Supporting Information for "A global flood risk modeling framework built with climate models and machine learning"

David A. Carozza[1], Mathieu Boudreault[1]

[1]Département de mathématiques, Université du Québec à Montréal, Montréal, QC, Canada

Contents of this file

1. Text S1 to S3
2. Figures S1 to S11

Introduction

The texts below present details on the neural network model topologies considered (S1), the validation techniques used (S2), and the interpolations and extrapolations applied to the population and GDP per capita data sets (S3). The supporting figures provide a variety of content such as maps of dataset test sample sizes (Figures S1 and S2), a map of population exposure (Figure S3), validations of globally and regionally fitted occurrence models (Figures S4 and S5), maps of aggregated occurrence model validations for the logistic regression and neural network models (Figures S6 and S7), aggregated validations of the globally and regionally fitted impact models (Figures S8 and S9), and analyses of the standard deviation correction applied to the impact model (Figures S10 and S11).

Text S1 Neural Network Model Topology

For the global occurrence model, we examined 3 topologies for the NN model: 1. a single hidden layer with 24 neurons (a 38-24-2 network), 2. two hidden layers with the same total of hidden neurons (a 38-12-12-2 network), and 3. two hidden layers each with 24 neurons (a 38-24-2 network). The choice of 24 neurons for the hidden layers is based on a commonly used rule of 2/3 as the ratio of hidden to input neurons. We consider a maximum of 2000 iterations to achieve convergence of the learning algorithm. Due to the smaller sample sets for the fits over HS2 watersheds, we consider a single hidden layer with 10 neurons (a 38-10-2 network).

For the global impact model, we examined 3 topologies for the NN impact model: 1. a single hidden layer with 26 neurons (a 41-26-1 network), 2. two hidden layers with the same total of hidden neurons (a 41-13-13-1 network), and 3. two hidden layers each with 26 neurons (a 41-26-26-1 network). NN models are fit with a single hidden layer with 13 neurons in the regional fits due to smaller datasets in the HS2 watersheds.

Text S2 Validation

In addition to the ROC curve and the AUC, we here plot the precision-recall curve, which is the positive predictive value (TP/(TP+FP)) versus the true positive rate (TP/(TP+FN)), and the F_1 Score, the harmonic mean of positive predictive value and true positive rate as a function of the cutoff probability that defines a prediction as a "flood". Since 17% of the flood occurrence observations are "flood", one could simply always predict "no flood" and be correct 83% of the time. The precision, however, focuses on the predictions of "flood" by considering the proportion of predicted floods that are truly floods.

Text S3 Population and GDP per capita data

Population from the GPW database is available from 2000-2020. Since growth rates from the early 2000's are not appropriate to back-extrapolate as far back as the 1980's, we use the HYDE3.2 database gridded population from 1980 to relate the GPW population in 2000 to that in 1980. For each grid point (i, j), we assume that

$$\mathrm{GPW}_{i,j,1980} = \frac{\mathrm{HYDE}_{i,j,1980}}{\mathrm{HYDE}_{i,j,2000}} \mathrm{GPW}_{i,j,2000}, \tag{1}$$

which simply implies that the GPW population in 1980 ($\mathrm{GPW}_{i,j,1980}$) is linearly scaled from that of year 2000 by the HYDE3.2 population ratio between 1980 and 2000.

GDP per capita data from the GDPHDI database is available from 1990-2015. For the years 2015-2020, we forward extrapolate at each grid point assuming a continuation of exponential growth based on years 2010-2015. For the years 1980-1990, we back-extrapolate assuming exponential growth at each grid point by applying the same parameters as fitted for 1990-1995.

