables to develop multiple linear regression and
373 random forest models to predict ion concentrations and geochemical gradients. The model results generally reflected the
374 measured geochemical variability across the region. Test datasets from the Shackleton and Darwin Glacier regions showed
375 that the random forest models typically outperformed the multiple linear regression models when correlating measured and
376 predicted values, especially for the Darwin region. Though most correlations did not exhibit a 1:1 relationship and had
377 varying slopes, the random forest models were able to adequately predict geochemical gradients, as demonstrated by
378 moderate to high R^2 values between measured and model predicted concentrations. As terrestrial Antarctic geochemical
379 databases expand and are included in the random forest model training dataset, we anticipate the model's predictive
380 capabilities will expand and improve as well. While these results are currently most applicable for Central Transantarctic
381 Mountain soils, similar techniques can be applied to other hyper-arid environments (e.g. Namib and Atacama Deserts, Mars)
382 to inform patterns of biodiversity and biogeography.



383 **Author Contributions**

384 The project was designed and funded by BJA, DHW, IDH, NF, and WBL. Fieldwork was conducted by BJA, DHW, IDH,
385 NF, and MAD. CBG, SAW, and MAD prepared and analyzed the samples for water-soluble ion and nutrient analyses. WAJ
386 prepared and analyzed the samples for ClO_4^- and ClO_3^- . MAD generated the scripts and performed the analyses for the IDW
387 interpolations, multiple linear regression, and random forest models. MAD wrote the article with contributions and edits
388 from all authors.

389 **Data Availability Statement**

390 The datasets generated for this study are included in the article or supplementary materials.

391 **Competing Interests**

392 The authors declare that they have no conflict of interest.

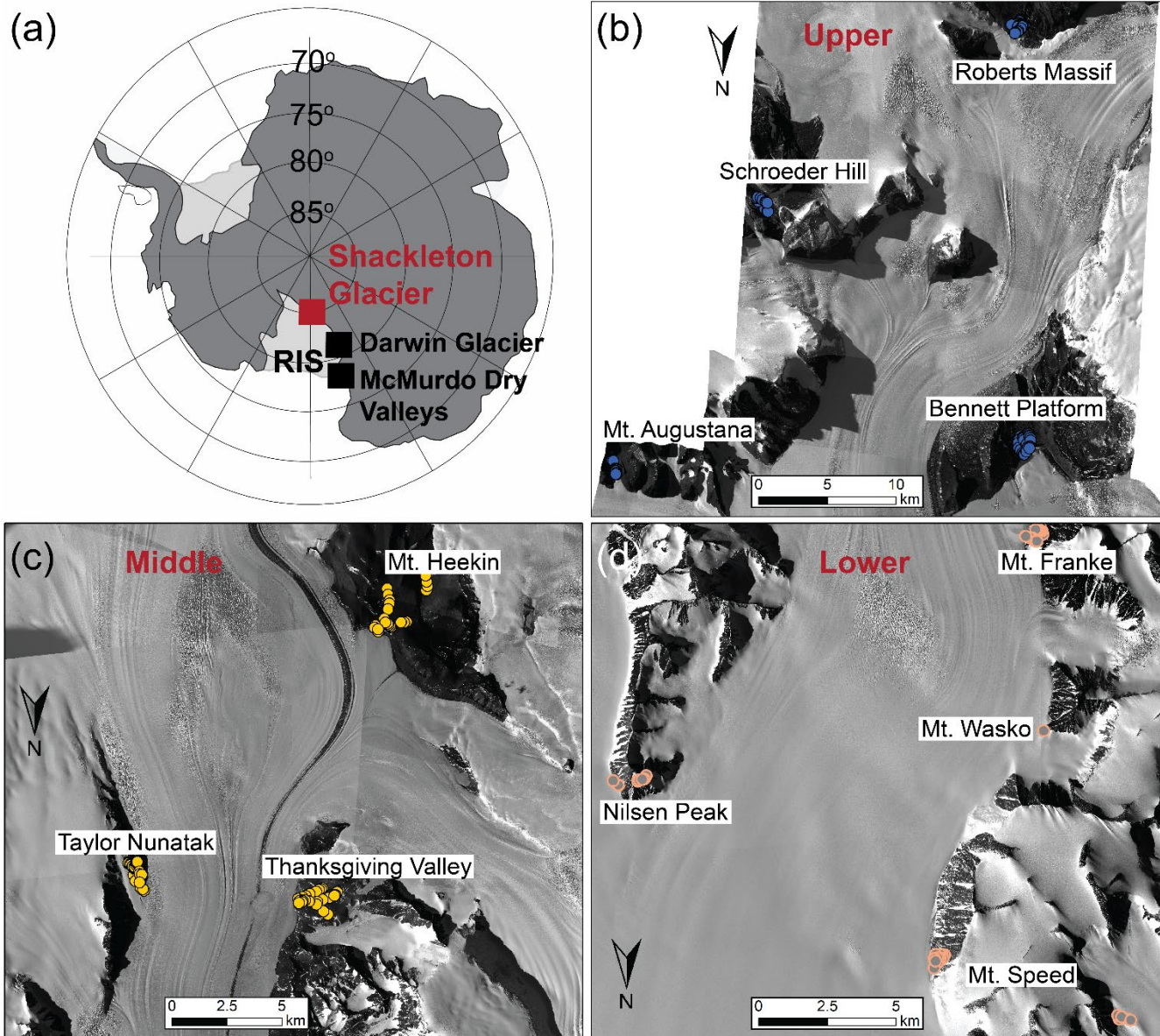
393 **Acknowledgments**

394 We thank the United States Antarctic Program (USAP), Antarctic Science Contractors (ASC), Petroleum Helicopters Inc.
395 (PHI), and Marci Shaver-Adams for logistical and field support. Additionally, we thank Daniel Gilbert for help with initial
396 laboratory analyses at The Ohio State University. This work was supported by NSF OPP grants 1341631 (WBL), 1341618
397 (DHW), 1341629 (NF), 1341736 (BJA), and NSF GRFP fellowship 60041697 (MAD). Geospatial support for this work
398 provided by the Polar Geospatial Center under NSF OPP grants 1043681 and 1559691.

