Supplement

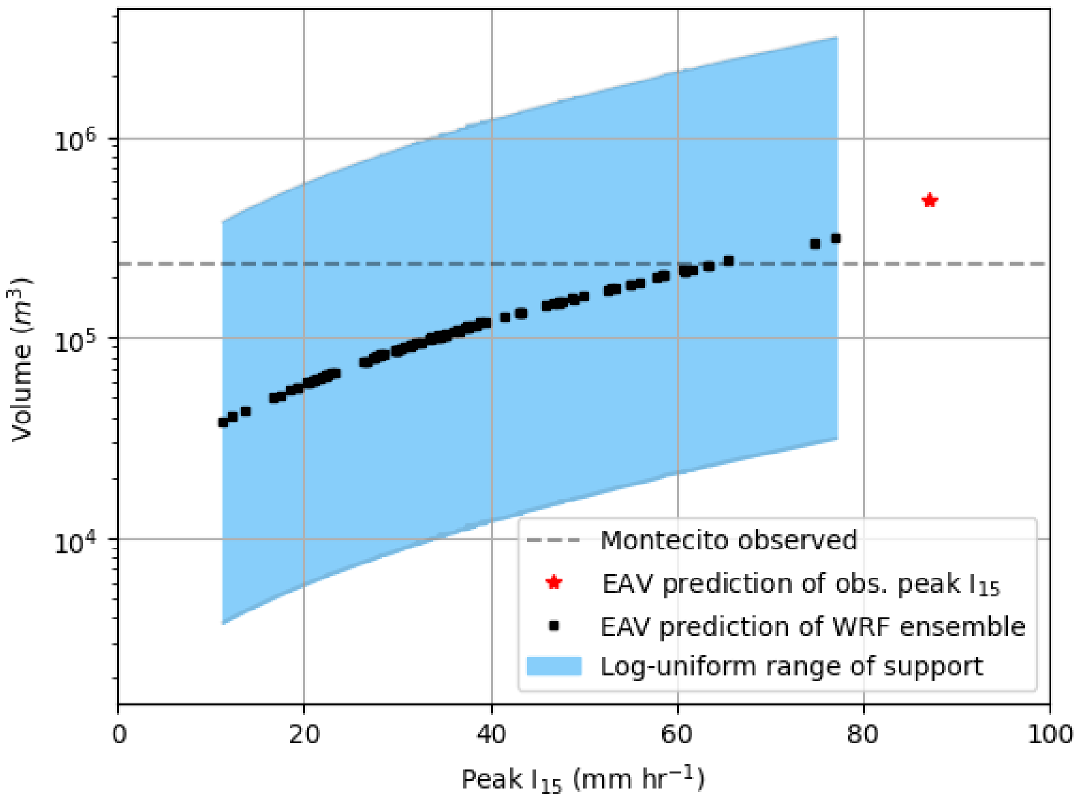

Figure S1: Volume samples used in the forecast simulation were drawn from log-uniform distributions, shown here for Montecito Creek with Weather Research and Forecast (WRF) ensemble members (Skamarock et al., 2021) sorted by increasing peak I_{15}. The emergency assessment volume (EAV) model produced a prediction of debris-flow volume from each ensemble member's prediction of peak 15-minute rainfall intensity (I_{15}), from which a log-uniform distribution centered on the EAV prediction was defined with a range of support 10x above and below.