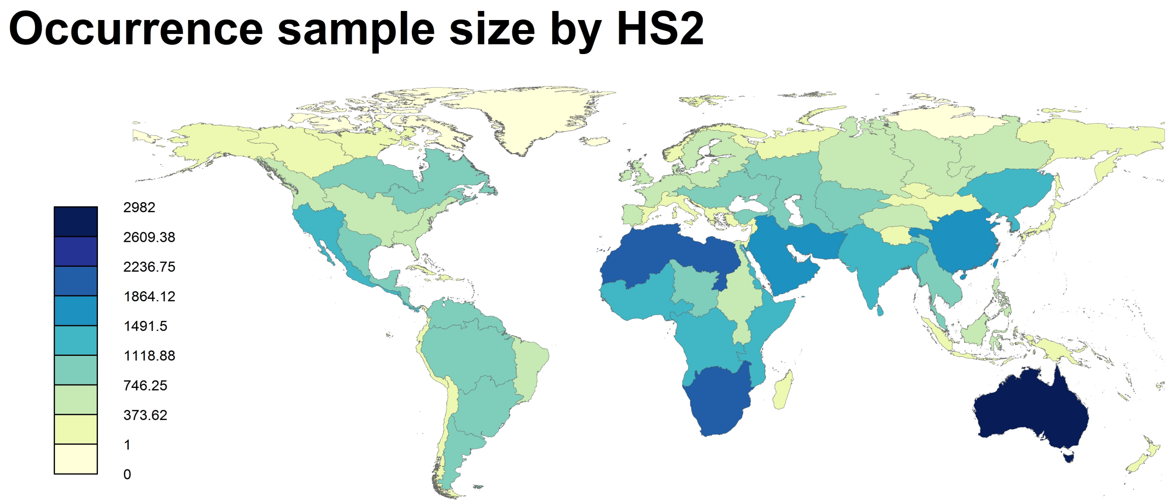

Figure S1. World map of sample size of occurrence model used in model comparison (test sets) aggregated over level 2 watersheds.

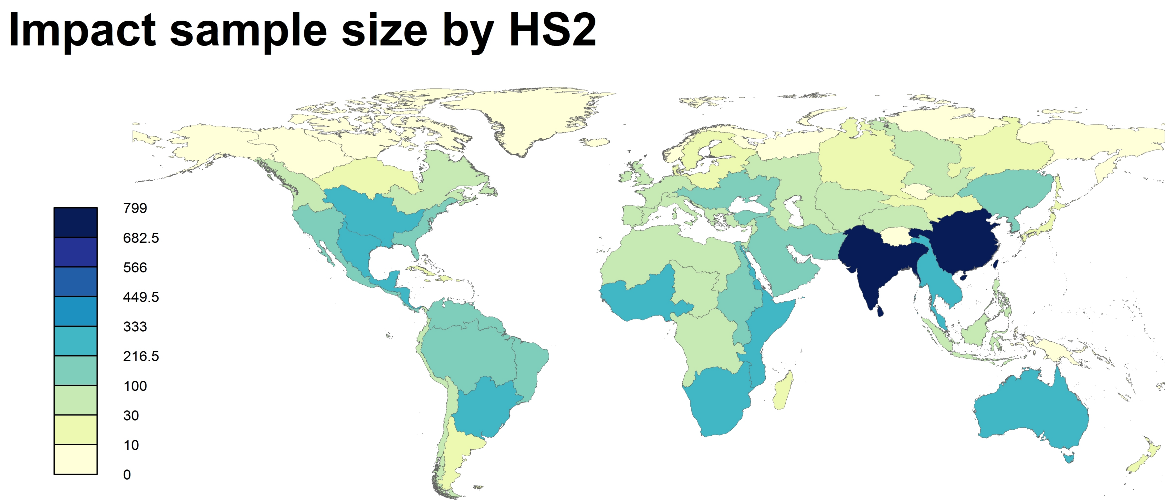

Figure S2. World map of sample size of impact model used in model comparison (test sets) aggregated over level 2 watersheds.

