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

Melisa A. Diaz[1,2†*], Christopher B. Gardner[1,2], Susan A. Welch[1,2], W. Andrew Jackson[3], Byron J. Adams[4], Diana H. Wall[5], Ian D. Hogg[6,7], Noah Fierer[8], W. Berry Lyons[1,2]

[1]School of Earth Sciences, The Ohio State University, Columbus, OH, USA
[2]Byrd Polar and Climate Research Center, The Ohio State University, Columbus, OH, USA
[3]Department of Civil, Environmental, & Construction Engineering, Texas Tech University, Lubbock, TX, USA
[4]Department of Biology, Evolutionary Ecology Laboratories, and Monte L. Bean Museum, Brigham Young University, Provo, UT, USA
[5]Department of Biology and School of Global Environmental Sustainability, Colorado State University, Fort Collins, CO, USA
[6]Canadian High Arctic Research Station, Polar Knowledge Canada, Cambridge Bay, Nunavut, Canada
[7]School of Science, University of Waikato, Hamilton, New Zealand
[8]Department of Ecology and Evolutionary Biology and Cooperative Institute for Research in Environmental Science, University of Colorado Boulder, Boulder, CO, USA

[†]Now at Departments of Geology and Geophysics and Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, MA, USA

Correspondence to: Melisa A. Diaz (mdiaz@whoi.edu)

Abstract. Previous studies have established links between biodiversity and soil geochemistry in the McMurdo Dry Valleys, Antarctica, where environmental gradients are important determinants of soil biodiversity. However, these gradients are not well established in the Central Transantarctic Mountains, which are thought to represent some of the least hospitable Antarctic soils. We analyzed 220 samples from 11 ice-free areas along the Shackleton Glacier (~85 °S), a major outlet glacier of the East Antarctic Ice Sheet. We established three zones of distinct geochemical gradients near the head of the glacier (upper), central (middle), and at the mouth (lower). The upper zone had the highest water-soluble salt concentrations with total salt concentrations exceeding 80,000 µg g^{-1}, while the lower zone had the lowest water-soluble N:P ratios, suggesting that, in addition to other parameters (such as proximity to water/ice), the lower zone likely represents the most favorable ecological habitats. Given the strong dependence of geochemistry on geographic parameters, we developed multiple linear regression and random forest models to predict soil geochemical trends given latitude, longitude, elevation, distance from the coast, distance from the glacier, and soil moisture (variables which can be inferred from remote measurements). Confidence in our random forest model predictions was moderately high, with R^2 values for total water-soluble salts, water-soluble N:P, ClO_4^-, and ClO_3^- of 0.81, 0.88, 0.78, and 0.74, respectively. These modeling results can be used to predict geochemical gradients and estimate salt concentrations for other Transantarctic Mountain soils, information that can ultimately be used to better predict distributions of soil biota in this remote region.

1. Introduction

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

In the McMurdo Dry Valleys (MDV), organic matter and salt concentrations influence soil communities, where soils with higher amounts of organic carbon, lower water-soluble N:P ratios, and lower total water-soluble salt concentrations generally harbor the greatest biomass and biodiversity (Barrett et al., 2006; Bottos et al., 2020; Caruso et al., 2019; Magalhães et al., 2012). These Antarctic ecosystems are relatively simple and are among few known soil systems where nematodes and microarthropods (Collembola, Acari) are at the top of the food chain (Freckman and Virginia, 1998; Hogg and Wall, 2012). Studies of soils in the MDV and Transantarctic Mountains (TAM) have been key to understanding ecosystem structure and function in extreme terrestrial environments (e.g. Caruso et al., 2019; Collins et al., 2019, 2020; Convey and McInnes, 2005; Freckman and Virginia, 1998; Hodgson et al., 2010).

Biological processes in Antarctic soils are largely dependent on the availability, duration, and proximity of soils to liquid water (Barrett et al., 2006). Due to the seasonality of thawing events, liquid water acts as a pulse to the ecosystem, providing water for organisms, but also wetting surface soils and dissolving soluble salts (Webster-Brown et al., 2010; Zeglin et al., 2009). Experiments of salt thresholds on Antarctic nematodes found that no individuals survived in highly saline soils over ~2,600 mg L^{-1} TDS (Nkem et al., 2006). Concentrations of soluble salts exist at these concentrations or higher at high elevation and inland locations in the TAM (Bockheim, 2008; Lyons et al., 2016). Additionally, studies on TAM soils have found that increased salt concentrations lead to a decrease in soil biodiversity in older soils compared to younger soils (Magalhães et al., 2012). Yet, despite these inhospitable conditions (e.g. high salt concentrations and glacial advance and retreat), some organisms are postulated to have found suitable refugia in TAM soils and persisted in isolation for millions of years and through glacial cycles (Beet et al., 2016; Collins et al., 2019, 2020; Stevens et al., 2006; Stevens and Hogg, 2003).

It is generally accepted that habitat suitability for invertebrate species in Antarctic soils is driven by a combination of geochemical, geographic, hydrologic, and geomorphic variables (Bottos et al., 2020; Courtright et al., 2001; Freckman and Virginia, 1998; Magalhães et al., 2012). Geographic variables, such as elevation, can be measured with advanced mapping tools and satellite imagery; however, surface exposure ages, soil geochemistry and nutrient content require extensive logistical support and resource allocation for sample collection and analysis. A better understanding of the

relationship between geographic variables and on-the-ground measurements is needed to aid in our ability to understand and predict habitat suitability for invertebrates throughout the TAM.

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

2. Study sites

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 East Antarctic Ice Sheet (EAIS) located to the west of the Beardmore Glacier and flows through the Queen Maud Mountains (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 coast towards the Polar Plateau. Long-term climate data are not yet available, but the Shackleton Glacier region is a polar desert regime, similar to the Beardmore Glacier region, with average annual temperatures well below freezing and little precipitation (LaPrade, 1984).

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

99   The soils contain a range of water-soluble salts derived primarily from atmospheric deposition and chemical
100   weathering (Claridge and Campbell, 1968; Diaz et al., 2020b). The major salts are typically nitrate and sulfate salts,
101   especially at higher elevations and further inland from the coast of the Ross Sea (Diaz et al., 2020b). The solubilities of the
102   salts vary, but nitrate salts are highly soluble and their occurrence at high elevation and inland locations suggests that those
103   soils have maintained persistent arid conditions.

104   **3. Methods**

105   3.1. Sample collection and preparation

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

115   Prior to geochemical analysis, the samples were dried at 50 °C for at least 72 hours with the loss in mass attributed
116   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
117   μm Nucleopore membrane filters (Diaz et al., 2018, 2020b; Nkem et al., 2006). Due to the low sediment to water ratio, this
118   leaching technique only dissolves the more water-soluble salts (Toner et al., 2013). These include salts with ClO_4^-, NO_3^-, Cl^-,
119   SO_4^{2-}, ClO_3^-, and $CO_3^{2-} + HCO_3^-$. Process blanks were generated and analyzed to account for any contamination from the
120   leaching process.

121   3.2. Analytical analysis of water-soluble anions, cations, and nutrients

122   The analytical techniques used here are similar to those reported by Diaz et al. (2020b). In brief, the analytes
123   included anions (F^-, Cl^-, Br^-, and SO_4^{2-}) which were measured on a Dionex ICS-2100 ion chromatograph, cations (K^+, Na^+,
124   Ca^{2+}, Mg^{2+}, and Sr^{2+}) which were measured on a PerkinElmer Optima 8300 Inductively Coupled Plasma-Optical Emission
125   Spectrometer (ICP-OES), and nutrients ($NO_3^- + NO_2^-$, PO_4^{3-}, H_4SiO_4, and NH_3) which were measured on a Skalar San++
126   Automated Wet Chemistry Analyzer at The Ohio State University. Perchlorate (ClO_4^-) and chlorate (ClO_3^-) were measured
127   using an ion chromatograph-tandem mass spectrometry technique (IC-MS/MS) at Texas Tech University (Jackson et al.,
128   2012, 2015). All analytes are reported as listed. Total water-soluble salt concentration was calculated as the sum of all
129   measured cations and anions. The precision of replicated check standards and samples was typically better than 10% for all
130   major anions, cations and nutrients, and better than 20% for perchlorate and chlorate. Accuracy was typically better than 5%

131 for all major anions, cations, and nutrients, as determined by the NIST 1643e external reference standard and the 2015 USGS

132 interlaboratory calibration standard (M-216), and better than 10% for perchlorate and chlorate, as determined by spike

133 recoveries. Precision and accuracy for individual analytes are located in Table S1. Detection limits for the analytes have been

134 previous reported (Diaz et al., 2018; Jackson et al., 2012).

135 3.3. Data interpolation and machine learning

136 Inverse distance weighted (IDW) interpolations were performed for Bennett Platform, Thanksgiving Valley, and

137 Roberts Massif using the Geostatistical Analyst tool in ArcMap 10.3. Since IDW is a deterministic method where unknown

138 values are predicted based on proximity to known values, we chose those three sites as they had the most defined transects

139 and relatively higher sample density. The interpolation parameters were constant with a power of 4, maximum neighbors of

140 15, minimum neighbors of 5, and 4 sectors, and a variable search radius. These parameters were chosen such that they

141 optimize for the lowest mean absolute error.

142 Multiple linear regressions were generated for all geochemical analytes, except H_4SiO_4 (total of 15 dependent

143 variables), with latitude, longitude, elevation, distance from the coast, distance from the glacier, and soil moisture as

144 independent variables using built-in functions in R 3.6.3 (R Core Team, 2020). Random forest regression models were

145 similarly generated using the randomForest library. The random forest model is a machine learning algorithm that utilizes

146 supervised learning algorithms to predict values given input predictor variables (Breiman, 2001). Multiple decision trees are

147 run in parallel with a randomized subset of predictor variables, and the aggregate result of each tree is used to generate a

148 predicted outcome. Since each tree is generated using a random sample and random predictor variables, the random forest

149 model is effective at minimizing overfitting and handling outliers (Breiman, 2001). For both models, all geochemical data

150 were log-transformed to ensure the data were normally distributed (verified using a Jarque-Bera normality test). Missing

151 values were input as NA.

152 Machine learning algorithms are widely used in variety of disciplines from finance (Patel et al., 2015) to ecology

153 (Davidson et al., 2009; Peters et al., 2007; Prasad et al., 2006), for both data prediction (regression) and classification.

154 Recently, these techniques have been used for Earth Science applications, including geologic mapping (Heung et al., 2014;

155 Kirkwood et al., 2016), air quality monitoring (Stafoggia et al., 2019), and water contaminant tracing (Tesoriero et al., 2017).

156 We developed a novel application of machine learning to predict concentrations and gradients of water-soluble salts in

157 Antarctic soils, given set geographic parameters, similar to the approaches developed for stock market and real estate

158 predictions (Antipov and Pokryshevskaya, 2012; Patel et al., 2015).

