

<sup>1</sup> A  $V_{s30}$  Map for New Zealand based on Geologic  
<sup>2</sup> and Terrain Proxy Variables and Field  
<sup>3</sup> Measurements

<sup>4</sup> Kevin Foster<sup>†\*</sup>, Brendon Bradley<sup>†</sup>, Chris McGann<sup>†</sup>, and Liam Wotherspoon<sup>‡</sup>

<sup>5</sup> \*Corresponding author. [kevin.foster@pg.canterbury.ac.nz](mailto:kevin.foster@pg.canterbury.ac.nz)

<sup>6</sup> <sup>†</sup>Department of Civil & Natural Resources engineering, University of Canterbury,  
<sup>7</sup> 69 Creyke Road, Christchurch, New Zealand

<sup>8</sup> <sup>‡</sup>University of Auckland, Auckland, New Zealand

<sup>9</sup> A time-averaged 30-meter depth shear wave velocity ( $V_{s30}$ ) map is developed for New  
<sup>10</sup> Zealand as a weighted combination of a geology-based and a terrain-based model. A  
<sup>11</sup> Bayesian updating process allows local  $V_{s30}$  measurements to control model estimates  
<sup>12</sup> where data exist, and uses model estimates developed for other parts of the world  
<sup>13</sup> where local data are sparse or nonexistent. Geostatistical interpolation is performed  
<sup>14</sup> on the geology- and terrain-based models using local  $V_{s30}$  measurements to constrain  
<sup>15</sup> the model in the vicinity of data. Conventional regression kriging is compared with a  
<sup>16</sup> flexible multivariate normal (MVN) approach that allows for arbitrary assumptions  
<sup>17</sup> regarding measurement uncertainty at each data location. A modification to the  
<sup>18</sup> covariance structure in the MVN application allows for more realistic estimates by  
<sup>19</sup> reducing undesirable extrapolation across geologic boundaries. The results of kriging  
<sup>20</sup> and MVN approaches are compared. The geology- and terrain-based MVN models  
<sup>21</sup> are combined to produce a final model suitable for engineering applications. 100m  
<sup>22</sup> resolution map outputs are publicly available.

## INTRODUCTION

24  $V_{s30}$ , the time-averaged vertical shear wave velocity in uppermost 30 metres, is widely used in  
25 earthquake engineering research and practice. Owing to the expense and difficulty of measuring  
26  $V_{s30}$  directly, recent work has been devoted to developing  $V_{s30}$  models using  $V_{s30}$ -correlated proxy  
27 data. The work presented herein represents an attempt to incorporate the best recent practices  
28 for modelling  $V_{s30}$  (*e.g.*, [Yong et al. \(2012\)](#); [Thompson et al. \(2014\)](#); [Parker et al. \(2017\)](#); [Ahdi et al. \(2017b\)](#)) using both available proxy data and direct  $V_{s30}$  measurements (with arbitrarily  
29 specified measurement uncertainty), within a statistical framework that treats uncertainty as an  
30 essential model component and allows consistent incremental improvements as new  $V_{s30}$  data  
31 are collected.

33 Others have applied methods similar to those presented here for mapping  $V_{s30}$  by proxy methods,  
34 where continuous and readily available proxy data are correlated with  $V_{s30}$  measurements and  
35 used for estimating  $V_{s30}$  in regions where direct measurements have not been taken. Generally  
36 the proxy data include geology and topographic data. First studies from around the world are  
37 summarized and then prior work in New Zealand (NZ) is highlighted, with a focus on the  
38 additional contributions to be presented in the work herein.

39 [Wills and Clahan \(2006\)](#) developed a  $V_{s30}$  map for California based on surface geology. Geology-  
40 based  $V_{s30}$  mapping is confounded by the fact that mapped geologic units are frequently less  
41 than 30 metres thick. [Wills and Clahan \(2006\)](#) handled this problem for alluvial basins by  
42 assigning locations in California to “shallow” or “deep” basin categories on the basis of areal  
43 extent: narrow valleys and locations near to the base of steep mountains were assigned “shallow”  
44 categories whereas extensive basins were assigned “deep” status. [Wills and Clahan \(2006\)](#)  
45 also separated alluvial units into fine, coarse and mixed/undifferentiated bins for better  $V_{s30}$   
46 discrimination. [Wills and Gutierrez \(2009\)](#) introduced a slope-dependent component to  $V_{s30}$   
47 modeling. Topographic slope is slightly correlated with  $V_{s30}$ , at least for soil deposits, owing  
48 to depositional characteristics. For example, the size of particles dropping out of suspension in

49 an alluvial depositional environment is correlated with flow velocity and, in turn, the slope of  
50 the landscape. The range and complexity of surficial rock depositional mechanisms renders this  
51 correlation virtually nonexistent for rock.

52 Building on the [Wills and Clahan \(2006\)](#) study, [Thompson et al. \(2014\)](#) developed a regional  
53  $V_{s30}$  map for California that added topographic slope based modifications ([Wills and Gutier-](#)  
54 [rez, 2009](#); [Allen and Wald, 2007](#)) to soil categories. Additionally, the well-known regression  
55 kriging approach was applied to the mapping, bringing model predictions into agreement with  
56 measurements locally. The kriging uncertainty in this study can be evaluated alongside the  
57 model predictions for a first-order, low-end assessment of expected disparity between model  
58 and reality. Importantly, kriging uncertainty is entirely a product of measurement locations  
59 and “globally” derived variogram, meaning that the mapped uncertainty is not conditioned on  
60 individual geology-based polygons.

61 [Lee and Tsai \(2008\)](#) generated a  $V_{s30}$  map of Taiwan in a two-step process, first generating a  
62 correlation-based model for  $V_{s30}$  using SPT (standard penetration testing) blow counts, then  
63 applying this model to a larger dataset. The resulting model was kriged. [Wald et al. \(2011\)](#)  
64 proposed a generalized framework for  $V_{s30}$  mapping, recommending a hierachal model with  
65 topographic slope (being globally available) as the primary proxy variable, and generating  
66 individual linear slope- $V_{s30}$  relations for significant geology groups. They demonstrated the  
67 method by producing a map of Taiwan and employing kriging to honor local observations.

68 [Vilanova et al. \(2018\)](#) developed a  $V_{s30}$  map for Portugal based on geology. Statistical testing was  
69 performed on candidate geologic units to optimize discrimination. Spatial declustering methods  
70 were evaluated but found to have non-negligible impact in only one of six candidate geologic  
71 units. After eliminating non-discriminating groups only three distinct geologic categories were  
72 used in the final model.

73 Ahdi et al. developed geology- and topography-based  $V_{s30}$  models for Alaska ([Ahdi et al., 2017b](#))  
74 and the Pacific Northwest ([Ahdi et al., 2017a](#)). The methods employed were similar to [Thompson](#)

<sup>75</sup> et al. (2014), although did not incorporate any geostatistical methods. Ahdi reevaluated the  
<sup>76</sup> geologic classification process, and hence the final product employs geologic categories that are  
<sup>77</sup> distinct and generally more discriminating than the Thompson et al. (2014) categories. Parker  
<sup>78</sup> et al. (2017) developed  $V_{s30}$  estimates for Central and Eastern North America (CENA) based on  
<sup>79</sup> geology, topographic slope, and indicator variables to indicate significant sedimentary basins  
<sup>80</sup> and the extent of Wisconsin glaciation. Basins and past glaciation were noted as having a strong  
<sup>81</sup> correlation with observed  $V_{s30}$  values.

<sup>82</sup> Yong et al. (2012) employed an automated, DEM (digital elevation model)-based 16-category  
<sup>83</sup> terrain classification scheme (Iwahashi and Pike, 2007) to the problem of  $V_{s30}$  modelling. The  
<sup>84</sup> terrain classification technique bins a problem domain into slope bins, then subdivides each bin  
<sup>85</sup> into further categories on the basis of local convexity and a local roughness measure. The results  
<sup>86</sup> were used to generate worldwide  $V_{s30}$  maps. A later comprehensive review of proxy-based  $V_{s30}$   
<sup>87</sup> measurements (Yong, 2016) examined the performance of this technique using newly available  
<sup>88</sup>  $V_{s30}$  data in California. Another  $V_{s30}$  model for Taiwan (Kwok et al., 2018) also used terrain-  
<sup>89</sup> based classifications, as well as the more common geology classifications, and combined the two  
<sup>90</sup> models in a weighted fashion.

<sup>91</sup> In NZ, Cousins et al. (1996) compiled geotechnical site class estimates for all strong-motion sta-  
<sup>92</sup> tions. The site class estimates were based on NZ Standards Association (1992). The geotechnical  
<sup>93</sup> data used to develop these estimates, however, was sparse and/or tentative at best (e.g. based on  
<sup>94</sup> geologists' estimates without any invasive explorations). Most station information, moreover,  
<sup>95</sup> was assumed based solely on regional geology maps.

<sup>96</sup> Destegul et al. (2009) developed a site amplification map for NZ based on geology maps. The  
<sup>97</sup> map was then validated by comparing map polygons against 687 accelerograph locations where  
<sup>98</sup> site class was available.<sup>a)</sup> The Destegul et al. (2009) map was developed before completion

---

<sup>a)</sup>The method of site class assessment for these accelerograph locations in Destegul et al. (2009) was unclear. The work is attributed to Cousins and is unpublished; it may be a continuation of the Cousins et al. (1996) report wherein most site class estimates were derived from geology maps.

99 of the QMAP project (most recent all-NZ geologic map: [GNS Science, 2016](#)) and therefore  
100 based on geologic data from several scales ranging from 1:25,000 to 1:1,000,000. [Destegul et al.](#)  
101 ([2009](#)) also made pragmatic assumptions about the nature of soils confined in narrow valleys:  
102 class C was assigned wherever “soil units were narrow and bounded on one or both sides by  
103 weak rock.” [Destegul et al. \(2009\)](#) also “buffered” polygons with abrupt changes, *e.g.* inserting a  
104 narrow region of class C between polygons of classes B and D.

105 [Perrin et al. \(2015\)](#) published the first  $V_{s30}$  map for NZ by assigning NZS 1170.5 ([King et al.,](#)  
106 [2004](#)) site class categories using QMAP ([GNS Science, 2016](#)) categories. The resulting map  
107 provided distinct ranges of  $V_{s30}$  corresponding to the well-known site categorization scheme.  
108 The [Perrin et al. \(2015\)](#) map also took steps to account for varying  $V_{s30}$  at the edge of geologic  
109 basins, applying ad-hoc assumptions about lateral extents of basin edges and the dip angle of  
110 rock beneath basins. Such decisions were fairly arbitrary by necessity; *e.g.*, dip angle assumption  
111 was driven in part by data resolution. The work drew extensively on data for geology correlations  
112 with  $V_{s30}$  based on the detailed California data in [Borcherdt \(1994\)](#).

## 113 APPROACH AND WORK FLOW

114 The present study incorporates many of the techniques applied by others in the prior work  
115 discussed above including geology-, slope-, and terrain-based proxy variables, and geostatistics  
116 for honoring local  $V_{s30}$  data. Our work builds on these methods but additionally adheres to a  
117 Bayesian framework for unifying first-order (prior)  $V_{s30}$  and uncertainty ( $\sigma$ ) estimates gleaned  
118 from models for other regions with field data in NZ. We also apply more advanced geostatistical  
119 methods ([Worden et al., 2018](#)) than the typical regression kriging approach. Compared with  
120 previous  $V_{s30}$  models for NZ which have only used geology proxy variables, the present study  
121 adds slope and terrain proxy variables, Bayesian statistics and geostatistics.

122 The work flow is summarized in Figure 1 with a box for each incremental model developed.  
123 Every model assumes lognormal  $V_{s30}$  and is therefore completely specified by two maps, a

<sup>124</sup> median  $V_{s30}$  map and a lognormal standard deviation ( $\sigma$ ) map.

<sup>125</sup> A geology-based model is generated by generating a simplified geologic map with categories  
<sup>126</sup> consistent with Ahdi et al. (2017b) and applying their  $V_{s30}$  estimates from Alaska. A terrain-  
<sup>127</sup> based model is developed based on DEM-based “terrain categories,” derived from first-order  
<sup>128</sup> properties of the DEM (local slope, convexity and texture) (Iwahashi and Pike, 2007). Terrain  
<sup>129</sup> categories are generated for NZ and assigned  $V_{s30}$  values using the terrain-based estimates of  
<sup>130</sup> Yong et al. (2012). Each of these two “prior” models are updated with  $V_{s30}$  data for NZ in a  
<sup>131</sup> standard Bayesian framework with lognormal conjugate priors and unknown standard deviation  
<sup>132</sup> (Gelman et al., 2014, section 3.3) . The geology-based model is modified to incorporate weak  
<sup>133</sup> correlations between topographic slope and  $V_{s30}$ , similarly to Thompson et al. (2014); Ahdi et al.  
<sup>134</sup> (2017b); Parker et al. (2017).

<sup>135</sup> Sample variograms are generated using the residuals corresponding to each of the two constituent  
<sup>136</sup> models, and theoretical variograms are fitted for forward prediction. Two geostatistical interpola-  
<sup>137</sup> tion methods are applied to each of the constituent models: conventional regression kriging (RK)  
<sup>138</sup> and a multivariate approach (Worden et al., 2018) (“MVN”) with a novel additional contribution  
<sup>139</sup> to automatically and intuitively reduce undesired extrapolation across geologic discontinuities.  
<sup>140</sup> Finally, a “weighted” model is generated as a statistically weighted combination of the MVN  
<sup>141</sup> models.

<sup>142</sup> [Figure 1 about here.]

## <sup>143</sup> DATA SOURCES

<sup>144</sup> The information sources used for model development include a NZ geologic map, topographic  
<sup>145</sup> map, and  $V_{s30}$  data, as described below.

146 SURFACE GEOLOGY AND TOPOGRAPHIC DATA

147 The geologic map, also referred to as QMAP (“quarter-million” for the map scale), is a recent  
148 compilation of existing surface geologic maps for regions across NZ ([GNS Science, 2016](#)). The  
149 data is in vector form, with coordinates indicating each corner of approximately 55,000 polygons,  
150 each with several metadata fields. Geologic boundaries (polygon edges) are accurate to  $\pm 250$   
151 metres. The topographic data source is a digital elevation model (DEM) developed by LRIS  
152 NZ ([Landcare Research New Zealand, 2010; Barringer et al., 2002](#)) and is available at 25m  
153 resolution. We use the DEM resampled to more manageable 100m and 270m resolutions for the  
154 terrain and slope proxy variables, respectively.

155  $V_{s30}$  DATA

156  $V_{s30}$  data are from three datasets as shown in Figure 2: [Kaiser et al. \(2017\)](#); [McGann et al.](#)  
157 ([2017](#)) and a compilation of recent surface-wave-based field investigations in Canterbury. The  
158 word “data,” as opposed to “measurements,” is deliberate; most  $V_{s30}$  input data are not from  
159 direct measurements. The following paragraphs describe each dataset in turn, followed by the  
160 assumed measurement uncertainties. A complete tabular listing of input  $V_{s30}$  data is provided as  
161 an electronic supplement.

162 [Kaiser et al. \(2017\)](#) compiled a list of  $V_{s30}$  values and/or estimates for NZ strong motion stations.  
163 Along with  $V_{s30}$  measurements, qualitative quality rankings (ranging from Q1=best to Q3=worst)  
164 were provided for each station. Q1 data include both well-constrained surface wave-based  
165 methods and invasive methods. Q2 data may be based on well-established local correlations,  
166 similar nearby sites, and/or well-constrained near-surface  $V_s$  profiles that do not necessarily reach  
167 30m depth. Q3 values are based solely on estimates, either from preexisting national scale maps  
168 ([Perrin et al., 2015](#)) or from geologists’ estimates. We do not use Q3 data for  $V_{s30}$  modelling  
169 here. A number of possible duplicate observations were encountered in the [Kaiser et al. \(2017\)](#)  
170 dataset, so a preliminary screening was implemented where duplicates (identified as  $V_{s30}$  points

<sup>171</sup> within two metres of another observation) were removed.