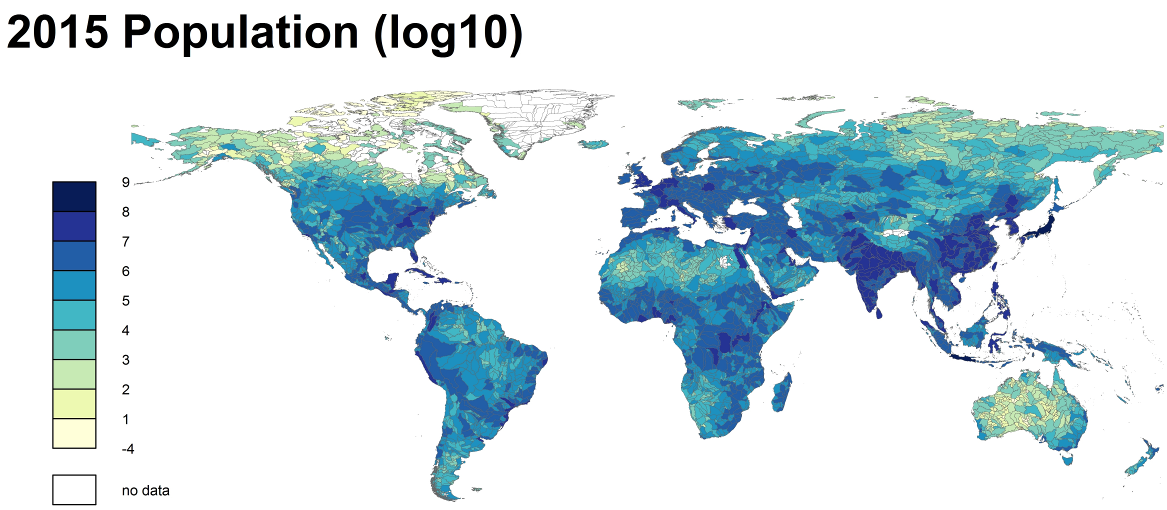

Figure S3. Population (Gridded Population of the World) in 2015 by HydroBASINS level 5 watershed.