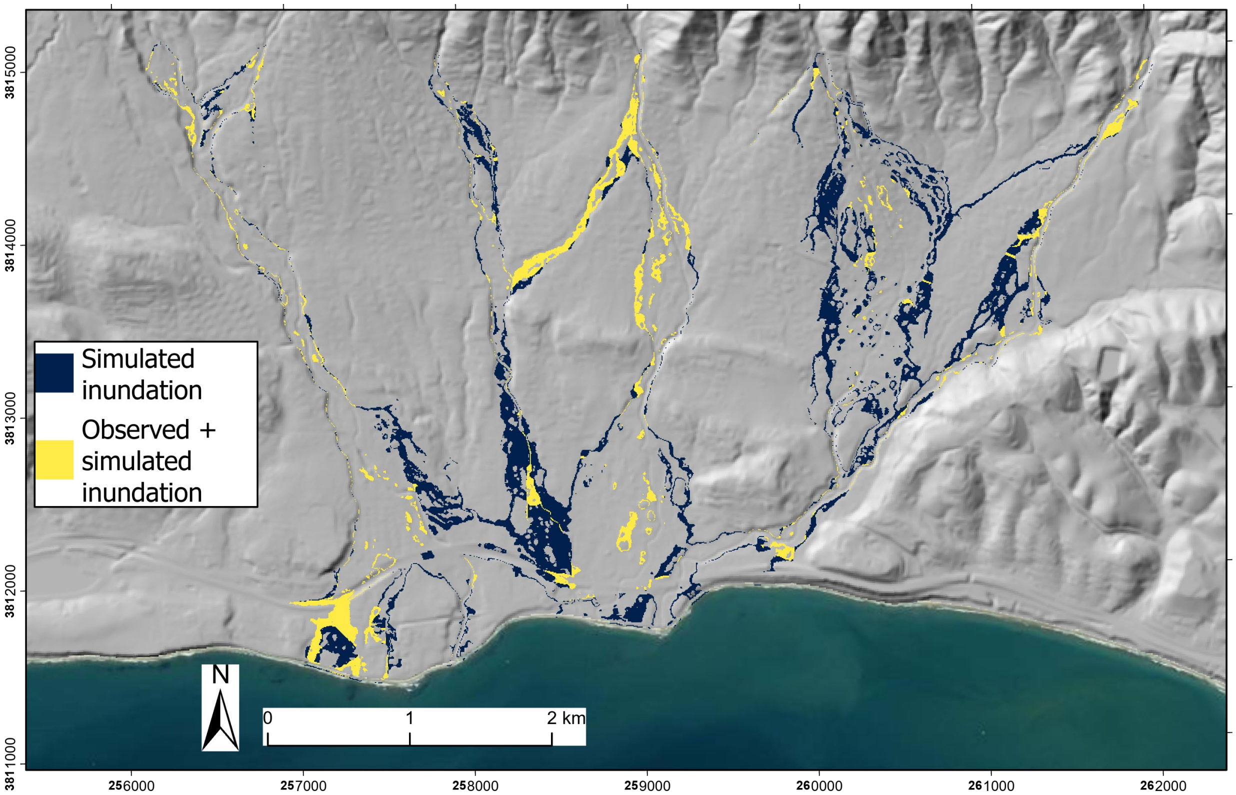

Figure S2: This map demonstrates the quantities used in construction of the calibration curve, a component of the reliability diagram (Section 3.6). The areas with blue or yellow color show all grid cells with a forecast inundation probability between 10 and 20%. The yellow color indicates that debris-flow inundation was actually observed in that grid cell. This data are used to generate the second-to-left-most point in the WRF ensemble forecast reliability diagram (Fig. 4a), where the mean simulated probability of inundation across all cells is 14.2% and the observed relative frequency is 29.1%. Ticks along the boundaries of map give coordinates in NAD 1983 UTM zone 11N.

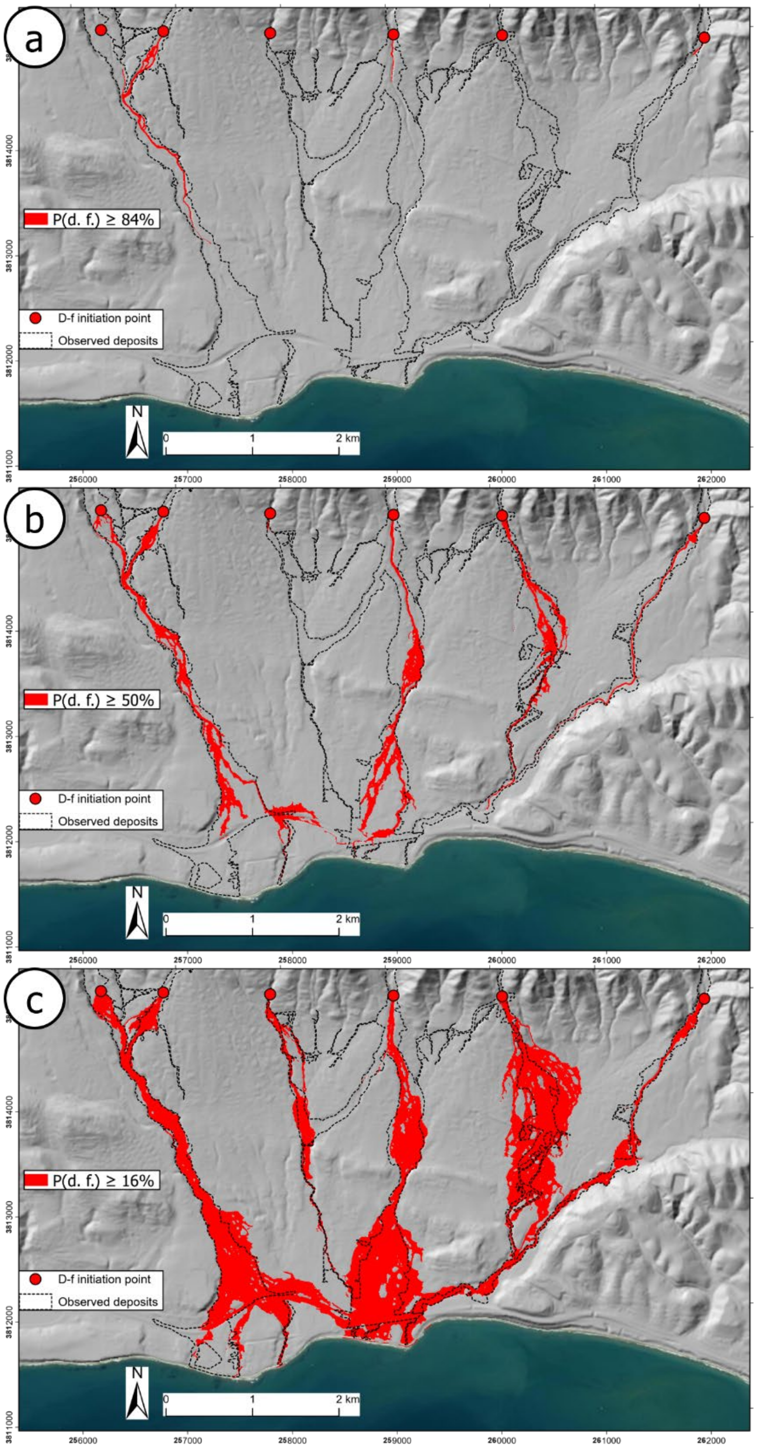

Figure S3: Binary maps of area inundated created by thresholding simulated debris-flow (d. f.) inundation probability (P) at different values. Inundated area increases as the probability threshold is decreased from (a) 84% to (b) 50% to (c) 16%. Similarity indices for each of the binary inundation maps are -0.95, -0.51, and -0.02, respectively.Ticks along the boundaries of each map give coordinates in NAD 1983 UTM zone 11N.