399

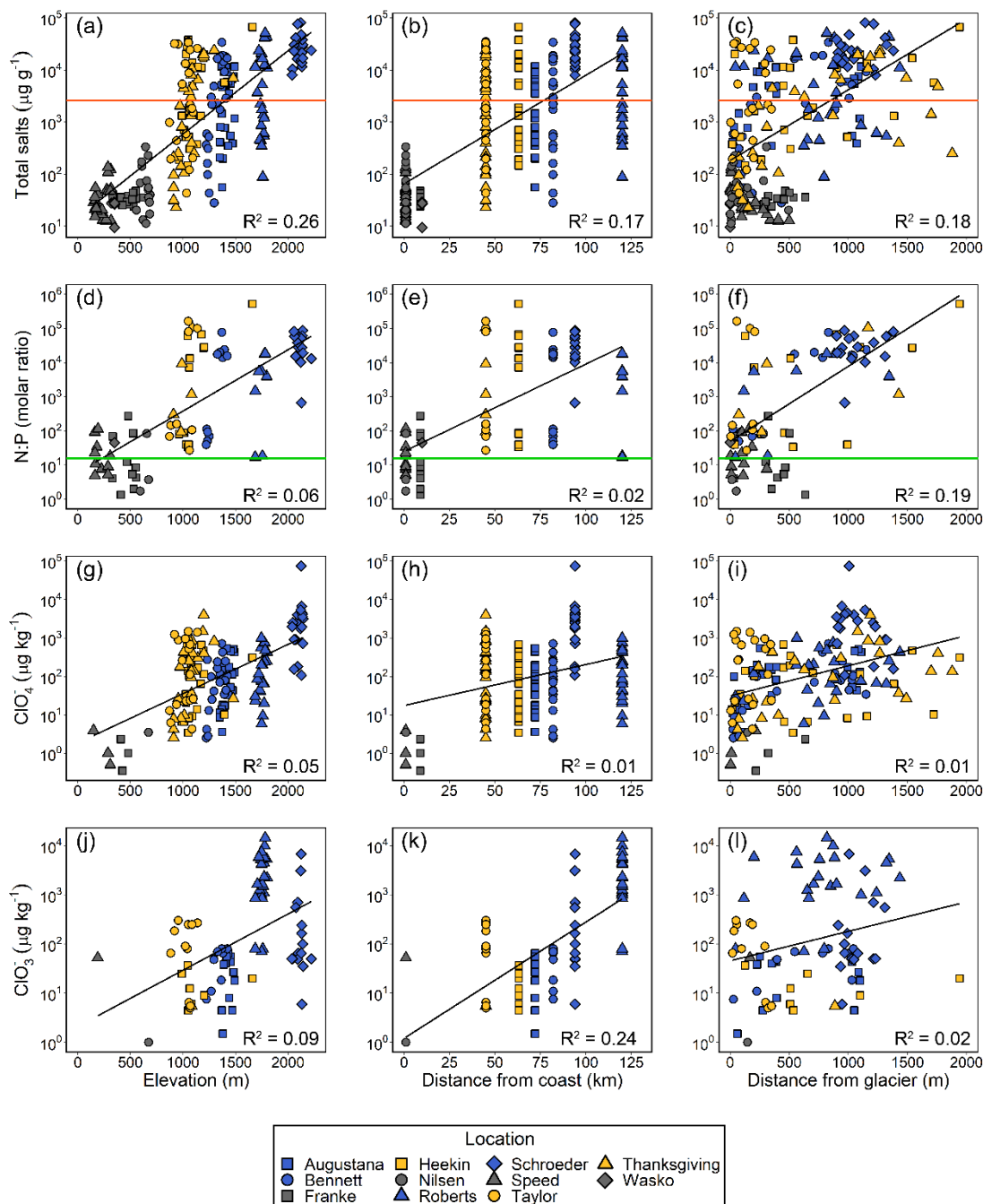


400 **Figures**



401

402 Figure 1. Samples were collected and analyzed from the exposed soils along the Shackleton Glacier, a major outlet glacier of
403 the EAIS (a), in three zones. The upper zone (b) was located at the head of Shackleton Glacier, the middle zone (c) was the
404 central portion, and the lower zone (d) was at the mouth of the glacier where it drains into the Ross Sea. Satellite images
405 were provided courtesy of the Polar Geospatial Center (PGC).
406



407

408

409

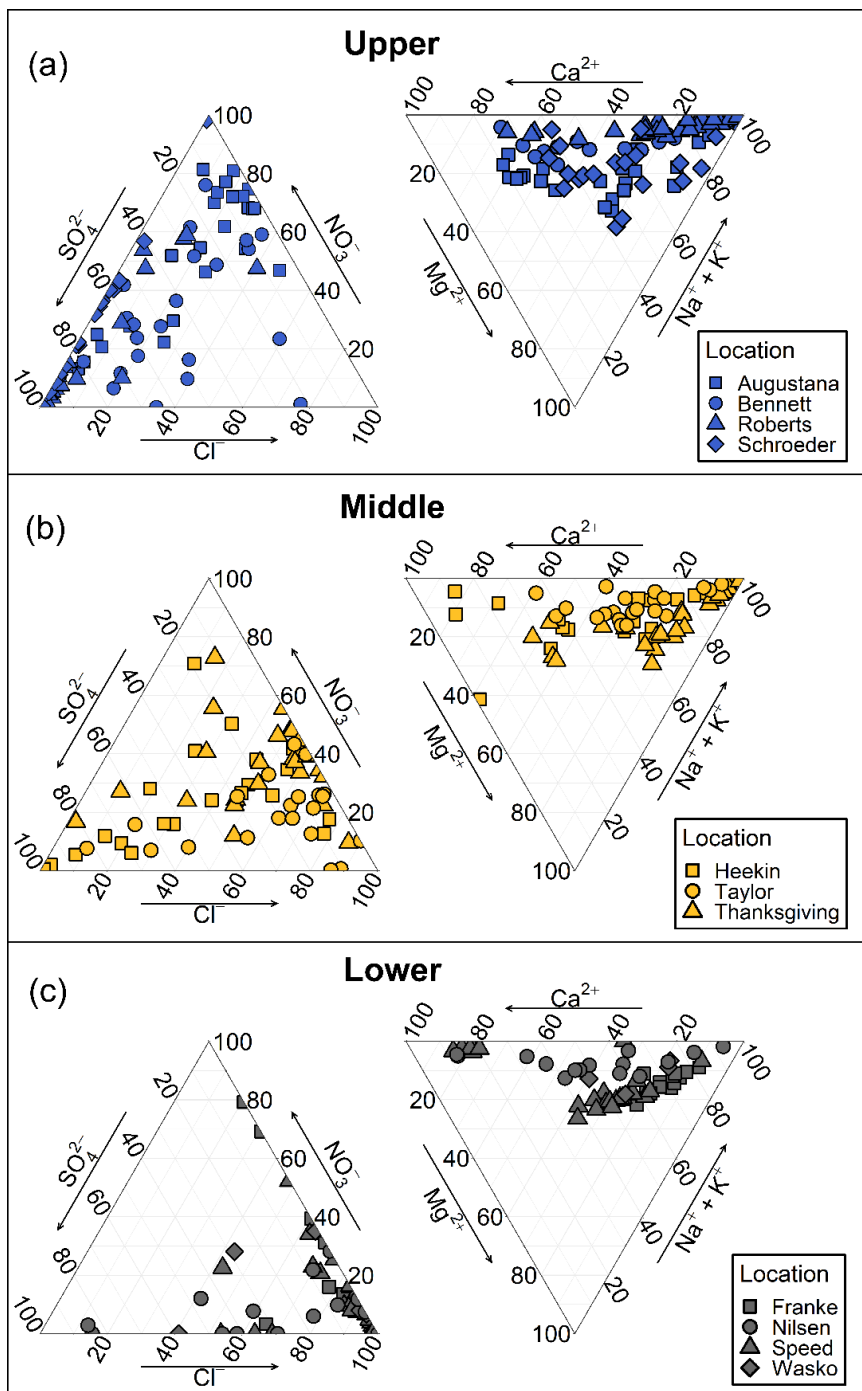
410

411

412

413

Figure 2. Total water-soluble salts, water-soluble N:P molar ratio, and ClO₄⁻ and ClO₃⁻ concentrations (log scale) were compared to elevation, distance from the coast, and distance from the glacier for samples from the three geographic zones. Linear regression lines are plotted and R² values are reported for each relationship. The horizontal orange line represents nematode salt tolerance of ~2,600 (Nkem et al., 2006) and the green line represents the Redfield ratio, N:P = 16 for phytoplankton in the ocean.