159 For our random forest models, any sparse missing values in Table S2 were estimated by averaging the geochemistry

160 of the samples collected immediately before and after in the same transect. Missing values due to concentrations below the

161 detection limit were input as NA. The new imputed dataset was split into a training set representing 86% of the data (n =

162 189, Table S3) and a testing set representing the remaining 14% (n = 31, Table S4), based on ideal model parameters

163 described by Breiman (2001). The training dataset was used to generate the random forest models for each analyte. Each of

164 the models were run with 2000 decision trees (ntree = 2000) to minimize the mean squared error. The number of random

165 variables used for each node split in the decision trees was set to the recommended regression default of variables/3 to

166 optimize the model randomness, which in our case was 2 (mtry = 2), following parameters described previously (Breiman,

167 2001). The scripts developed for both the multiple linear regression and random forest models are included in the

168 supplementary materials.

4. Results

169 **4. Results**

170 4.1. Geochemistry of upper, middle, and lower zones

171 The maximum, minimum, mean, standard deviation and coefficient of variation are reported in Table 1 for the

172 measured geographic and geochemical data. Concentrations of water-soluble ions span up to five orders of magnitude and

173 are variable across the region. Elevation, distance from the coast, distance from the glacier, and soil moisture are also

174 variable and span up to three orders of magnitude. The highest elevation samples (> 2,000 m.a.s.l.) were collected from

175 Schroeder Hill and the greatest soil moisture content is from Mt. Wasko at 12.3%, with a mean of 2.1% for all samples.

176 Shackleton Glacier region surface soils can be separated into three zones based on their water-soluble geochemistry:

177 an upper zone near the Polar Plateau, a middle zone near the center of the glacier, and a lower zone where the glacier flows

178 into the Ross Sea (Figs. 1; 2). The upper zone samples are characterized by the highest total water-soluble salt

179 concentrations, with the highest values greater than 80,000 µg g^{-1} at Schroeder Hill, while the lower zone samples have the

180 lowest total salt concentrations, with the lowest values near 10 µg g^{-1} at Mt. Wasko (Fig. 2a-c). The middle zone has

181 intermediate values. Water-soluble N:P molar ratios generally follow a similar trend (Fig. 2d-f). The lowest N:P ratios are in

182 the lower zone soils, while the middle and upper zones have more variable values. Concentrations of ClO_4^- and ClO_3^- follow

183 similar trends as the total salts, with less distinction between middle and upper zones, though most concentrations in the

184 lower zone are below the detection limit (Fig. 2g-l; Table S2).

185 Observed trends between the zones appear to be driven, at least partially, by geography. Regressions of total water-

186 soluble salt concentration, water-soluble N:P ratio, and ClO_3^- concentration with elevation, distance from the coast, and

187 distance from the glacier are all positive (Fig. 2). The strongest relationships are between total salts and elevation, and N:P

188 ratio and elevation, with R^2 values of 0.59 and 0.52, respectively, and p-values < 0.001 with a Bonferroni Correction, which

189 was applied to minimize the familywise type 1 error rate associated with multiple comparisons (Fig. 2a;2d). The weakest

190 relationships are between ClO_4^- and distance from the coast , and ClO_3^- and distance from the glacier, with R^2 values of 0.11

191 and 0.06, respectively (Fig. 2h; 2i). Distance from the glacier varies widely between individual zones with frequent overlaps,

192 but there appears to be a moderate relationship with N:P ratio and total salts (Fig. 2c; 2f). Overall, total salt concentration has

193 the strongest relationship with geography and ClO_4^- has the weakest relationships.

194   Ternary diagrams highlight the specific geochemical gradients within and between the zones. The anion ternary
195 diagram only includes SO_4^{2-}, NO_3^-, and Cl^-, which are the major water-soluble salts in the region (Claridge and Campbell,
196 1968; Diaz et al., 2020b). Though carbonate and bicarbonate salts have been identified in both lacustrine sediments and soils
197 in Antarctica, previously measured concentrations in the Shackleton Glacier region were low, ranging from 0.07 to 2.5%,
198 and bicarbonate salts were not identified in the highest elevation and furthest inland soils (Claridge and Campbell, 1968;
199 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
200 total anion budget in some Schroeder Hill and Roberts Massif samples, though other locations are dominated by NO_3^- (Fig.
201 3). The anions are more evenly distributed in the middle zone, though the majority of samples are most abundant in NO_3^- and
202 Cl^-. The lower zone has much lower SO_4^{2-} fractions than the upper zone and the dominant anion is generally Cl^-. The cation
203 distribution is very similar for all three zones (Fig. 3). $Na^+ + K^+$ is the most abundant cation pair representing over 90% of
204 the total cations for many upper and middle zone samples, while Ca^{2+} is the second most abundant. In general, Mg^{2+} is the
205 least abundant cation across all sampling locations.

206 4.2. Statistical geochemical variability

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

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

227 4.3. Spatial interpolation and machine learning model performance

228   The total salt concentrations of individual samples at Bennett Platform produce the most defined interpolation

229 gradient from the glacier front to further inland compared to Roberts Massif and Thanksgiving Valley (Fig. 6). Bennett

230 Platform also has the smoothest salt concentration contours suggesting that the interpolation model is the strongest and most

231 robust at this location. The second strongest interpolation is Thanksgiving Valley. Contrary to the measurements at Bennett

232 Platform, Thanksgiving Valley has the highest salt concentrations in the center of the valley, with lower concentrations to

233 both the east and west. The lowest concentration contours are closest to the glacier for both Bennett Platform and

234 Thanksgiving Valley, which is likely related to glacial history since the soils near the glacier are relatively younger than

235 those further inland based on meteoric ^{10}Be data (Diaz et al., 2020a). The interpolation from Roberts Massif does not have a

236 distinguishable spatial trend.

237   The multiple linear regression and random forest models vary in their strength for the individual analytes. The

238 highest R^2 value from the linear regression is 0.65 for Na^+, while total water-soluble salts, water-soluble N:P ratio, ClO_4^-, and

239 ClO_3^- have values of 0.61, 0.60, 0.44, and 0.55, respectively (Table 2). The lowest R^2 value is for PO_4^{3-} at 0.17. The p-values

240 for all analytes are <<0.001, even with a Bonferroni Correction. The highest out-of-the-box explained variance values from

241 the random forest models are for total salts and ClO_3^- at 75% and 63%, respectively. The lowest explained variance is for

242 Sr^{2+} at 37%. Values N:P ratio and ClO_4^- are 52% and 48%, respectively. We also evaluated the most important and least

243 important variables in the random forest models based on node purity. The most important variable for the majority of

244 analytes is elevation, while distance from the glacier is most important for N:P ratio and latitude for ClO_3^- (Table 2). The

245 least important variables are distance from the coast and latitude for every analyte, except ClO_3^-, for which distance from the

246 glacier is least important.

247 **5. Discussion**

248 5.1. Implications for ecological habitat suitability

249   By establishing geochemical zones for the Shackleton Glacier region, we can better understand the relationship

250 between geochemistry and geography, and ultimately biogeography. As stated in the introduction, we focused particularly on

251 total water-soluble salt concentrations, water-soluble N:P ratios, and ClO_4^- and ClO_3^- concentrations.

252 5.1.1. Elevation and moisture controls on total water-soluble salt gradients

253   The elevational trends of total salt concentrations at the Shackleton Glacier are similar to those previously described

254 in the TAM, where soils from higher elevation sites typically have higher salt concentrations (Bottos et al., 2020; Lyons et

255 al., 2016; Magalhães et al., 2012). Our results are also consistent with those from Scarrow et al. (2014) who found that salt

256 concentrations typically decreased with distance from the glacier in the Beardmore and Lennox King Glacier regions. Our

257 total water-soluble salt interpolation maps highlight the spatial variability in Shackleton Glacier region soils (Fig. 6). The

258 most spatially variable location is Robert Massif, which does not appear to follow local elevational, latitudinal, and/or

259 distance inland gradients. This heterogeneity is not necessarily due to currently active soil leaching, as the soil moisture

260 values are not drastically different between the samples (Table S2). Though the variability in cation concentrations is likely

261 due to weathering of tills, scree, and bedrock (Claridge and Campbell, 1968), recent work on the isotopic composition of

262 water-soluble nitrate and sulfate, the major anions in the upper zone, suggests a common, atmospheric source (Diaz et al.,

263 2020b).

264 We argue that the heterogeneity in the total salt concentrations at Roberts Massif (Figs. 2; 6) is probably related to

265 different and complex wetting history, where seasonal snow patch melt may pool in local depressions, transporting water-

266 soluble salts from slightly higher elevations and/or from saline wet-patches (Levy et al., 2012). This is demonstrated on a

267 larger scale at Thanksgiving Valley, a glacially carved valley, where the higher concentrations of salts in the center of the

268 valley are likely due to the transport of salts from nearby higher elevation slopes during melting events. This is further

269 evidenced by the presence of two small, closed-basin ponds in the center of the valley, which likely formed from glacial melt

270 and may have been larger in size in the recent past (Diaz et al., 2019). Similarly, streams and meltwater tracks in the MDV

271 leach soils and carry salts into closed basin, brackish to hyper-saline lakes, where salts are cryoconcentrated over time

272 (Lyons et al., 1998). Our results suggest that elevation and wetting history are important contributors to total salt gradients in

273 the Shackleton Glacier region, as they influence the accumulation of salts and subsequent leaching from soils.

274 5.1.2. Influence of glacial history on water-soluble N:P ratios

275 Stoichiometric dependencies have been identified for Antarctic terrestrial organisms, where nutrient concentrations,

276 in addition to soil aridity, limit ecosystem development (Nkem et al., 2006). Since nitrate is primarily derived from

277 atmospheric deposition and phosphorus is primarily liberated from minerals by chemical weathering in the CTAM, many

278 inland and higher elevation soils have accumulated high concentrations of NO_3^-, resulting in stoichiometric imbalance with

279 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

280 MDV, younger and coastal soils at lower elevations in the Shackleton Glacier region have the lowest water-soluble N:P

281 ratios, driven by relatively low concentrations of NO_3^- and high concentrations of PO_4^{3-} due to an increase in moisture

282 content and chemical weathering (Heindel et al., 2017) (Fig. 2; 4). It is not surprising that life was conspicuous in these soils,

283 with thick lichen growth on several rocks and the presence of both Collembola and mites at Mt. Speed and Mt. Wasko (Fig.

284 S1). However, despite overall elevational and latitudinal gradients, some inland locations in the middle and upper zones have

285 water-soluble N:P ratios near those from the lower zone (Fig. 2).

