Supporting Information for "Predicting Patagonian Landslides: Roles of Forest Cover and Wind Speed"

Eric Parra[1,2], Christian Mohr[1], Oliver Korup[1,2]

[1]Institute of Environmental Science and Geography, University of Potsdam, Germany

[2]Institute of Geosciences, University of Potsdam, Germany

Contents of this file

1. Text S1
2. Figures S1 to S4

Introduction

This file contains supporting text devoted to outline the model setup and additional figures that expand on the results documented in the main manuscript.

Text S1. We used a Bayesian robust regression to predict the posterior probability $P(L)$ at which a given location y_i in our study areas is classified as part of a landslide source, transport, or deposition area. We write this as $P(L) = P(y_i = 1)$ and denote the inverse probability of classifying a location without a landslide as $P(y_i = 0) = 1 - P(y_i = 1)$. The index i refers to the ith location, i.e. raster value, out of n observations in given study area. We can write this model with a Bernoulli likelihood conditional on the observed data in general form as:

$$P(y_i = 1|\mathbf{X}_i, \mathbf{w}) = \frac{1}{1 + \mathrm{e}^{-\mathbf{X}_i\mathbf{w}}}, \tag{1}$$

where $\mathbf{X}_i$ is the ith row in a $n \times m$ design matrix with n observations and m columns. The first column consists of 1's for the intercept, while the remaining $m-1$ columns collect the predictor (and possible interaction) values. The $m \times 1$ column vector $\mathbf{w}$ contains the m model coefficients, i.e. the intercept, the predictor weights, and weights for interaction terms, if any. The matrix-vector product $\mathbf{X}_i\mathbf{w}$ constitutes the linear predictor of the model.

We acknowledge structure in our data by admitting $j = \{1, ..., 10\}$ different grouping levels l_j to the model. These levels represent ten landform types that we classified from the Topographic Position Index. We standardised the input values to zero means and unit standard deviations and designate standardised parameters with an asterisk (*). In this hierarchical (or multilevel) model structure, we use as inputs standardised crown openness χ^* and standardised wind speed u^*, as well as their interaction $\chi^*.u^*$ and write this for our specific case:

$$P(y_i = 1|\mathbf{X}_i, \mathbf{w}) = \left(1 + \mathrm{e}^{-w_0[j] + w_1\chi_i^* + w_2u_i^* + w_3\chi_i^*.u_i^*}\right)^{-1}, \tag{2}$$

where we allow the intercept $w_0[j]$ to vary by landform type j as:

$$w_0[j] \sim \mathcal{N}(0, \sigma_j^2), \tag{3}$$

where σ_j^2 is the variance of group-level intercepts, i.e. whether and by how much the log-odds ratios for average predictor inputs vary by landform type. Note that these intercepts represent log-odds ratios of classifying location y_i as part of a landslide for zero (i.e. average) predictor values. Positive log-odds ratios are equivalent to values of $P(L) > 0.5$, while negative log-odds ratios are equivalent to $P(L) < 0.5$. The zero mean of the Gaussian distribution in Equation 3 is the intercept of the model pooled over all data regardless of landform type.

We specify our prior assumptions using the following distributions:

$$w_0[j] \sim \mathcal{T}(\nu = 3, \mu = 0, \sigma = 2.5), \tag{4}$$

$$w_1 \sim \mathcal{T}(\nu = 3, \mu = 0, \sigma = 2.5), \tag{5}$$

$$w_2 \sim \mathcal{T}(\nu = 3, \mu = 0, \sigma = 2.5), \tag{6}$$

$$w_3 \sim \mathcal{T}(\nu = 3, \mu = 0, \sigma = 2.5), \tag{7}$$

These independent prior distributions are weakly informative and based on the assumption that all model weights are from a Student's t-distribution with location $\mu = 0$ and scale $\sigma = 2.5$. This choice of prior means that we surmise that regression weights may be equally likely positive or negative, though symmetrically distributed and within the interval $[-7.96, 7.96]$ with 95% probability. Such extreme regression weights are unlikely for standardised predictors and underline our weakly informative choice.