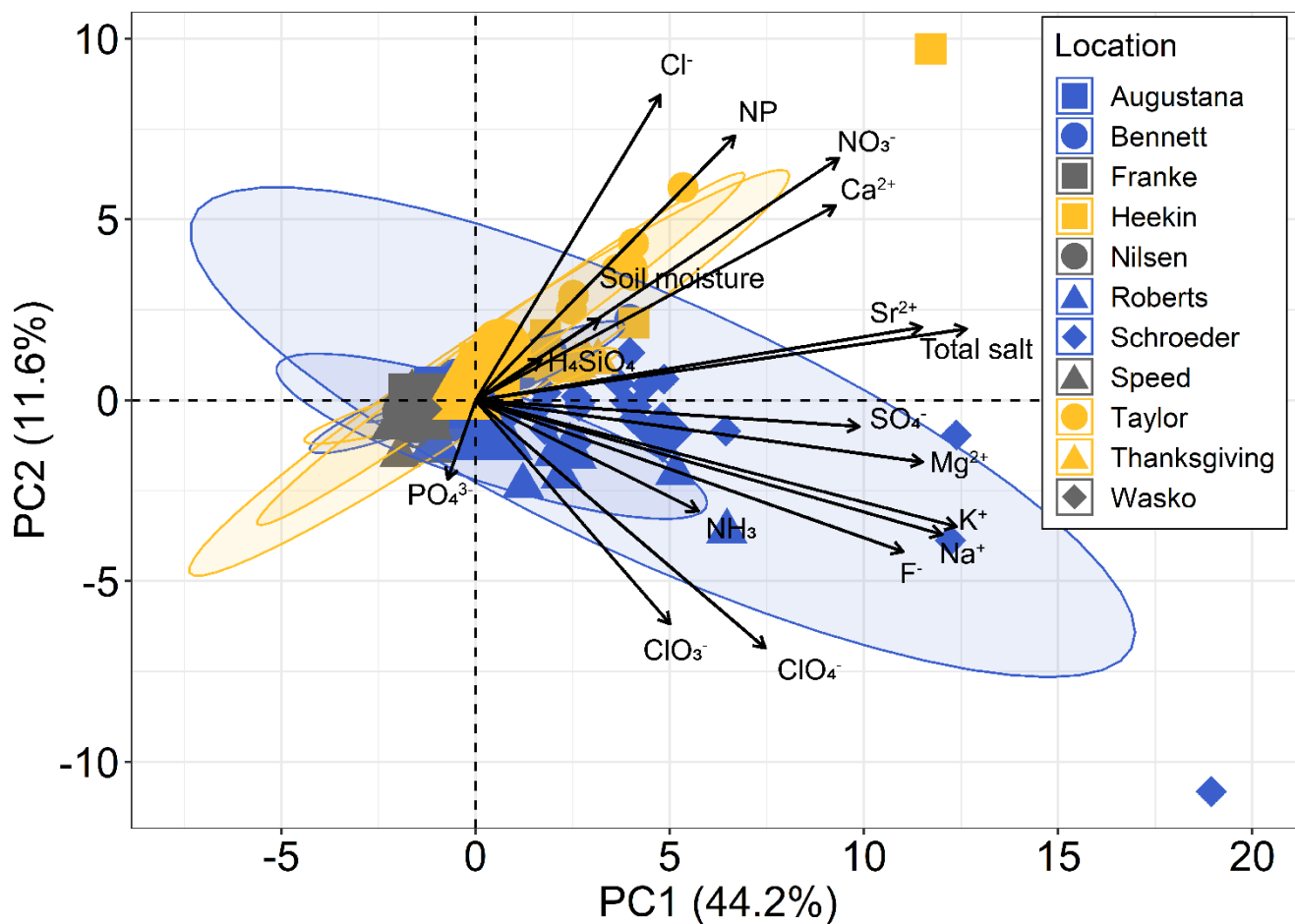


414

415

416

Figure 3. Anion and cation ternary diagrams for the three geographic zones.