286 The interpolation model from Bennett Platform shows that some locations near the glacier have lower total water-

287 soluble salt concentrations (Fig. 6), similar to soils surveyed in the MDV (Bockheim, 2002). However, the samples near the

288 glacier at Bennett Platform not only have lower total salt concentrations, they also have lower N:P ratios than samples

289 collected further inland. This is also the case for the middle zone locations (Fig. 2f). We argue this is due to differences in

290 glacial history between the locations. Our previous work showed that soils near the glacier are younger than soils further

291 inland in the Shackleton Glacier region (Diaz et al., 2020a). These soils are shielded from nitrate accumulation during glacial

292 periods, and the recently exposed rocks likely serve as fresh mineral weathering material for PO_4^{3-} mobilization (Heindel et

293 al., 2017). Recently exposed and relatively nutrient rich soils might be important refugia for invertebrates. Previous

294 hypotheses have suggested that organisms may have persisted at higher elevations during glacial periods (Bennett et al.,

295 2016; Stevens and Hogg, 2003). However, abiotic gradients in the Beardmore Glacier region suggest that higher elevation

296 soils have salt concentrations that would classify them as unsuitable habitats (Lyons et al., 2016). If few organisms survived

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

298 glaciation.

299 5.1.3. High and variable ClO_4^- and ClO_3^- concentrations

300 Our ClO_4^- and ClO_3^- concentrations include the highest measured in Antarctica to date and are comparable to

301 concentrations from the Atacama and Mojave Deserts (Jackson et al., 2015). Though not a strong correlation, the highest

302 elevation samples (upper zone) have the highest ClO_4^- and ClO_3^- concentrations (Fig. 2g; 2j). Similar to NO_3^-, ClO_4^- and

303 ClO_3^- are derived from atmospheric deposition and because of their high solubilities, their accumulations are related to

304 wetting and glacial histories (Jackson et al., 2016, 2015). Therefore, soils which have been exposed for long periods of time

305 and have not experienced snow or ice melt, such as those from Schroeder Hill and Roberts Massif, are able to accumulate

306 high concentrations of ClO_4^- and ClO_3^-. Interestingly, our ClO_4^- concentrations are lower (maximum of ~1.9 g L^{-1}) than the

307 highest recorded tolerance (1.1M (~130 g L^{-1}) $NaClO_4$) for the extremotolerant bacteria *Planococcus halocryophilus*, yet a

308 recent study shows no detectable biomass for Schroeder Hill samples (Dragone et al., 2020). (Per)chlorates are strong

309 oxidizers and are well established as toxic, thus the concentrations of ClO_4^- and ClO_3^- might be additional, crucial indicators

310 of habitat suitability. However, the concentrations are highly heterogenous across our sampled locations (Fig. 2k-l), and

311 unlike ClO_3^-, neither the multiple linear regression nor random forest models were able to adequately capture the variability

312 in ClO_4^- concentrations (Table 2).

313 5.2. Machine learning as a tool to predict soil geochemical trends

314 We sought to evaluate our multiple linear regression and random forest models using a testing dataset from the

315 Shackleton Glacier region (n = 31) and a second dataset from the Darwin Mountains (~80°S) (n = 10) (Magalhães et al.,

316 2012). Few published/available TAM dataset include sample GPS coordinates, soil moisture, and water-soluble ion

317 geochemistry. As stated in Section 3.3, the Shackleton Glacier region test data were not included in the random forest model

318 generation so we could evaluate our models with an independent dataset. For the Darwin dataset, distance from the glacier,

319 distance from the coast, and elevation were determined using the Reference Elevation Model of Antarctica (REMA), while

320 location, soil moisture and geochemistry were retrieved from the literature (Howat et al., 2019; Magalhães et al., 2012). We

321 evaluated all 15 analytes from the original models with the Shackleton dataset and, due to a lack of data, only evaluated 7

322 analytes from the Darwin soils (Figure 7).

323   Both the multiple linear regression and random forest model outputs are moderately well-correlated for the

324   Shackleton dataset, as determined by Pearson correlations between the measured and predicted values (Fig. 7a; Table 3). The

325   random forest models outperform the linear regression models for nearly every analyte, with the exception of Sr^{2+}, NH_3, and

326   PO_4^{3-}, and nearly all p-values are <0.001. For Cl^-, in particular, the random forest model significantly outperforms the

327   multiple linear regression model, with R^2 values of 0.67 and 0.16, respectively. N:P molar ratio is the most accurately

328   predicted analyte, with R^2 values of 0.88 and 0.59 for the random forest and linear regression models, respectively. However,

329   the highest R^2 value for the multiple linear regression model is for Na^+ at 0.64 (Table 3). In terms of our analytes of interest

330   regarding habitat suitability, total salts have the second strongest correlation (following N:P ratio) with the random forest

331   model ($R^2 = 0.81$), followed by ClO_4^- ($R^2 = 0.78$), and ClO_3^- ($R^2 = 0.74$). Mean absolute error (MAE) and root mean squared

332   error (RMSE) values indicate that the random forest models also have a smaller error compared to the multiple linear

333   regression models (Table 4). MAE values are lower than RMSE values for both models, indicating the strong influence of

334   outliers in the testing dataset. This is unsurprising as the standard deviation and coefficient of variation values for the entire

335   dataset are relatively large for all analytes. Additionally, the outliers are likely one reason why the random forest models are

336   stronger than the multiple linear regression models.

337   Similar to the model performance in the Shackleton Glacier region, the water-soluble ion predictions for the Darwin

338   Glacier region are more strongly correlated with measured values in the random forest models compared to the multiple

339   linear regressions (Fig. 7b). In fact, the linear regression models fail for nearly all the Darwin samples and most

340   concentration outputs are negative, which is likely due to overfitting during model generation. Here, Ca^{2+} and K^+ are

341   exceptions and the multiple linear regression models outperform the random forest models in both cases. MAE and RSME

342   values for both models are higher than those for the Shackleton dataset (Table 4). On the other hand, the random forest

343   models perform particularly well for some analytes. Though a small sample size, the R^2 values for N:P molar ratio and Ca^{2+}

344   are 0.68 and 0.66, respectively, with p-values <<0.001. Total salts is moderately correlated ($R^2 = 0.47$). It is unclear why

345   some analytes, such as N:P molar ratio, are the most accurately predicted, though we suspect that this is due to 1) weathering

346   trends of local lithology across the TAM since chemical weathering is probably the major source of these ions, and 2)

347   deposition and accumulation of atmospherically-derived ions at higher elevations (Diaz et al., 2020b).

348   It should be noted that the R^2 values simply measure the strength of the correlations between the measured and

349   predicted values. We performed slope tests by fitting bivariate lines using the standardized major axis (SMA) to further

350   understand the relationship between the two values using the smatr library in R (Warton et al., 2012). For this test, we

351   specifically evaluated the null hypothesis (H_0) where slope = 1, which would indicate whether an ideal, direct 1:1

352   relationship exists between the measured and predicted values. Test statistic values (t) were used to measure the sample

353   correlation between the residuals and fitted values (Warton et al., 2012). Test statistic values near 1 indicate that we reject

354   the null hypothesis. In other words, higher absolute test statistic values indicate a slope other than 1. Of the 15 analytes in the

355   Shackleton dataset, 5 analytes have slopes near 1 for the multiple linear regression models and 11 for the random forest, as

356 indicated by test statistic values less than 0.5. For the Darwin, no analytes have test statistic values less than 0.5 (Fig. 7;
357 Table 3).

358   These data indicate that while some analytes have high correlations between measured and predicted values, the
359 models perform best with the Shackleton Glacier region soils. Additionally, though the relationship may not be 1:1, the
360 random forest models are effective at predicting the measured geochemical gradients. For example, similar to our data, the
361 Darwin Glacier samples generally have greater water-soluble N:P ratios and total water-soluble salt concentrations further
362 from the glacier and at higher elevations (Magalhães et al., 2012), a trend that is reflected by our model results despite offset
363 values. Additionally, corrections for the offset of the model from a slope = 1 (i.e. multiplying the model output value by the
364 regression slope) can be made to better estimate specific concentrations, though the difference between modeled and
365 measured values can still be up to 2x greater. Our sample size for building the multiple linear regression and random forest
366 models is small. We anticipate that, as more data are collected throughout the CTAM, these data can be added to the model
367 training dataset, expanding our prediction capabilities and increasing model reliability.

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

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

384   Five geographic variables (latitude, longitude, elevation, distance from the coast, and distance from the glacier) and
385 soil moisture were correlated with soil geochemistry. We used these variables to develop multiple linear regression and
386 random forest models to predict ion concentrations and geochemical gradients. The model results generally reflected the
387 measured geochemical variability across the region. Test datasets from the Shackleton and Darwin Glacier regions showed

388   that the random forest models typically outperformed the multiple linear regression models when correlating measured and

389   predicted values, especially for the Darwin region. Though most correlations did not exhibit a 1:1 relationship and had

390   varying slopes, the random forest models were able to adequately predict geochemical gradients, as demonstrated by

391   moderate to high R^2 values between measured and model predicted concentrations. As terrestrial Antarctic geochemical

392   databases expand and are included in the random forest model training dataset, we anticipate the model's predictive

393   capabilities will expand and improve as well. While these results are currently most applicable for Central Transantarctic

394   Mountain soils, similar techniques can be applied to other hyper-arid environments (e.g. Namib and Atacama Deserts, Mars)

395   to inform patterns of biodiversity and biogeography.

Author Contributions

396
397   The project was designed and funded by BJA, DHW, IDH, NF, and WBL. Fieldwork was conducted by BJA, DHW, IDH,

398   NF, and MAD. CBG, SAW, and MAD prepared and analyzed the samples for water-soluble ion and nutrient analyses. WAJ

399   prepared and analyzed the samples for ClO_4^- and ClO_3^-. MAD generated the scripts and performed the analyses for the IDW

400   interpolations, multiple linear regression, and random forest models. MAD wrote the article with contributions and edits

401   from all authors.

Data Availability Statement

402
403   The datasets generated for this study are included in the article or supplementary materials.

Competing Interests

404
405   The authors declare that they have no conflict of interest.

Acknowledgments

406
407   We thank the United States Antarctic Program (USAP), Antarctic Science Contractors (ASC), Petroleum Helicopters Inc.