Our prior distribution of the variance of the group-level intercept is a standard exponential. This reflects our assumption that the spread of log-odds of classifying landslides is more likely near zero, hence more likely to differ less with landform type:

$$\sigma_j \sim \exp(\lambda = 1), \tag{8}$$

This prior specifies that σ_j lies in the interval $[0, 3]$ with 95% probability, with larger spread becoming exponentially less likely.

We ran our model with different parameter choices for these priors and observed only minute changes in the posterior distributions, given the large number of observations in each study area. Although using the three study areas as another grouping level in the model would be possible, we preferred using the data from the different regions as testing data for models trained elsewhere.

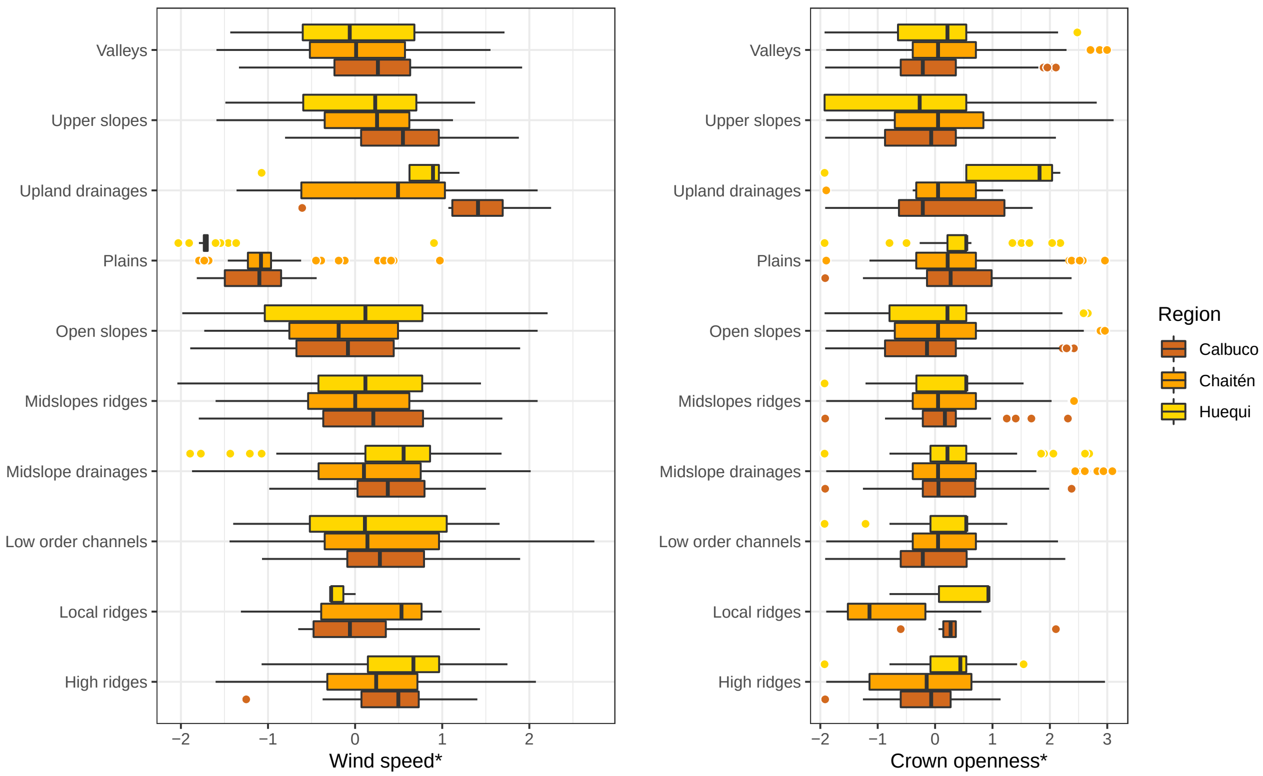

Figure S1. Distributions of standardised wind speed and standardised crown openness per landform type in temperate rainforests in three study areas of south-central Chile. Boxes encompass the interquartile range, and box centres are medians; whiskers span 1.5 times the interquartile range and points are outliers.

