Maricar Rabonza (Dr)[1,2], Michele Nguyen (Dr.)[1], Sebastien Biass (Dr)[3], Susanna Jenkins (Asst Prof)[1,2], Benoit Taisne (Assoc Prof)[1,2], David Lallemant (Asst Prof)[1,2]
[1] Earth Observatory of Singapore, [2] Asian School of the Environment Nanyang Technological University Singapore, [3] Department of Earth Sciences, University of Geneva Switzerland

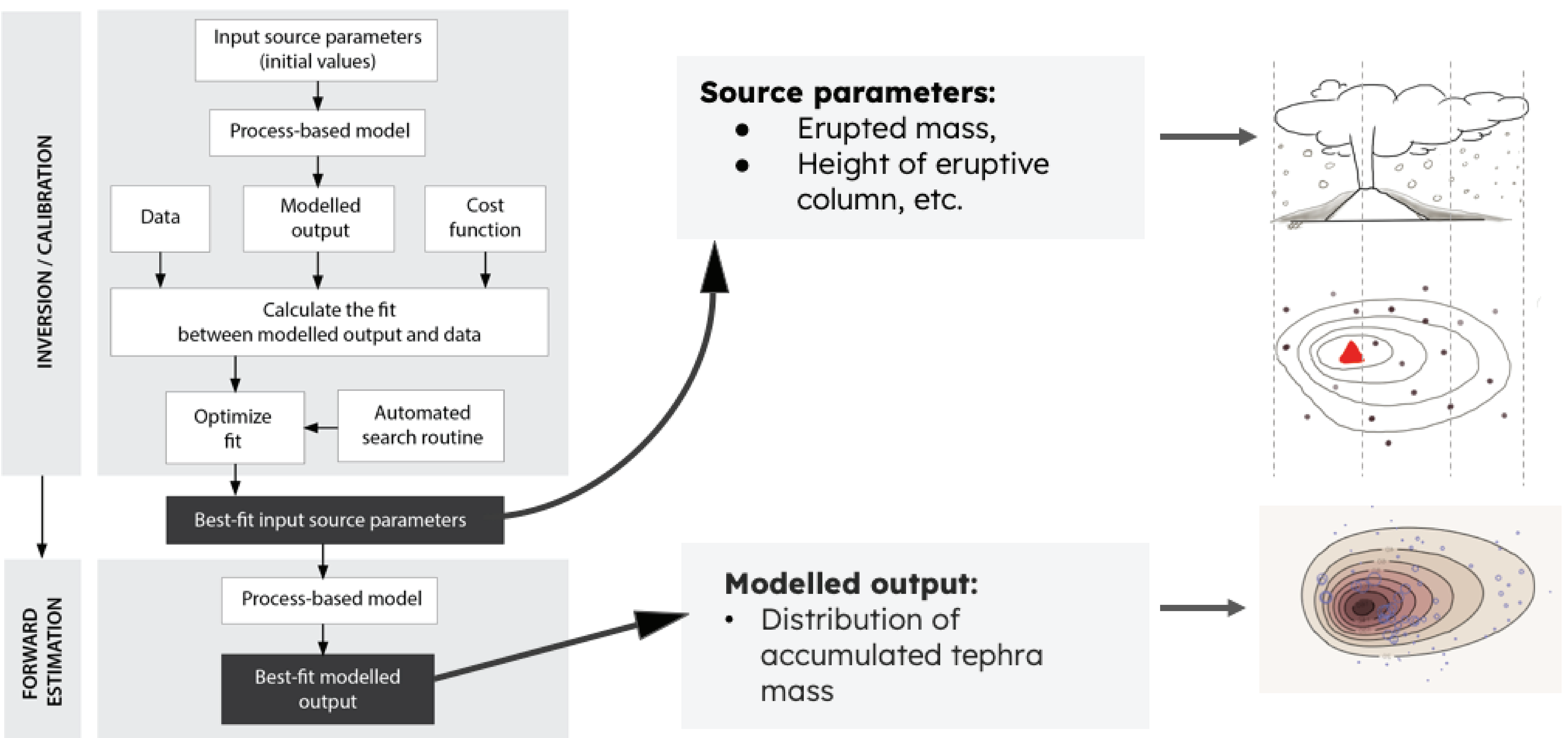
Abstract ID: 1384223
Paper Number: H21W-1652
AGU 2023, San Francisco - 11-15 Dec 2023

Leveraging spatial information and data-related uncertainty in process-based model inversion and forward estimation