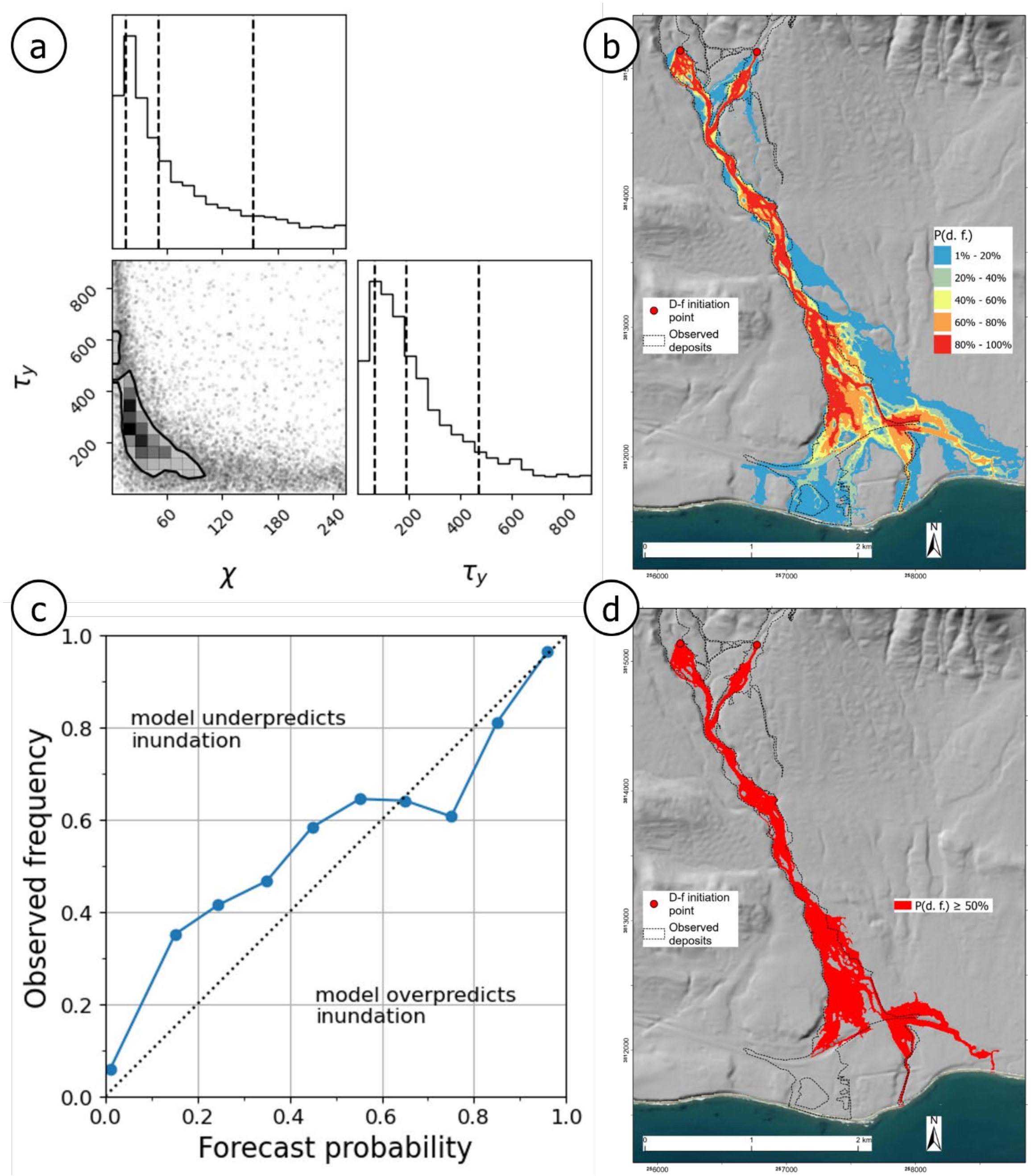

Figure S4: The Markov Chain Monte Carlo (MCMC) calibrated posterior distribution was tested on the Montecito Creek Basin. (a) The two-dimensional and one-dimensional histograms of flow-mobility parameters sampled from the MCMC calibrated joint posterior distribution. The top subplot shows the histogram of χ samples, and the right subplot shows the histogram of τ_y samples (generated with the *corner* Python package; Foreman-Mackey, 2016); (b) probabilistic map of debris-flow inundation on Montecito Creek with n=10,000 samples drawn from the MCMC posterior; (c) reliability curve of the inundation map; (d) inundation binary classified with a threshold probability of 50% (similarity index -0.047). Ticks along the boundaries of (b) and (d) give coordinates in NAD 1983 UTM zone 11N.