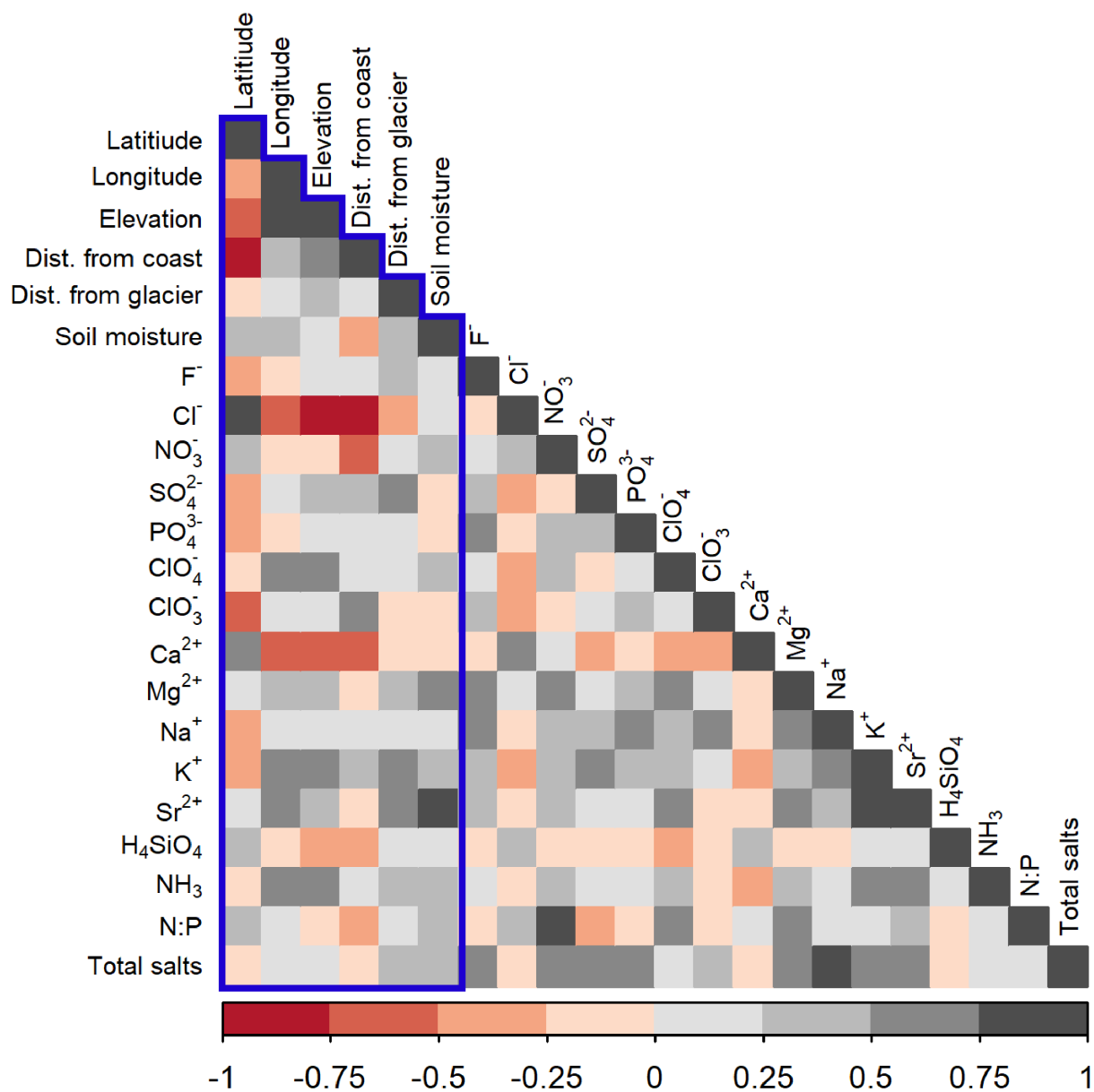


417

418 Figure 4. Principal component analysis (PCA) biplot generated in R using factoextra and built in software libraries with all
419 anions, cations, nutrients, and soil moisture for the three geographic zones. Principal component 1 and principal component 2
420 are plotted on the x and y axes, respectively. Shaded ellipses represent 95% confidence intervals.
421

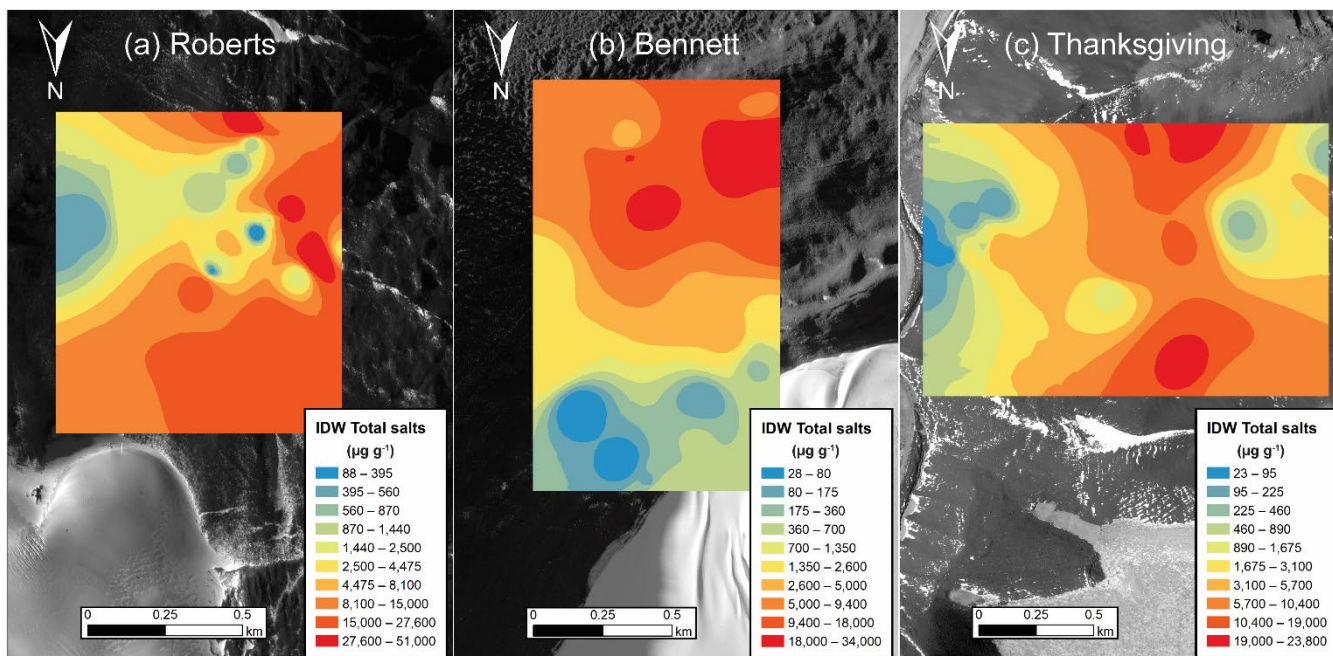


422



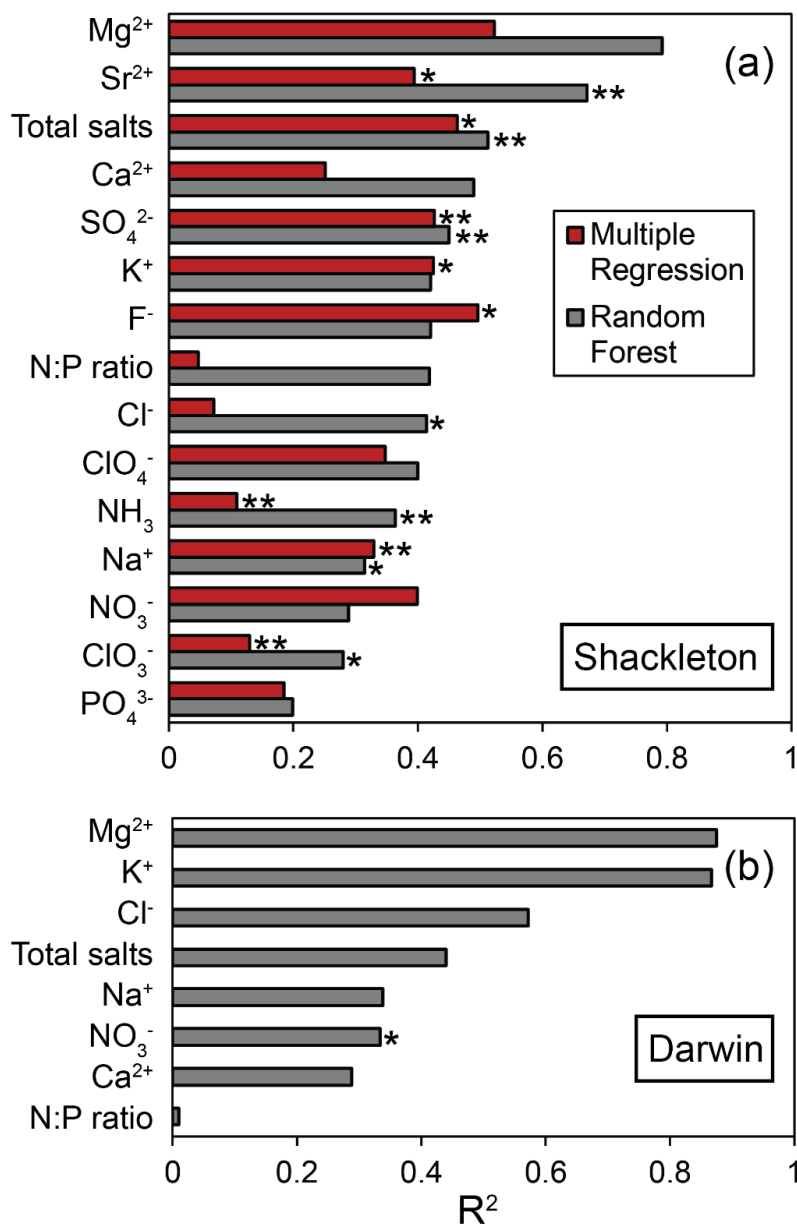
423

424 Figure 5. Spearman's rank correlation matrix generated in R using the corrplot library. The colors represent correlation
425 coefficients, indicating the strength and magnitude of the correlation. The blue box indicates the geographic variables and
426 soil moisture, which were variables used in the multiple linear regression and random forest models.
427