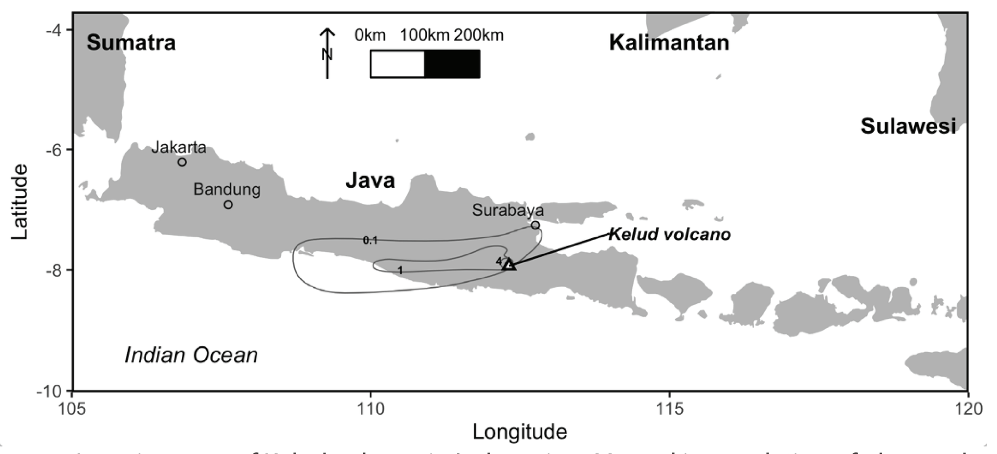
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE · EARTH OBSERVATORY OF SINGAPORE · DISASTER ANALYTICS FOR SOCIETY LAB @NTU

How can we make the most of limited and uncertain spatial data in inversion and forward estimation with process-based hazard models?

Inversion-foward estimation framework with a tephra fallout model (Tephra2)



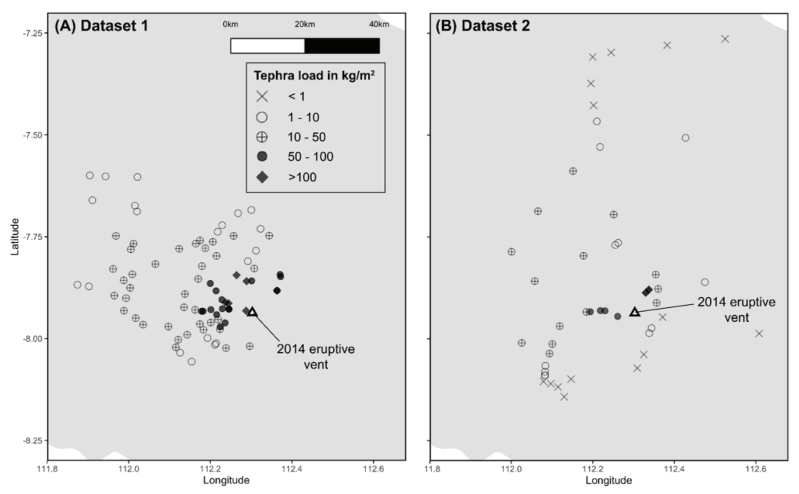
The test case: Modelling the tephra fallout from the 2014 Kelud volcano eruption



Location map of Kelud volcano in Indonesia + Manual interpolation of observed data from Maeno (2019)

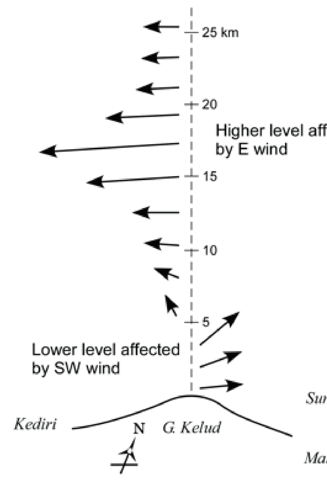
Tephra fall damage to residential roofs at the volcano's vicinity. Damage assessments can benefit most if spatial predictions reflect building-scale observations.

Dataset 1: Obtained 2-3 days post-eruption



Dataset 2: Obtained a month after the eruption

Dataset 1 has less uncertainty than Dataset 2 due to delay in the field data acquisition between datasets



The wind pattern above Kelud volcano during the 2014 eruption varies at different elevations

3 areas of contribution:

Full paper: Environ. Model. Soft. Journal

iPoster:

The choice of cost function in the calibration process

What is the impact of this choice?

Different cost functions treat data differently.

Sensitive to outliers	Treat small and large values equally	Treat under/overestimation differently
Squared error	Relative square error	Logarithmic error
$(y_i - x_i)^2$	$\frac{(y_i - x_i)^2}{x_i}$	$\log \frac{y_i}{x_i}$

Prediction (y_i), Observation (x_i)

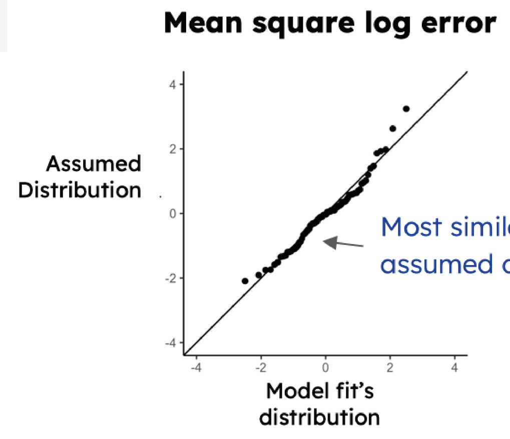
We propose to use cost functions that treat thin and thick deposits equally.