<sup>172</sup> McGann et al. (2017) used CPT (cone penetration testing) correlation-derived  $V_{s30}$  values  
<sup>173</sup> (McGann et al., 2015) to produce a regional  $V_{s30}$  map for the Christchurch area. For our purposes  
<sup>174</sup> these data were too numerous (by comparison with the data available for the rest of NZ), and  
<sup>175</sup> without downsizing would overwhelm the impact of data elsewhere. Accordingly, the McGann  
<sup>176</sup> et al. (2017) data were resampled from 7,402 points to 280 points by overlaying an arbitrary 1km  
<sup>177</sup> grid and selecting the nearest McGann data point to each gridpoint (Figure 2). This decision was  
<sup>178</sup> arbitrary but reasonable for the goals of the model development.

<sup>179</sup> [Figure 2 about here.]

<sup>180</sup> The “surface waves” dataset refers to a compilation of  $V_{s30}$  data from several surface wave  
<sup>181</sup> analysis-based site investigations performed following the 2010-2011 Canterbury earthquake  
<sup>182</sup> sequence (Cox et al., 2011; Wood et al., 2011; Wotherspoon et al., 2013; Van Houtte et al., 2014;  
<sup>183</sup> Wotherspoon et al., 2016; Wood et al., 2017; Teague et al., 2018). Cox et al. (2011) compiled a  
<sup>184</sup> rapid preliminary report of surface wave testing (MASW) performed in Christchurch after the  
<sup>185</sup> 2010-2011 earthquake sequence along with inverted site profiles from which  $V_{s30}$  estimates were  
<sup>186</sup> obtained. Wood et al. (2011) and Wotherspoon et al. (2013) performed passive and active surface  
<sup>187</sup> wave testing at 13 strong motion stations in and near Christchurch city to obtain  $V_s$  profiles from  
<sup>188</sup> which  $V_{s30}$  were obtained.  $V_{s30}$  values reported in Wood et al. (2011) have been adjusted in the  
<sup>189</sup> present work based on updated analysis of the dispersion curves. Van Houtte et al. (2014) used  
<sup>190</sup>  $V_{s30}$  values for seven hard sites in an investigation about the role of local seismic attenuation  
<sup>191</sup> ( $\kappa$ ). The sites’  $V_{s30}$  were inferred from surface wave-based methods using frequencies  $\geq 14$   
<sup>192</sup> Hz, meaning that  $V_{s30}$  profiles could not be developed down to 30m and assumptions were  
<sup>193</sup> used. Wotherspoon et al. (2016) report seven  $V_{s30}$  values derived from surface wave based site  
<sup>194</sup>  $V_s$  profiles measured in conjunction with work by Cox et al. (2014). Deschenes et al. (2018)  
<sup>195</sup> provide these seven plus two additional  $V_s$  profiles. Wood et al. (2017) evaluated liquefaction  
<sup>196</sup> case histories at a number of sites with well-characterized  $V_s$  profiles obtained from surface wave

197 based methods. Not all  $V_{s30}$  profiles extended to 30m depth; in a few cases where the profiles  
198 were within a few metres of 30m, Wotherspoon (pers. comm.) extended the profiles for  $V_{s30}$   
199 estimates. Teague et al. (2018) developed  $V_s$  profiles at 14 Christchurch sites with deep and  
200 complex interbedded geology. Many candidate profiles were generated from the experimental  
201 dispersion curves to study epistemic uncertainty associated with the inversions.

202 **MEASUREMENT UNCERTAINTY ( $\sigma_{\text{MEAS.}}$ )**

203 In the subsequent  $V_{s30}$  model development we use measurement uncertainty quantified as  $\sigma_{\text{meas.}}$ ,  
204 assuming lognormal distributions. Kaiser et al. (2017) give approximate subjective uncertainty  
205 quantities of 10% and 20% for Q1 and Q2 data respectively. Accordingly, we assigned lognor-  
206 mal standard deviations for measurement uncertainty ( $\sigma_{\text{meas.}}$ ) of 0.1 and 0.2 respectively. All  
207 other data are assigned  $\sigma_{\text{meas.}} = 0.2$ . These uncertainties are broadly consistent with both the  
208 uncertainties provided by Kaiser et al. (2017), and with the general findings of Moss (2008) for  
209  $V_{s30}$  determined from surface wave-based measurements and geology correlations.

210 **GEOLOGY- AND TERRAIN-BASED MODELS**

211 The development of the geology- and terrain-based models (in the context of Figure 1) is  
212 summarized in Figures 3 and 4 ( $V_{s30}$  and  $\sigma$ , respectively). The individual maps/panes shown  
213 in these figures are recalled and discussed in detail in the subsequent subsections as they arise.  
214 These maps show only the Canterbury region for clarity; the models for all of NZ are available  
215 in the electronic supplement.

216 [Figure 3 about here.]

217 [Figure 4 about here.]

218 **GEOLOGY-BASED PRIOR MODEL**

219 To assign geology-based median  $V_{s30}$  and uncertainty estimates, all map locations first need to  
220 be assigned one of a finite number of geologic categories. This is accomplished by simplifying  
221 or “flattening” the underlying geology map—which is itself a compilation of the work of many  
222 geologists, containing more than 55,000 individual polygons, each with several dimensions of  
223 textual metadata—into a simpler, “one-dimensional” map (Figure 5a). The geology categories of  
224 Ahdi et al. (2017a,b) are selected so that direct comparisons can be made, for straightforward  
225 Bayesian updating, and because the geology categories chosen by Ahdi are more specific and  
226 discriminating than in many similar studies (and hence more flexible for our application). Ahdi  
227 et al. generated  $V_{s30}$  maps for the Pacific Northwest region of North America (2017a) and Alaska  
228 (2017b). The categories are enumerated in Table 1.

229 The 18 geology categories of Ahdi et al. are followed with two exceptions. Firstly, groups 2 and  
230 3 (Fraser river) are discarded because of their regional geologic specificity. Secondly, groups 7  
231 and 13 (fine and coarse floodplain deposits, respectively) are merged because of the difficulty  
232 in distinguishing between the two on the basis of the NZ surface geology metadata. Group 7 is  
233 not populated, and all floodplain deposits are lumped together into category 13. Merging these  
234 categories is likely more appropriate than retaining the fine/coarse distinction, even if it were  
235 practical: Wills and Gutierrez (2009) discuss the reasons that grain size at the ground surface is  
236 not generally correlated with grain size over the entire uppermost 30 metres for young alluvial  
237 deposits, meaning that for these depositional environments, little or no correlation is expected  
238 between fine and coarse surface deposits.

239 [Figure 5 about here.]

240 The categorization process is implemented using primarily text-based search. Extensive and  
241 iterative examination of the metadata fields is performed manually, and five metadata categories  
242 are chosen based on the richness of information they carry related to  $V_{s30}$ . The first-order

243 geology-based  $V_{s30}$  model—using  $V_{s30}$  and  $\sigma$  values from Ahdi et al. (2017b)—is shown in  
244 Figure 3a ( $V_{s30}$ ) and 4a ( $\sigma$ ). The values assigned to these models correspond to the “prior”  
245 columns in Table 1.

246 **TERRAIN-BASED PRIOR MODEL**

247 Iwahashi and Pike (2007) proposed a method of using a DEM to automatically generate a map of  
248 terrain categories that roughly correlate with surface geology and/or geomorphology. The method  
249 relies on successive discrimination based upon three spatial fields, respectively: topographic  
250 slope, local convexity, and texture. The approach follows a “nested-means” logic (*i.e.* the  
251 categories are divided first on the basis of comparing local slope to mean slope for the entire  
252 map domain, then subdivided by convexity and texture respectively.) It can be used to generate  
253 terrain classes comprised of 8, 12, or 16 unique categories, with 12- and 16-class discrimination  
254 achieved by repeated subdivision of the regions with low topographic slope. We followed Yong  
255 et al. (2012) in choosing the 16-category implementation; this decision is justified presently  
256 in the context of the Bayesian updating step. One advantage of the Iwahashi and Pike (2007)  
257 method (in contrast to other similar classification schemes) is that it is “unsupervised,” *i.e.* no  
258 decisions need be made regarding the values (slope, convexity, texture) defining boundaries  
259 between terrain classes. One downside of the approach is that these boundary values are based  
260 on computing the mean for the map domain, yielding different maps for different domains. This  
261 restriction implies that application-specific calibrations are important.

262 We generated rasters for the Iwahashi and Pike (2007) categories in NZ using the DEM resampled  
263 at 100 metres (our final target map resolution).<sup>b)</sup> By contrast, the worldwide classification by  
264 Iwahashi and Pike (2007) used the 1 km resolution Shuttle Radar Topography Mission digital  
265 elevation model (SRTM30 DEM) (Farr et al., 2007). Because of both the DEM resolution and the  
266 domain-dependent nature of the algorithm, our categories do not precisely match the Iwahashi

<sup>b)</sup>The resolution of the final product is not 100m in the strictest sense, but in a "neighborhood average" sense. Slope and convexity are both computed for a 9-cell moving window, and texture is computed using a 10-cell radius.

267 and Pike (2007) SRTM30-derived categories for NZ. We expect this to be of trivial consequence,  
268 both because of the roughly scale-invariant (fractal) features of landforms, and because of  
269 subsequent reductions in epistemic uncertainty owing to the Bayesian updating process in the  
270 next step (*i.e.* the application-specific calibration).

271 The newly generated 100m-resolution terrain category map is shown in Figure 5(b). The Iwahashi  
272 and Pike (2007) categories are listed with short descriptions in Table 2. It must be emphasized  
273 that these are mnemonic labels affixed to quantities that are purely statistical in nature; the labels  
274 do not carry strong geologic meaning. The mnemonic labels are borrowed directly from Yong  
275 et al. (2012).  $V_{s30}$  and  $\sigma$  values from Yong et al. (2012) (see “prior” columns in Table 2) are  
276 assigned to their corresponding categories from Figure 5(b), analogously to the geology-based  
277 model, to produce the prior terrain-based model (Figures 3b and 4b).

278 [Table 1 about here.]

279 [Table 2 about here.]

## 280 BAYESIAN UPDATING USING NZ DATA

281 This section describes the development of posterior models as shown in the second row of  
282 Figure 1. Among the desirable aspects of Bayesian analysis are its amenability to simple verbal  
283 descriptions, its agnosticism of the distinction between subjective beliefs and objective data  
284 (allowing for a crude “better-than-nothing” formulation of prior beliefs encoded as wide or  
285 noninformative prior distributions), and its tendency to demand an explicit accounting of all  
286 statistical assumptions being made. Thorough discussions of the semantics of Bayesian theory  
287 and analysis are found in D’Agostini (2003); McElreath (2015). In this work Bayesian updating  
288 is performed on both the geology- and terrain-based models presented in the previous section.  
289 Since Bayesian updating is intuitive, the updating is described here at a high level first. The next  
290 section discusses some important implementation details.

291 **Bayesian updating summary**

292 To facilitate comparison with the prior, the ratio of the posterior to prior models'  $V_{s30}$  estimates  
293 is presented in Figure 3 (e,f). <sup>c)</sup> The greatest changes to the geology model resulting from the  
294 updating process are in the floodplain deposits (group 13) beneath Christchurch city, where  
295  $V_{s30}$  data are abundant and generally lower than for the comparable for soil units from Ahdi  
296 et al. (2017b) in Alaska. The  $\sigma$  maps corresponding to the posterior geology- and terrain-based  
297 models are shown in Figures 4c and 4d.

298 The median and standard deviation of each group's prior and posterior  $V_{s30}$  distributions are  
299 consisely summarized alongside the data in Figure 6 (geology) and 7 (terrain). Here the gray  
300 circles represent the NZ  $V_{s30}$  data and are plotted with transparency and a small horizontal "jitter"  
301 to reduce overplotting. The red lines and dots to the left of each group indicate the prior model  
302 median and +/- 1 standard deviation. The blue lines and dots to the right of each group represent  
303 the posterior model median and +/- 1 standard deviation. Note that the prior and posterior median  
304 ( $V_{s30}$ ) and standard deviation ( $\sigma$ ) correspond to the  $V_{s30}$  and  $\sigma$  values in the maps in Figures 3  
305 and 4.

306 [Figure 6 about here.]

307 [Figure 7 about here.]

308 The general behaviour illustrated in Figures 6 and 7 can be summarized as follows: If the prior  
309 model and the data are in stark disagreement, the posterior model predicts a "compromise"  
310 between the data and prior, and a larger posterior  $\sigma$  reflects this underlying uncertainty. By  
311 contrast, if the NZ  $V_{s30}$  data are tightly clustered around the prior prediction, the posterior  $\sigma$  will  
312 tend to be lower than the prior  $\sigma$  to reflect the additional confidence conferred on the model by the  
313 data. This can be seen in Figure 4 (e and f). While much of the map area is negative (indicating a

---

<sup>c)</sup>Panes c and e in Figure 3 refer to the posterior geology with slope modification, discussed subsequently; this is nearly the same as the posterior geology without slope modification which is omitted to save space. The difference between the two is trivial enough to disregard momentarily.

314 reduction in uncertainty from the prior to posterior models), not every category shows a reduced  
315  $\sigma$ . These are categories where NZ data were sparse, and/or not in good agreement with the prior  
316  $V_{s30}$  values from Ahdi et al. (2017b) and/or Yong et al. (2012).

317 The primary difference between panes e and f in Figures 3 and 4 is the lower absolute value  
318 in pane f of both figures. This shows that prior distributions were—at least in the Canterbury  
319 region—in better agreement with NZ data for the terrain-based (Yong et al., 2012)  $V_{s30}$  model  
320 than for the geology-based (Ahdi et al., 2017b) model. This might suggest that the geology  
321 of California is more similar to NZ than the geology of Alaska. Alternately, or additionally, it  
322 might simply demonstrate that terrain-based  $V_{s30}$  estimation is inherently less subjective than  
323 geology-based estimation.

324 In the course of developing the terrain-based model, the question arises as to whether the 16-  
325 category implementation of Iwahashi and Pike (2007) is superior to the 12- or 8-category options.  
326 In particular, there is the concern of whether or not a 16-category-model is “overfitting” available  
327 data by comparison with a comparable 12- or 8-category model. We decided to use 16 categories  
328 for two reasons. First, this choice simplifies the updating process. Because Yong et al. (2012)  
329 used 16 categories, deviating from this choice would require us to “lump” the Yong categories  
330 together, introducing another form of epistemic uncertainty related to the categories’ spatial  
331 distributions and sampling distributions in California and NZ. The second reason for using 16  
332 categories is that it offers the best predictive power given the quantity of data available and its  
333 distribution across terrain categories. The motivation for using less than 16 categories would arise  
334 only if the 8- and 12-category schemes result in fewer sparsely populated categories. But on the  
335 contrary, this is not the case. The Iwahashi and Pike (2007) categories are numbered in order of  
336 decreasing topographic slope, and—due to various clustering influences already discussed—the  
337 steep categories (generally mountainous/rocky) tend to be more sparsely populated than the  
338 shallow categories (generally low-lying/soil). The fortuitous result of the interaction of these  
339 various statistical biases is that the 16-category implementation gives superior discrimination in

340 flatter regions, where  $V_{s30}$  data are plentiful, without “splitting” the data within any bins that are  
341 sparsely populated in the 8- or 12-category versions.

342 **Bayesian implementation details**

343 The preceding section presented the Bayesian updating process at a high level; this section  
344 contains details of the implementation that are useful for replicating our model development,  
345 including our application-specific implementation choices. This discussion adheres to the  
346 notation in [Gelman et al. \(2014, pp. 67-69\)](#).

347 The updating presumes normally distributed data (*i.e.*  $\ln(V_{s30})$  is normal) with a conjugate prior  
348 distribution and unknown variance ( $\sigma^2$ ). Two application-specific decisions are included in the  
349 discussion below. The first relates to the relative weighting between the prior and the data; the  
350 second relates to an arbitrary minimum threshold  $\sigma$  applied before the updating process in order  
351 to avoid overfitting caused by clustered observational data.

352 A transformation of variables is performed to perform lognormal updating using the procedure  
353 for normal distributions from [Gelman et al. \(2014\)](#). Each lognormal distribution is completely  
354 specified by  $\mu$  (the mean of  $\ln V_{s30}$ ) and  $\sigma$  (lognormal standard deviation).  $\sigma$  is updated assuming  
355 a scaled inverse-chi-squared (Inv- $\chi^2$ ) marginal posterior density.