428

429 Figure 6. Inverse distance weighted (IDW) interpolations of total salt concentration for Roberts Massif (a), Bennett Platform
430 (b), and Thanksgiving Valley (c). The color scale represents the 10 natural breaks in the data. Interpolations were created and
431 mapped using the Geostatistical Analyst tool in ArcMap 10.3.
432



433

434 Figure 7. R² values for the multiple linear regression and random forest model predicted and measured values for the
 435 different analytes (Table 3). Test datasets include the Shackleton Glacier region (n = 31) and the Darwin Glacier region (n =
 436 10) (Magalhães et al., 2012). Analytes with slopes near 1, indicating good agreement between measured and predicted
 437 values, are indicated (* t < 0.5; ** t < 0.20).
 438



439
440

Table 1. Overview of geography, soil moisture, and water-soluble ions from the Shackleton Glacier region. The minimum values reported are those within the detection limits. Individual sample concentrations are detailed in Table S2.

	Max	Min	Mean	STD	CV
Elevation (m)	2,220	150	1,130	551	48
Distance from coast (km)	120	1	55	38	68
Distance from glacier (m)	1,940	1	519	472	90
Soil moisture (%)	12.3	0.1	2.1	2.1	102
F ⁻ (µg g ⁻¹)	120	0.39	8.87	11.78	133
Cl ⁻ (µg g ⁻¹)	13,600	1.59	615	1,780	289
NO ₃ ⁻ (µg g ⁻¹)	38,400	0.10	1,470	3,450	235
SO ₄ ²⁻ (µg g ⁻¹)	55,300	0.08	4,390	8,080	184
PO ₄ ³⁻ (µg kg ⁻¹)	4,200	76.09	381	560	147
ClO ₄ ⁻ (µg kg ⁻¹)	75,000	0.35	985	6,020	611
ClO ₃ ⁻ (µg kg ⁻¹)	14,500	1.00	1,170	2,500	214
Ca ²⁺ (µg g ⁻¹)	4,400	0.55	839	1,160	139
Mg ²⁺ (µg g ⁻¹)	6,280	0.12	293	705	240
Na ⁺ (µg g ⁻¹)	25,300	0.39	1,140	2,880	252
K ⁺ (µg g ⁻¹)	440	0.86	28.31	51.61	182
Sr ²⁺ (µg g ⁻¹)	46.61	0.01	8.63	10.31	119
H ₄ SiO ₄ (µg g ⁻¹)	60.78	1.14	21.78	11.03	50.67
NH ₃ (µg kg ⁻¹)	5,080	18.85	324	587	181
N:P ratio (molar)	526,000	0.29	23,600	62,700	266
Total salt (µg g ⁻¹)	80,500	9.46	7,932	13,300	167
STD, standard deviation; CV, coefficient of variation					

441



442

Table 2. Out-of-the-bag multiple linear regression and random forest model statistics generated in R.

	Multiple regression		Random forest		
	R ²	p-value	Variance explained (%)	Most important variable	Least important variable
F ⁻	0.27	<<0.001	36	Elevation	Distance from coast
Cl ⁻	0.05	0.082	20	Elevation	Distance from coast
NO ₃ ⁻	0.18	<<0.001	-4	Distance from glacier	Distance from coast
SO ₄ ²⁻	0.37	<<0.001	44	Elevation	Distance from coast
PO ₄ ³⁻	0.16	0.017	-7	Latitude	Distance from coast
ClO ₄ ⁻	0.1	0.010	-3	Elevation	Distance from coast
ClO ₃ ⁻	0.33	<<0.001	43	Latitude	Distance from glacier
Ca ²⁺	0.26	<<0.001	46	Soil moisture	Distance from coast
Mg ²⁺	0.29	<<0.001	22	Elevation	Distance from coast
Na ⁺	0.21	<<0.001	38	Elevation	Distance from coast
K ⁺	0.4	<<0.001	62	Elevation	Distance from coast
Sr ²⁺	0.55	<<0.001	62	Elevation	Distance from coast
NH ₃	0.29	<<0.001	54	Elevation	Distance from glacier
N:P	0.37	<<0.001	-3	Distance from glacier	Distance from coast
Total salts	0.37	<<0.001	45	Elevation	Distance from coast

443

444



445
 446
 447
 448
 449

Table 3. Multiple linear regression and random forest statistics between predicted and measured concentrations from the Shackleton and Darwin Glacier regions. R^2 and p-values are reported for the correlations between measured and predicted concentrations. Regression slopes and test statistic values (t) were calculated using the smatr library (Warton et al., 2012) in R to evaluate the null hypothesis (H_0) of slope = 1. Higher test statistic values (closer to one) indicate that we reject the null hypothesis.

Analyte	Multiple Linear Regression				Random Forest			
	R^2	p-value	Reg. slope	Test statistic (t) for H_0 slope = 1	R^2	p-value	Reg. slope	Test statistic (t) for H_0 slope = 1
Shackleton								
Mg ²⁺	0.52	<<0.001	0.52	-0.711	0.79	<<0.001	0.58	0.780
Sr ²⁺	0.39	<0.001	1.22	0.247*	0.67	<<0.001	0.91	-0.166**
Total salts	0.46	<<0.001	0.76	-0.343*	0.51	<<0.001	0.93	-0.107**
Ca ²⁺	0.25	0.004	0.42	-0.747	0.49	<<0.001	0.61	-0.586
SO ₄ ²⁻	0.43	<<0.001	1.07	0.093**	0.45	<<0.001	1.10	0.130**
K ⁺	0.42	<<0.001	1.54	0.504*	0.42	<<0.001	1.79	0.629
F ⁻	0.50	<<0.001	1.22	0.267*	0.42	<0.001	1.78	0.617
N:P ratio	0.05	0.241	0.59	-0.517	0.42	<<0.001	0.35	-0.867
Cl ⁻	0.07	0.144	0.28	-0.867	0.41	<<0.001	0.70	-0.424*
ClO ₄ ⁻	0.35	<0.001	2.01	0.685	0.40	<0.001	3.40	0.897
NH ₃	0.11	0.070	1.04	0.037**	0.36	<0.001	1.09	0.106**
Na ⁺	0.33	<0.001	0.91	-0.112**	0.31	0.001	1.54	0.473*
NO ₃ ⁻	0.40	<0.001	0.47	-0.725	0.29	0.002	0.56	-0.594
ClO ₃ ⁻	0.13	0.043	1.20	0.197**	0.28	0.002	0.71	-0.382*
PO ₄ ³⁻	0.18	0.016	0.50	-0.645	0.20	0.022	0.15	-0.967
Darwin								
Mg ²⁺	-	-	-	-	0.87	<<0.001	0.39	-0.948
K ⁺	-	-	-	-	0.87	<<0.001	0.49	-0.895
Cl ⁻	-	-	-	-	0.57	0.011	0.13	-0.984
Total salts	-	-	-	-	0.44	0.001	3.25	0.940
Na ⁺	-	-	-	-	0.34	0.078	0.23	-0.931
NO ₃ ⁻	-	-	-	-	0.33	0.080	0.65	-0.476*
Ca ²⁺	-	-	-	-	0.29	0.110	0.17	-0.961