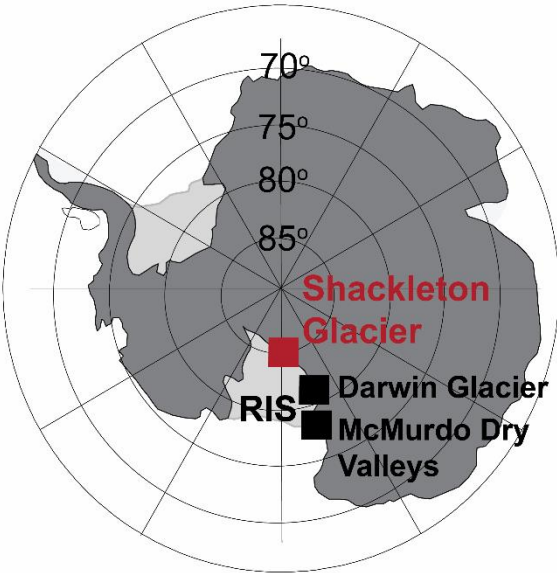
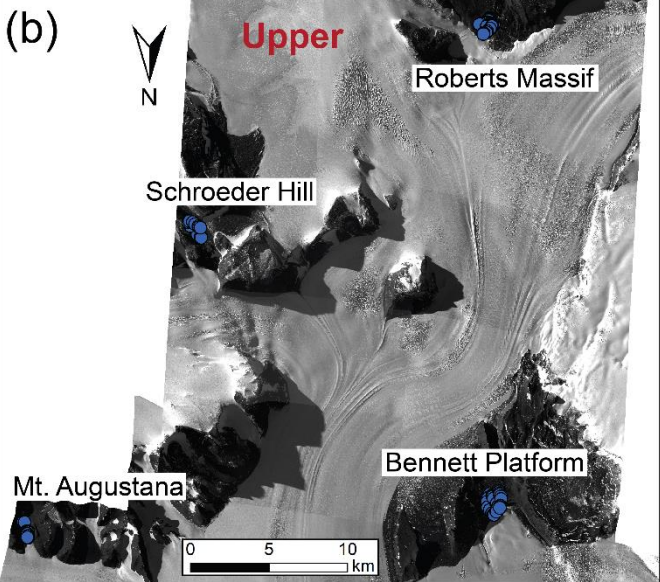
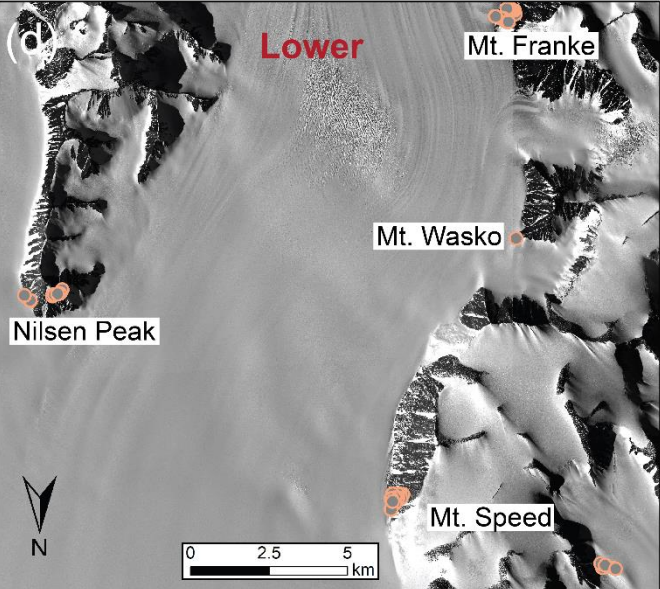
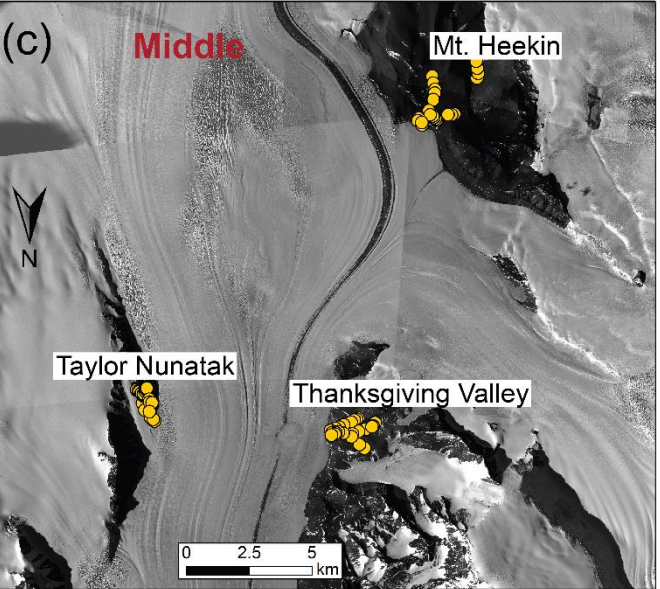
408   (PHI), and Marci Shaver-Adams for logistical and field support. Additionally, we thank Daniel Gilbert for help with initial

409   laboratory analyses at The Ohio State University. We are grateful to Dr. Peter Convey and Dr. Natasja van Gestel for

410   thoughtful reviews which strengthened our results and broader implications. This work was supported by NSF OPP grants

411   1341631 (WBL), 1341618 (DHW), 1341629 (NF), 1341736 (BJA), and NSF GRFP fellowship 60041697 (MAD).

412   Geospatial support for this work provided by the Polar Geospatial Center under NSF OPP grants 1043681 and 1559691.

413

415

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

420

421

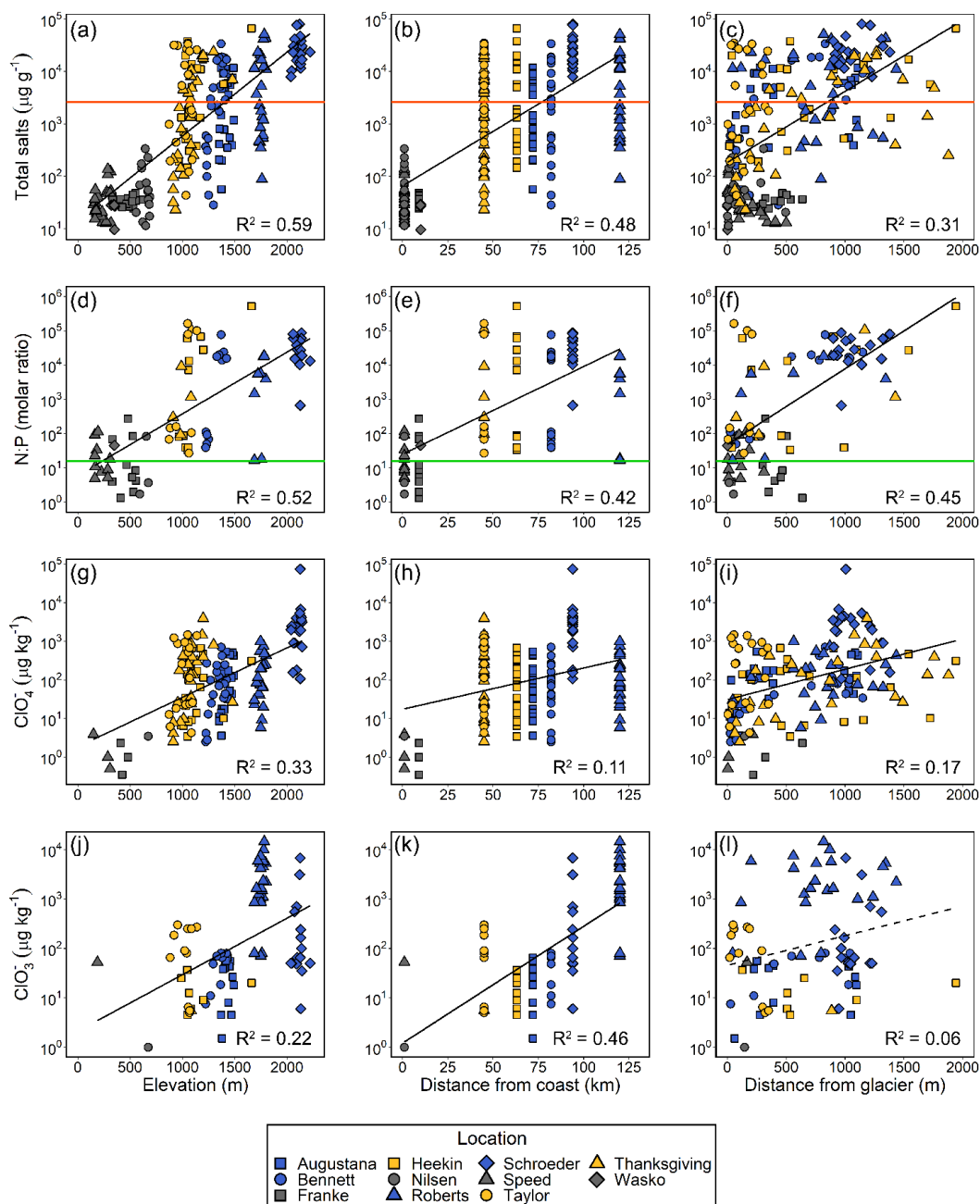
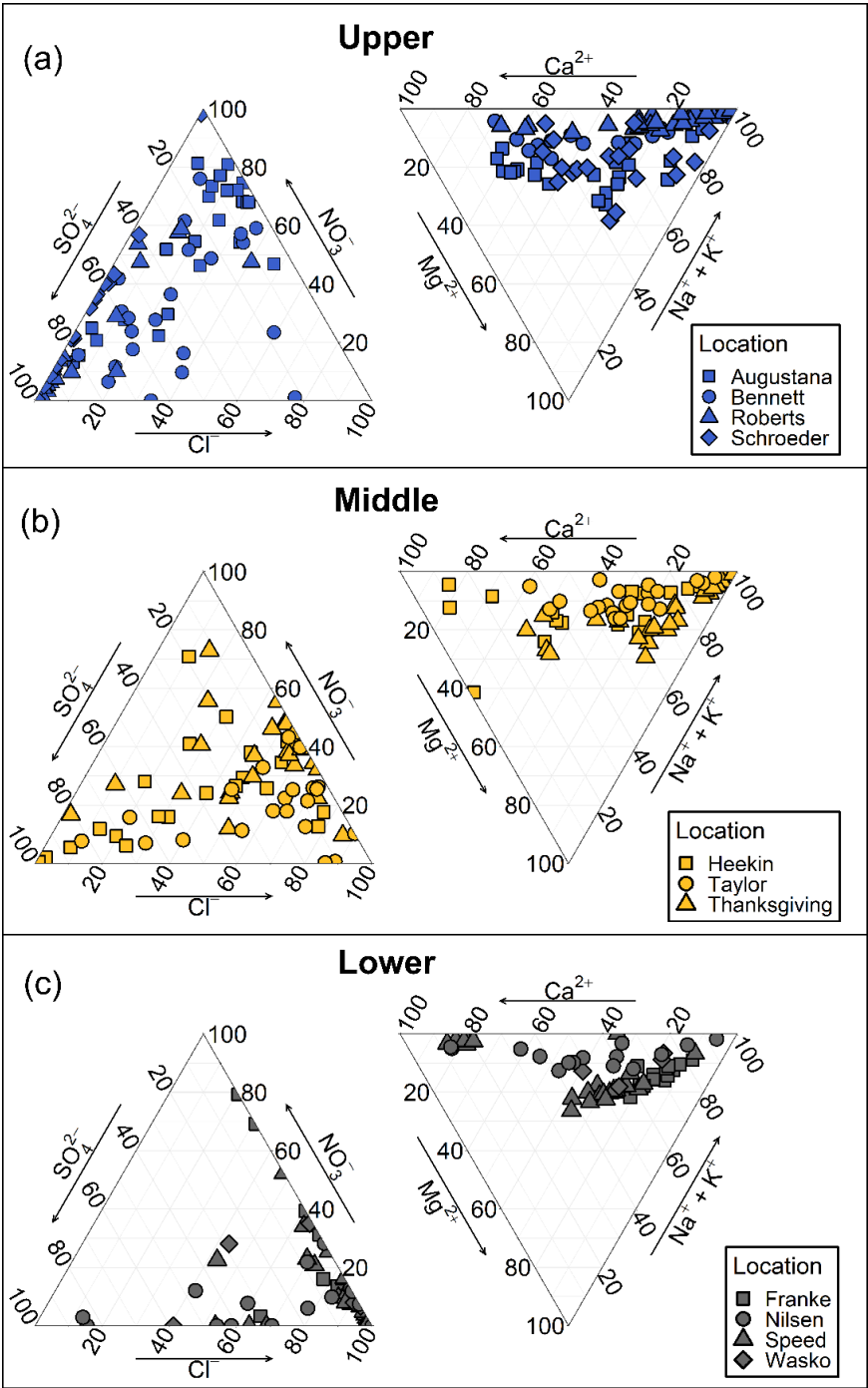