356 As discussed, prior distributions are selected for each group using the  $\mu$  and  $\sigma$  values reported  
357 in [Ahdi et al. \(2017b\)](#) for Alaska and in [Yong et al. \(2012\)](#) for California (Tables 1 and 2,  
358 respectively). That is to say, we view the conclusions of [Ahdi et al. \(2017b\)](#); [Yong et al. \(2012\)](#)  
359 (and  $V_{s30}$  data from Alaska and California) as sound bases for first-order estimates of  $V_{s30}$  in NZ,  
360 given no other data.

361 For brevity we have relegated a summary of the mathematical implementation ([Gelman et al.,](#)  
362 [2014](#)) of the approach to an electronic appendix. In our implementation we made application-  
363 specific assumptions about the relative weighting between the prior and data. These assumptions  
364 are reflected in the choice of initial values for the integer-valued counter variables  $v_0$  and  $\kappa_0$  (we

365 set  $v_i = \kappa_i$  which is common but not required.) In Gelman et al. (2014)  $\kappa_0 = v_0$  represent the  
366 “number of observations” contained in the prior. Quotation marks are used to emphasize that  
367 the prior is not generally a uniform dataset but may be a degree of belief or a combination of  
368 qualitative and quantitative elements. The meaning of  $\kappa_0$  and  $v_0$  is a step removed from reality:  
369 these parameters do not reflect the process applied by Ahdi et al. (2017b) in its entirety, nor do  
370 they directly represent the number of physical  $V_{s30}$  measurements in their work. (Indeed, the  
371 clustering of  $V_{s30}$  data may mean that unique values of  $\kappa_0$  and  $v_0$  should be chosen for each  
372 geologic group, but this is not explored further.) In any case, the choice of  $\kappa_0$  and  $v_0$  represents a  
373 subjective decision about the appropriateness of applying Ahdi et al. (2017b) models in a NZ  
374 context. This “appropriateness” reflects issues of an epistemic nature such as (*e.g.*) the quality of  
375  $V_{s30}$  data used in the prior study, the authors’ choices of geologic grouping criteria, the degree to  
376 which NZ and California geologic deposits are essentially similar or different, and regional or  
377 discipline-specific differences in the naming and classification of various geologic deposits by  
378 various geologists.

379 We choose  $\kappa_0 = v_0 = 3$  for this application. A geologic category with  $n = 3$  datapoints in NZ  
380 would therefore give a posterior distribution based on equal weighting between the data and the  
381 prior distribution. Considering the available data, this choice yields a reasonable compromise  
382 between the prior distributions and the data (Figures 6 and 7).

383 A second application-specific decision is made regarding the  $\sigma$  values in the priors. Given that  
384 the observational data are sparse and clustered, there is the risk that posterior  $\sigma$  is artificially low  
385 because of the influence of clustered  $V_{s30}$  data when the data may not be representative of the  
386 entire map domain. To address this, we impose an arbitrary minimum value of 0.5 (natural log  
387 scale) on the model priors before applying the updating algorithm. The intent of this threshold  $\sigma$   
388 constraint is to avoid overfitting in categories where  $V_{s30}$  data are few, but of similar value (*e.g.*  
389 terrain category T05 in Figure 7). The results of this can be seen in category groupings that are  
390 poorly constrained, such as in geology groups G08 and G12 in Figure 6, as well as terrain groups

391 T01 through T05 and T12 in Figure 7. (Prior distributions shown are those from Tables 1 and 2;  
392 the minimum  $\sigma = 0.5$  modification is only visible in the posterior distributions).

393 **TOPOGRAPHIC SLOPE-BASED MODIFICATION TO POSTERIOR GEOLOGY MODEL**

394 Following Thompson et al. (2014); Ahdi et al. (2017b); Parker et al. (2017), the geology-  
395 based  $V_{s30}$  model can be refined by capturing any slope-dependence in the observed  $V_{s30}$  of the  
396 various geologic groupings. To examine the slope- $V_{s30}$  correlations,  $V_{s30}$  values were plotted  
397 against topographic slope, computed (Horn, 1981) at both 9 and 30 arcsecond resolutions for  
398 direct comparison with Thompson et al. (2014).<sup>d)</sup> Thompson and Wald (2012) found that the  
399 correlations associated with the coarser 30-arcsecond resolution were slightly better than for  
400 9 arcseconds, and Thompson et al. (2014) postulated that this might arise because of spurious  
401 elevation correlations from nongeomorphic features (vegetation and built infrastructure) resolved  
402 at the finer resolution in California. By contrast, in NZ the finer resolution correlations are the  
403 same or slightly better than those for the coarser resolution, and we postulate that the scale  
404 of NZ's built environment is unlikely to impact these correlations in the same way as (*e.g.*)  
405 California. Our slope- $V_{s30}$  correlation is based on 9-arcsecond slopes.

406 Figure 8 shows the correlations of 9 arcsecond (270m) slope with  $V_{s30}$  for four geology groups.  
407 Standard least-squares fitting and likelihood testing yield linear relationships along with the  $\pm 2$   
408 standard deviations (95% confidence) bounds shown as dotted lines. Only the geology groups  
409 with definitively positive trends are shown, and only these groups are selected for slope-dependent  
410 modification (*i.e.* the remaining geology groups are modeled with no slope dependence).

411 [Figure 8 about here.]

412 The limits of the linear fits in Figure 8 are shown in Table 3. These upper and lower slope ( $\nabla$ )  
413 limits were selected to define continuous piecewise-linear functions (after log transformation)

---

<sup>d)</sup>Actual resolutions of the 9 and 30 arcsecond maps are about 270 and 900 meters respectively; north-south arc-seconds and east-west arc-seconds are not similar in NZ.

<sup>414</sup> defined by  $(\nabla_0, V_{s30,0})$  and  $(\nabla_1, V_{s30,1})$ :

$$\ln V_{s30} = \begin{cases} \ln V_{s30,0} & \nabla \leq \nabla_0 \\ \ln V_{s30,0} + \frac{\ln(\nabla/\nabla_0)}{\ln(\nabla_1/\nabla_0)} \ln \left( \frac{V_{s30,1}}{V_{s30,0}} \right) & \nabla_0 < \nabla \leq \nabla_1 \\ \ln V_{s30,1} & \nabla_1 < \nabla \end{cases} \quad (1)$$

<sup>415</sup> from which any value of slope yields a single  $V_{s30}$  value that is constrained not to extrapolate  
<sup>416</sup> beyond the range of observed data. A more sophisticated approach might use logistic curves  
<sup>417</sup> rather than piecewise-linear, and might expend more effort in determining whether the highest  
<sup>418</sup> and lowest observed slopes are the “best” places to define the inflection points of the function,  
<sup>419</sup> but we view our approach as simple, objective and effective.

<sup>420</sup> [Table 3 about here.]

<sup>421</sup> In Bayesian parlance the application of slope correlations to the model is another instance of  
<sup>422</sup> “updating,” where a model is improved based on the use of data that were not accounted for  
<sup>423</sup> beforehand. The posterior  $V_{s30}$  predictions are given by the piecewise-linear trends.

<sup>424</sup> The results of slope-based correction represent a small change from the posterior geology model  
<sup>425</sup> (by comparison with the prior geology model; compare  $\sigma$  values in Table 3 to those in Table  
<sup>426</sup> 1). The slope-based  $V_{s30}$  and  $\sigma$  maps are visually indistinguishable from the posterior geology  
<sup>427</sup> model, so only the log of the ratio of the two models’ median  $V_{s30}$  predictions is presented here  
<sup>428</sup> (Figure 9). The slope-updated geology model is shown in Figure 3c ( $V_{s30}$ ) and Figure 4c ( $\sigma$ ).  
<sup>429</sup> Hereafter, we refer to this model as simply the “geology-based” model.

<sup>430</sup> [Figure 9 about here.]

#### <sup>431</sup> COMPARING GEOLOGY- AND TERRAIN-BASED MODELS

<sup>432</sup> Quantitative comparisons of the two posterior model predictions are presented in Figures 10  
<sup>433</sup> (mapped prediction comparison for all NZ) and 11 (per-datapoint residual comparison). In the

434 map-based prediction (Figure 10), it is evident that the places where the models disagree most  
435 (e.g., the volcanic Taupo & Rotorua area in central North Island, indicated by a box in Figure  
436 10) are places characterized by unique surface geology and geomorphology, and where there is  
437 consequently a unique pairing of geology and terrain categories. The model can benefit from  
438 data in these locations.

439 Figure 11 shows normalized geology and terrain model residuals  $\zeta$ ,

$$\zeta_i = \frac{\ln V_{s30\text{obs},i} - \ln V_{s30\text{pred},i}}{\sigma_i} \quad (2)$$

440 on the  $x$ - and  $y$ -axes respectively.  $\sigma_i$  values correspond to the values in Figure 4 (c and d). The  
441 residuals are roughly symmetrically scattered across the 1:1 line, suggesting that there is no  
442 systematic advantage of the geology-based model over the terrain-based model, or vice-versa,  
443 after Bayesian updating. Coloring of the points is by measured  $V_{s30}$  and highlights some biases  
444 in individual categories, particularly rock categories which are poorly represented.

445 [Figure 10 about here.]

446 [Figure 11 about here.]

## 447 GEOSTATISTICS APPLICATION TO POSTERIOR MODELS

448 The geology and terrain models developed in the previous section predict  $V_{s30}$  as functions of  
449 surface geology, local topographic slope, and/or terrain proxy variables. A shortcoming of these  
450 models is that they do not offer improved predictions in the vicinity of existing measurements.  
451 The tool for addressing this issue is geostatistics. Two geostatistical approaches are applied to  
452 the geology and terrain models in this section. The first and simpler of the two is regression  
453 kriging (RK). The second is the so-called “multivariate normal” (MVN) method (Worden et al.,  
454 2018). Both methods are presented and discussed for comparative purposes; the MVN approach  
455 is ultimately chosen as the superior method for reasons discussed presently.

456 In the following sections, first a concise review is given of the broad aspects of geostatistical  
457 approaches, wherein quantifiable geospatial phenomena are modeled as stationary random  
458 processes, and the parameterization (*i.e.* variance and autocorrelation) of the stationary random  
459 processes are inferred from variogram analysis. Next the aspects of the MVN approach that  
460 depart from kriging, and its advantages for this work, are highlighted. Aspects of variogram  
461 and correlation function selection are discussed in the context of the  $V_{s30}$  map. The results of  
462 applying both RK and MVN are presented. The problem of interpolating/extrapolating residuals  
463 for a lognormal process such as  $V_{s30}$ , which can result in unreasonably high  $V_{s30}$  estimates in  
464 some situations, is discussed. Finally, we present two unique approaches we implemented to  
465 ameliorate these issues in a more-or-less automated way that requires few subjective decisions  
466 for implementation.

467 **PRELIMINARIES**

468 This discussion follows notation from [Diggle and Ribeiro \(2007\)](#). The most common assumption  
469 underlying geostatistical methods is that spatial fluctuations in an earth science system (*e.g.*  $V_{s30}$ )  
470 can be modeled as a Gaussian random process,  $S(x)$ , a 2-dimensional function wherein a set  
471 of observations  $S(x_1, \dots, x_n)$  for  $n$  locations  $x_1, \dots, x_n$  is assumed to be drawn from a multivariate  
472 Gaussian (normal) distribution. A Gaussian random process is defined completely by its mean  
473 function  $\mu(x) = E[S(x)]$  and its covariance function  $\gamma(x, x') = \text{Cov}\{S(x), S(x')\}$ . If the mean is  
474 constant and the covariance structure is formulated solely as a function of distance (*i.e.*  $\mu(x) = \mu$   
475 and  $\gamma(x, x') = \gamma(u)$  where  $u = x - x'$ ), then the process is known as the special case of a stationary  
476 random function (SRF). Usually SRF are also assumed isotropic, *i.e.*  $\gamma(u) = \gamma(\|u\|) = \gamma(u)$   
477 where  $\|\cdot\|$  designates Euclidean distance. The variance of an SRF is constant:  $\sigma^2 = \gamma(0)$ .  
478 Hereafter we discuss only functions of scalar  $u$  rather than vector difference  $u$  as our application  
479 assumes isotropy as is typical for regional applications (*e.g.* [Thompson et al., 2014](#)).  
480 For the stationary case, the variogram function is an alternative representation of the covariance

481 function:  $V(x, x') = \frac{1}{2}\text{Var}\{S(x) - S(x')\}$ . This reduces to  $V(u) = \sigma^2\{1 - \rho(u)\}$ , where  $\rho(u) =$   
 482  $\frac{\gamma(u)}{\sigma^2}$  is the correlation function, again for the stationary case. The correlation function is 1 for  
 483  $u = 0$  (for the typical application where the nonspatial component of randomness is zero) and  
 484 decreases monotonically to approach zero asymptotically with increasing  $u$ .

485 The variogram is useful for both for interpreting observed spatial processes, and for generating  
 486 predictions using models with SRF. The formulation above is known as the theoretical variogram.  
 487 For using observational data to parameterize a geostatistical model, the sample variogram  
 488 or empirical variogram can be obtained and then used to guide selection of the theoretical  
 489 variogram. Observational data  $Y$  are assumed to be of the form  $Y_i = S(x_i) + Z_i$ , where  $Z_i$  are  
 490 mutually independent and identically distributed with zero mean and variance  $\tau^2$ . The sample  
 491 variogram is

$$V_Y(u_{ij}) = \frac{1}{2}\text{E}[(Y_i - Y_j)^2], \quad (3)$$

492 and the functional form for fitting a theoretical variogram is

$$V_Y(u) = \tau^2 + \sigma^2\{1 - \rho(u)\}. \quad (4)$$

493 This formulation is more general than the one introduced above, with the addition of the  $\tau^2$  term.  
 494 The intercept,  $\tau^2$ , is known as the nugget variance. It represents the nonspatial component  
 495 of randomness in the process—requiring a dual interpretation of physical meaning, which we  
 496 discuss presently to contrast conventional kriging with MVN.  $\sigma^2$  is the signal variance. The  
 497 asymptotic value,  $\tau^2 + \sigma^2$ , is known as the sill and is the variance of the observed process  $Y(x)$ .  
 498 For the common special case of  $\tau^2 = 0$  the sill is equivalent to the signal variance. The range of  
 499 the variogram is the distance  $u$  beyond which there is no change in  $V(u)$ .

## 500 EMPIRICAL VARIOGRAMS

501 Two theoretical variograms, one each for the geology- and terrain-based models, are selected by  
 502 fitting to empirical data. One of the simplest and most common functional forms is chosen for

503 the theoretical variograms, the exponential model

$$V(u) = \exp\left(-\frac{u}{\phi}\right) \quad (5)$$

504 where  $\phi$ , the range or shape parameter, has units of distance. The practical range, the value of  $u$   
505 for which  $\rho(u) = 0.05$ , is approximately  $3\phi$  (Diggle and Ribeiro, 2007).

506 The empirical variograms here are produced with normalized residuals ( $\zeta_i$ , Equation 2) in  
507 lieu of observations  $S(x_i)$  above. Normalization ensures homoscedacity of the residuals (e.g.,  
508 Jayaram and Baker, 2009) which is necessary to ensure the geostatistical assumption of Gaussian  
509 stationarity.

510 Because of the highly clustered nature of the available  $V_{s30}$  data, we explore several subsets of  
511 the data to assess the sensitivity of the fitted theoretical variogram to the data underlying the  
512 empirical variogram. The selection of subsets is motivated by a hypothesis that the large scale  
513 and relative spatial uniformity of the Canterbury plains—the largest alluvial deposit in NZ—may  
514 yield a variogram with a range that is longer than appropriate for other geographic regions. Thus  
515 we examine subsets with various degrees of exclusion of Christchurch and/or Canterbury  $V_{s30}$   
516 data. These are summarized in Table 4.

517 [Table 4 about here.]