N:P ratio	-	-	-	-	0.01	0.765	8.04	0.970
* $t < 0.5$; ** $t < 0.20$								

450



451 Table 4. Multiple linear regression and random forest model mean absolute error (MAE) and root mean square error (RMSE)

Analyte	Multiple Linear Regression		Random Forest	
	MAE	RMSE	MAE	RMSE
Shackleton				
Mg ²⁺	300	461	204	347
Sr ²⁺	3.74	4.96	1.83	2.90
Total salts	5,640	7,070	4,400	7,030
Ca ²⁺	797	1,100	554	912
SO ₄ ²⁻	3,310	3,890	2,200	3,780
K ⁺	15.86	21.16	13.48	25.61
F ⁻	3.14	4.19	3.13	6.31
N:P ratio	39,700	59,300	7,310	17,210
Cl ⁻	936	1,540	658	1,240
ClO ₄ ⁻	1,180	1,560	875	2,960
NH ₃	214	301	158	244
Na ⁺	883	1,170	918	1,730
NO ₃ ⁻	1,200	1,910	1,130	2,040
ClO ₃ ⁻	1,110	1,630	343	1,050
PO ₄ ³⁻	428	690	261	742
Darwin				
Mg ²⁺	6,300	6,320	302	475
K ⁺	1,060	1,060	13.33	15.84
Cl ⁻	206,000	206,000	2,140	3,330
Total salts	215,000	215,000	5,540	7,590
Na ⁺	8,330	8,530	1,500	2,600
NO ₃ ⁻	128,000	128,000	3,260	4,870
Ca ²⁺	70,300	70,300	1,410	2,070
N:P ratio	18,100,000	18,100,000	18,700	46,900
MAE, mean absolute error; RMSE, root mean squared error				

452



- 453 **References**
- 454 Antipov, E. A. and Pokryshevskaya, E. B.: Mass appraisal of residential apartments: An application of Random forest for
455 valuation and a CART-based approach for model diagnostics, *Expert Syst. Appl.*, 39(2), 1772–1778,
456 doi:10.1016/j.eswa.2011.08.077, 2012.
- 457 Ball, B. A., Adams, B. J., Barrett, J. E., Wall, D. H. and Virginia, R. A.: Soil biological responses to C, N and P fertilization
458 in a polar desert of Antarctica, *Soil Biol. Biochem.*, 122, 7–18, doi:10.1016/J.SOILBIO.2018.03.025, 2018.
- 459 Balter, A., Bromley, G., Balco, G., Thomas, H. and Jackson, M. S.: A 14.5 million-year record of East Antarctic Ice Sheet
460 fluctuations from the central Transantarctic Mountains, 2 constrained with cosmogenic, *Cryosph. Discuss.*, in review,
461 doi:10.5194/tc-2020-57, 2020.
- 462 Barrett, J. E., Virginia, R. A., Wall, D. H., Cary, S. C., Adams, B. J., Hacker, A. L. and Aislabie, J. M.: Co-variation in soil
463 biodiversity and biogeochemistry in northern and southern Victoria Land, Antarctica, *Antarct. Sci.*, 18(4), 535–548,
464 doi:10.1017/S0954102006000587, 2006.
- 465 Barrett, J. E., Virginia, R. A., Lyons, W. B., McKnight, D. M., Priscu, J. C., Doran, P. T., Fountain, A. G., Wall, D. H. and
466 Moorhead, D. L.: Biogeochemical stoichiometry of Antarctic Dry Valley ecosystems, *J. Geophys. Res.*, 112(G1), G01010,
467 doi:10.1029/2005JG000141, 2007.
- 468 Beet, C. R., Hogg, I. D., Collins, G. E., Cowan, D. A., Wall, D. H., Adams, B. J., Beet, C., Hogg, I., Collins, G., Cowan, D.
469 and Adams, B.: Genetic diversity among populations of Antarctic springtails (Collembola) within the Mackay Glacier
470 ecotone 1, *Genome*, 59, 762–770, doi:10.1139/gen-2015-0194, 2016.
- 471 Bennett, K. R., Hogg, I. D., Adams, B. J. and Hebert, P. D. N.: High levels of intraspecific genetic divergences revealed for
472 Antarctic springtails: evidence for small-scale isolation during Pleistocene glaciation, *Biol. J. Linn. Soc.*, 119(1), 166–178,
473 doi:10.1111/bij.12796, 2016.
- 474 Bockheim, J. G.: Landform and Soil Development in the McMurdo Dry Valleys, Antarctica: A Regional Synthesis, *Arctic,
475 Antarct. Alp. Res.*, 34(3), 308–317, doi:10.1080/15230430.2002.12003499, 2002.
- 476 Bockheim, J. G.: Functional diversity of soils along environmental gradients in the Ross Sea region, Antarctica, *Geoderma*,
477 144(1–2), 32–42, doi:10.1016/j.geoderma.2007.10.014, 2008.
- 478 Bottos, E. M., Laughlin, D. C., Herbold, C. W., Lee, C. K., McDonald, I. R. and Cary, S. C.: Abiotic factors influence
479 patterns of bacterial diversity and community composition in the Dry Valleys of Antarctica, *FEMS Microbiol. Ecol.*, 96(5),
480 42, doi:10.1093/femsec/fiaa042, 2020.
- 481 Breiman, L.: Random forests, *Mach. Learn.*, 45(1), 5–32, doi:10.1023/A:1010933404324, 2001.
- 482 Caruso, T., Hogg, I. D., Nielsen, U. N., Bottos, E. M., Lee, C. K., Hopkins, D. W., Cary, S. C., Barrett, J. E., Green, T. G.
483 A., Storey, B. C., Wall, D. H. and Adams, B. J.: Nematodes in a polar desert reveal the relative role of biotic interactions in
484 the coexistence of soil animals, *Commun. Biol.*, 2(63), doi:10.1038/s42003-018-0260-y, 2019.
- 485 Claridge, G. G. C. and Campbell, I. B.: Soils of the Shackleton glacier region, Queen Maud Range, Antarctica, *New Zeal. J.
486 Sci.*, 11(2), 171–218, 1968.
- 487 Collins, G. E., Hogg, I. D., Convey, P., Barnes, A. D. and McDonald, I. R.: Spatial and temporal scales matter when
488 assessing the species and genetic diversity of springtails (Collembola) in Antarctica, *Front. Ecol. Evol.*, 7, 76,
489 doi:10.3389/fevo.2019.00076, 2019.
- 490 Courtright, E. M., Wall, D. H. and Virginia, R. A.: Determining habitat suitability for soil invertebrates in an extreme
491 environment: The McMurdo Dry Valleys, Antarctica, *Antarct. Sci.*, 13(1), 9–17, doi:10.1017/S0954102001000037, 2001.
- 492 Davidson, A. D., Hamilton, M. J., Boyer, A. G., Brown, J. H. and Ceballos, G.: Multiple ecological pathways to extinction in