Figure 2. Total water-soluble salts, water-soluble N:P molar ratio, and ClO_4^- and ClO_3^- concentrations (log scale) were compared to elevation, distance from the coast, and distance from the glacier for samples from the three geographic zones (blue for upper, yellow for middle, grey for lower zones). Linear regression lines are plotted, where dashed lines represent regressions where $p > 0.05$ with a Bonferroni Correction, and R^2 values are reported for each relationship. The horizontal orange lines represent nematode salt tolerance of ~2,600 (Nkem et al., 2006) and the green lines represent the Redfield ratio, N:P = 16 for phytoplankton in the ocean.

428

429
430   Figure 3. Anion and cation ternary diagrams for the three geographic zones.
431

432

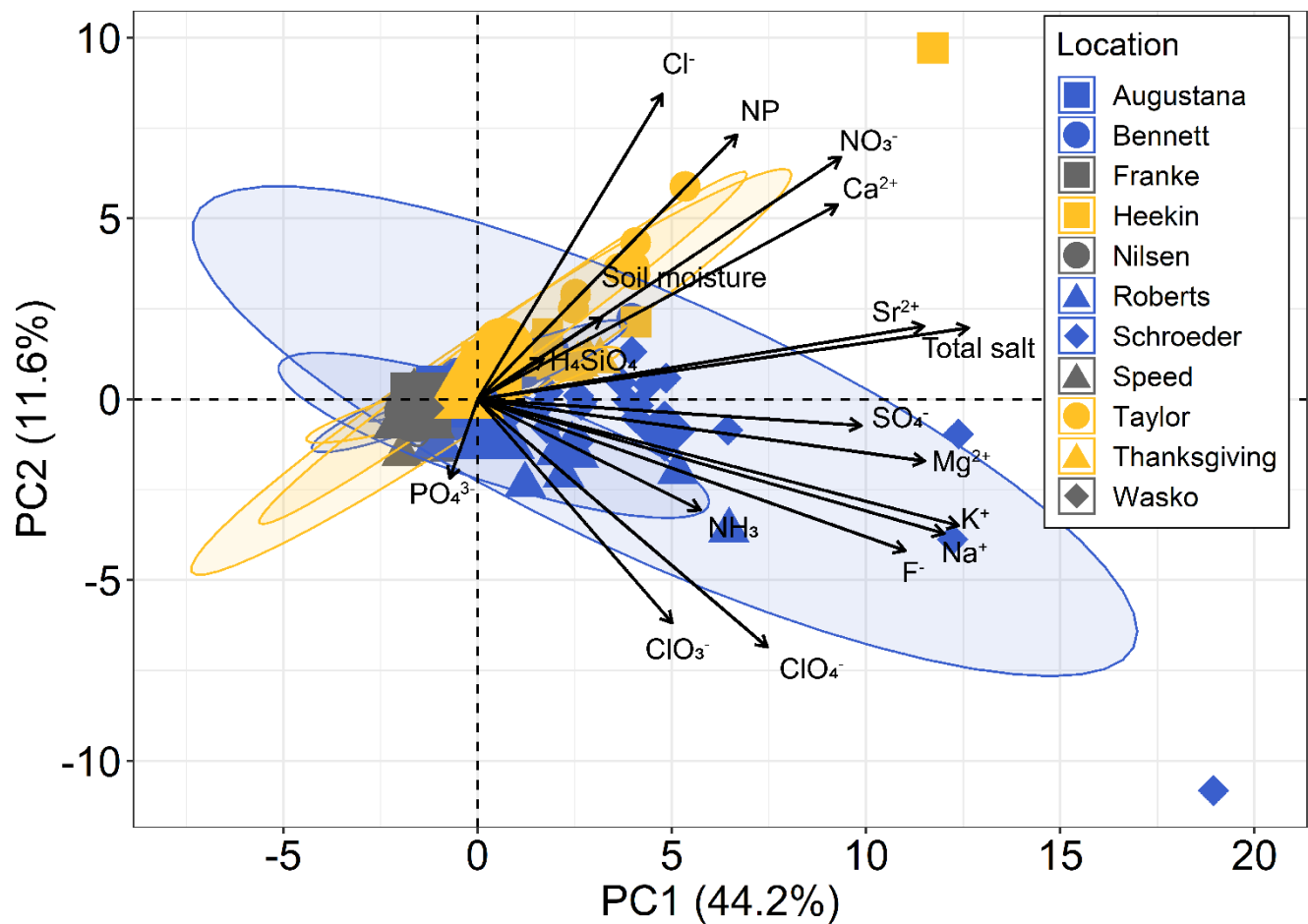



Figure 4. Principal component analysis (PCA) biplot generated in R using factoextra and built in R software libraries with all
anions, cations, nutrients, and soil moisture for the three geographic zones. The PCA is based on the correlation matrix (i.e.
scale = TRUE). Principal component 1 and principal component 2 are plotted on the x and y axes, respectively. Shaded
ellipses represent 95% confidence intervals.

437

439

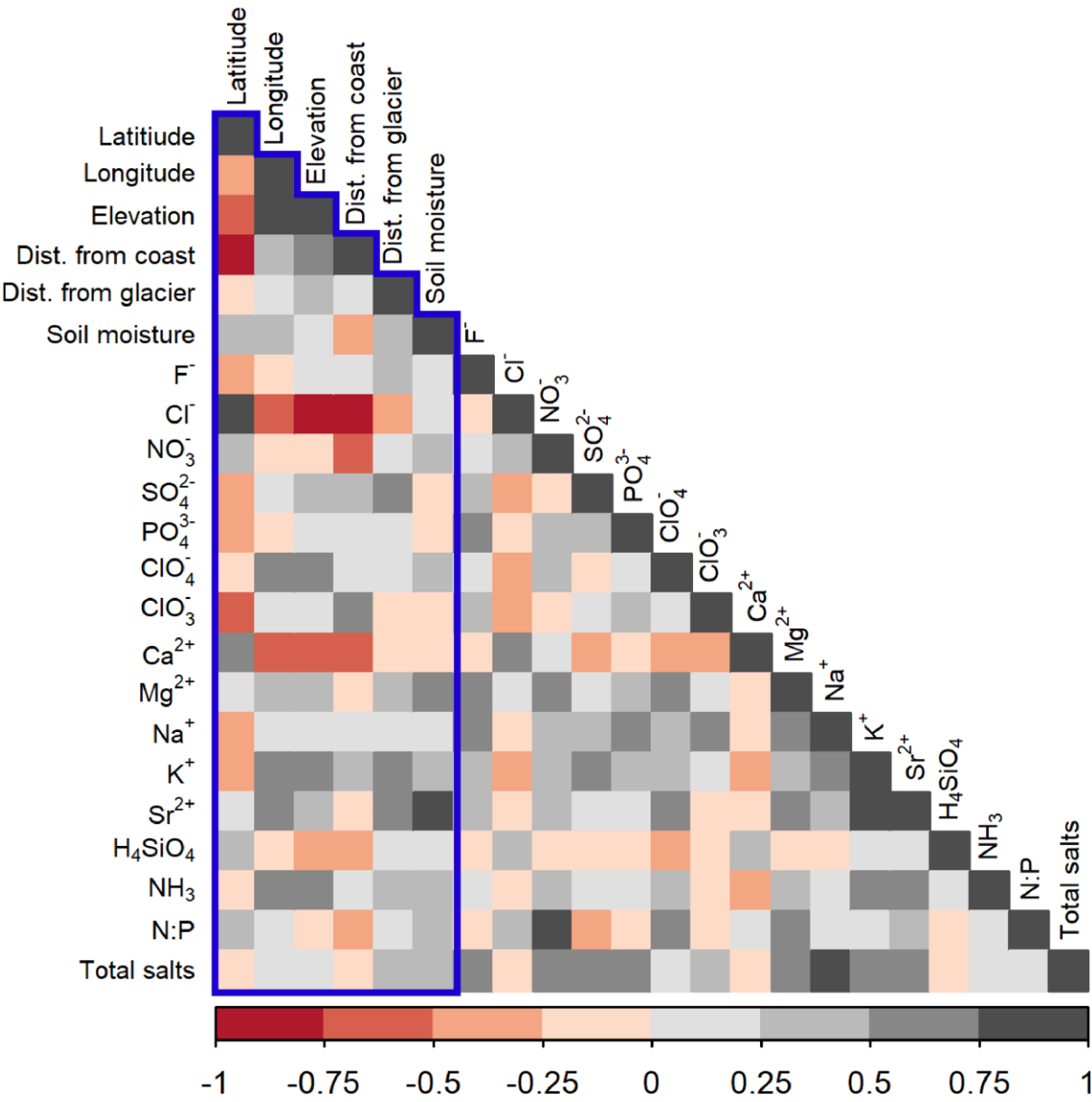



Figure 5. Spearman's rank correlation matrix generated in R using the corrplot library. The colors represent correlation coefficients, indicating the strength and magnitude of the correlation. The blue box indicates the geographic variables and soil moisture, which were variables used in the multiple linear regression and random forest models. Familywise type 1 error corrections were not applied for this analysis.