Mean square log

$$\frac{1}{n}\sum_{i=1}^{n}\left(\log \frac{y_i}{\boxed{x_i}}\right)^2$$

observation

Cost functions imply a certain distribution on the errors



Takeaways:

1. The selection of the cost function needs to be a conscious choice in the inversion process.
2. Each cost function comes with assumptions regarding how the model would fit the data.
3. The best-fit source parameters are influenced by this choice.

Making use of multiple data sources with varying uncertainty

How can we account for varying uncertainty?

Extended the Tephra2 inversion algorithm to account for varying uncertainty across data points

Weighted Mean Square Log Error

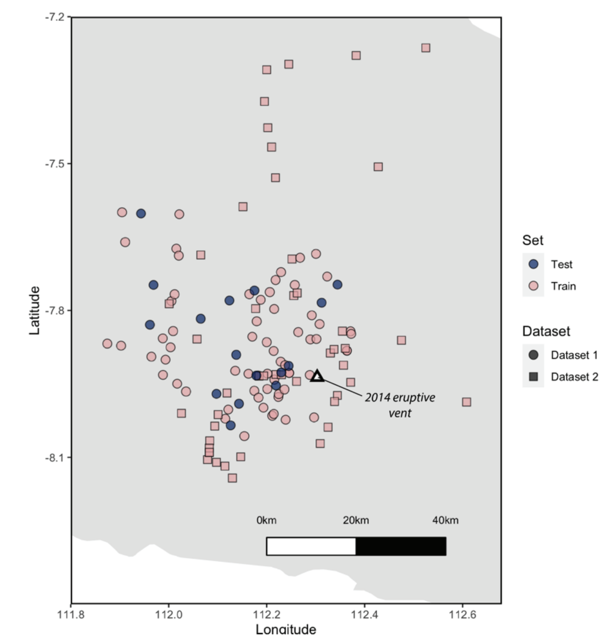
$$\frac{1}{n}\sum_{i=1}^{n}\boxed{w_i}\left(\log \frac{y_i}{x_i}\right)^2$$

Uncertainty weight

Adding weights results to better predictive performance on unseen data

Errors on data points not used in the calibration

Test error (Standard MSLE)	→	Test error (Weighted MSLE)
0.05		0.03



Blue points were not included in the calibration (i.e. test points)

Takeaways:

1. The approach fits the model to the more reliable data, while still considering the information from the less reliable data.
2. Using weights -> lower test error -> better predictive performance

Combining model estimates and data in the forward-model spatial estimate

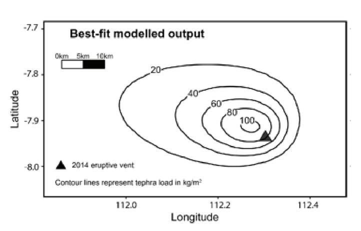
How can we further improve the map of tephra accumulation?

Modelled outputs may diverge from observations in a spatially structured way due to model approximations, and unaccounted processes in the model.

We propose a model-data fusion approach that combines the forward model output and the data to improve estimates of the spatial distribution of tephra fall load.

The weighted fusion shows better out-of-sample performance from a cross validation (from 0.040 to 0.033).

Step 1: Inversion and forward modelling

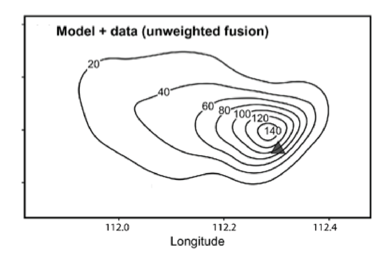


Step 2: Predict residuals at all sites using **kriging interpolation:**

$$\hat{Z}(s_o) = \sum_{i=1}^{N} \lambda_i Z(s_i)$$

Residual at any location ($\hat{Z}(s_o)$); Residual at observed location ($Z(s_i)$); **Kriging weight** (λ_i) (Captures the spatial relationships in the data)

Step 3: Add modelled estimates from Step 1 to the predicted residuals



Takeaways:

1. The fusion results to an improved modelled output that better reflects the data and its spatial characteristics
2. Incorporating uncertainty-based weights to the fusion improves predictive performance.

Presenter: Dr. Maricar L. Rabonza (maricar.rabonza@ntu.edu.sg). This poster is based on our paper entitled "Inversion and forward estimation with process-based models: An investigation into cost functions, uncertainty-based weights and model-data fusion" published in the Environmental Modelling & Software. Funded by NRF Singapore and EOS.