

*Peer Review Framework for Predictive Analytics
in Humanitarian Response*

MODEL REPORT: Typhoon impact model

510 - An initiative of the Netherlands Red Cross

August 2022

OCHA CENTRE FOR HUMANITARIAN DATA



OCHA

centre for humdata

Model Report:

Typhoon impact model

1. Background

This document summarises the documentation and the finding of the peer review of the trigger used in OCHA's Philippines Anticipatory Action framework for typhoons. The impact prediction model used as the basis for the trigger was developed by the Netherlands Red Cross 510 data and digital team, on behalf of the IFRC network and its partners. The model uses typhoon track and rainfall forecasts to predict the percentage of houses that will be severely damaged per municipality in the Philippines. OCHA Philippines used this model as a basis to begin a response before a typhoon hit, based on the predicted number of houses damaged in the regions of Bicol and Eastern Visayas. To find out more about this application, please see the [Anticipatory Action Framework for the Philippines](#).

This peer review was conducted between September 2021 and July 2022.

2. Main Findings and Recommendations

The documentation regarding the model, its application and the review process can be found at the following links:

- The [Model Card](#) describes version 1.0 of the model and was completed in September 2021.
- The [Model Evaluation Matrix](#) was completed in October 2021 by a technical expert from Leiden University, and in February 2022 by the client.
- The [Implementation Plan](#) was completed in October 2021. It summarises the concrete actions that the model will trigger or inform as part of the 2021-2022 OCHA anticipatory action framework in the Philippines
- The [Ethical Matrix](#) aims to identify all stakeholders and potential issues regarding the intended use of the model. It was completed in November 2021 by an expert from the London School of Economics; and updated in July 2022.

A summary of the main findings and recommendations is provided below.

2.1 Technical Review

Intended Use

- The target of the model is P (#damaged houses > threshold t). Several combinations of Probability P and threshold t are used as a trigger for early action. Make a clear statement under which circumstance which combination is used. This will help with later evaluation of the model and transparency of its use.

Model Development

- As mentioned in the ethical review, more information about the sources of the input data is desirable. The model should include information on how often the data sets are updated, and how often the model is retrained to reflect these updates.

Model Evaluation

- The model is compared against a damage curve that is currently used to estimate damage. Adjust this benchmark to whichever model or algorithm is used by the Philippine authorities to estimate damage.
- Specify how much variance in model performance is acceptable, i.e., with the model working on average better than the benchmark, how much extreme over / underestimation is acceptable for field use. More specifically, indicate how robust the model performance is over different typhoons.

Operational Readiness

- For every model version, make a model release and keep an updated list of associated input data. This will make reproducibility of earlier results easier and help keep track of the data sources (and whose responsibility those are).

2.2 Ethical Review

Inaccuracy

The model performance scores for 2021 Typhoon Surigae are not reported. The rationale for using the given error / accuracy tool and not others is missing. The model does not perform well at predicting really low and really high damage.

- Recommendation: (1) Report latest error scores; (2) Contextualise the quality of the error scores in light of existing and relevant research; (3) Include the rationale for using a particular method of calculating error score as opposed to others.

False Negatives

In the case of Typhoon Goni in 2020, the model did not predict an extreme event until only a few hours before landfall.

- Recommendation: If the model does have 100% weightage in the activation of the trigger, then the possibility of a false negative should be clear operationally, e.g., in the case of Typhoon Goni. This is of particular importance because we know that the model does perform well in cases of rapid intensification or high damage scenarios.

Missing Attributes

The implementation model mentions a priority list based on poverty, gender, and building material, while the model uses poverty, hazard data, and building type where poverty is not disaggregated unless it follows some national standard of income levels to estimate poverty lines.

- Recommendation: Clarity is required on the relative importance of variables used, and consistency between the model and the implementation plan. There needs to be a distinction drawn between the model and how the outputs of the model are used. Clear delineation between input data (inclusion / exclusion criteria etc.); modelling process, output, and use; and validation / verification, with a clear trail of documentation.

Lead Times

Lead time is the length of time between when a forecast is output and the occurrence of the shock that was forecasted. It is unclear the impact that forecast running time has on lead times for implementation. In the case of Typhoon Ursula (Phanfone) in 2019, the trigger was reached only 12 hours prior to landfall for which no humanitarian action could be planned.

- Recommendation: In the implementation plan, more clarity is required on potential operational impact of long lead times, especially with 100% weightage of model output for trigger activation. The effect of different lead times is unclear for the implementation. To quote one of the documents: 'It takes several hours for the forecast to run so the initial conditions are already often outdated'.

Decision Support

Decision support concerns the extent to which action based on model output.

- Recommendation: Clarity is required in the implementation plan about weightage of model output on trigger activation, i.e., how much of the decision to act depends on the model output. If this is not 100%, then the other considerations should be described.
- Recommendation: Given the possibility of a false negative, operational readiness in case of model failure should be detailed in the implementation plan.

Statistical Bias

Statistical bias occurs when a model or statistic is not representative of the underlying population. In this model, different input datasets are used to model different typhoon scenarios, social vulnerability / poverty data is not sufficiently disaggregated for evaluation, and the relative importance / weightage of the variables in the model is unclear.

- Recommendation: The model requires clear documentation of the different datasets used, particularly the rationale behind parameters used and their disaggregation, e.g., poverty index. Discrepancies on this front exist between the model card and implementation plan, and the model would benefit from their harmonisation. It would also be useful to include these in an overview of the model development process, i.e., the relative weighting and importance of these variables, and how the model changes when they change.

Lack of Transparency

The model is open source and has undergone technical review but is missing some documentation.

- Recommendation: The model would benefit from better documentation of the modelling process, rationale for the algorithm selection, its fitness for purpose, inclusion/ exclusion criteria for variables, etc.

Systemic Bias

Systemic bias is the inherent tendency of a process to support specific outcomes. In this mode, social vulnerability and poverty data are not substantially disaggregated. The move from 'common dataset' to 'composite dataset' involved ranking of variables.

- Recommendation: Clarity, rationale and disaggregation of parameters like social vulnerability and poverty would be beneficial. Clarity required in relative weights of variables when ranking.