- 493 mammals., 2009.
- 494 Diaz, M. A., Adams, B. J., Welch, K. A., Welch, S. A., Opiyo, S. O., Khan, A. L., McKnight, D. M., Cary, S. C. and Lyons,
495 W. B.: Aeolian Material Transport and Its Role in Landscape Connectivity in the McMurdo Dry Valleys, Antarctica, J.
496 Geophys. Res. Earth Surf., 123(12), 3323–3337, doi:10.1029/2017JF004589, 2018.
- 497 Diaz, M. A., Corbett, L. B., Bierman, P. R., Adams, B. J., Wall, D. H., Hogg, I. D., Fierer, N. and Lyons, W. B.: Relative
498 terrestrial exposure ages inferred from meteoric ¹⁰Be and NO₃⁻ concentrations in soils along the Shackleton Glacier,
499 Antarctica, Earth Surf. Dyn., in review, doi:<https://doi.org/10.5194/esurf-2020-50>, 2020a.
- 500 Diaz, M. A., Li, J., Michalski, G., Darrach, T. H., Adams, B. J., Wall, D. H., Hogg, I. D., Fierer, N., Welch, S. A., Gardner, C.
501 B. and Lyons, W. B.: Stable isotopes of nitrate, sulfate, and carbonate in soils from the Transantarctic Mountains, Antarctica:
502 A record of atmospheric deposition and chemical weathering, Front. Earth Sci., 8(341), doi:10.3389/feart.2020.00341,
503 2020b.
- 504 Dragone, N. B., Diaz, M. A., Hogg, I., Lyons, W. B., Jackson, W. A., Wall, D. H., Adams, B. J. and Fierer, N.: Exploring the
505 boundaries of microbial habitability in soil, bioRxiv, doi:<https://doi.org/10.1101/2020.08.03.234583>, 2020.
- 506 Freckman, D. W. and Virginia, R. A.: Soil Biodiversity and Community Structure in the McMurdo Dry Valleys, Antarctica,
507 in Ecosystem dynamics in a polar desert; the McMurdo dry valleys, Antarctica, edited by J. C. Priscu, pp. 323–335,
508 American Geophysical Union (AGU), 1998.
- 509 Golledge, N. R. and Levy, R. H.: Geometry and dynamics of an East Antarctic Ice Sheet outlet glacier, under past and
510 present climates, J. Geophys. Res., 116(F3), F03025, doi:10.1029/2011JF002028, 2011.
- 511 Golledge, N. R., Fogwill, C. J., Mackintosh, A. N. and Buckley, K. M.: Dynamics of the last glacial maximum Antarctic ice-
512 sheet and its response to ocean forcing, Proc. Natl. Acad. Sci. U. S. A., 109(40), 16052–16056, doi:10.1073/pnas.1, 2012.
- 513 Golledge, N. R., Levy, R. H., McKay, R. M., Fogwill, C. J., White, D. A., Graham, A. G. C., Smith, J. A., Hillenbrand, C.
514 D., Licht, K. J., Denton, G. H., Ackert, R. P., Maas, S. M. and Hall, B. L.: Glaciology and geological signature of the Last
515 Glacial Maximum Antarctic ice sheet, Quat. Sci. Rev., 78, 225–247, doi:10.1016/j.quascirev.2013.08.011, 2013.
- 516 Heindel, R. C., Spickard, A. M. and Virginia, R. A.: Landscape-scale soil phosphorus variability in the McMurdo Dry
517 Valleys, Antarct. Sci., 29(3), 252–263, doi:10.1017/S0954102016000742, 2017.
- 518 Heung, B., Bulmer, C. E. and Schmidt, M. G.: Predictive soil parent material mapping at a regional-scale: A Random Forest
519 approach, Geoderma, 214–215, 141–154, doi:10.1016/j.geoderma.2013.09.016, 2014.
- 520 Hogg, I. D. and Wall, D. H.: Polar deserts, in Life at Extremes: Environments, Organisms and Strategies for Survival, edited
521 by E. M. Bell, pp. 176–195, CAB International., 2012.
- 522 Howat, I. M., Porter, C., Smith, B. E., Noh, M.-J. and Morin, P.: The Reference Elevation Model of Antarctica, Cryosph.,
523 13, 665–674, doi:10.5194/tc-13-665-2019, 2019.
- 524 Jackson, A., Davila, A. F., Böhlke, J. K., Sturchio, N. C., Sevanthi, R., Estrada, N., Brundrett, M., Lacelle, D., McKay, C. P.,
525 Poghosyan, A., Pollard, W. and Zacny, K.: Deposition, accumulation, and alteration of Cl⁻, NO₃⁻, ClO₄⁻ and ClO₃⁻ salts in
526 a hyper-arid polar environment: Mass balance and isotopic constraints, Geochim. Cosmochim. Acta, 182, 197–215,
527 doi:10.1016/j.gca.2016.03.012, 2016.
- 528 Jackson, M. S., Hall, B. L. and Denton, G. H.: Asynchronous behavior of the Antarctic Ice Sheet and local glaciers during
529 and since Termination 1, Salmon Valley, Antarctica, Earth Planet. Sci. Lett., 482, 396–406, doi:10.1016/j.epsl.2017.11.038,
530 2018.
- 531 Jackson, W. A., Davila, A. F., Estrada, N., Lyons, W. B., Coates, J. D. and Priscu, J. C.: Perchlorate and chlorate
532 biogeochemistry in ice-covered lakes of the McMurdo Dry Valleys, Antarctica, Geochim. Cosmochim. Acta, 98, 19–30,