518 Preliminary examination of the empirical variograms led us to conclude that the geology model  
519 yielded fairly “well-behaved” variograms while the terrain model variograms were considerably  
520 less smooth. We postulated that since the geology model incorporates continuously varying  
521 slope correction functions for a few hand-selected geologic categories, whereas the terrain  
522 model handles subdivides regions based on discrete slope bins using the Iwahashi and Pike  
523 (2007) method, the latter might be showing some arbitrary stochasticity in the pairwise empirical  
524 variogram ordinates, particularly in categories T15 & T16 (which are the most prevalent in  
525 the Canterbury plains and other wide alluvial basins). The terrain subsets were modified by

526 removing all points from terrain categories T15 & T16 and the resulting theoretical fits were  
527 improved.

528 The final selections for the geology- and terrain-based model variograms are summarized in  
529 Figure 12. For each of the five subsets, a logarithmically-spaced binning scheme was chosen and  
530 pairwise variogram ordinates were averaged within each bin to generate an unbiased estimator  
531 of the theoretical variogram. 95% confidence intervals obtained by bootstrapping are shown  
532 for each bin. The judgment-based weighting factors shown in Table 4 were used to fit one final  
533 exponential variogram each to the geology- and terrain-based models by weighted least-squares  
534 minimization (Pebesma, 2014). Notably, inspection of the plots reveals that the variogram is fairly  
535 insensitive to the data subsetting schemes we evaluated. The effective ranges for the geology  
536 and terrain variograms are 4.2 and 3.0 km, respectively. We note that since the variograms are  
537 calibrated to the normalized model residuals their differences are not necessarily attributable to  
538 physical meaning.

539 [Figure 12 about here.]

## 540 KRIGING

541 “Predicting” unobserved values of a spatial process on the basis of geostatistics entails interpolat-  
542 ing the observed values with interpolation weights derived from the variogram, and assuming that  
543 the mean of the process tracks the observed values. This process is known as kriging. Regression  
544 kriging (RK, *e.g.* Thompson et al., 2014) is an approach that makes the stationarity assumption  
545 useful for a wider variety of models where  $\mu = \mu(x)$  and is not constant across the problem  
546 domain. For RK, consistent with variogram development, the kriging is done on normalized  
547 observation residuals,  $\zeta_i$ , assuming that the residual surface is the mean surface of a SRF.

548 The kriged geology- and terrain-based  $V_{s30}$  maps are shown in Figure 13. While the new  $V_{s30}$   
549 estimates appear reasonable in general, we wish to highlight one location where this is not  
550 the case, a datapoint in Rakaia (identified by a box in Figure 13a). Here the geology model

551 predicts unrealistically high  $V_{s30}$  values because a high-valued observation (presumably reflective  
552 of loess deposits, category G11) appears to be located near a geologic boundary on a lower-  
553 valued polygon (G06 or G13) yielding a higher normalized residual than may be appropriate.  
554 The smoothed residual surface is multiplicative rather than additive, owing to the lognormal  
555 assumption, and consequently exerts a great influence over the area. (A tiny region in the vicinity  
556 of this error is colored gray indicating the color scale is “clipped,” *i.e.* the model predicts a  
557 median  $V_{s30}$  in excess of 1000 m/s.)

558 [Figure 13 about here.]

559 This issue is exemplary of geostatistical problems that may occur wherever observational data  
560 happen to be located on the “wrong side” of a high contrast geologic boundary (between cemented  
561 loess and unconsolidated sediments in this example, or between rock and soil generally). In  
562 the next section we propose a novel method of handling such errors that is effective at reducing  
563 undesirable cross-boundary extrapolation in lieu of the time-consuming and subjective alternative  
564 approach (*i.e.* manually identifying and relocating problematic observations or map boundaries  
565 wherever this issue arises).

566 The kriged uncertainty ( $\sigma$ ) maps are shown in Figure 14 and, as expected, approach zero near  
567 measurements because the nugget of the variogram is zero. The “background”  $\sigma$ , in regions  
568 where there are no nearby observations, reverts to the input  $\sigma$  (*i.e.* Figure 4).

569 [Figure 14 about here.]

570 The kriged residual in Figure 14 suggests that the baseline model may systematically underpredict  
571  $V_{s30}$  in the Canterbury plains, west of Christchurch city. This is not unexpected; in the simplified  
572 geology categories (Figure 5a) there is no discrimination between the surface geology beneath  
573 Christchurch and in the plains to the west. But the geology of the coastal basin consists of  
574 interbedded layers from alternating deposits of river/alluvial and marine deposits which may  
575 suggest that surface geology alone (presuming it is representative of less than 30m depth) is

576 inadequate as a proxy for  $V_{s30}$ . This dilemma with young alluvial deposits is common for  $V_{s30}$   
577 estimates that rely on surface geology (e.g. [Wills and Gutierrez, 2009](#)).

578 Related to this issue is the dominance of the Christchurch city data. An alternative approach  
579 would be to generate another model by choosing sparser sampling distance for the Christchurch  
580 data and thereby decreasing the weight accorded to urban data. Ultimately this decision represents  
581 a tradeoff that should be informed by application. Our view is that  $V_{s30}$  data density is generally  
582 correlated with population density for good reason, and that for engineering applications it is  
583 acceptable to allow data to drive the model, irrespective of the unavoidable spatial clustering.

#### 584 MVN (MULTIVARIATE NORMAL) METHOD

585 [Worden et al. \(2018\)](#) present a more generalized geostatistical approach than kriging, which we  
586 refer to as the “multivariate normal” (MVN) method. The primary advantage of this method in  
587 our work (compared with RK) is that it allows for assumptions about measurement uncertainty  
588 to be applied on a location-by-location basis.

589 In conventional kriging, if the nonspatial component of variance—the nugget—is zero, then the  
590 interpolated mean surface is constrained so as to match the observational values at their respective  
591 locations. A nonzero nugget means the interpolated surface need not honor the data precisely.  
592 The nonzero nugget has a dual physical meaning: On the one hand, the tendency of the modeled  
593 system to exhibit spatial discontinuities; and on the other hand, measurement uncertainty. In  
594 practice these causes are rarely disentangled, because many collocated measurements would be  
595 required. A nonzero nugget in the variogram is formally equivalent to modeling the measurement  
596 process as a Gaussian spatial process with a discontinuity at the origin of the correlation function.  
597 But importantly, in conventional kriging, a nonzero nugget is the same across the problem  
598 domain. By contrast, in the MVN approach, explicit assumptions about measurement uncertainty  
599 are enforced via “correlation adjustment factors” that can be assigned to measurements on an  
600 individual basis. The output  $\sigma$  can be different at different data points, in accordance with the

601 respective input measurement uncertainties. Correspondingly, individual observations exert a  
602 “pull” on the interpolated surface that is inversely correlated with measurement uncertainty. The  
603 variogram nugget is thus effectively localized for data from different sources. (The MVN results  
604 reduce to being equivalent to the kriging results with zero nugget for the special case where all  
605 measurement uncertainties are set to zero).

606 We assigned each input datum a lognormal measurement uncertainty,  $\sigma_{\text{meas.}}$ . These values were  
607 chosen following the discussion in Section 3 and are tabulated in the electronic supplement.  
608 Assigning  $\sigma_{\text{meas.}}$  is not trivial or objective (*e.g.*, McElreath, 2015). However, given that measure-  
609 ment uncertainty is a “nuisance parameter” in the parlance of Gelman et al. (2014, pp. 63-64), it  
610 is intuitive to expect that the chosen value of  $\sigma_{\text{meas.}}$  becomes insignificant in regions of dense  
611 data (*e.g.* Christchurch city) and therefore impacts the model most strongly in regions with little  
612 data. Moreover, this framework allows for future model refinements on the basis of more and  
613 better observational data.

#### 614 **Correcting overpredictions from cross-boundary extrapolation**

615 The issue of “cross-boundary extrapolation” was discussed earlier as it pertains to regression  
616 kriging. Here we propose a novel solution that entails modifying the correlation function  $\rho(u)$   
617 for every unique pairwise combination of locations in the problem domain. Toward this end, a  
618 coefficient is introduced, the “covariance reduction factor” (CRF) with a value between 0 and  
619 1. The CRF is a function of the “difference” between two points of interest. Qualitatively, in  
620 this context “difference” implies the appropriateness in general of assuming nonzero correlation  
621 between two locations, irrespective of their separation distance. Quantitatively,  $0 \leq \text{CRF} \leq 1$ .

622 A proxy variable needs to be selected for choosing CRF given two locations,  $x_1$  and  $x_2$ . We chose  
623 the ratio of the two corresponding model  $V_{s30}$  predictions, since this reflects the most pertinent  
624 information contained in both the geology- and terrain-based models. Alternatives could be  
625 chosen, for example the ratio of slopes alone, or a more complex assessment of the geologic map.

626 The function for CRF is chosen as

$$\text{CRF} = \exp \left( -a \left| \ln \frac{V_{s30_1}}{V_{s30_2}} \right| \right) \quad (6)$$

627 which ensures  $\text{CRF} = 1$  for  $V_{s30_1} = V_{s30_2}$  with exponential decay as the ratio of the baseline  
628  $V_{s30}$  estimates decreases. We selected  $a = 1.5$  based on the heuristic that the decay function  
629 ought to yield  $\text{CRF} \approx 0.1$  (*i.e.* almost no correlation) for the ratio  $\frac{V_{s30_1}}{V_{s30_2}} \approx 5$ ; see Figure 15. This  
630 corresponds to an arbitrary reference hypothetical where  $V_{s30_1} = 1000\text{m/s}$  and  $V_{s30_2} = 200\text{m/s}$ ,  
631 *i.e.* “rock” and “soil.” This heuristic is necessarily arbitrary and reflects an intuition about the  
632 complexity of geologic processes that depart from idealized assumptions underlying geostatistical  
633 methods. While there are alternative approaches to representing this complexity, the proposed  
634 method appears to handle the issues well globally and relies on the selection of only a single  
635 parameter. This is desirable for updating the model in a more or less automated fashion in the  
636 future, as additional data are incorporated into the model.

637 [Figure 15 about here.]

638 To apply the above modification to the MVN method, CRF is computed in accordance with  
639 Equation 6. Then Equation 7 from Worden et al. (2018) is adjusted by defining a modified  
640 correlation coefficient:

$$\rho'_{Y_i Y_j} = \rho_{Y_i Y_j} \text{CRF}_{Y_i Y_j} \quad (7)$$

## 641 Results

642 The result of the covariance-weighted MVN applications are presented in Figure 16. The ratios  
643 of Figure 16 to the regression kriging models (Figure 13) are shown in Figure 17.  $\sigma$  for the  
644 covariance-weighted MVN model is shown in Figure 18.

645 Three items are noteworthy in comparing the performance of the RK and MVN methods (Figure  
646 17):

- 647 1. In regions where data are dense, such as Christchurch city, the models yield virtually the

648 same predictions. The collocation of many  $V_{s30}$  datapoints results in a smoothing effect  
649 whose details are largely unchanged by the addition of measurement uncertainty.

- 650 2. In regions where data are sparse, such as the rural Canterbury plains, the models are  
651 slightly different in the vicinity of data because of the way RK and MVN differ in handling  
652 measurement uncertainty. The baseline  $V_{s30}$  estimate (geology or terrain model) has no  
653 effect at the precise location of an observation for RK, whereas the observation and baseline  
654 model estimates are combined in the MVN method, with the proportional influence being  
655 a function of the assumed measurement uncertainty.
- 656 3. The localized overprediction in the Rakaia area, noted previously, has been resolved by the  
657 MVN method. The different degrees of covariance applied across the geologic boundaries  
658 in this area are clear in Figure 17(a).

659 [Figure 16 about here.]

660 [Figure 17 about here.]

661 [Figure 18 about here.]

## 662 MERGING GEOLOGY & TERRAIN MODELS

663 To summarize the preceding discussion, the MVN method is more general and more statistically  
664 sound than regression kriging because it allows pointwise assignment of measurement uncertainty,  
665 albeit at the expense of longer computation time. The “covariance reduction” method proposed  
666 above is a transparent and automated way of handling potentially misleading extrapolation of  
667 normalized residuals across geologic boundaries. Consequently we present the MVN model  
668 versions as superior to the more conventional regression kriging results. In this section we  
669 combine the geology- and terrain-based models into a final model.

670 Having conditioned each constituent model on available data, and applied geostatistical inter-  
671 polation, we see no reason to favor one model over the other and assign each a weighting of

672 0.5. The decision to weight the two models equally, rather than favoring one over the other,  
 673 is motivated by the assumption that aspects of the geology- and terrain-based approaches are  
 674 mutually complementary. For example, the geology categories can convey information pertaining  
 675 to  $V_{s30}$  that may not be readily discriminated by the terrain categories, but the terrain data are  
 676 more spatially uniform, less subject to human interpretation, and may indicate finer detail in local  
 677  $V_{s30}$  variations and/or more accurate velocity contrast boundaries, particularly in areas where the  
 678 geology map is derived from digitizing older regional-scale maps.

679 The weighting is done as

$$V_{s,30\text{final}} = \exp \left( \sum_{i=1}^n w_i \ln V_{s,30_i} \right) \quad (8)$$

680 where in general  $\sum_i w_i = 1$  and in the present application (combining two equally-weighted  
 681 models),  $n = 2$ ,  $w_1 = w_2 = 0.5$ . The combining of uncertainty represents a “mixture problem:” If  
 682 the constituent models predict similar  $V_{s30}$  and have small  $\sigma$  then the resulting  $\sigma$  is a compromise  
 683 between the two input  $\sigma$  values. On the other hand if the two models predict significantly  
 684 different  $V_{s30}$  values then the combined uncertainty is high, even if the constituent models have  
 685 low  $\sigma$ :

$$\sigma_{\text{final}}^2 = \sum_{i=1}^n w_i \left( (\ln V_{s,30_i} - \ln V_{s,30\text{final}})^2 + \sigma_i^2 \right) \quad (9)$$

686 where again  $n = 2$ ,  $w_1 = w_2 = 0.5$ . Once more, this is functionally equivalent to Bayesian  
 687 updating with equal weighting of “prior” (geology-based model) and “data” (terrain-based  
 688 model) (e.g., [McElreath, 2015](#)). The weighted, final model median  $V_{s30}$  estimate and lognormal  
 689  $\sigma$  are presented in Figure 19.

690 [Figure 19 about here.]

## 691 CONCLUSION

692 A  $V_{s30}$  model for NZ has been developed. The salient features of the model include a fine  
 693 (100m) resolution; making use of both geology and terrain covariates; a consistent, local

694  $V_{s30}$  data inclusion via transparent and readily updateable Bayesian framework; lognormal  
695 standard deviations alongside median  $V_{s30}$  estimates; and a novel modification to the MVN  
696 method (Worden et al., 2018) that reduces covariance for observation-prediction pairs that  
697 cross geologic/geomorphic boundaries, yielding heuristically sensible  $V_{s30}$  estimates near these  
698 boundaries. The model represents an improvement over recent  $V_{s30}$  models for New Zealand  
699 which have used geology proxy variables but have not quantified uncertainty or employed  
700 geostatistical methods.

701 The model can be updated relatively easily and it is expected that recent field work performed in  
702 Nelson, Auckland and Wellington will be incorporated into an incremental update in the near  
703 future.

704 The code repository for this work is available on GitHub at <https://github.com/fostergeotech/>  
705 [Vs30\\_NZ](#).

## 706 DATA AND RESOURCES

707 All data used in this paper came from published sources listed in the references, with the  
708 exception of the  $V_{s30}$  values from Wood et al. (2011) which, as noted previously, have been  
709 adjusted in the present work based on review of the dispersion curve interpretation which is  
710 unpublished. The tabulated  $V_{s30}$  values in the electronic supplement corresponding to the Wood  
711 et al. (2011) work are not the same as the original published work.