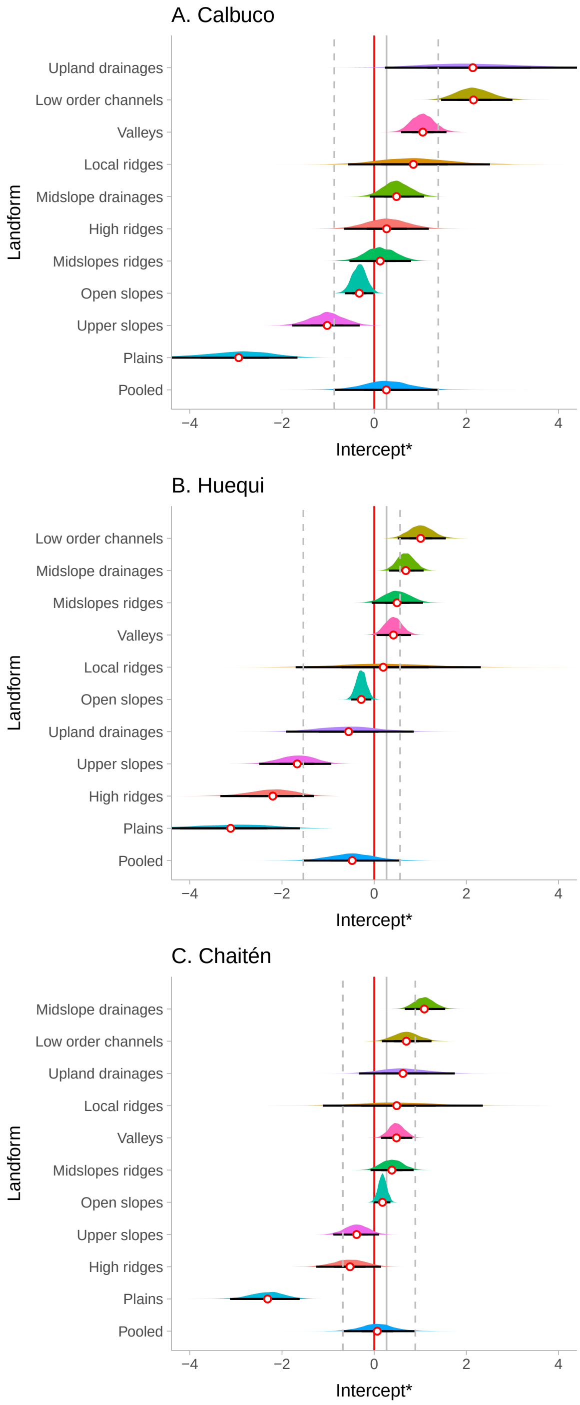

Figure S2. Posterior distributions of standardised model intercepts by landform type. Black horizontal lines are 95% highest density intervals (HDIs), and white circles are posterior means. Vertical grey solid (dashed) lines are pooled means (95% HDIs). Intercepts are the log-odds ratios of classifying a pixel with average crown openness and wind speed as part of a landslide.