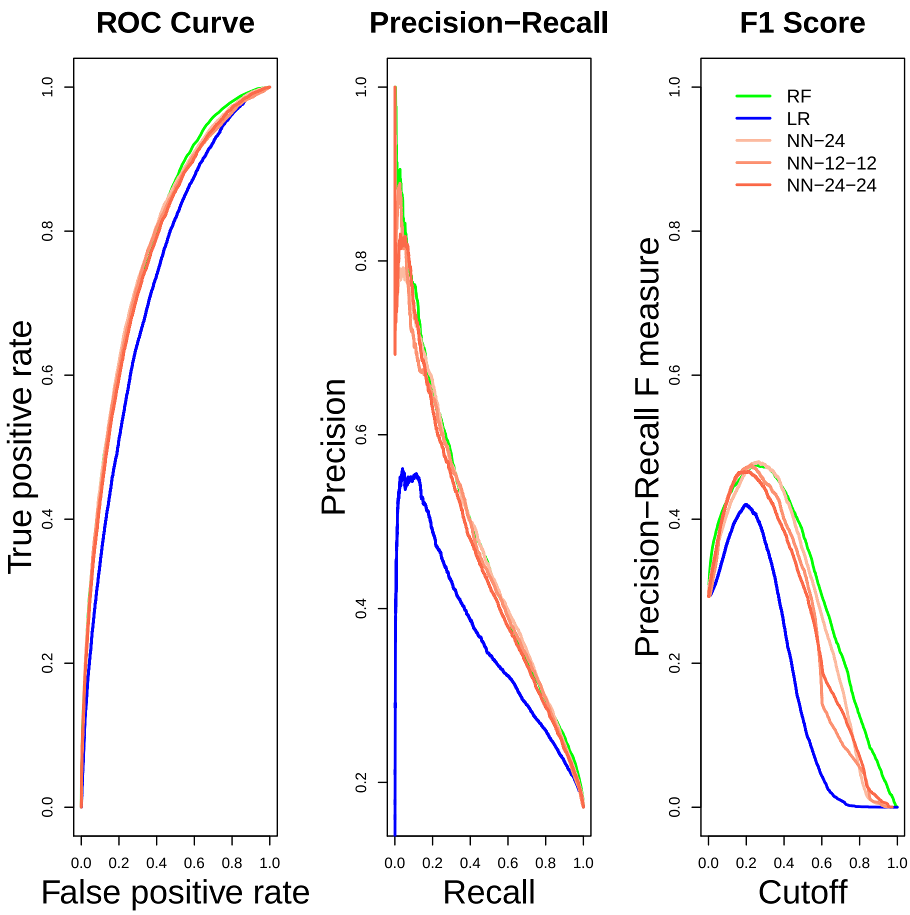

Figure S4. Occurrence validation curves fitted globally. ROC curves (left), Precision-recall curves (middle), F1 scores (right) for globally fitted models.

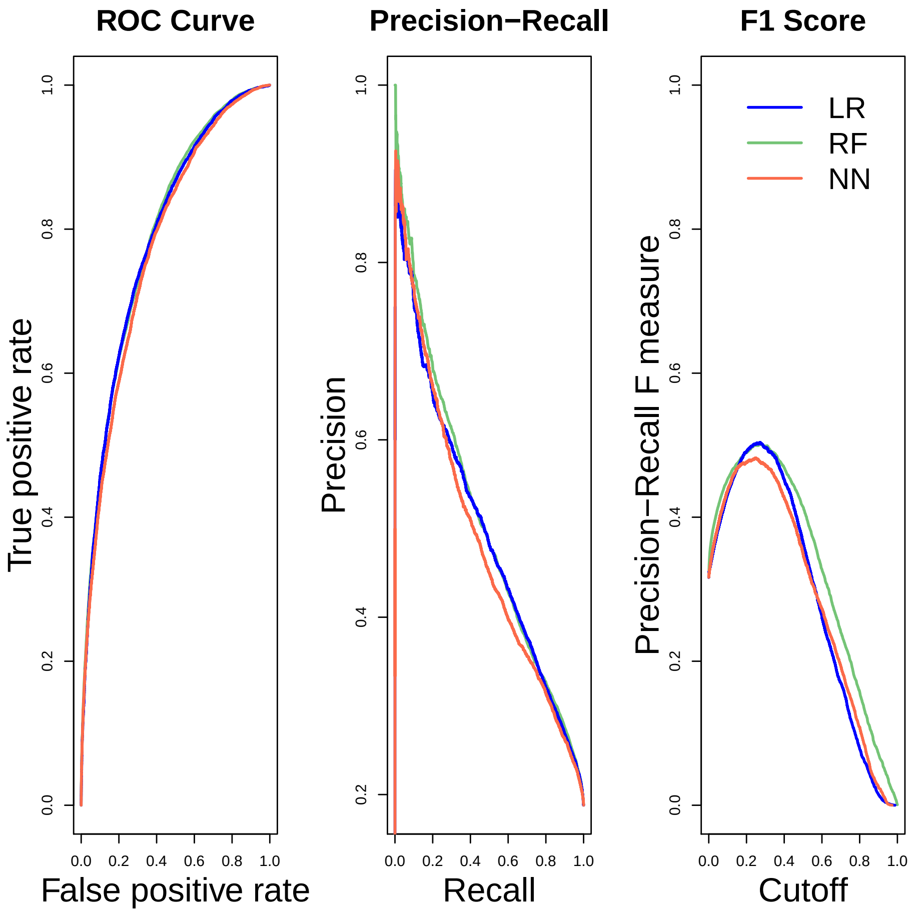

Figure S5. Occurrence validation curves fitted regionally. ROC curves (left), Precision-recall curves (middle), F1 scores (right) for models fitted over HS2 watersheds.