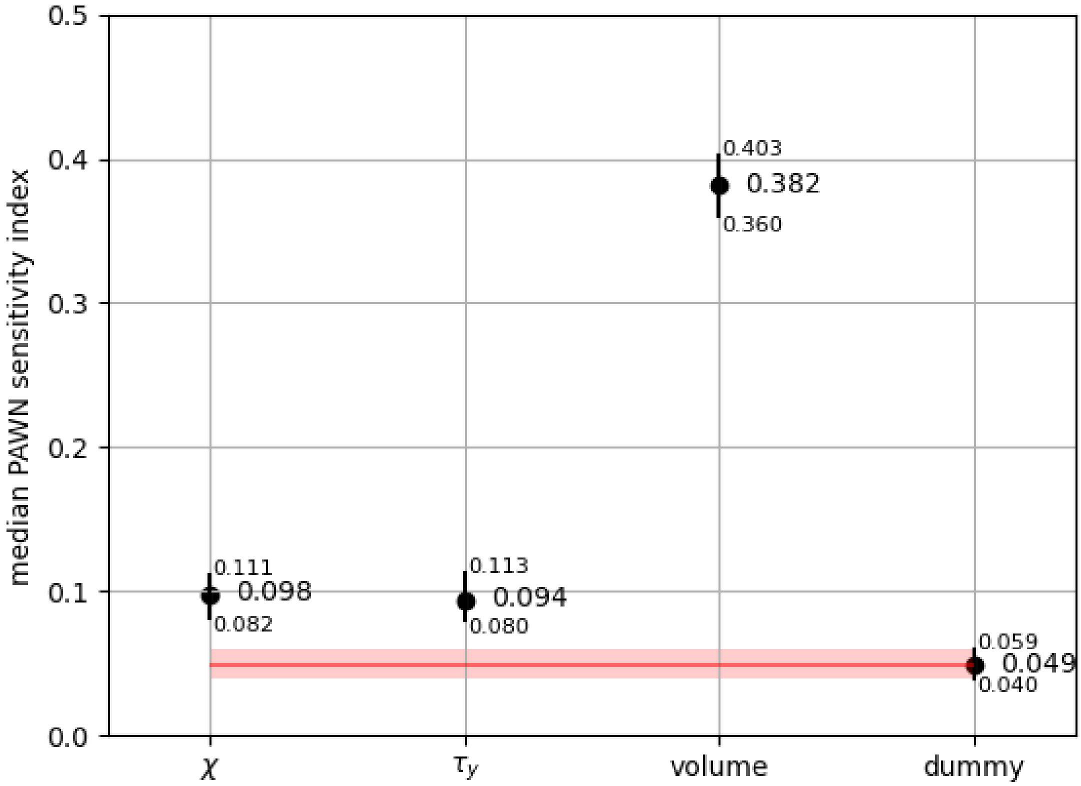

Figure S5: PAWN sensitivity indices (Pianosi and Wagener, 2018) with confidence intervals from bootstrapping (n=50 iterations). Circles show the mean of the 50 median sensitivity index values, and vertical lines show the two-sided 95% confidence interval. The horizontal red line and shaded area show the mean and 95% confidence interval of the dummy sensitivity index, respectively.

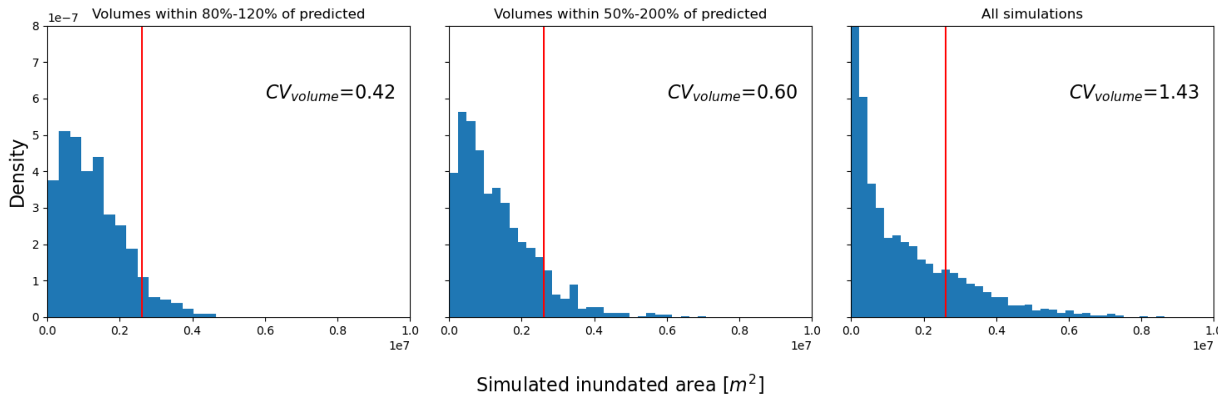

Figure S6: Accounting for the input volume uncertainty is essential to capturing low-probability inundation scenarios. This three-panel figure shows the normalized histograms of simulated inundated area when considering (left) only the simulations whose sampled volume is within 80-120% of predicted volume, (center) only simulations with volume within 50-200% of the predicted volume, and (right) all simulations. Vertical red lines show the observed inundated area, and the text insets show the coefficient of variation (CV) of the volume samples. An increase in the volume CV drives increased variability in the simulated inundation output, as the tails of the inundated area distribution are only resolved with the order-of-magnitude envelope around the predicted volumes (right).