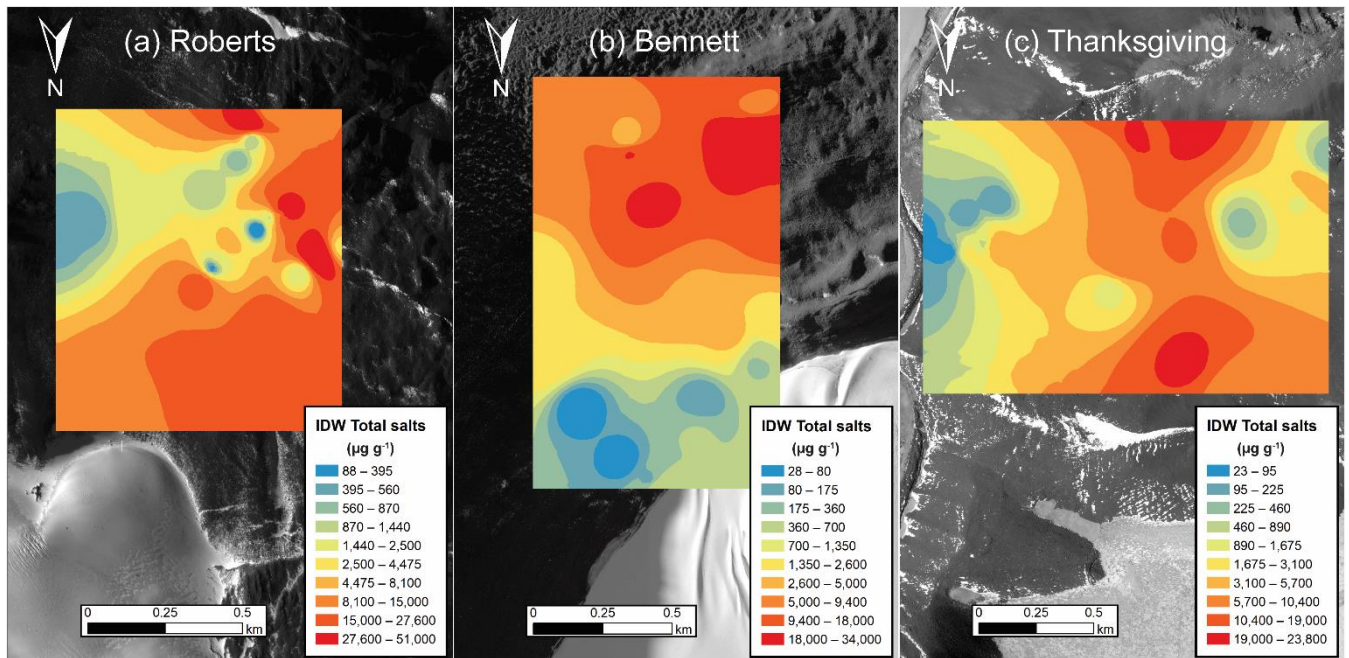
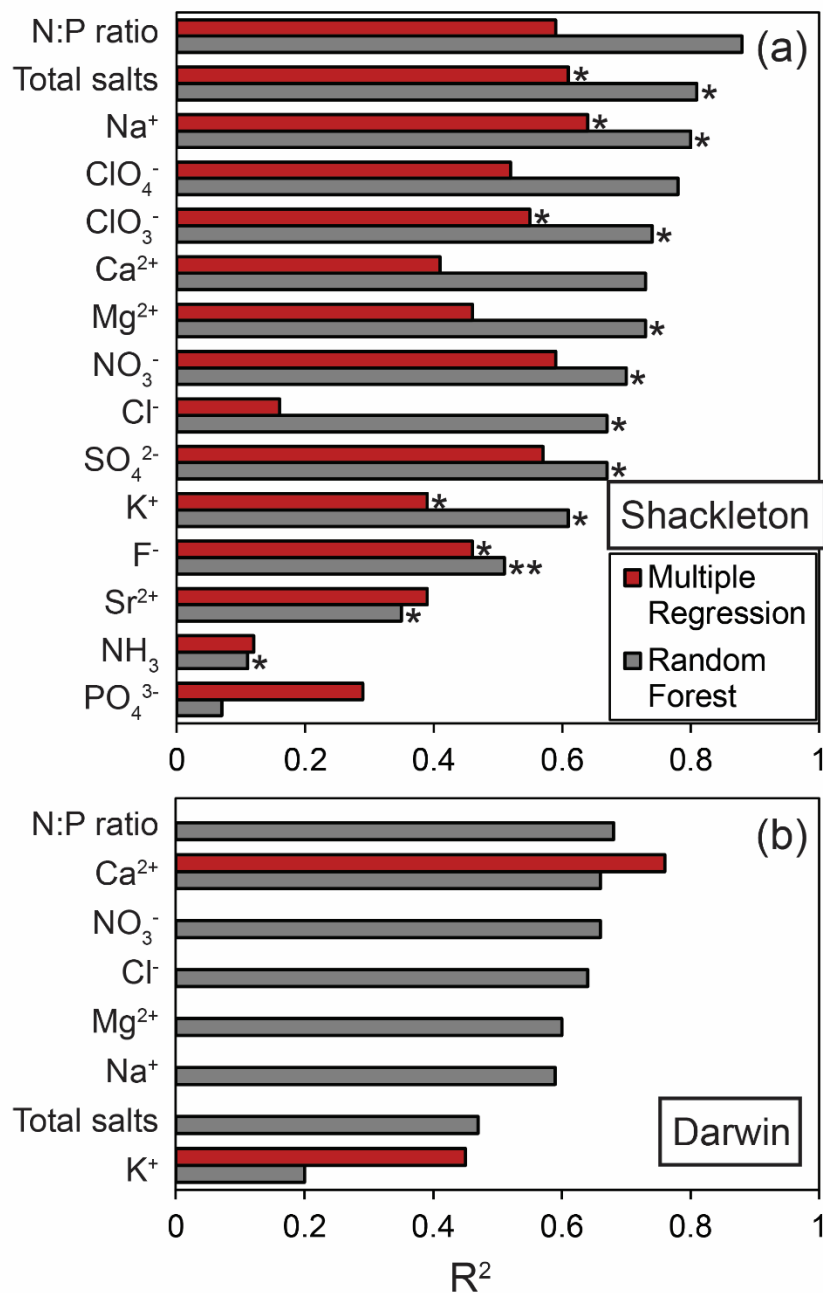



Figure 6. Inverse distance weighted (IDW) interpolations of total salt concentration for Roberts Massif (a), Bennett Platform (b), and Thanksgiving Valley (c). The color scale represents the 10 natural breaks in the data. Interpolations were created and mapped using the Geostatistical Analyst tool in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

451 Figure 7. R^2 values for the multiple linear regression and random forest model predicted and measured values for the
452 different analytes (Table 3). Test datasets include the Shackleton Glacier region (n =31) and the Darwin Glacier region (n =
453 10) (Magalhães et al., 2012). Analytes with slopes near 1, indicating good agreement between measured and predicted
454 values, are indicated (* t < 0.5; ** t < 0.20).

450

455

456 Table 1. Overview of geography, soil moisture, and water-soluble ions from the Shackleton Glacier region. The minimum
457 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^- ($\mu g\ g^{-1}$)	120	0.39	8.87	11.78	133
Cl^- ($\mu g\ g^{-1}$)	13,600	1.59	615	1,780	289
NO_3^- ($\mu g\ g^{-1}$)	38,400	0.10	1,470	3,450	235
SO_4^{2-} ($\mu g\ g^{-1}$)	55,300	0.08	4,390	8,080	184
PO_4^{3-} ($\mu g\ kg^{-1}$)	4,200	76.09	381	560	147
ClO_4^- ($\mu g\ kg^{-1}$)	75,000	0.35	985	6,020	611
ClO_3^- ($\mu g\ kg^{-1}$)	14,500	1.00	1,170	2,500	214
Ca^{2+} ($\mu g\ g^{-1}$)	4,400	0.55	839	1,160	139
Mg^{2+} ($\mu g\ g^{-1}$)	6,280	0.12	293	705	240
Na^+ ($\mu g\ g^{-1}$)	25,300	0.39	1,140	2,880	252
K^+ ($\mu g\ g^{-1}$)	440	0.86	28.31	51.61	182
Sr^{2+} ($\mu g\ g^{-1}$)	46.61	0.01	8.63	10.31	119
H_4SiO_4 ($\mu g\ g^{-1}$)	60.78	1.14	21.78	11.03	50.67
NH_3 ($\mu g\ kg^{-1}$)	5,080	18.85	324	587	181
N:P ratio (molar)	526,000	0.29	23,600	62,700	266
Total salt ($\mu g\ g^{-1}$)	80,500	9.46	7,932	13,300	167
STD, standard deviation; CV, coefficient of variation					

458

459 Table 2. Out-of-the-box multiple linear regression and random forest model statistics generated in R. All geochemical data
460 were log-transformed.

	Multiple regression		Random forest		
	R^2	p-value	Variance explained (%)	Most important variable	Least important variable
F^-	0.47	<<0.001	57	Elevation	Distance from coast
Cl^-	0.19	<<0.001	60	Elevation	Longitude
NO_3^-	0.52	<<0.001	60	Elevation	Longitude
SO_4^{2-}	0.53	<<0.001	62	Elevation	Longitude
PO_4^{3-}	0.17	<<0.001	4	Elevation	Distance from coast
ClO_4^-	0.44	<<0.001	48	Elevation	Distance from coast
ClO_3^-	0.55	<<0.001	63	Latitude	Distance from glacier
Ca^{2+}	0.44	<<0.001	60	Elevation	Distance from coast
Mg^{2+}	0.49	<<0.001	61	Elevation	Longitude
Na^+	0.65	<<0.001	75	Elevation	Longitude
K^+	0.48	<<0.001	60	Elevation	Distance from coast
Sr^{2+}	0.34	<<0.001	37	Elevation	Distance from coast
NH_3	0.29	<<0.001	38	Elevation	Distance from coast
N:P	0.60	<<0.001	52	Distance from glacier	Longitude
Total salts	0.61	<<0.001	75	Elevation	Longitude

461

462

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. All geochemical data were log-transformed.