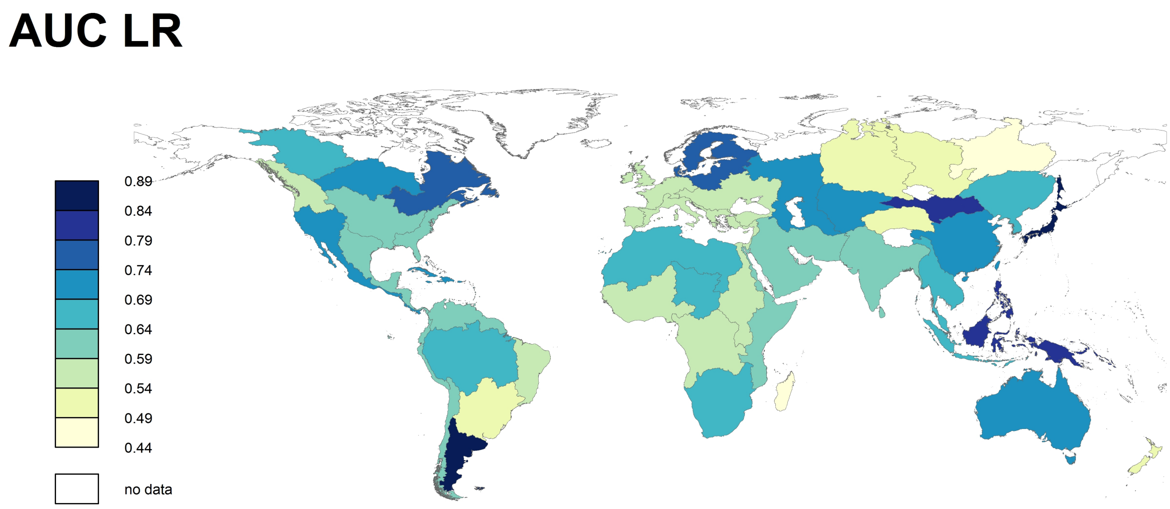

Figure S6. World map of the area under the receiver receiver operating characteristic curve, which measures probability of a model to differentiate between "flood" and "no flood", evaluated at each level 2 watershed (HS2) for the globally fitted logistic regression model.

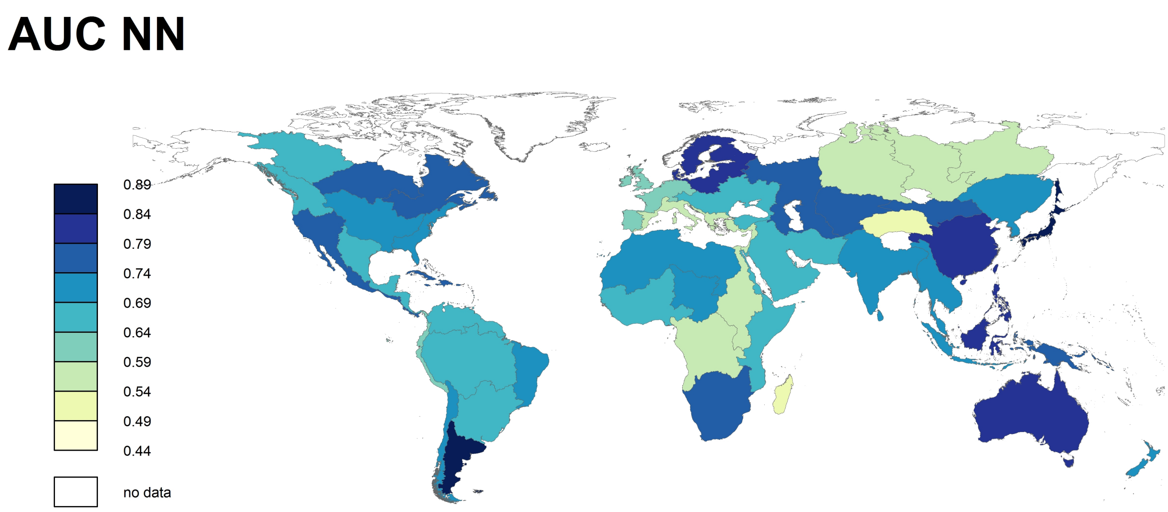

Figure S7. World map of the area under the receiver receiver operating characteristic curve, which measures probability of a model to differentiate between "flood" and "no flood", evaluated at each level 2 watershed (HS2) for the globally fitted neural network model.