## 712 ACKNOWLEDGMENTS

713 The authors are grateful to the University of Canterbury's QuakeCoRE ground motion simulation  
714 computational team for helpful advice and support, especially Daniel Lagrava, Sung Bae and  
715 Viktor Polak. We appreciate useful feedback from Seokho Jeong, Alan Yong, David Wald, Sean  
716 Ahdi, Chris Wills, Bruce Worden and Eric Thompson. This work was supported by a Marsden  
717 Fund research grant and by QuakeCoRE, a New Zealand Tertiary Education Commission-funded

718 Centre. This is QuakeCoRE publication number 0346.

719

## REFERENCES

- 720 Ahdi, S. K., Stewart, J. P., Ancheta, T. D., Kwak, D. Y., and Mitra, D., 2017a. Development of Vs  
721 Profile Database and Proxy-Based Models for Vs30 Prediction in the Pacific Northwest Region of  
722 North America. *Bulletin of the Seismological Society of America* **107**, 1781–1801.
- 723 Ahdi, S. K., Stewart, J. P., Kwak, D. Y., Ancheta, T. D., and Mitra, D., 2017b. Proxy-Based Vs30  
724 Prediction in Alaska Accounting for Limited Regional Data. *PBD-III, Performance Based Design in*  
725 *Geotechnical Earthquake Engineering, Vancouver, British Columbia* .
- 726 Allen, T. I. and Wald, D. J., 2007. *Topographic slope as a proxy for seismic site-conditions (Vs30)*  
727 *and amplification around the globe. Open-File Report 2007-1357*, United States Geological Survey  
728 (USGS).
- 729 Barringer, J., Pairman, D., and McNeill, S., 2002. *Development of a high resolution digital elevation*  
730 *model for New Zealand. Landcare Research Contract Report LC0102/170*, Landcare Research.
- 731 Borcherdt, R. D., 1994. Estimates of site-dependent response spectra for design (methodology and  
732 justification). *Earthquake spectra* **10**, 617–653.
- 733 Cousins, W. J., Perrin, N., McVerry, G., Herreford, R., and Porritt, T., 1996. *Ground conditions at*  
734 *strong-motion recording sites in New Zealand*, vol. 96. Institute of Geological & Nuclear Sciences  
735 Ltd. 244pp.
- 736 Cox, B. R., Wood, C. M., Deschenes, R., and Pearson, M., 2011. University of Arkansas Preliminary  
737 Data Report for: Surface Wave Testing in Christchurch, New Zealand Affected by the New Zealand  
738 Earthquakes. Work Performed in Conjunction with the NSF RAPID Proposal: Liquefaction and its  
739 Effects on Buildings and Lifelines in the February 22, 2011 Christchurch, New Zealand Earthquake.
- 740 Cox, B. R., Wood, C. M., and Teague, D. P., 2014. Synthesis of the UTexas1 Surface Wave Dataset  
741 Blind-Analysis Study: Inter-Analyst Dispersion and Shear Wave Velocity Uncertainty. In *Geo-*  
742 *Congress 2014 Technical Papers: Geo-Characterization and Modeling for Sustainability (GSP 234)*.  
743 ASCE (American Society of Civil Engineers). doi:10.1061/9780784413272.083.
- 744 D'Agostini, G., 2003. *Bayesian reasoning in data analysis: A critical introduction*. World Scientific,  
745 329 pp.
- 746 Deschenes, M. R., Wood, C. M., Wotherspoon, L. M., Bradley, B. A., and Thomson, E., 2018. Develop-  
747 opment of Deep Shear Wave Velocity Profiles in the Canterbury Plains, New Zealand. *Earthquake*  
748 *Spectra* **34**, 1065–1089. doi:10.1193/122717EQS267M.
- 749 Destegul, U., Dellow, G., and Heron, D., 2009. A ground shaking amplification map for New Zealand.  
750 *Bulletin of the New Zealand Society for Earthquake Engineering* **42**, 122.
- 751 Diggle, P. J. and Ribeiro, P. J., Jr, 2007. *Model-based geostatistics*. Springer Series in Geostatistics.

- 752 Springer, 228 pp.
- 753 Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Ro-  
754 driguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank,  
755 D., and Alsdorf, D., 2007. The Shuttle Radar Topography Mission. *Reviews of Geophysics* **45**.  
756 doi:10.1029/2005RG000183.
- 757 Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B., 2014. *Bayesian  
758 data analysis*. 3rd edition edn. CRC Press, Boca Raton, FL, 639 pp.
- 759 GNS Science, 2016. 1:250000 Geological Map of New Zealand (QMAP). Last accessed October 2018.
- 760 Horn, B. K., 1981. Hill shading and the reflectance map. *Proceedings of the IEEE* **69**, 14–47.
- 761 Iwashashi, J. and Pike, R. J., 2007. Automated classifications of topography from DEMs by an unsuper-  
762 vised nested-means algorithm and a three-part geometric signature. *Geomorphology* **86**, 409–440.
- 763 Jayaram, N. and Baker, J. W., 2009. Correlation model for spatially distributed ground-motion intensi-  
764 ties. *Earthquake Engineering & Structural Dynamics* **38**, 1687–1708.
- 765 Kaiser, A., Van Houtte, C., Perrin, N., Wotherspoon, L., and McVerry, G., 2017. Site characterisation of  
766 GeoNet stations for the New Zealand strong motion database. *Bull. New Zeal. Soc. Earthq. Eng.* **50**,  
767 39–49.
- 768 King, A., Bull, D., Clifton, C., Dowrick, D., Jury, R., McVerry, G., Moss, P., and O’Leary, A., 2004.  
769 NZS1170.5: Structural design actions - Part 5: Earthquake actions - New Zealand.
- 770 Kwok, O. L. A., Stewart, J. P., Kwak, D. Y., and Sun, P.-L., 2018. Taiwan-Specific Model for Vs30  
771 Prediction Considering Between-Proxy Correlations. *Earthquake Spectra* .
- 772 Landcare Research New Zealand, 2010. New Zealand Digital Elevation Model 25-Metre.
- 773 Lee, C.-T. and Tsai, B.-R., 2008. Mapping Vs30 in Taiwan. *Terrestrial, Atmospheric & Oceanic Sciences*  
774 **19**.
- 775 McElreath, R., 2015. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. CRC  
776 Press, Boca Raton, FL.
- 777 McGann, C., Bradley, B., Taylor, M., Wotherspoon, L., and Cubrinovski, M., 2015. Development of an  
778 empirical correlation for predicting shear wave velocity of Christchurch soils from cone penetration  
779 test data. *Soil Dynamics and Earthquake Engineering* **75**, 66–75. doi:10.1016/j.soildyn.2015.03.023].
- 780 McGann, C. R., Bradley, B. A., and Cubrinovski, M., 2017. Development of Regional  $V_{s30}$  Model  
781 and Typical  $V_s$  Profiles for Christchurch, New Zealand from CPT Data and Region-Specific CPT- $V_s$   
782 Correlation. *Soil Dynamics and Earthquake Engineering* **95**, 48–60.
- 783 Moss, R. E. S., 2008. Quantifying measurement uncertainty of thirty-meter shear-wave velocity. *Bulletin  
784 of the Seismological Society of America* **98**, 1399–1411.
- 785 NZ Standards Association, 1992. New Zealand Standard NZS4203: Code of practice for general struc-  
786 tural design and design loadings for buildings.

- 787 Parker, G. A., Harmon, J. A., Stewart, J. P., Hashash, Y. M., Kottke, A. R., Rathje, E. M., Silva, W. J.,  
788 and Campbell, K. W., 2017. Proxy-based Vs30 estimation in central and eastern North America.  
789 *Bulletin of the Seismological Society of America* **107**, 117–131.
- 790 Pebesma, E. J., 2014. Gstat User's Manual (Version 2.5.1).
- 791 Perrin, N., Heron, D., Kaiser, A., and Van Houtte, C., 2015. Vs30 and NZS 1170.5 site class maps of  
792 New Zealand. In *2015 NZSEE Conference*, pp. 10–12. New Zealand Society for Earthquake Engi-  
793 neering.
- 794 Teague, D., Cox, B., Bradley, B., and Wotherspoon, L., 2018. Development of Deep Shear Wave Ve-  
795 locity Profiles with Estimates of Uncertainty in the Complex Inter-Bedded Geology of Christchurch,  
796 New Zealand. *Earthquake Spectra* **34**, 639–672. doi:10.1193/041117EQS069M.
- 797 Thompson, E. and Wald, D., 2012. Developing Vs30 site-condition maps by combining observations  
798 with geologic and topographic constraints. In *Proc. 15th World Conf. on Earthq. Eng., Lisbon, Por-  
799 tugal*, pp. 24–28.
- 800 Thompson, E. M., Wald, D. J., and Worden, C. B., 2014. A Vs30 Map for California with Geologic  
801 and Topographic Constraints. *Bulletin of the Seismological Society of America* **104**, 2313–2321. doi:  
802 10.1785/0120130312.
- 803 Van Houtte, C., Ktenidou, O.-J., Larkin, T., and Holden, C., 2014. Hard-site  $\kappa_0$  (kappa) calculations for  
804 Christchurch, New Zealand, and comparison with local ground-motion prediction models. *Bulletin of  
805 the Seismological Society of America* **104**, 1899–1913.
- 806 Vilanova, S. P., Narciso, J., Carvalho, J. P., Lopes, I., Quinta-Ferreira, M., Pinto, C. C., Moura, R.,  
807 Borges, J., and Nemser, E. S., 2018. Developing a Geologically Based VS 30 Site-Condition Model  
808 for Portugal: Methodology and Assessment of the Performance of Proxies. *Bulletin of the Seismolog-  
809 ical Society of America* **108**, 322–337.
- 810 Wald, D. J., McWhirter, L., Thompson, E., and Hering, A. S., 2011. A new strategy for developing  
811 VS30 maps. In *Proc. 4th Int. Effects of Surface Geology on Seismic Motion Symposium*.
- 812 Wills, C. and Clahan, K., 2006. Developing a map of geologically defined site-condition categories for  
813 California. *Bulletin of the Seismological Society of America* **96**, 1483–1501.
- 814 Wills, C. and Gutierrez, C., 2009. Investigation of Geographic Rules for Improving Site-Conditions  
815 Mapping. In *Digital Mapping Techniques '08—Workshop Proceedings U.S. Geological Survey Open-  
816 File Report 2009-1298*, pp. 205–216. U.S. Geological Survey.
- 817 Wood, C. M., Cox, B. R., Green, R. A., Wotherspoon, L. M., Bradley, B. A., and Cubrinovski, M.,  
818 2017. Vs-Based Evaluation of Select Liquefaction Case Histories from the 2010-2011 Canterbury  
819 Earthquake Sequence. *Journal of Geotechnical and Geoenvironmental Engineering* **143**, 04017066.  
820 doi:10.1061/(ASCE)GT.1943-5606.0001754.
- 821 Wood, C. M., Cox, B. R., Wotherspoon, L. M., and Green, R. A., 2011. Dynamic site characterization of  
822 Christchurch strong motion stations. *Bulletin of the New Zealand Society for Earthquake Engineering*

- 823      **44**, 195–204.
- 824      Worden, C. B., Thompson, E. M., Baker, J. W., Bradley, B. A., Luco, N., and Wald, D. J., 2018. Spatial  
825      and Spectral Interpolation of Ground-Motion Intensity Measure Observations. *Bulletin of the*  
826      *Seismological Society of America* **108**, 866–875.
- 827      Wotherspoon, L., Bradley, B., Thomson, E., Wood, C., Deschenes, M., and Cox, B., 2016. *Dynamic*  
828      *Site Characterisation of Canterbury Strong Motion Stations using Active and Passive Surface Wave*  
829      *Testing - Version 1.0 - June 2016. Earthquake Commission Report Project No. 14/663*, University of  
830      Auckland.
- 831      Wotherspoon, L., Orense, R., Bradley, B., Cox, B., Wood, C., and Green, R., 2013. *Geotechnical*  
832      *characterization of Christchurch strong motion stations - Version 1.0 - September 2013. Earthquake*  
833      *Commision Report Project No. 12/629*, University of Auckland, Faculty of Engineering.
- 834      Yong, A., 2016. Comparison of Measured and Proxy-Based Vs30 Values in California. *Earthquake*  
835      *Spectra* **32**, 171–192.
- 836      Yong, A., Hough, S. E., Iwahashi, J., and Braverman, A., 2012. A terrain-based site-conditions map of  
837      California with implications for the contiguous United States. *Bulletin of the Seismological Society*  
838      *of America* **102**, 114–128.

## AUTHOR CONTACT

- 840 Kevin Foster, Brendon Bradley, Chris McGann:  
841 Room E312, E304, E445 (respectively)  
842 Department of Civil and Natural Resources Engineering  
843 University of Canterbury  
844 69 Creyke Road  
845 Christchurch  
846 New Zealand  
847 Liam Wotherspoon:  
848 Engineering Block 1 - Bldg 401 Level 11, Room 1105 20 Symonds Street Auckland 1010 New  
849 Zealand

**List of Tables**

851      1	Geology categories with $n$ =number of observations per category, and prior and posterior (slope-adjusted) $V_{s30}$ and $\sigma$ values for each. . . . .	36
852		
853      2	Terrain categories with $n$ =number of observations per category, and prior and posterior (slope-adjusted) $V_{s30}$ and $\sigma$ values for each. . . . .	37
854		
855      3	Slope adjustment details for categories G05, G06, G07, G09. . . . .	38
856		
856      4	Data subsets used for variogram fitting. . . . .	39

**Table 1.** Geology categories with  $n$ =number of observations per category, and prior and posterior (slope-adjusted)  $V_{s30}$  and  $\sigma$  values for each.

ID	Description	$n$	$V_{s30}$ (m/s)		$\sigma$	
			(pri.)	(post.)	(pri.)	(post.)
G01	peat	9	161	163	0.52	0.30
G04	artificial fill	11	198	273	0.31	0.28
G05	fluvial & estuarine deposits	11	239	200	0.87	0.44
G06	alluvium & valley sediments	25	323	271	0.36	0.24
G08	lacustrine (incl. glaciolacustrine)	0	326	326	0.14	0.50
G09	beach, bar, dune deposits	70	339	204	0.65	0.23
G10	fan deposits	5	360	247	0.34	0.34
G11	loess	4	376	473	0.38	0.35
G12	glacigenic sediments (drift & outwash)	0	399	399	0.30	0.50
G13	flood deposits	252	448	197	0.43	0.20
G14	glacial moraines & till	0	453	453	0.51	0.51
G15	undifferentiated sediments & sedimentary rocks	0	455	455	0.55	0.55
G16	terrace deposits & old alluvium	2	458	335	0.76	0.60
G17	volcanic rocks & deposits	0	635	635	0.99	0.99
G18	crystalline rocks (igneous & metamorphic)	4	750	691	0.64	0.45

**Table 2.** Terrain categories with  $n$ =number of observations per category, and prior and posterior (slope-adjusted)  $V_{s30}$  and  $\sigma$  values for each.