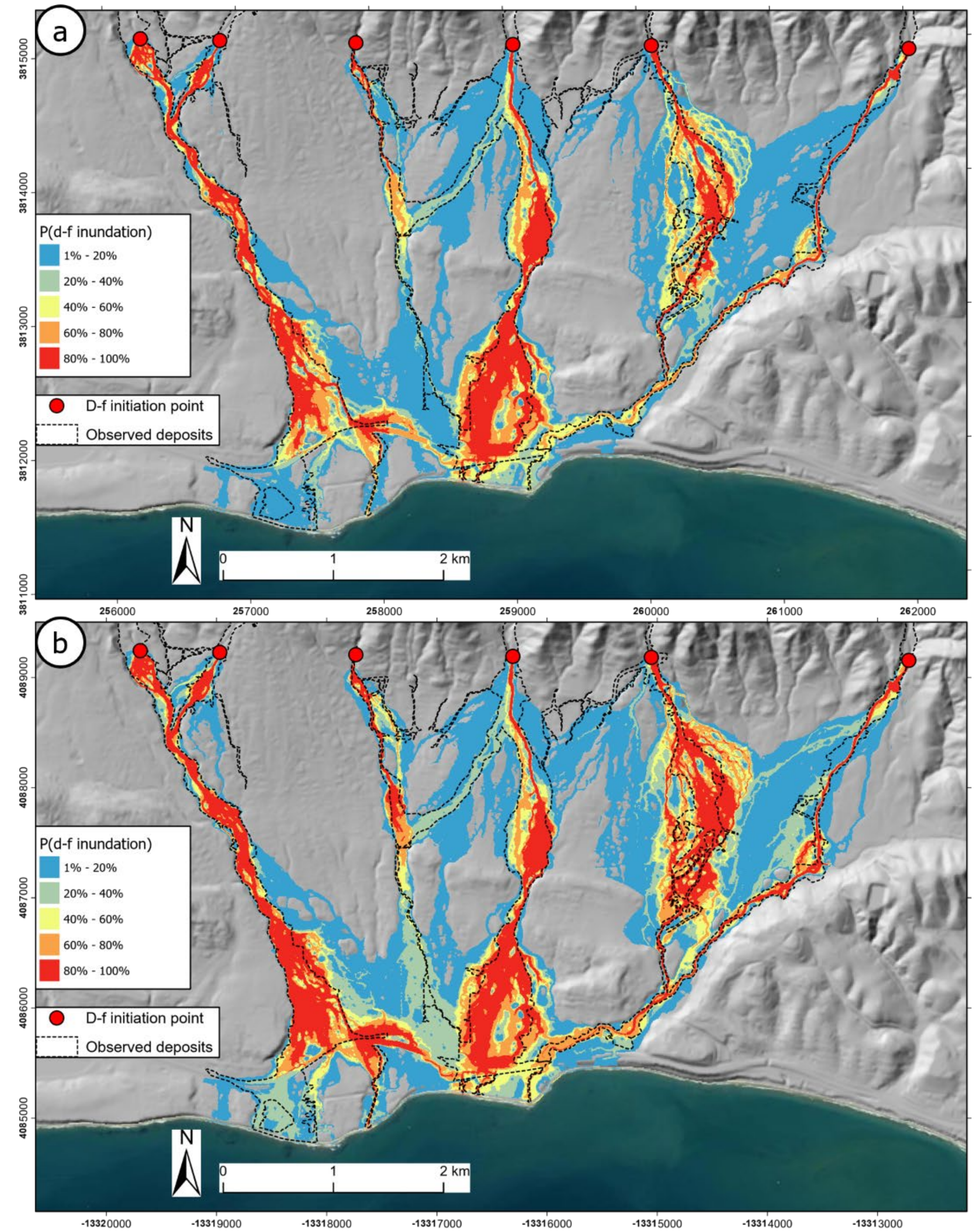

Figure S7: Inundation probability maps resulting from simulation scenarios A and B; (a) scenario A, constant measured input debris-flow volumes (Kean et al., 2019); (b) scenario B, constant input debris-flow volumes predicted with the EAV model from observed I_{15} values. I_{15} values at each initiation point were computed with inverse distance weighting of the observed rainfall rates at the KTYD and Doulton Tunnel rain gauges (78 and 105 mm/hr, respectively; Kean et al., 2019). Ticks along the boundaries of each map give coordinates in NAD 1983 UTM zone 11N.