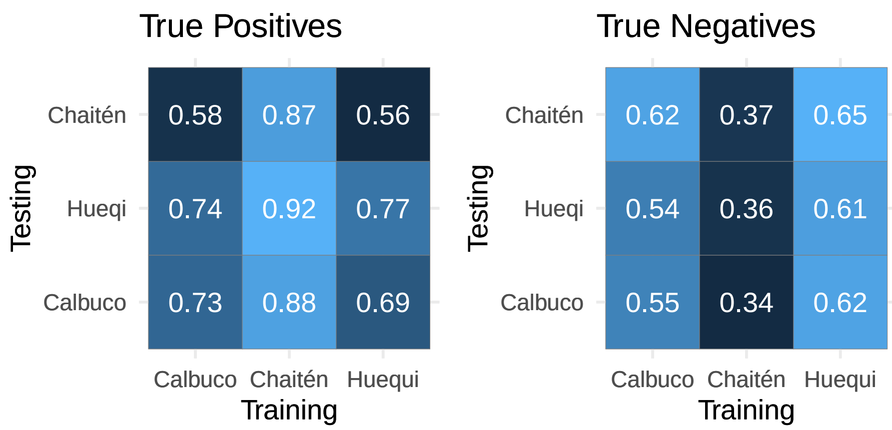

Figure S3. Performance of Bayesian robust logistic regression for predicting landslide terrain from crown openness and wind speed. True positive (and negative) rates refer to the fraction of correctly predicted landslides (and their absence) for a given training and testing dataset. All rates refer to the pooled model averaged over all landform types. The average true positive rate is 0.75, whereas the average true negative rate is 0.52.

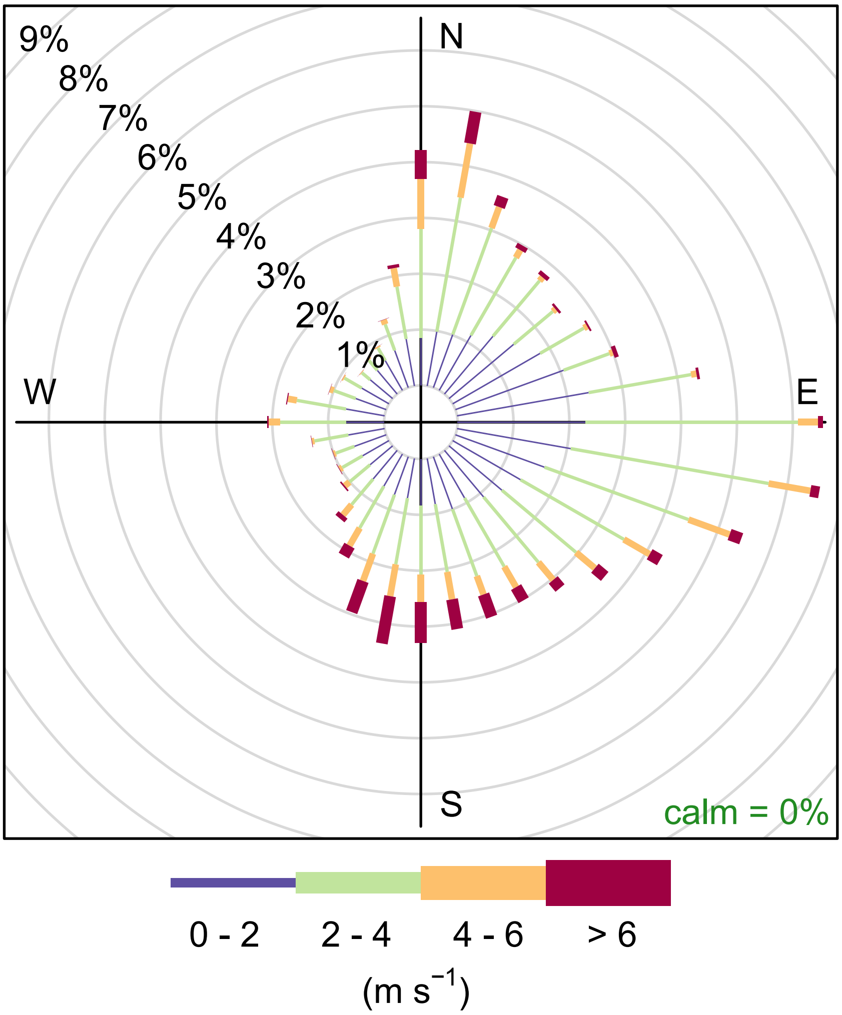

Figure S4. Rose diagram of hourly wind speed and direction for south-central Chile (40–43°S, 72–73°W) from January 2013 to December 2017; grey circles are the frequency of observations. The overall mean wind speed was 2.9 m s^{-1}, mostly from a W to WNW direction and with negligible amount of calm conditions. Data are freely available from the Global Forecast System (GFS) of the U.S. National Weather Service (NWS) at `https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/globalforcast-system-gfs`.