Supplementary Information: Hazard area performance with optimised, global and hold-back parameters

At each of the six study sites we optimise the two parameters in the SHALRUN-EQ model required to predict hazard area, initiation angle (θ_m) and stopping angle (θ_s), by sampling values for each parameter uniformly in 1 degree increments over the range [20,70] for initiation angle and [0,50] for stopping angle, and imposing the requirement: $\theta_m > \theta_s$. For our objective function we use the area under the receiver operating characteristic (ROC) curve, comparing landslide hazard derived from hazard area to the inventory of observed landslides at each site. The optimisation surfaces are shown in Figure S1. To generalise our results, we then take arithmetic means of the optimum initiation and stopping angles, to generate hazard area predictions using a single 'global' rule averaged over all six inventories (Table S1). To remove the influence of test data on the test itself, we re-run the hazard area prediction for each inventory as a hold back-test, in which we re-calculate the initiation and stopping parameters excluding the optimised values from that inventory and using only the remaining five inventories.

We find that the differences in ROC curves (Figure S2) and area under the curve values (Table S1) are fairly subtle. Hazard area with global averaged parameters performs well overall, with AUC values that range from 0.78 to 0.86. Hazard area with parameters that are optimised for each inventory offers only a slight further improvement, with AUC increased by <3% in each case (Table S1). Optimised initiation and stopping angles can differ quite radically between sites, ranging from 31-45˚ for initiation angle and from 3-19˚ for stopping angle. This might signal cause for concern about how feasible it is to find a single general rule given such variability in optimum parameters between sites. However, hazard area skill is relatively insensitive to parameter variation close to the optimum parameters, as indicated by the relatively smooth and gentle peaks of the optimisation surfaces in Figure S1. Thus, the (sometimes large) differences between global and optimised parameter values do not translate into large performance differences between hazard area predictions using global or optimised parameters. The use of hold-back rather than global parameters results in an even smaller difference in performance; AUC values are reduced by <1% for every inventory and hazard area is still the best metric at all sites. For this reason, we include hold-back tests here but report results from global average parameters rather than hold-back parameters in the paper for simplicity. It is these global average parameters (initiation angle of 39° and stopping angle of 10°) that form the basis of our simple rule, and that we would recommend when applying the SHALRUN-EQ approach to a new location (in the absence of a landslide inventory with which to test and calibrate the parameters).

Table S1: Parameter values and areas under the ROC curve for the six inventories

	Parameters		Area Under ROC Curve		
	Initiation slope θ_i (˚)	Stopping slope θ_s (˚)	Hazard area optimised	Hazard area global	Hazard area holdback
Finisterre	34	19	0.91	0.89	0.88
Northridge	41	19	0.80	0.79	0.78
Chichi	44	4	0.80	0.80	0.79
Wenchuan	39	3	0.78	0.78	0.78
Haiti	31	9	0.88	0.86	0.85
Gorkha	45	6	0.89	0.88	0.88
Average	**39**	**10**	**0.84**	**0.83**	**0.83**
1σ	**6**	**7**	**0.1**	**0.1**	**0.1**

Figure S1: Model predictive skill for SHALRUN-EQ for each of the six landslide inventories across reasonable ranges for the two parameters, initiation angle (θ_m) and stopping angle (θ_s). Predictive skill is quantified using area under the receiver operating characteristic curve. The six inventories are: a) Finisterre, b) Northridge, c) Chi-Chi, d) Wenchuan, e) Haiti, f) Gorkha. Symbols show the parameter combinations from site specific optimisation, global average and hold-back average.

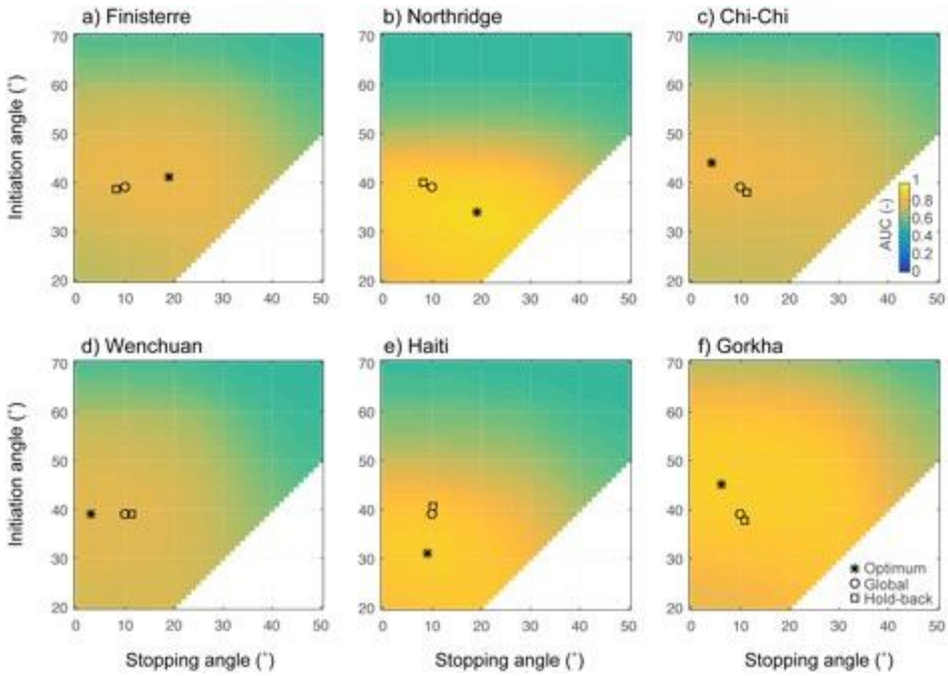

Figure S2: Receiver operating characteristic (ROC) curves for the six landslide inventories and five metrics examined here, as shown in Figure 6 of the paper, but with the addition of ROC curves for SHALRUN-EQ with site-specific optimised parameters and hold-back parameters (i.e., global averages from five sites excluding the test site). The six inventories are: a) Finisterre, b) Northridge, c) Chi-Chi, d) Wenchuan, e) Haiti, f) Gorkha. False positive rate is given by the number of false positives divided by the sum of false positives and true negatives. True positive rate is given by the number of true positives divided by the sum of true positives and false negatives. The 1:1 line represents the naïve (random) case. Curves plotting closer to the top left corner of each panel represent better model performance.