ID	Description	$n$	$V_{s30}$ (m/s)		$\sigma$	
			(pri.)	(post.)	(pri.)	(post.)
T01	Well dissected alpine summits, mountains, etc.	0	519	519	0.35	0.50
T02	Large volcano, high block plateaus, etc.	0	393	393	0.42	0.50
T03	Well dissected, low mountains, etc.	0	547	547	0.47	0.50
T04	Volcanic fan, foot slope of high block plateaus, etc.	0	459	459	0.35	0.50
T05	Dissected plateaus, etc.	2	402	324	0.31	0.41
T06	Basalt lava plain, glaciated plateau, etc.	7	345	301	0.28	0.31
T07	Moderately eroded mountains, lava flow, etc.	3	388	536	0.42	0.38
T08	Desert alluvial slope, volcanic fan, etc.	3	374	515	0.32	0.38
T09	Well eroded plain of weak rocks, etc.	4	497	284	0.35	0.36
T10	Valley, till plain, etc.	6	349	317	0.28	0.33
T11	Eroded plain of weak rocks, etc.	2	328	267	0.27	0.40
T12	Desert plain, delta plain, etc.	0	297	297	0.29	0.50
T13	Incised terrace, etc.	15	500	217	0.50	0.25
T14	Eroded alluvial fan, till plain, etc.	8	209	242	0.17	0.31
T15	Dune, incised terrace, etc.	166	363	199	0.28	0.21
T16	Fluvial plain, alluvial fan, low-lying flat plains, etc.	170	246	202	0.22	0.21

**Table 3.** Slope adjustment details for categories G05, G06, G07, G09.

ID	$\nabla_0$	$\nabla_1$	$V_{s30,0}$	$V_{s30,1}$	$\sigma_{f(\nabla)}$
G04	0.0141	0.0596	242	418	0.12
G05	0.0020	0.0452	171	228	0.11
G06	0.0004	0.1316	252	275	0.38
G09	0.0003	0.1171	183	239	0.15

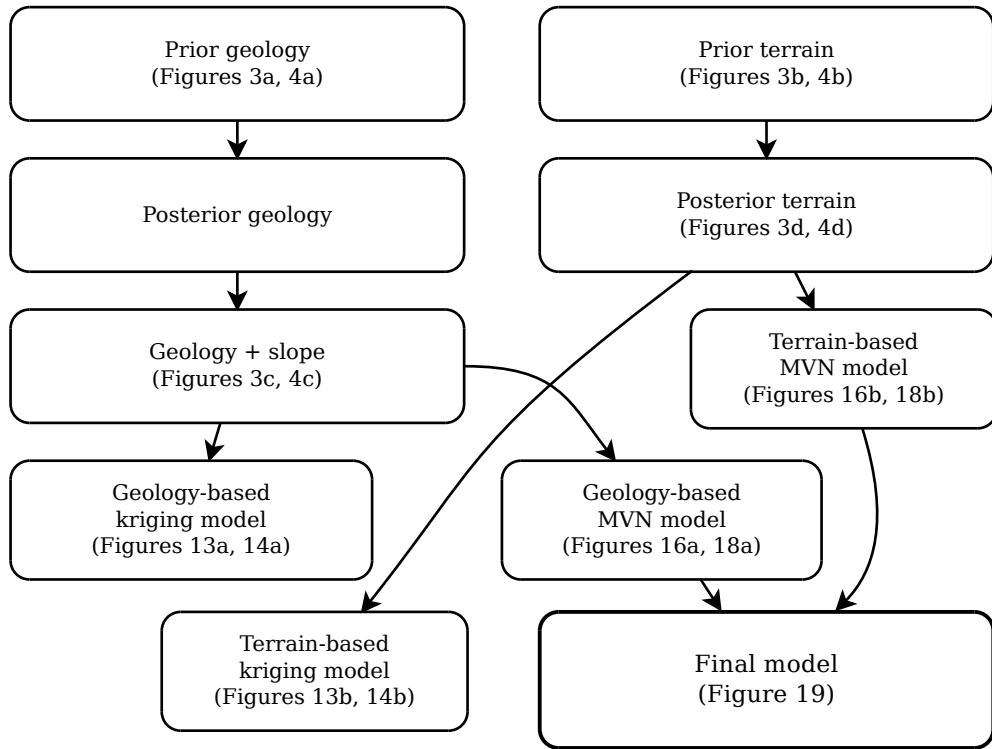
**Table 4.** Data subsets used for variogram fitting.

Subset	Description	n	Weighting
GS1	Kaiser et al. (2017) Q1 & Q2; no Canterbury data	49	0.5
GS2	Kaiser et al. (2017) Q1 & Q2 and surface-wave-based data	127	0.25
GS3	McGann et al. (2017) data, (resampled to 1km)	266	0.25
TS1	Same as GS1 but with terrain categories T15 & T16 removed	29	0.5
TS2	Same as GS2 but with terrain categories T15 & T16 removed	42	0.5

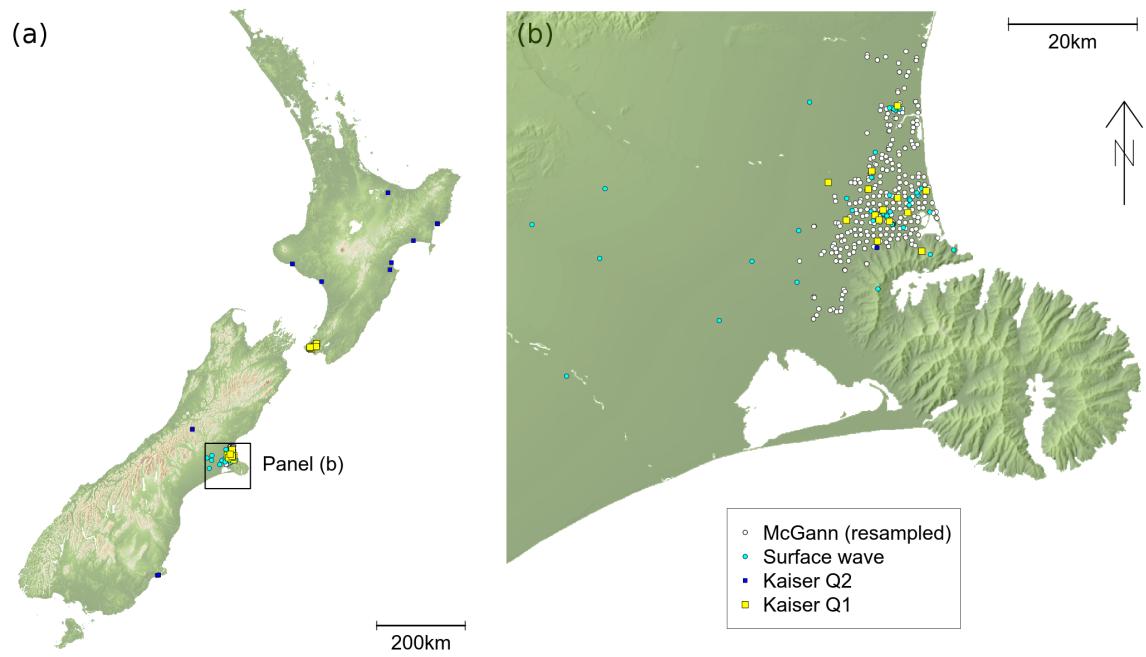
## List of Figures

859	1	Relationships among various model components. Arrows indicate incremental model refinements. . . . .	42
860	2	(a) Data locations for all of NZ and (b) inset showing Canterbury region for detail. McGann et al. (2017) data were downsampled based on closest proximity to an arbitrary 1km grid (white circles). Data labeled “Surface wave” (cyan circles) comprise several Canterbury-area surface wave investigations enumerated in the text. The majority of Kaiser et al. (2017) are “Q3” and are not used herein. Kaiser et al. (2017) Q2 (medium quality) and Q1 (highest quality) data are shown as small blue and large yellow squares, respectively. . . . .	43
861	3	Illustrative regional (Canterbury) $V_{s30}$ maps for various stages of model devel- opment: (a) Prior geology model; (b) prior terrain model; (c) posterior geology model with slope-based adjustment; (d) posterior terrain model. Comparison of prior to posterior models: (e) ratio of pane (c) to pane (a); (f) ratio of pane (d) to pane (b). Point overlays in panes (c) and (d) show $V_{s30}$ measurements used for updating. . . . .	44
862	4	$\sigma$ (sigma) maps for various stages of model development: (a) Prior geology model; (b) prior terrain model; (c) posterior geology model with slope-based adjustment; (d) posterior terrain model. Comparison of prior to posterior models: (e) difference between panes (c) and (a); (f) difference between panes (d) and (b). . . . .	45
863	5	(a) Map of geology categories from Ahdi et al. (2017) as applied to NZ. (b) Map of terrain categories from Iwahashi and Pike (2007) as applied to NZ. Terrain colors are selected for direct comparison with Yong et al. (2012). . . . .	46
864	6	Bayesian updating for geology-based model. The colored dots and lines represent median and +/- 1 standard deviation for lognormal prior (left, red) and posterior (right, blue) distributions. The prior is based on Ahdi et al. (2017b). The minimum $\sigma = 0.5$ criterion is applied for groups G08 and G12. Gray dots represent data used for updating. . . . .	47
865	7	Bayesian updating for terrain-based model. The colored dots and lines represent median and +/- 1 standard deviation for lognormal prior (left, red) and posterior (right, blue) distributions. The prior is based on Yong et al. (2012). The minimum $\sigma = 0.5$ criterion is applied for groups T01, T02, T03, T04, and T12. Gray dots represent data used for updating. . . . .	48

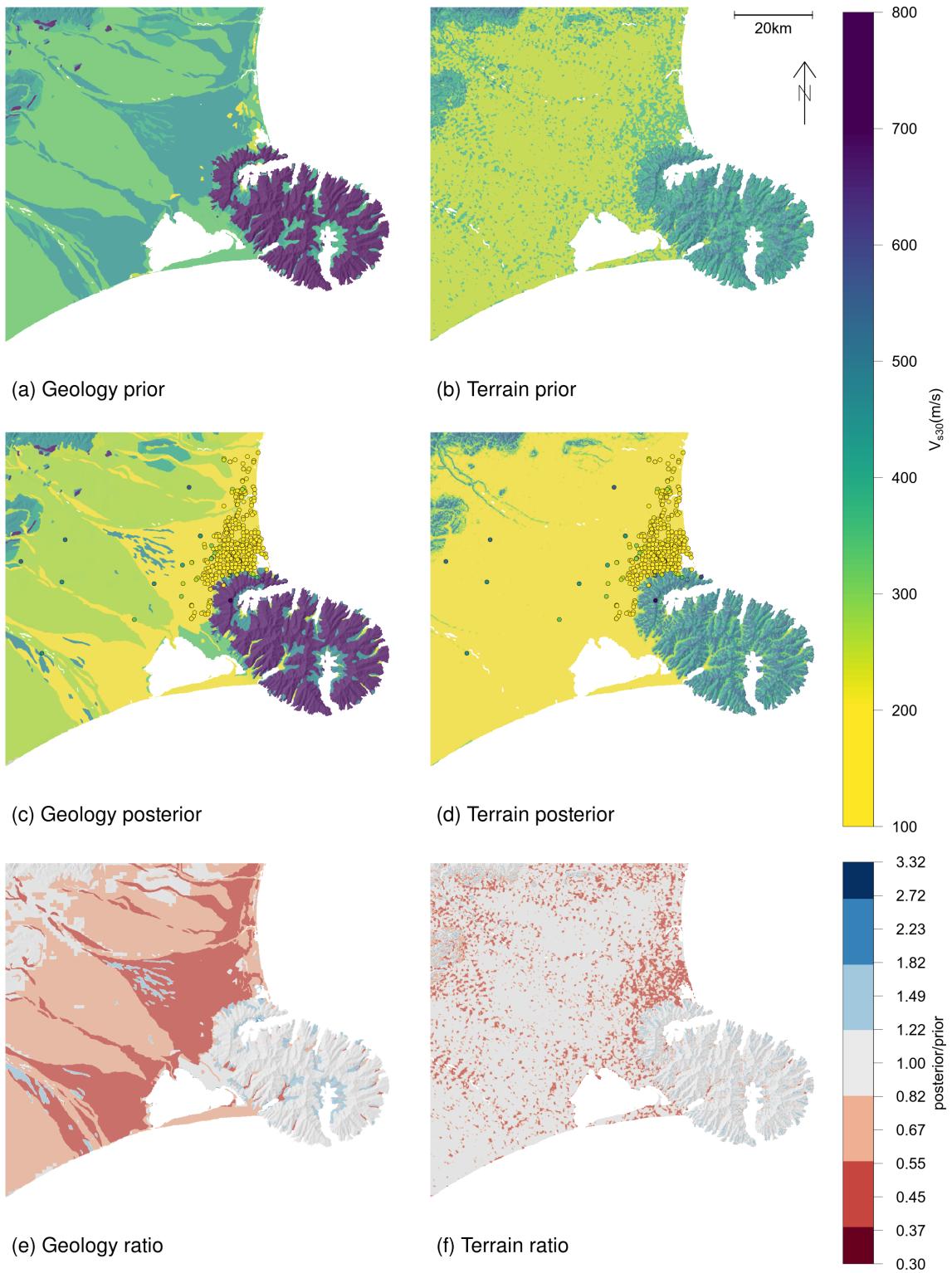
891	8	Slope dependence of geology-based $V_{s30}$ model. Slope resolved from 9 arcsecond 892 ( $\sim 270\text{m}$ ) digital elevation model. Only the four geology categories with positive 893 trends are shown. Dotted lines indicate $\pm 2$ standard deviations (95% confidence 894 bounds) on slope fit. . . . .	49
895	9	Ratio of slope+geology model (Figure 3c) to posterior geology model (not 896 shown) median estimates. Note the small magnitude on the graphic scale indicating 897 much smaller model change by comparison with the posterior geology 898 update. . . . .	50
899	10	Map comparing median $V_{s30}$ for posterior geology- and terrain-based models. 900 The Taupo and Rotorua region (boxed) is an example of a region with unique 901 interaction between the geology- and terrain-based categories that may merit 902 prioritization for future field investigations. . . . .	51
903	11	Scatterplot comparing $V_{s30}$ residuals for posterior geology- and terrain-based 904 models. Colored by $V_{s30}$ . . . . .	52
905	12	Theoretical variogram selection for (a) geology- and (b) terrain-based models. .	53
906	13	The kriged $V_{s30}$ maps for the (a) geology- and (b) terrain-based models. An 907 example of potentially inappropriate extrapolation across geologic boundaries is 908 visible in the Rakaia area (boxed). . . . .	54
909	14	The kriged $\sigma$ maps for the (a) geology and (b) terrain-based models. Lower 910 uncertainty corresponds to locations of $V_{s30}$ data (Figure 2). . . . .	55
911	15	Several example covariance reduction functions. . . . .	56
912	16	Covariance-adjusted MVN model $V_{s30}$ estimates for (a) geology- and (b) terrain- 913 based models. The overprediction noted in the regression kriging (Rakaia area, 914 Figure 13) is eliminated. . . . .	57
915	17	Covariance-adjusted MVN model $V_{s30}$ predictions compared with regression 916 kriging (RK) predictions for (a) geology- and (b) terrain-based models. . . . .	58
917	18	Covariance-adjusted MVN model $\sigma$ for (a) geology- and (b) terrain-based mod- 918 els. . . . .	59
919	19	(a) $V_{s30}$ and (b) $\sigma$ for the final weighted model. . . . .	60



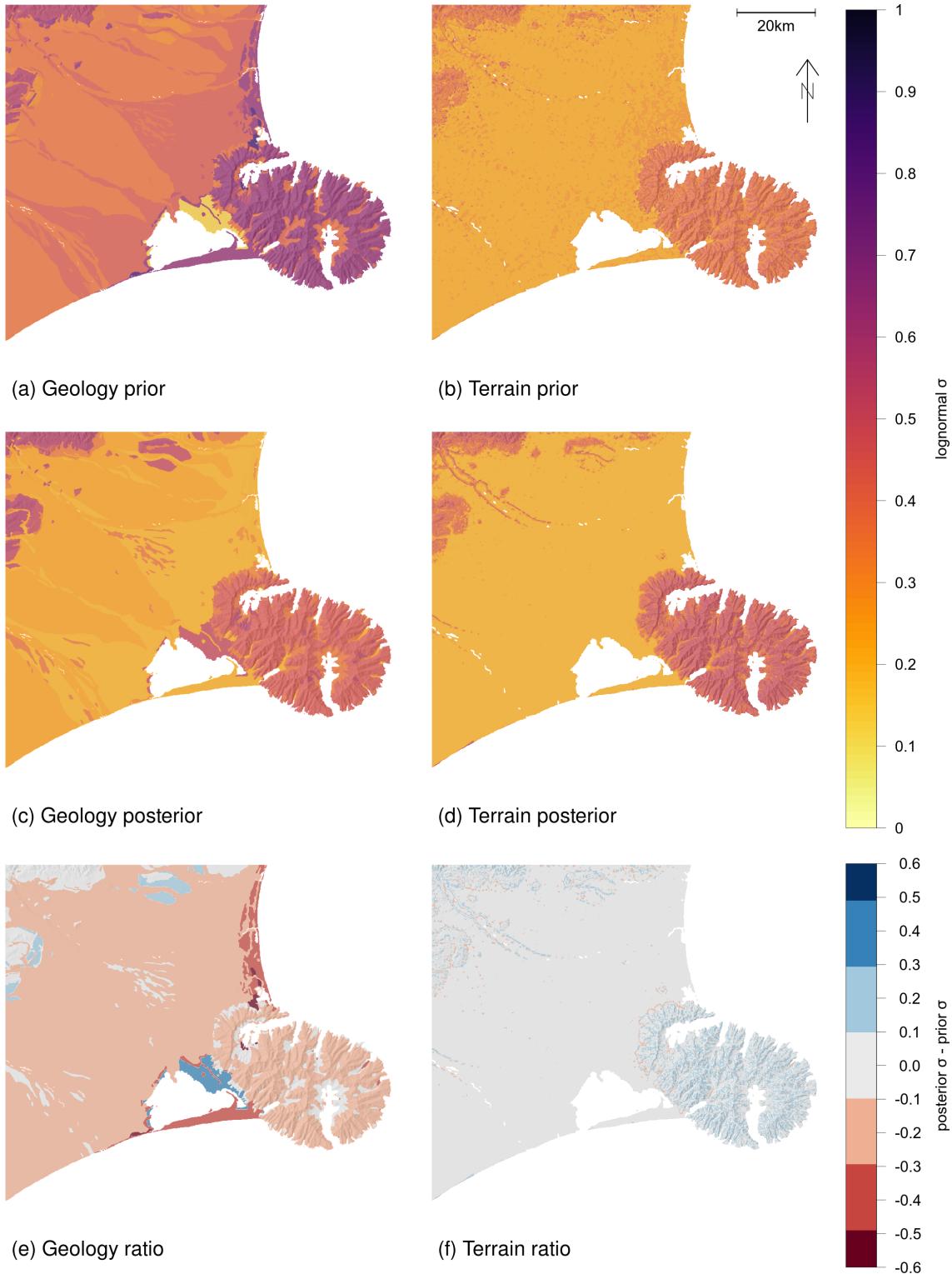
**Figure 1.** Relationships among various model components. Arrows indicate incremental model refinements.