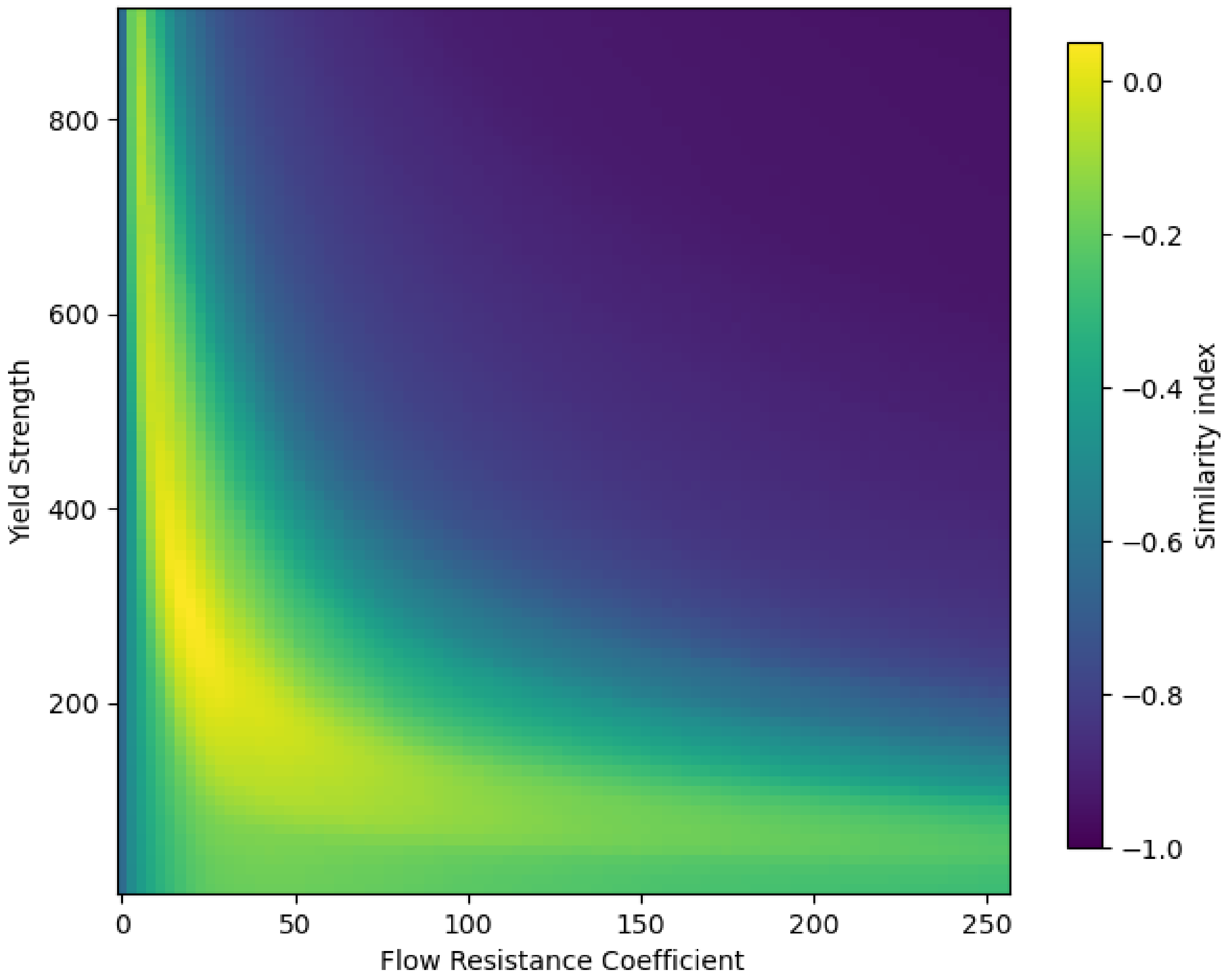

Figure S8: Similarity index of Progressive Debris-Flow routing and inundation model (ProDF) (Gorr et al., 2022) simulations on a discretized grid over the flow mobility parameters. This response surface was created from simulations on the Oak, San Ysidro, Buena Vista, and Romero Creek drainages, and it was used in the objective function of the MCMC calibration process.