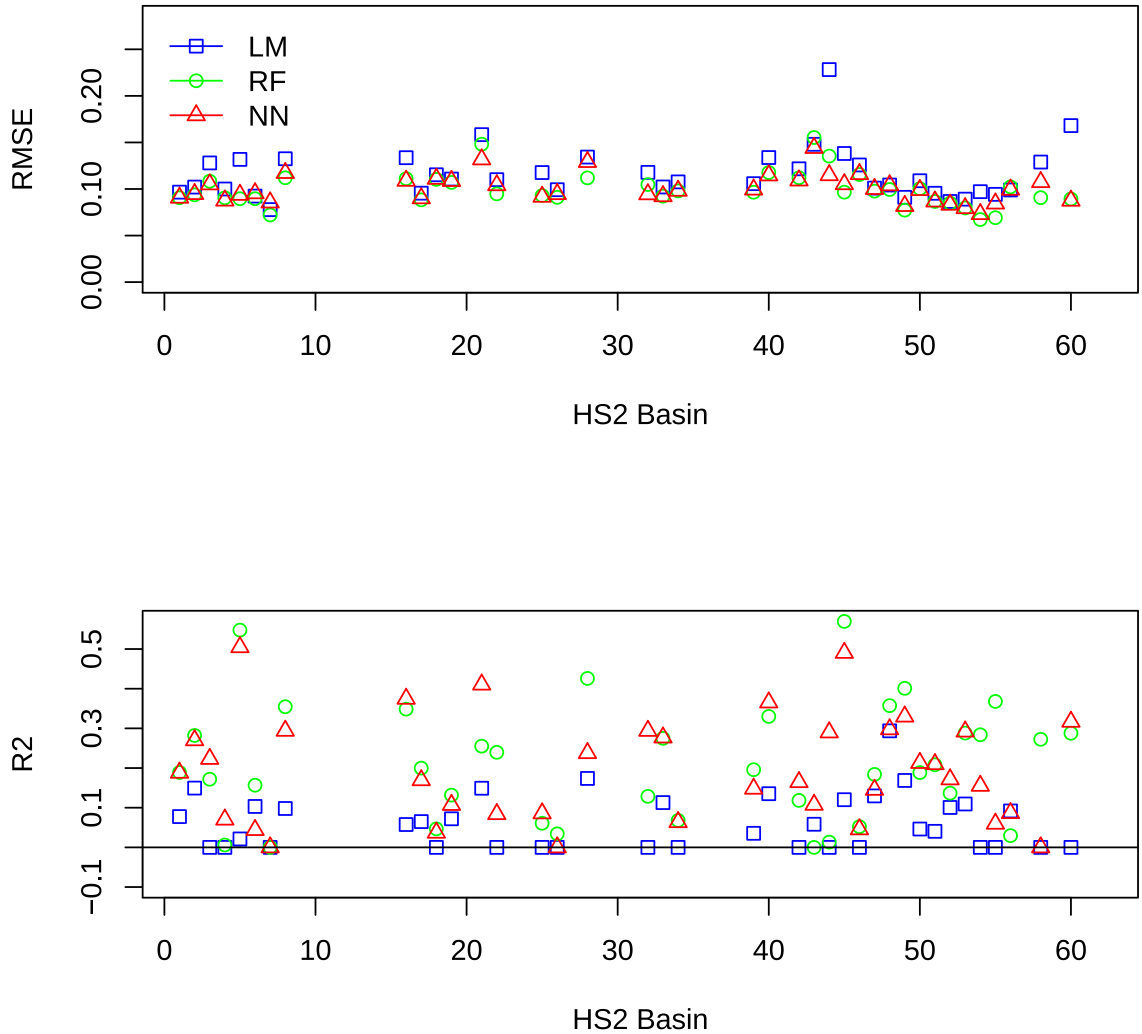

Figure S8. RMSE (top) and R-squared (bottom) by HS2 watershed for globally fitted models (LM, RF and NN).