- 533 doi:10.1016/j.gca.2012.09.014, 2012.
- 534 Jackson, W. A., Böhlke, J. K., Andraski, B. J., Fahlquist, L., Bexfield, L., Eckardt, F. D., Gates, J. B., Davila, A. F., McKay,
535 C. P., Rao, B., Sevanthi, R., Rajagopalan, S., Estrada, N., Sturchio, N., Hatzinger, P. B., Anderson, T. A., Orris, G.,
536 Betancourt, J., Stonestrom, D., Latorre, C., Li, Y. and Harvey, G. J.: Global patterns and environmental controls of
537 perchlorate and nitrate co-occurrence in arid and semi-arid environments, *Geochim. Cosmochim. Acta*, 164, 502–522,
538 doi:10.1016/J.GCA.2015.05.016, 2015.
- 539 Kassambara, A. and Mundt, F.: Package “factoextra,” *Extr. Vis. results Multivar. data Anal.*, 76 [online] Available from:
540 <https://github.com/kassambara/factoextra/issues> (Accessed 11 August 2020), 2017.
- 541 Kirkwood, C., Cave, M., Beamish, D., Grebby, S. and Ferreira, A.: A machine learning approach to geochemical mapping, *J.*
542 *Geochemical Explor.*, 167, 49–61, doi:10.1016/j.gexplo.2016.05.003, 2016.
- 543 LaPrade, K. E.: Climate, geomorphology, and glaciology of the Shackleton Glacier area, Queen Maud Mountains,
544 Transantarctic Mountains, Antarctica, *Antarct. Res. Ser.*, 36(9), 163–196, doi:10.1029/ar036p0163, 1984.
- 545 Levy, J. S., Fountain, A. G., Welch, K. A. and Lyons, W. B.: Hypersaline “wet patches” in Taylor Valley, Antarctica,
546 *Geophys. Res. Lett.*, 39(5), n/a-n/a, doi:10.1029/2012GL050898, 2012.
- 547 Lyons, W. B., Welch, K. A., Neumann, K., Toxey, J. K., McArthur, R., Williams, C., McKnight, D. M. and Moorhead, D.
548 L.: Geochemical linkages among glaciers, streams and lakes within the Taylor Valley, Antarctica, *Ecosyst. Dyn. a polar*
549 *desert; McMurdo dry Val. Antarct.*, 72, 77–92, doi:10.1029/AR072p0077, 1998.
- 550 Lyons, W. B., Deuerling, K., Welch, K. A., Welch, S. A., Michalski, G., Walters, W. W., Nielsen, U., Wall, D. H., Hogg, I.
551 and Adams, B. J.: The Soil Geochemistry in the Beardmore Glacier Region, Antarctica: Implications for Terrestrial
552 Ecosystem History, *Sci. Rep.*, 6, 26189, doi:10.1038/srep26189, 2016.
- 553 Magalhães, C., Stevens, M. I., Cary, S. C., Ball, B. A., Storey, B. C., Wall, D. H., Türk, R. and Ruprecht, U.: At Limits of
554 Life: Multidisciplinary Insights Reveal Environmental Constraints on Biotic Diversity in Continental Antarctica, edited by F.
555 de Bello, *PLoS One*, 7(9), e44578, doi:10.1371/journal.pone.0044578, 2012.
- 556 Nakada, M. and Lambeck, K.: The melting history of the late Pleistocene Antarctic ice sheet., 1988.
- 557 Nkem, J. N., Virginia, A. R. A., Barrett, A. J. E., Wall, D. H. and Li, A. G.: Salt tolerance and survival thresholds for two
558 species of Antarctic soil nematodes, *Polar Biol.*, 29, 643–651, doi:10.1007/s00300-005-0101-6, 2006.
- 559 Patel, J., Shah, S., Thakkar, P. and Kotecha, K.: Predicting stock market index using fusion of machine learning techniques,
560 *Expert Syst. Appl.*, 42(4), 2162–2172, doi:10.1016/j.eswa.2014.10.031, 2015.
- 561 Peters, J., De Baets, B., Verhoest, N. E. C., Samson, R., Degroeve, S., Becker, P. De and Huybrechts, W.: Random forests as
562 a tool for ecohydrological distribution modelling, *Ecol. Modell.*, 207(2–4), 304–318, doi:10.1016/j.ecolmodel.2007.05.011,
563 2007.
- 564 Prasad, A. M., Iverson, L. R. and Liaw, A.: Newer classification and regression tree techniques: Bagging and random forests
565 for ecological prediction, *Ecosystems*, 9(2), 181–199, doi:10.1007/s10021-005-0054-1, 2006.
- 566 R Core Team: R: A language and environment for statistical computing, 2020.
- 567 Stafoggia, M., Bellander, T., Bucci, S., Davoli, M., de Hoogh, K., de’ Donato, F., Gariazzo, C., Lyapustin, A., Michelozzi,
568 P., Renzi, M., Scortichini, M., Shtein, A., Viegi, G., Kloog, I. and Schwartz, J.: Estimation of daily PM10 and PM2.5
569 concentrations in Italy, 2013–2015, using a spatiotemporal land-use random-forest model, *Environ. Int.*, 124, 170–179,
570 doi:10.1016/j.envint.2019.01.016, 2019.
- 571 Stevens, M. I. and Hogg, I. D.: Long-term isolation and recent range expansion from glacial refugia revealed for the endemic



- 572 springtail *Gomphiocephalus hodgsoni* from Victoria Land, Antarctica, *Mol. Ecol.*, 12(9), 2357–2369, doi:10.1046/j.1365-
573 294X.2003.01907.x, 2003.
- 574 Stevens, M. I., Greenslade, P., Hogg, I. D. and Sunnucks, P.: Southern Hemisphere Springtails: Could Any Have Survived
575 Glaciation of Antarctica?, *Mol. Biol. Evol.*, 23(5), 874–882, doi:10.1093/molbev/msj073, 2006.
- 576 Talarico, F. M., McKay, R. M., Powell, R. D., Sandroni, S. and Naish, T.: Late Cenozoic oscillations of Antarctic ice sheets
577 revealed by provenance of basement clasts and grain detrital modes in ANDRILL core AND-1B, *Glob. Planet. Change*, 96–
578 97, 23–40, doi:10.1016/j.gloplacha.2009.12.002, 2012.
- 579 Tesoriero, A. J., Gronberg, J. A., Juckem, P. F., Miller, M. P. and Austin, B. P.: Predicting redox-sensitive contaminant
580 concentrations in groundwater using random forest classification, *Water Resour. Res.*, 53(8), 7316–7331,
581 doi:10.1002/2016WR020197, 2017.
- 582 Toner, J. D., Sletten, R. S. and Prentice, M. L.: Soluble salt accumulations in Taylor Valley, Antarctica: Implications for
583 paleolakes and Ross Sea Ice Sheet dynamics, *J. Geophys. Res. Earth Surf.*, 118(1), 198–215, doi:10.1029/2012JF002467,
584 2013.
- 585 Wall, D. H., Bardgett, R. D., Behan-Pelletier, V., Herrick, J. E., Jones, H., Ritz, K., Six, J., Strong, D. R. and van der Putten,
586 W. H., Eds.: *Soil Ecology and Ecosystem Services*, Oxford University Press, Oxford., 2012.
- 587 Warton, D. I., Duursma, R. A., Falster, D. S. and Taskinen, S.: smatr 3- an R package for estimation and inference about
588 allometric lines, *Methods Ecol. Evol.*, 3(2), 257–259, doi:10.1111/j.2041-210X.2011.00153.x, 2012.
- 589 Webster-Brown, J., Gall, M., Gibson, J., Wood, S. and Hawes, I.: The biogeochemistry of meltwater habitats in the Darwin
590 Glacier region (80°S), Victoria Land, Antarctica, *Antarct. Sci.*, 22(6), 646–661, doi:10.1017/S0954102010000787, 2010.
- 591 Wilson, D. J., Bertram, R. A., Needham, E. F., van de Flierdt, T., Welsh, K. J., McKay, R. M., Mazumder, A., Riesselman,
592 C. R., Jimenez-Espejo, F. J. and Escutia, C.: Ice loss from the East Antarctic Ice Sheet during late Pleistocene interglacials,
593 *Nature*, 561(7723), 383–386, doi:10.1038/s41586-018-0501-8, 2018.
- 594 Zeglin, L. H., Sinsabaugh, R. L., Barrett, J. E., Gooseff, M. N. and Takacs-Vesbach, C. D.: Landscape Distribution of
595 Microbial Activity in the McMurdo Dry Valleys: Linked Biotic Processes, Hydrology, and Geochemistry in a Cold Desert
596 Ecosystem, *Ecosystems*, 12(4), 562–573, doi:10.1007/s10021-009-9242-8, 2009.
- 597