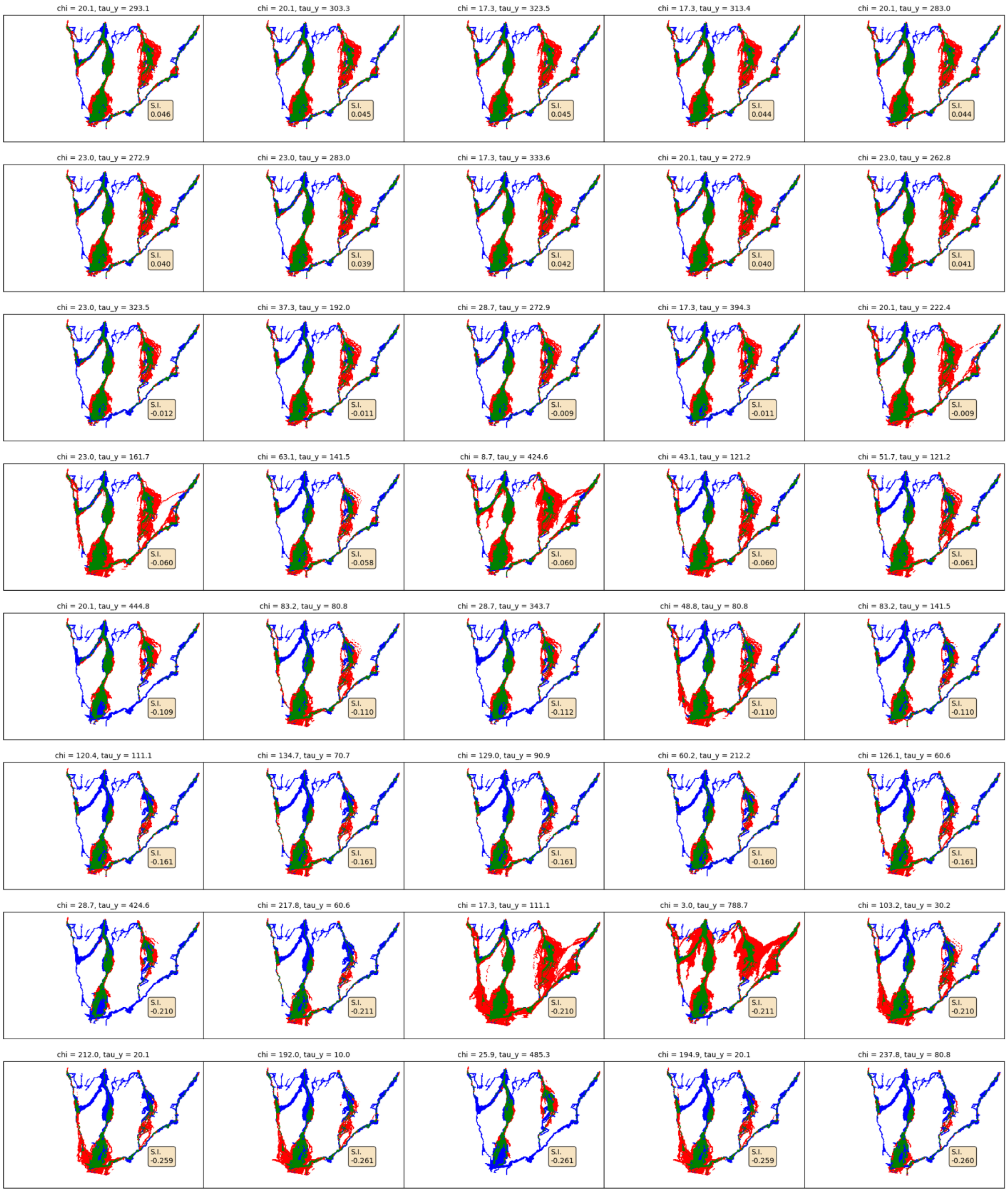

Figure S9: Inundation maps used by experts (see Text S1) to determine the similarity index contour level of half degraded performance. In each plot, green = True Positives, blue = False Negatives, and red = False Positives.

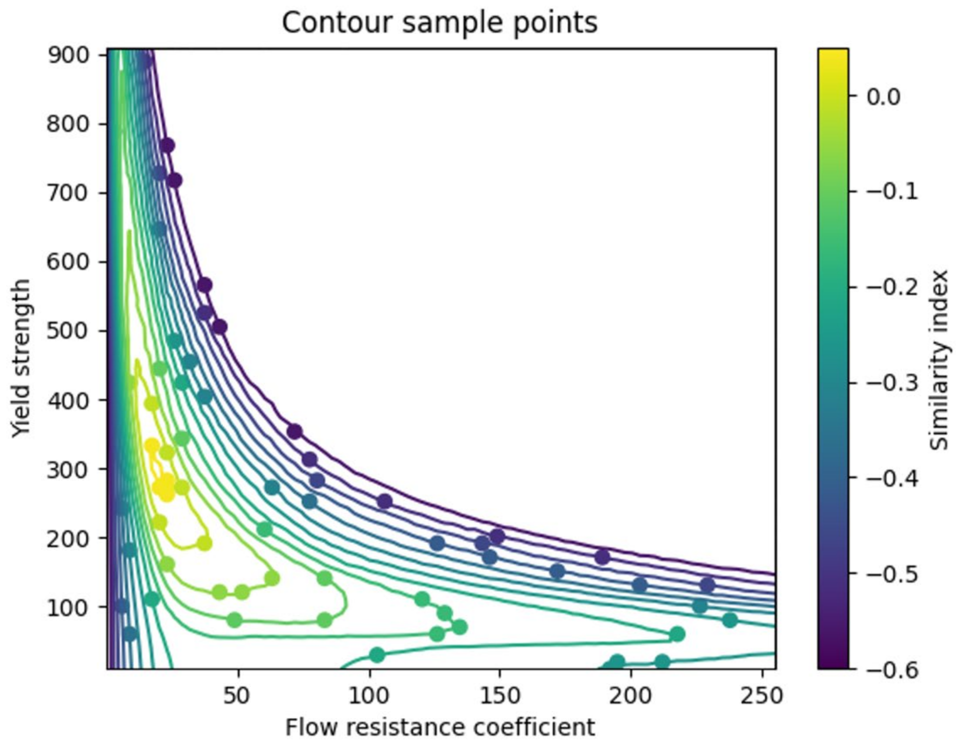

Figure S10: Location of the flow mobility samples that produce the maps used in the expert assessment of half degraded performance (Fig. S9). Contours and colors map the similarity index of the response surface displayed in Fig. S8.

Text S1: Obtaining and testing the posterior parameter distribution

Markov-Chain Monte Carlo (MCMC) sampling is useful for taking samples from generic distributions that do not have a closed-form representation (MacKay, 2003). In our case, we use the affine invariant MCMC ensemble method of Goodman and Weare (2010) implemented in the open-source *emcee* Python package (Foreman-Mackey et al., 2013) to sample the posterior distribution over the two flow mobility parameters in Progressive Debris-Flow routing and inundation model (ProDF). We then sub-sampled from the set of MCMC posterior samples (n=18,496) for all simulations used by ProDF in the debris-flow inundation forecast model and, in the test, performed on Montecito Creek. Rather than running ProDF for every parameter pair sampled by the MCMC walkers, we used linear interpolation on a statistical surrogate model because we had a large number of ProDF model samples and the objective function surface was smooth.