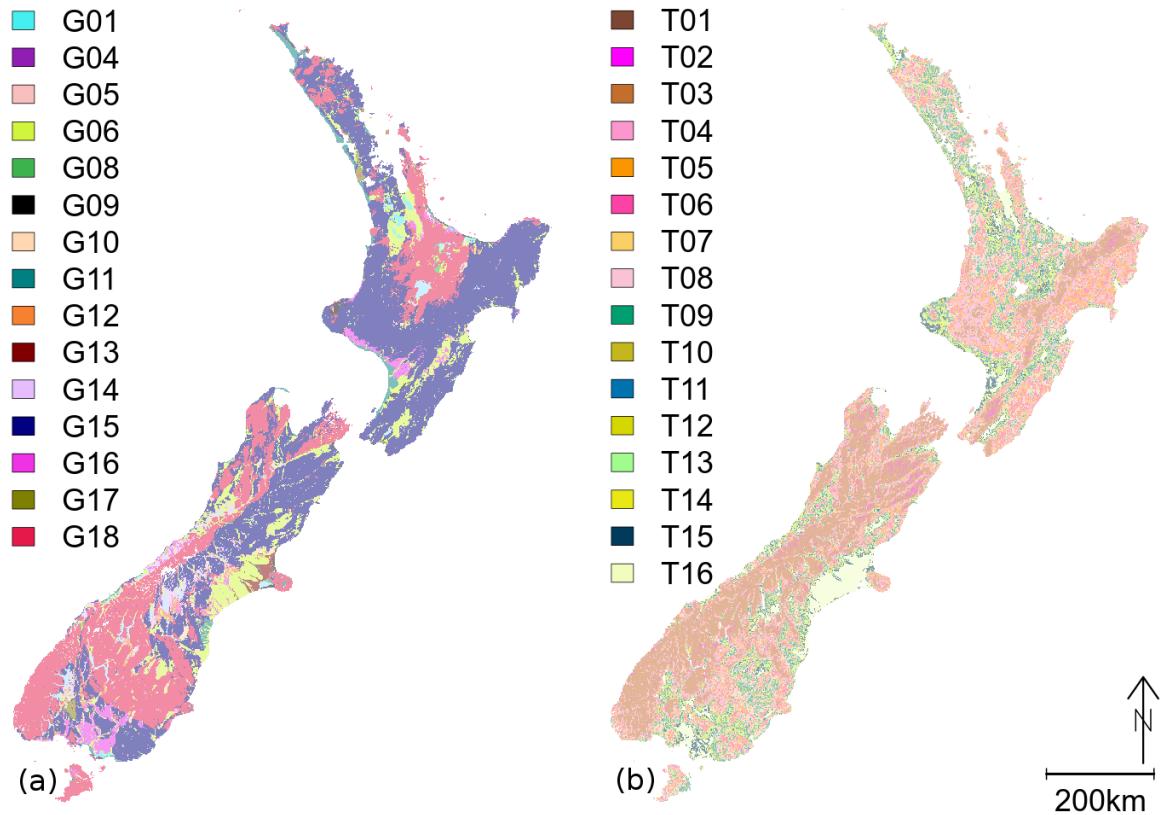
**Figure 2.** (a) Data locations for all of NZ and (b) inset showing Canterbury region for detail. [McGann et al. \(2017\)](#) data were downsampled based on closest proximity to an arbitrary 1km grid (white circles). Data labeled “Surface wave” (cyan circles) comprise several Canterbury-area surface wave investigations enumerated in the text. The majority of [Kaiser et al. \(2017\)](#) are “Q3” and are not used herein. [Kaiser et al. \(2017\)](#) Q2 (medium quality) and Q1 (highest quality) data are shown as small blue and large yellow squares, respectively.



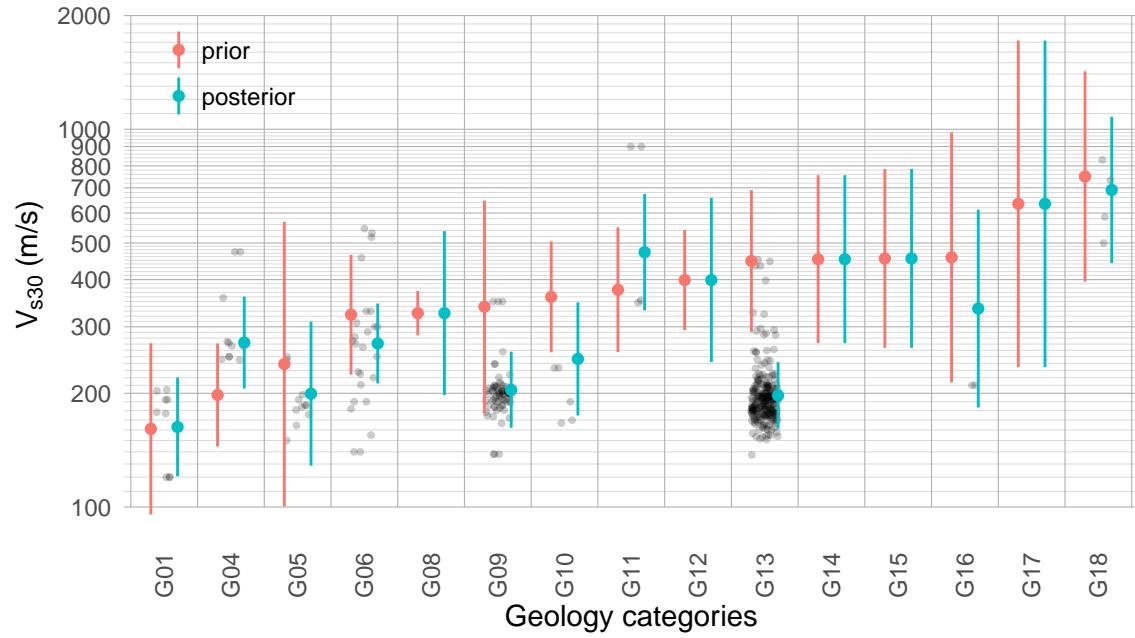
**Figure 3.** Illustrative regional (Canterbury)  $V_{30}$  maps for various stages of model development: (a) Prior geology model; (b) prior terrain model; (c) posterior geology model with slope-based adjustment; (d) posterior terrain model. Comparison of prior to posterior models: (e) ratio of pane (c) to pane (a); (f) ratio of pane (d) to pane (b). Point overlays in panes (c) and (d) show  $V_{30}$  measurements used for updating.



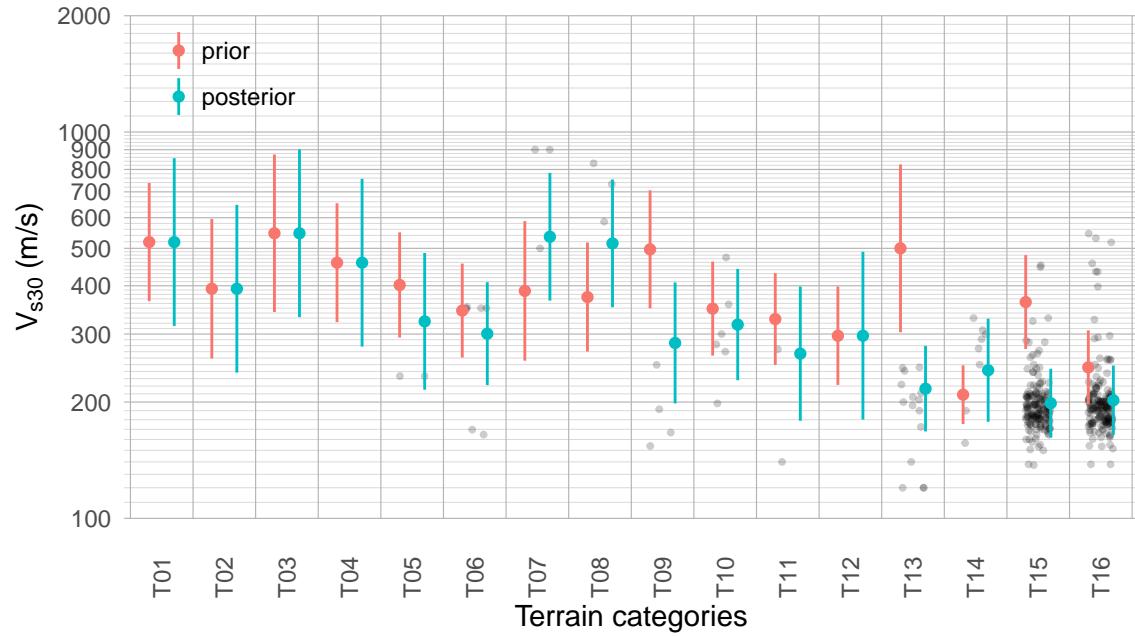
**Figure 4.**  $\sigma$  (sigma) maps for various stages of model development: (a) Prior geology model; (b) prior terrain model; (c) posterior geology model with slope-based adjustment; (d) posterior terrain model. Comparison of prior to posterior models: (e) difference between panes (c) and (a); (f) difference between panes (d) and (b).



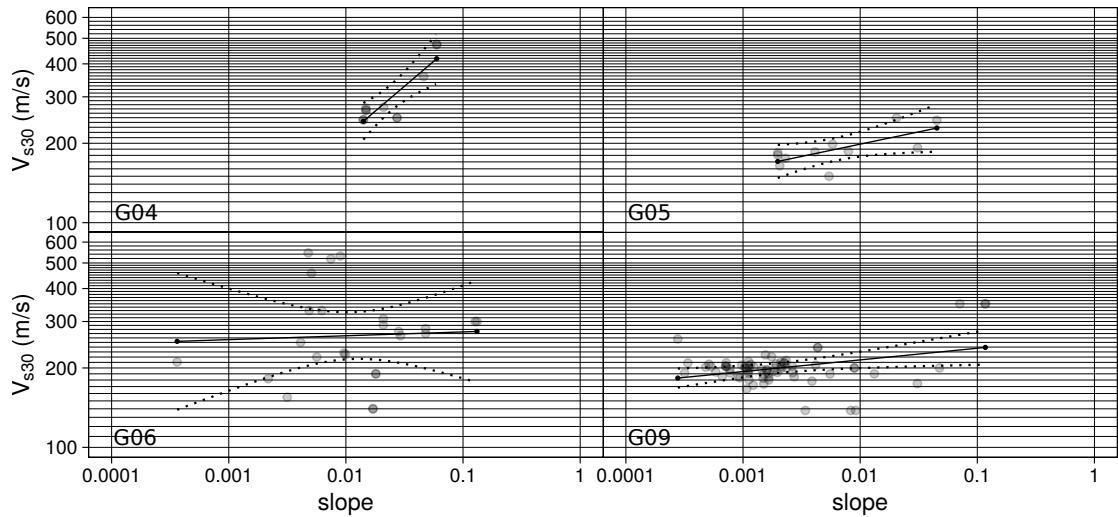
**Figure 5.** (a) Map of geology categories from Ahdi et al. (2017) as applied to NZ. (b) Map of terrain categories from Iwahashi and Pike (2007) as applied to NZ. Terrain colors are selected for direct comparison with Yong et al. (2012).



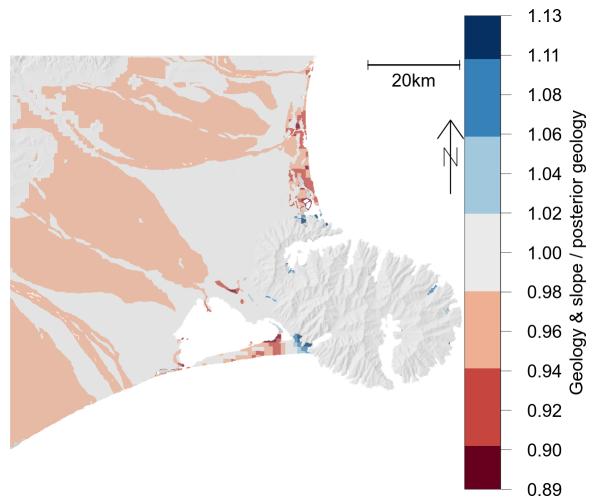
**Figure 6.** Bayesian updating for geology-based model. The colored dots and lines represent median and  $\pm 1$  standard deviation for lognormal prior (left, red) and posterior (right, blue) distributions. The prior is based on Ahdi et al. (2017b). The minimum  $\sigma = 0.5$  criterion is applied for groups G08 and G12. Gray dots represent data used for updating.



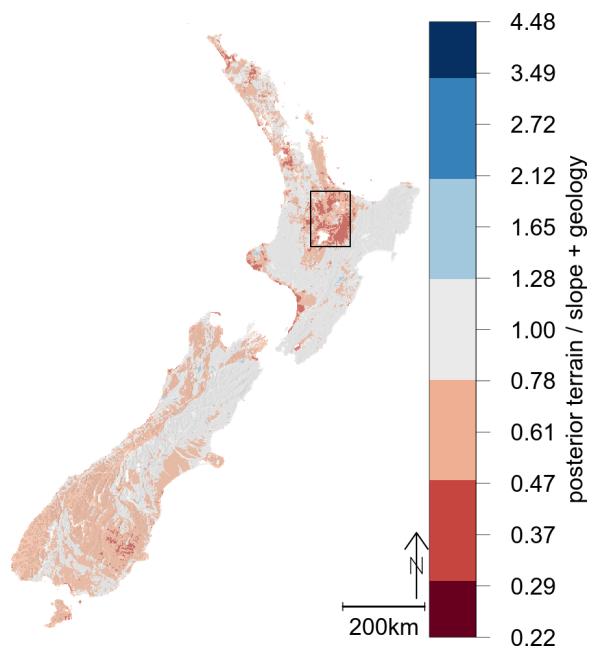
**Figure 7.** Bayesian updating for terrain-based model. The colored dots and lines represent median and  $\pm 1$  standard deviation for lognormal prior (left, red) and posterior (right, blue) distributions. The prior is based on [Yong et al. \(2012\)](#). The minimum  $\sigma = 0.5$  criterion is applied for groups T01, T02, T03, T04, and T12. Gray dots represent data used for updating.