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								
N:P ratio	0.59	<0.001	0.58	-0.720	0.88	<<0.001	0.64	-0.792
Total salts	0.61	<<0.001	0.71	-0.483*	0.81	<<0.001	0.86	-0.324*
Na^+	0.64	<<0.001	0.76	-0.424*	0.80	<<0.001	0.89	-0.262*
ClO_4^-	0.52	<0.001	0.60	-0.614	0.78	<<0.001	0.71	-0.590
ClO_3^-	0.55	0.009	0.72	-0.454*	0.74	<0.001	0.86	-0.284*
Mg^{2+}	0.46	<<0.001	0.63	-0.550	0.73	<<0.001	0.76	-0.469*
Ca^{2+}	0.41	<0.001	0.57	-0.613	0.73	<<0.001	0.74	-0.512
NO_3^-	0.59	<<0.001	0.62	-0.615	0.70	<<0.001	0.75	-0.465*
Sr^{2+}	0.35	0.026	0.54	-0.631	0.67	<0.001	0.82	-0.326*
SO_4^{2-}	0.57	<<0.001	0.63	-0.584	0.67	<<0.001	0.83	-0.310*
Cl^-	0.16	0.028	0.38	-0.773	0.67	<<0.001	0.76	-0.428*
K^+	0.39	<0.001	0.69	-0.429*	0.61	<<0.001	0.83	-0.291*
F^-	0.46	<0.001	0.76	-0.352*	0.51	<<0.001	0.91	-0.141**
NH_3	0.12	0.052	0.57	-0.528	0.11	0.068	0.61	-0.475*
PO_4^{3-}	0.29	0.070	0.32	-0.857	0.07	0.408	0.38	-0.764
Darwin								
N:P ratio	-	-	-	-	0.68	0.021	0.54	-0.765
Ca^{2+}	0.76	0.001	0.66	-0.645	0.66	0.004	0.50	-0.785
NO_3^-	-	-	-	-	0.66	0.004	0.49	-0.794
Cl^-	-	-	-	-	0.64	0.005	0.22	-0.962
Mg^{2+}	-	-	-	-	0.60	0.140	0.67	-0.544
Na^+	-	-	-	-	0.59	0.160	0.42	-0.836
Total salts	-	-	-	-	0.47	0.028	0.42	-0.802

K$^+$	0.45	0.070	0.62	-0.550	0.20	0.267	0.28	-0.882
* t < 0.5; ** t < 0.20								

468

469 Table 4. Multiple linear regression and random forest model mean absolute error (MAE) and root mean squared error
470 (RMSE). All geochemical data were log-transformed for the analysis.

Analyte	Multiple Linear Regression		Random Forest	
	MAE	RMSE	MAE	RMSE
Shackleton				
N:P ratio	2.19	2.73	1.75	2.11
Total salts	1.45	1.69	0.86	1.17
Na^+	1.23	1.52	0.83	1.13
ClO_4^-	1.33	1.62	0.91	1.12
ClO_3^-	1.07	1.67	1.01	1.26
Mg^{2+}	1.78	2.08	1.07	1.48
Ca^{2+}	1.84	2.21	1.18	1.53
NO_3^-	1.96	2.29	1.56	1.93
Sr^{2+}	1.05	1.17	0.59	0.82
SO_4^{2-}	1.58	1.94	1.35	1.67
Cl^-	2.11	2.39	1.07	1.5
K^+	0.73	0.89	0.56	0.72
F^-	0.48	0.6	0.46	0.58
NH_3	0.67	0.83	0.65	0.86
PO_4^{3-}	0.75	0.96	0.78	1.14
Darwin				
N:P ratio	267	267	1.48	1.73
Ca^{2+}	5.79	5.85	2.63	2.83
NO_3^-	261	261	3.19	3.52
Cl^-	372	372	2.99	3.32
Mg^{2+}	460	460	2.89	3.06
Na^+	245	245	2.57	2.88
Total salts	139	139	1.22	1.67
K^+	30.8	30.8	1.00	1.19
MAE, mean absolute error; RMSE, root mean squared error				

471

References

Antipov, E. A. and Pokryshevskaya, E. B.: Mass appraisal of residential apartments: An application of Random forest for valuation and a CART-based approach for model diagnostics, Expert Syst. Appl., 39(2), 1772–1778, doi:10.1016/j.eswa.2011.08.077, 2012.

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 in a polar desert of Antarctica, Soil Biol. Biochem., 122, 7–18, doi:10.1016/J.SOILBIO.2018.03.025, 2018.

Balter-Kennedy, A., Bromley, G., Balco, G., Thomas, H. and Jackson, M. S.: A 14.5-million-year record of East Antarctic Ice Sheet fluctuations from the central Transantarctic Mountains, constrained with cosmogenic 3He, 10Be, 21Ne, and 26Al, Cryosph., 14(8), 2647–2672, doi:10.5194/tc-2020-57, 2020.

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 biodiversity and biogeochemistry in northern and southern Victoria Land, Antarctica, Antarct. Sci., 18(4), 535–548, doi:10.1017/S0954102006000587, 2006.

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 Moorhead, D. L.: Biogeochemical stoichiometry of Antarctic Dry Valley ecosystems, J. Geophys. Res., 112(G1), G01010, doi:10.1029/2005JG000141, 2007.

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. and Adams, B.: Genetic diversity among populations of Antarctic springtails (Collembola) within the Mackay Glacier ecotone 1, Genome, 59, 762–770, doi:10.1139/gen-2015-0194, 2016.

Bennett, K. R., Hogg, I. D., Adams, B. J. and Hebert, P. D. N.: High levels of intraspecific genetic divergences revealed for Antarctic springtails: evidence for small-scale isolation during Pleistocene glaciation, Biol. J. Linn. Soc., 119(1), 166–178, doi:10.1111/bij.12796, 2016.

Bockheim, J. G.: Landform and Soil Development in the McMurdo Dry Valleys, Antarctica: A Regional Synthesis, Arctic, Antarct. Alp. Res., 34(3), 308–317, doi:10.1080/15230430.2002.12003499, 2002.

Bockheim, J. G.: Functional diversity of soils along environmental gradients in the Ross Sea region, Antarctica, Geoderma, 144(1–2), 32–42, doi:10.1016/j.geoderma.2007.10.014, 2008.

Bottos, E. M., Laughlin, D. C., Herbold, C. W., Lee, C. K., McDonald, I. R. and Cary, S. C.: Abiotic factors influence patterns of bacterial diversity and community composition in the Dry Valleys of Antarctica, FEMS Microbiol. Ecol., 96(5), 42, doi:10.1093/femsec/fiaa042, 2020.

Breiman, L.: Random forests, Mach. Learn., 45(1), 5–32, doi:10.1023/A:1010933404324, 2001.

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. A., Storey, B. C., Wall, D. H. and Adams, B. J.: Nematodes in a polar desert reveal the relative role of biotic interactions in the coexistence of soil animals, Commun. Biol., 2(63), doi:10.1038/s42003-018-0260-y, 2019.

Claridge, G. G. C. and Campbell, I. B.: Soils of the Shackleton glacier region, Queen Maud Range, Antarctica, New Zeal. J. Sci., 11(2), 171–218, 1968.

Collins, G. E., Hogg, I. D., Convey, P., Barnes, A. D. and McDonald, I. R.: Spatial and temporal scales matter when assessing the species and genetic diversity of springtails (Collembola) in Antarctica, Front. Ecol. Evol., 7, 76, doi:10.3389/fevo.2019.00076, 2019.

Collins, G. E., Hogg, I. D., Convey, P., Sancho, L. G., Cowan, D. A., Lyons, W. B., Adams, B. J., Wall, D. H. and Green, T. G. A.: Genetic diversity of soil invertebrates corroborates timing estimates for past collapses of the West Antarctic Ice Sheet, Proc. Natl. Acad. Sci. U. S. A., 117(36), 22293–22302, doi:10.1073/pnas.2007925117, 2020.

512 Convey, P. and McInnes, S. J.: Exceptional tardigrade-dominated ecosystems in Ellsworth Land, Antarctica, Ecology, 86(2),
513 519–527, doi:10.1890/04-0684, 2005.

514 Courtright, E. M., Wall, D. H. and Virginia, R. A.: Determining habitat suitability for soil invertebrates in an extreme
515 environment: The McMurdo Dry Valleys, Antarctica, Antarct. Sci., 13(1), 9–17, doi:10.1017/S0954102001000037, 2001.

516 Davidson, A. D., Hamilton, M. J., Boyer, A. G., Brown, J. H. and Ceballos, G.: Multiple ecological pathways to extinction in
517 mammals., 2009.

518 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,
519 W. B.: Aeolian Material Transport and Its Role in Landscape Connectivity in the McMurdo Dry Valleys, Antarctica, J.
520 Geophys. Res. Earth Surf., 123, 3323–3337, doi:10.1029/2017JF004589, 2018.

521 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
522 terrestrial exposure ages inferred from meteoric 10Be and NO3- concentrations in soils along the Shackleton Glacier,
523 Antarctica, Earth Surf. Dyn., in review, doi:https://doi.org/10.5194/esurf-2020-50, 2020a.

524 Diaz, M. A., Li, J., Michalski, G., Darrah, T. H., Adams, B. J., Wall, D. H., Hogg, I. D., Fierer, N., Welch, S. A., Gardner, C.
525 B. and Lyons, W. B.: Stable isotopes of nitrate, sulfate, and carbonate in soils from the Transantarctic Mountains, Antarctica:
526 A record of atmospheric deposition and chemical weathering, Front. Earth Sci., 8(341), doi:10.3389/feart.2020.00341,
527 2020b.

528 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
529 boundaries of microbial habitability in soil, bioRxiv, doi:https://doi.org/10.1101/2020.08.03.234583, 2020.

530 Freckman, D. W. and Virginia, R. A.: Soil Biodiversity and Community Structure in the Mcmurdo Dry Valleys, Antarctica,
531 in Ecosystem dynamics in a polar desert; the McMurdo dry valleys, Antarctica, edited by J. C. Priscu, pp. 323–335,
532 American Geophysical Union (AGU)., 1998.

533 Golledge, N. R. and Levy, R. H.: Geometry and dynamics of an East Antarctic Ice Sheet outlet glacier, under past and
534 present climates, J. Geophys. Res., 116(F3), F03025, doi:10.1029/2011JF002028, 2011.

535 Golledge, N. R., Fogwill, C. J., Mackintosh, A. N. and Buckley, K. M.: Dynamics of the last glacial maximum Antarctic ice-
536 sheet and its response to ocean forcing, Proc. Natl. Acad. Sci. U. S. A., 109(40), 16052–16056, doi:10.1073/pnas.1, 2012.

537 Golledge, N. R., Levy, R. H., McKay, R. M., Fogwill, C. J., White, D. A., Graham, A. G. C., Smith, J. A., Hillenbrand, C.
538 D., Licht, K. J., Denton, G. H., Ackert, R. P., Maas, S. M. and Hall, B. L.: Glaciology and geological signature of the Last
539 Glacial Maximum Antarctic ice sheet, Quat. Sci. Rev., 78, 225–247, doi:10.1016/j.quascirev.2013.08.011, 2013.