The input volumes at each of the training basin initiation points (Oak, San Ysidro, Buena Vista, and Romero; Fig. 1) were the observed volume from Kean et al. (2019, their Table 5). Following Gorr et al. (2022), we used the similarity index (SI; Heiser et al., 2017) to evaluate model performance for different pairings of the flow mobility parameters, which we pre-computed from ProDF simulations on a 90 by 90 discretization of χ - τ_y space (Fig. S8). The surrogate model was fit to this response surface. However, we faced a challenge in transforming the SI into a log-likelihood function for use by the MCMC sampler because the SI does not have the form or properties of a traditional weighted least squares objective function. We addressed this issue by positing a log-likelihood function, $F(\chi, \tau_y)$, in which the traditional weighted least squares objective function is replaced by the adjusted similarity index, $(1 - TS)$ (Barnhart et al., 2021):

$$F(\chi, \tau_y) \; = \; -0.5 * \; [(N_D + N_{PR}) * ln\,(2\pi) + f * (1 - TS)]$$

χ, τ_y = the ProDF flow mobility parameters
N_D = Number of data (3)
N_{PR} = Number of priors (0)
$1 - TS = 0.5 * (1 - SI) = g(\chi, \tau_y)$, the SI transformed so 0=worst and 1=best
f = factor that scales the objective function according to our expert consensus

We determined a value for f by identifying SI values with half degraded performance. We posit that the flow mobility parameters that produce the SI of half degraded performance have half the relative likelihood of those that produce the best-performing SI values. That is, we sought the factor such that:

$$e^{F(\chi,\tau_y)_{best}} / e^{F(\chi,\tau_y)_{half}} = 2$$

We determined the SI of half degraded performance through expert elicitation. All co-authors and two other individuals (one with a Ph.D. in mechanical engineering and another with a MS in statistics) independently determined the value of SI from binary inundation maps computed from simulations along SI contours (Fig. S9 and Fig. S10). All six participants determined the same SI range (less than -0.06 and more than -0.11). This corresponds to a factor between 18 and 26, and so we used a factor of 20.

We tested the MCMC calibrated poster distribution with simulations on the Montecito Creek Basin (Fig. S4). The input debris-flow volume for Montecito Creek was apportioned among the two initiation points according to the pre-event predicted volume for a design storm with 15-minute duration rainfall intensity of 24 mm/hr (Kean et al., 2019, their Table 1). Flow mobility parameters were sampled from the posterior distribution and used in ProDF simulations (n=10,000). The output depth maps were converted to inundation binary maps using a threshold depth of 0.1 m; the binary maps were averaged together with equal weights to produce a probabilistic inundation map; and this map was classified using a threshold probability level of 50% for the purpose of computing the similarity index performance metric.

The test of the calibrated flow mobility parameters at Montecito Creek yielded a similarity index of -0.047, similar to the optimal values reported in Gorr et al. (2022) and Barnhart et al. (2021). The reliability diagram showed a conditional bias in the forecast probabilities, under-forecasting the lower values and slightly over-forecasting the higher values (Fig. S4). The calibrated posterior distribution of the ProDF flow mobility parameters strongly resembled the pattern of model performance documented in Gorr et al. (2022).

References

Barnhart, K. R., Jones, R. P., George, D. L., McArdell, B. W., Rengers, F. K., Staley, D. M., and Kean, J. W.: Multi-model comparison of computed debris flow runout for the 9 January 2018 Montecito, California post-wildfire event, J. Geophys. Res.-Earth, 126(12), e2021JF006245, https://doi.org/10.1029/2021JF006245, 2021.

Foreman-Mackey, D.: Corner.py: Scatterplot matrices in Python, J. Open Source Softw., 1(2), 24, https://doi.org/10.21105/joss.00024, 2016.

Foreman-Mackey, D., Hogg, D. W., Lang, D., and Goodman, J.: Emcee: The MCMC Hammer, Publ. Astron. Soc. Pac., 125, 306–312, https://doi.org/10.1086/670067, 2013.

Goodman, J., and Weare, J.: Ensemble samplers with affine invariance, Comm. App. Math. Com. Sc., 5(1), 65–80, http://doi.org/10.2140/camcos.2010.5.65, 2010.

Gorr, A. N., McGuire, L. A., Youberg, A. M., and Rengers, F. K.: A progressive flow-routing model for rapid assessment of debris-flow inundation, Landslides, 19, 2055–2073, https://doi.org/10.1007/s10346-022-01890-y, 2022.

Heiser, M., Scheidl, C., and Kaitna, R.: Evaluation concepts to compare observed and simulated deposition areas of mass movements, Computat. Geosci., 21(3), 335–343, https://doi.org/10.1007/s10596-016-9609-9, 2017.

Kean, J. W., Staley, D. M., Lancaster, J. T., Rengers, F. K., Swanson, B. J., Coe, J. A., Hernandez, J. L., Sigman, A. J., Allstadt, K. E., Lindsay, D. N.: Inundation, flow dynamics, and damage in the 9 January 2018 Montecito debris-flow event, California, USA: Opportunities and challenges for post-wildfire risk assessment, Geosphere, 15(4), 1140–1163, https://doi.org/10.1130/GES02048.1, 2019.

MacKay, D. J. C.: Information Theory, Inference, and Learning Algorithms, Cambridge University Press, Cambridge, England, 640 pp., ISBN 9780521642989, 2003.

Pianosi, F., and Wagener, T.: Distribution-based sensitivity analysis from a generic input-output sample, Environ. Modell. Softw., 108, 197–207, https://doi.org/10.1016/j.envsoft.2018.07.019, 2018.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Liu, Z., Berner, J., Wang, W., Powers, J. G., Duda, M. G., Barker, D., and Huang, X.-Y.: A description of the advanced research WRF model version 4.3, National Center for Atmospheric Research, Boulder, Colorado, Technical Notes NCAR/TN-556+STR, 162 pp., https://doi.org/10.5065/1dfh-6p97, 2021.

Disclaimer. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.