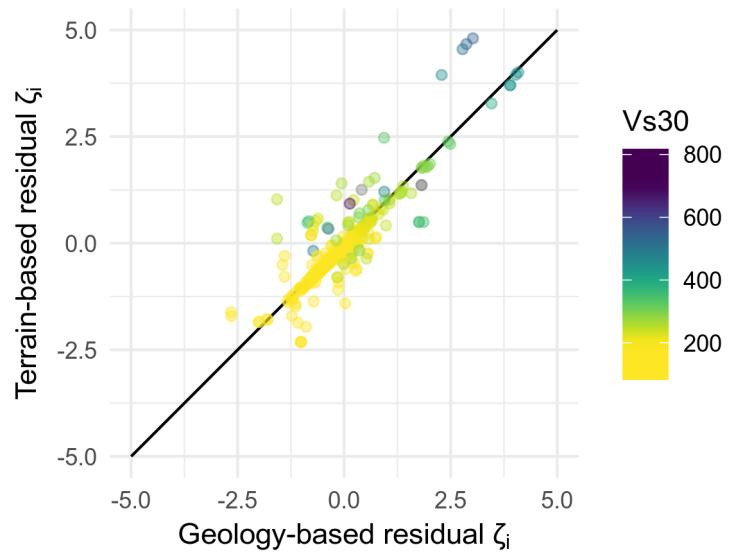
**Figure 8.** Slope dependence of geology-based  $V_{s30}$  model. Slope resolved from 9 arcsecond ( $\sim 270\text{m}$ ) digital elevation model. Only the four geology categories with positive trends are shown. Dotted lines indicate  $\pm 2$  standard deviations (95% confidence bounds) on slope fit.



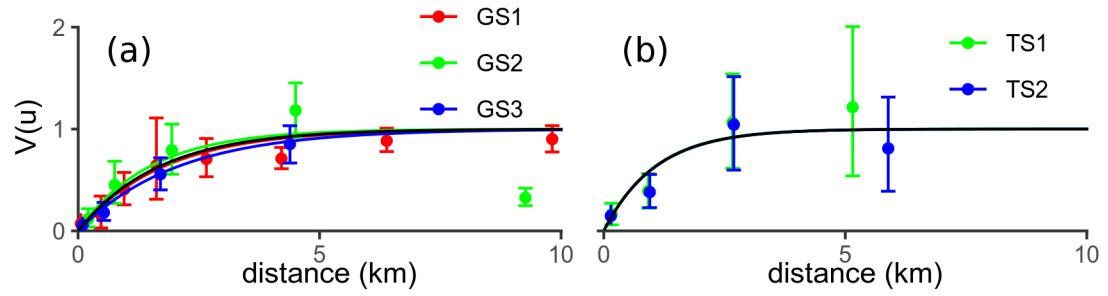
**Figure 9.** Ratio of slope+geology model (Figure 3c) to posterior geology model (not shown) median estimates. Note the small magnitude on the graphic scale indicating much smaller model change by comparison with the posterior geology update.



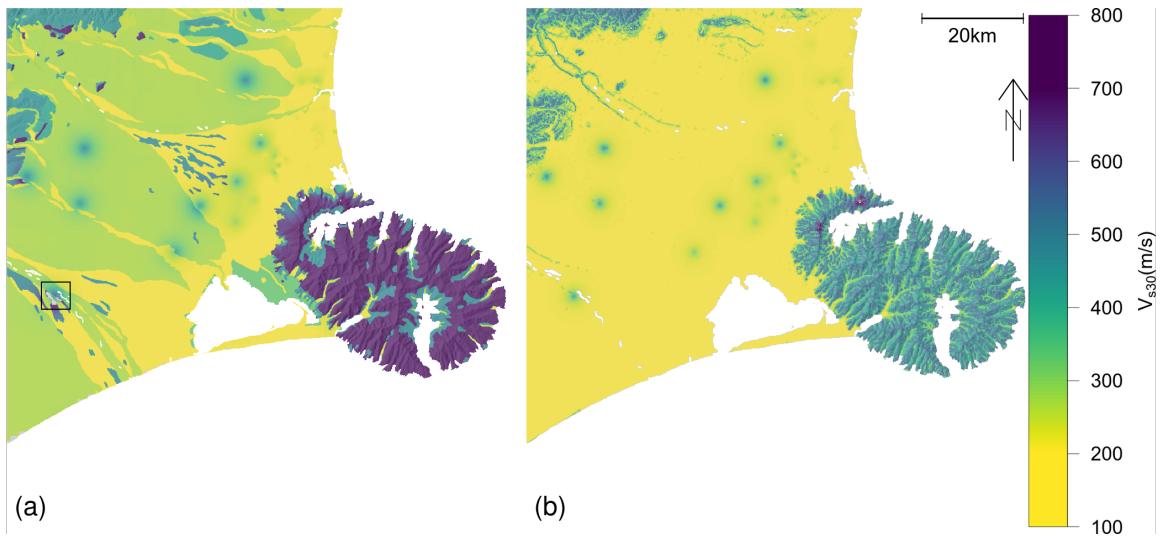
**Figure 10.** Map comparing median  $V_{s30}$  for posterior geology- and terrain-based models. The Taupo and Rotorua region (boxed) is an example of a region with unique interaction between the geology- and terrain-based categories that may merit prioritization for future field investigations.



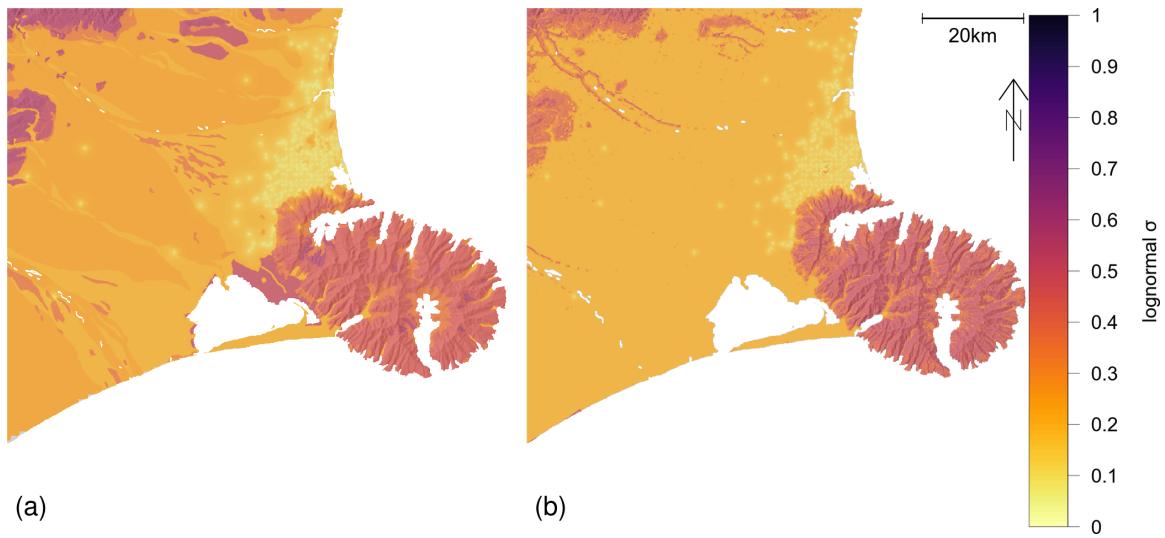
**Figure 11.** Scatterplot comparing  $V_{s30}$  residuals for posterior geology- and terrain-based models. Colored by  $V_{s30}$ .



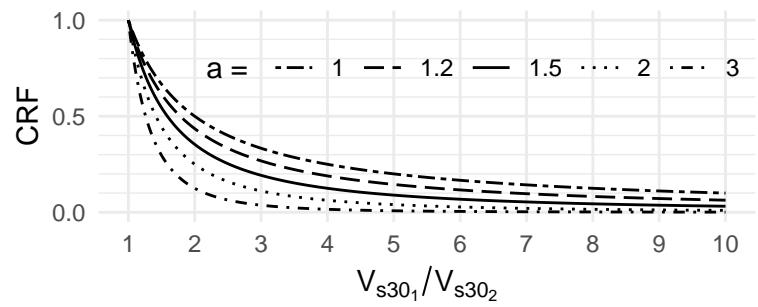
**Figure 12.** Theoretical variogram selection for (a) geology- and (b) terrain-based models.



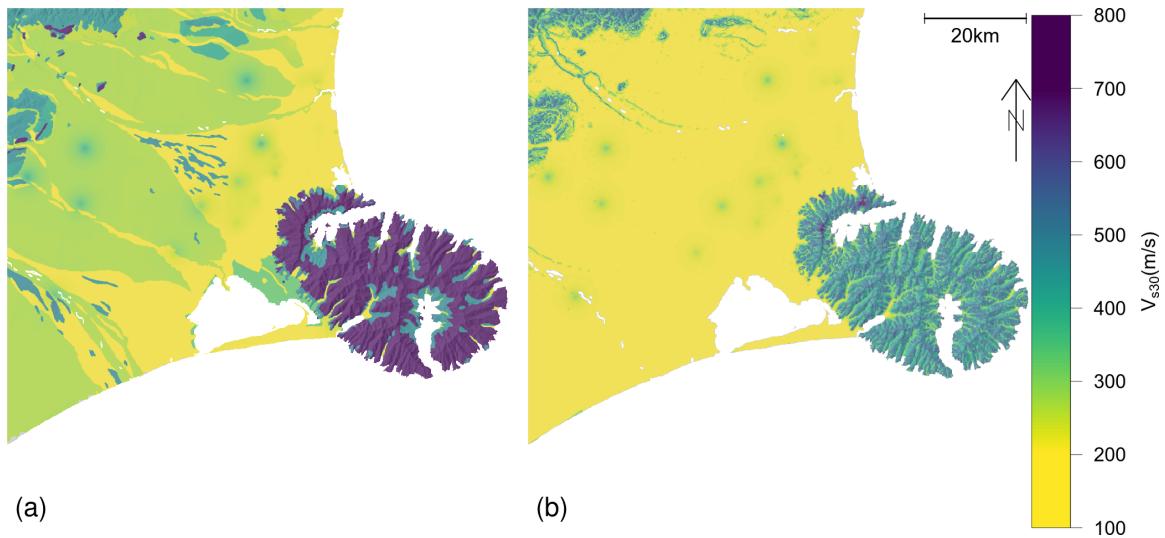
**Figure 13.** The kriged  $V_{s30}$  maps for the (a) geology- and (b) terrain-based models. An example of potentially inappropriate extrapolation across geologic boundaries is visible in the Rakaia area (boxed).



**Figure 14.** The kriged  $\sigma$  maps for the (a) geology and (b) terrain-based models. Lower uncertainty corresponds to locations of  $V_{s30}$  data (Figure 2).



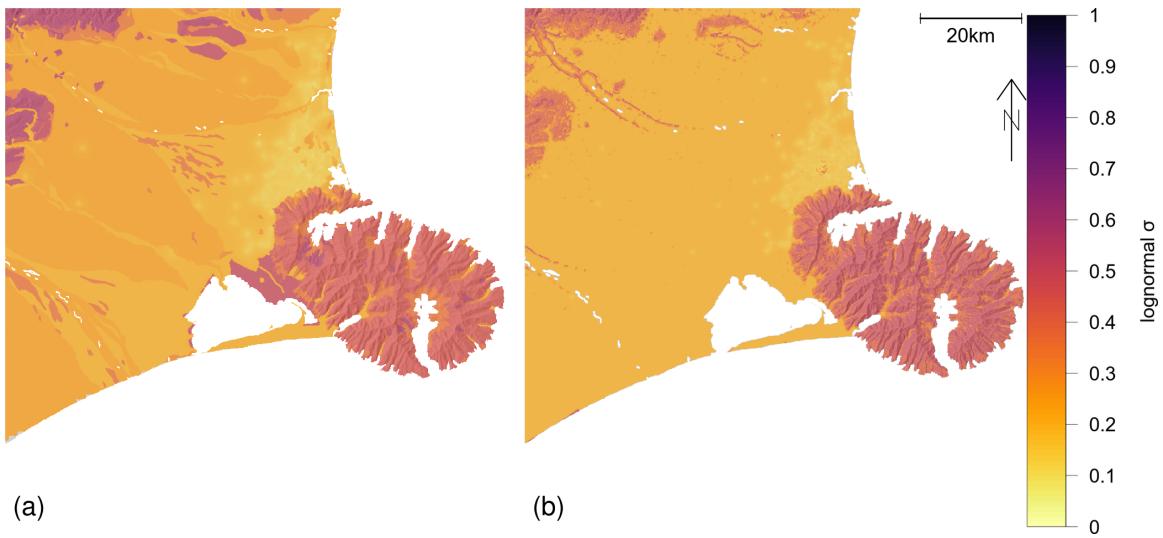
**Figure 15.** Several example covariance reduction functions.



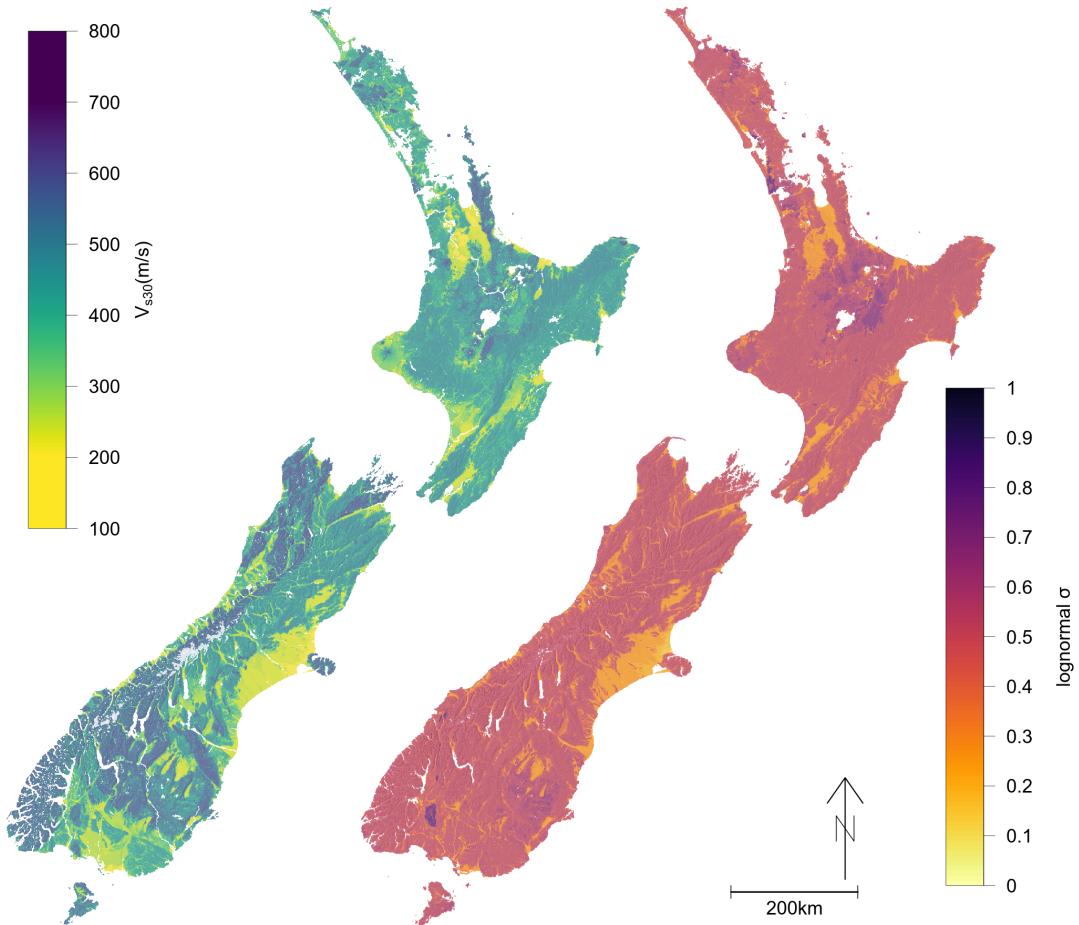
**Figure 16.** Covariance-adjusted MVN model  $V_{s30}$  estimates for (a) geology- and (b) terrain-based models. The overprediction noted in the regression kriging (Rakaia area, Figure 13) is eliminated.



**Figure 17.** Covariance-adjusted MVN model  $V_{30}$  predictions compared with regression kriging (RK) predictions for (a) geology- and (b) terrain-based models.



**Figure 18.** Covariance-adjusted MVN model  $\sigma$  for (a) geology- and (b) terrain-based models.



**Figure 19.** (a)  $V_{s30}$  and (b)  $\sigma$  for the final weighted model.