540 Heindel, R. C., Spickard, A. M. and Virginia, R. A.: Landscape-scale soil phosphorus variability in the McMurdo Dry
541 Valleys, Antarct. Sci., 29(3), 252–263, doi:10.1017/S0954102016000742, 2017.

542 Heung, B., Bulmer, C. E. and Schmidt, M. G.: Predictive soil parent material mapping at a regional-scale: A Random Forest
543 approach, Geoderma, 214–215, 141–154, doi:10.1016/j.geoderma.2013.09.016, 2014.

544 Hodgson, D. A., Convey, P., Verleyen, E., Vyverman, W., McInnes, S. J., Sands, C. J., Fernández-Carazo, R., Wilmotte, A.,
545 De Wever, A., Peeters, K., Tavernier, I. and Willems, A.: The limnology and biology of the Dufek Massif, Transantarctic
546 Mountains 82° South, Polar Sci., 4(2), 197–214, doi:10.1016/j.polar.2010.04.003, 2010.

547 Hogg, I. D. and Wall, D. H.: Polar deserts, in Life at Extremes: Environments, Organisms and Strategies for Survival, edited
548 by E. M. Bell, pp. 176–195, CAB International., 2012.

549 Howat, I. M., Porter, C., Smith, B. E., Noh, M.-J. and Morin, P.: The Reference Elevation Model of Antarctica, Cryosph.,
550 13, 665–674, doi:10.5194/tc-13-665-2019, 2019.

551 Jackson, A., Davila, A. F., Böhlke, J. K., Sturchio, N. C., Sevanthi, R., Estrada, N., Brundrett, M., Lacelle, D., McKay, C. P.,
552 Poghosyan, A., Pollard, W. and Zacny, K.: Deposition, accumulation, and alteration of Cl-, NO3-, ClO4- and ClO3- salts in
553 a hyper-arid polar environment: Mass balance and isotopic constraints, Geochim. Cosmochim. Acta, 182, 197–215,
554 doi:10.1016/j.gca.2016.03.012, 2016.

555 Jackson, M. S., Hall, B. L. and Denton, G. H.: Asynchronous behavior of the Antarctic Ice Sheet and local glaciers during
556 and since Termination 1, Salmon Valley, Antarctica, Earth Planet. Sci. Lett., 482, 396–406, doi:10.1016/j.epsl.2017.11.038,
557 2018.

558 Jackson, W. A., Davila, A. F., Estrada, N., Lyons, W. B., Coates, J. D. and Priscu, J. C.: Perchlorate and chlorate
559 biogeochemistry in ice-covered lakes of the McMurdo Dry Valleys, Antarctica, Geochim. Cosmochim. Acta, 98, 19–30,
560 doi:10.1016/j.gca.2012.09.014, 2012.

561 Jackson, W. A., Böhlke, J. K., Andraski, B. J., Fahlquist, L., Bexfield, L., Eckardt, F. D., Gates, J. B., Davila, A. F., McKay,
562 C. P., Rao, B., Sevanthi, R., Rajagopalan, S., Estrada, N., Sturchio, N., Hatzinger, P. B., Anderson, T. A., Orris, G.,
563 Betancourt, J., Stonestrom, D., Latorre, C., Li, Y. and Harvey, G. J.: Global patterns and environmental controls of
564 perchlorate and nitrate co-occurrence in arid and semi-arid environments, Geochim. Cosmochim. Acta, 164, 502–522,
565 doi:10.1016/J.GCA.2015.05.016, 2015.

566 Kassambara, A. and Mundt, F.: Package "factoextra," Extr. Vis. results Multivar. data Anal., 76 [online] Available from:
567 https://github.com/kassambara/factoextra/issues (Accessed 11 August 2020), 2017.

568 Kirkwood, C., Cave, M., Beamish, D., Grebby, S. and Ferreira, A.: A machine learning approach to geochemical mapping, J.
569 Geochemical Explor., 167, 49–61, doi:10.1016/j.gexplo.2016.05.003, 2016.

570 LaPrade, K. E.: Climate, geomorphology, and glaciology of the Shackleton Glacier area, Queen Maud Mountains,
571 Transantarctic Mountains, Antarctica, Antarct. Res. Ser., 36(9), 163–196, doi:10.1029/ar036p0163, 1984.

572 Levy, J. S., Fountain, A. G., Welch, K. A. and Lyons, W. B.: Hypersaline "wet patches" in Taylor Valley, Antarctica,
573 Geophys. Res. Lett., 39(5), n/a-n/a, doi:10.1029/2012GL050898, 2012.

574 Lyons, W. B., Welch, K. A., Neumann, K., Toxey, J. K., McArthur, R., Williams, C., McKnight, D. M. and Moorhead, D.
575 L.: Geochemical linkages among glaciers, streams and lakes within the Taylor Valley, Antarctica, Ecosyst. Dyn. a polar
576 desert; McMurdo dry Val. Antarct., 72, 77–92, doi:10.1029/AR072p0077, 1998.

577 Lyons, W. B., Deuerling, K., Welch, K. A., Welch, S. A., Michalski, G., Walters, W. W., Nielsen, U., Wall, D. H., Hogg, I.
578 and Adams, B. J.: The Soil Geochemistry in the Beardmore Glacier Region, Antarctica: Implications for Terrestrial
579 Ecosystem History, Sci. Rep., 6, 26189, doi:10.1038/srep26189, 2016.

580 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
581 Life: Multidisciplinary Insights Reveal Environmental Constraints on Biotic Diversity in Continental Antarctica, edited by F.
582 de Bello, PLoS One, 7(9), e44578, doi:10.1371/journal.pone.0044578, 2012.

583 Nakada, M. and Lambeck, K.: The melting history of the late Pleistocene Antarctic ice sheet., 1988.

584 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
585 species of Antarctic soil nematodes, Polar Biol., 29, 643–651, doi:10.1007/s00300-005-0101-6, 2006.

586 Patel, J., Shah, S., Thakkar, P. and Kotecha, K.: Predicting stock market index using fusion of machine learning techniques,
587 Expert Syst. Appl., 42(4), 2162–2172, doi:10.1016/j.eswa.2014.10.031, 2015.

588 Peters, J., De Baets, B., Verhoest, N. E. C., Samson, R., Degroeve, S., Becker, P. De and Huybrechts, W.: Random forests as
589 a tool for ecohydrological distribution modelling, Ecol. Modell., 207(2–4), 304–318, doi:10.1016/j.ecolmodel.2007.05.011,
590 2007.

591 Prasad, A. M., Iverson, L. R. and Liaw, A.: Newer classification and regression tree techniques: Bagging and random forests
592 for ecological prediction, Ecosystems, 9(2), 181–199, doi:10.1007/s10021-005-0054-1, 2006.

593 R Core Team: R: A language and environment for statistical computing, 2020.

594 Scarrow, J. W., Balks, M. R. and Almond, P. C.: Three soil chronosequences in recessional glacial deposits near the polar
595 plateau, in the Central Transantarctic Mountains, Antarctica, Antarct. Sci., 26(5), 573–583,
596 doi:10.1017/S0954102014000078, 2014.

597 Stafoggia, M., Bellander, T., Bucci, S., Davoli, M., de Hoogh, K., de' Donato, F., Gariazzo, C., Lyapustin, A., Michelozzi,
598 P., Renzi, M., Scortichini, M., Shtein, A., Viegi, G., Kloog, I. and Schwartz, J.: Estimation of daily PM10 and PM2.5
599 concentrations in Italy, 2013–2015, using a spatiotemporal land-use random-forest model, Environ. Int., 124, 170–179,
600 doi:10.1016/j.envint.2019.01.016, 2019.

601 Stevens, M. I. and Hogg, I. D.: Long-term isolation and recent range expansion from glacial refugia revealed for the endemic
602 springtail Gomphiocephalus hodgsoni from Victoria Land, Antarctica, Mol. Ecol., 12(9), 2357–2369, doi:10.1046/j.1365-
603 294X.2003.01907.x, 2003.

604 Stevens, M. I., Greenslade, P., Hogg, I. D. and Sunnucks, P.: Southern Hemisphere Springtails: Could Any Have Survived
605 Glaciation of Antarctica?, Mol. Biol. Evol., 23(5), 874–882, doi:10.1093/molbev/msj073, 2006.

606 Talarico, F. M., McKay, R. M., Powell, R. D., Sandroni, S. and Naish, T.: Late Cenozoic oscillations of Antarctic ice sheets
607 revealed by provenance of basement clasts and grain detrital modes in ANDRILL core AND-1B, Glob. Planet. Change, 96–
608 97, 23–40, doi:10.1016/j.gloplacha.2009.12.002, 2012.

609 Tesoriero, A. J., Gronberg, J. A., Juckem, P. F., Miller, M. P. and Austin, B. P.: Predicting redox-sensitive contaminant
610 concentrations in groundwater using random forest classification, Water Resour. Res., 53(8), 7316–7331,
611 doi:10.1002/2016WR020197, 2017.

612 Toner, J. D., Sletten, R. S. and Prentice, M. L.: Soluble salt accumulations in Taylor Valley, Antarctica: Implications for
613 paleolakes and Ross Sea Ice Sheet dynamics, J. Geophys. Res. Earth Surf., 118(1), 198–215, doi:10.1029/2012JF002467,
614 2013.

615 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,
616 W. H., Eds.: Soil Ecology and Ecosystem Services, Oxford University Press, Oxford., 2012.

617 Warton, D. I., Duursma, R. A., Falster, D. S. and Taskinen, S.: smatr 3- an R package for estimation and inference about
618 allometric lines, Methods Ecol. Evol., 3(2), 257–259, doi:10.1111/j.2041-210X.2011.00153.x, 2012.

619 Webster-Brown, J., Gall, M., Gibson, J., Wood, S. and Hawes, I.: The biogeochemistry of meltwater habitats in the Darwin
620 Glacier region (80°S), Victoria Land, Antarctica, Antarct. Sci., 22(6), 646–661, doi:10.1017/S0954102010000787, 2010.

621 Wilson, D. J., Bertram, R. A., Needham, E. F., van de Flierdt, T., Welsh, K. J., McKay, R. M., Mazumder, A., Riesselman,
622 C. R., Jimenez-Espejo, F. J. and Escutia, C.: Ice loss from the East Antarctic Ice Sheet during late Pleistocene interglacials,
623 Nature, 561(7723), 383–386, doi:10.1038/s41586-018-0501-8, 2018.

624 Zeglin, L. H., Sinsabaugh, R. L., Barrett, J. E., Gooseff, M. N. and Takacs-Vesbach, C. D.: Landscape Distribution of
625 Microbial Activity in the McMurdo Dry Valleys: Linked Biotic Processes, Hydrology, and Geochemistry in a Cold Desert
626 Ecosystem, Ecosystems, 12(4), 562–573, doi:10.1007/s10021-009-9242-8, 2009.

627