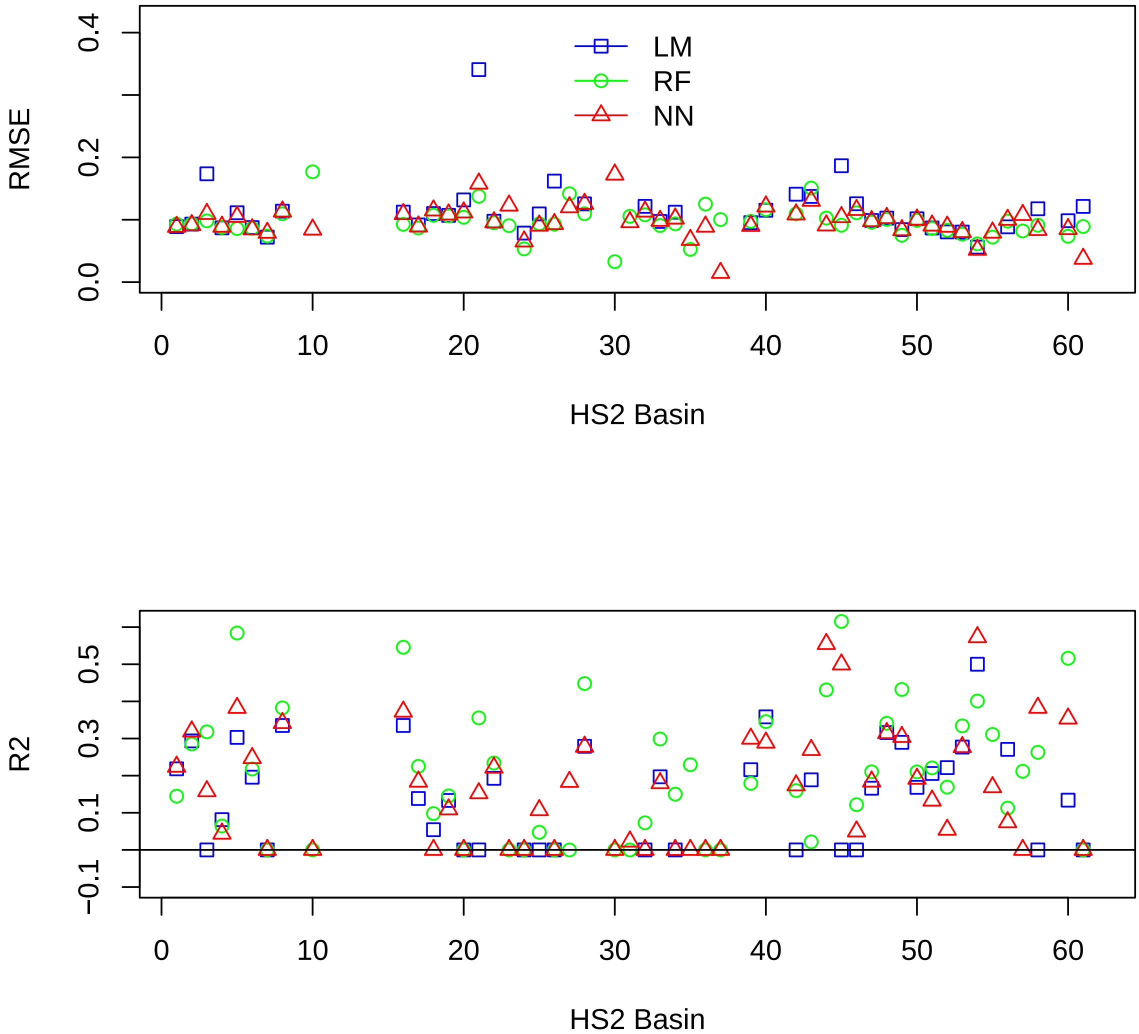

Figure S9. RMSE (top) and R-squared (bottom) by HS2 watershed for models (LM, RF and NN) fitted over subsets of watersheds.

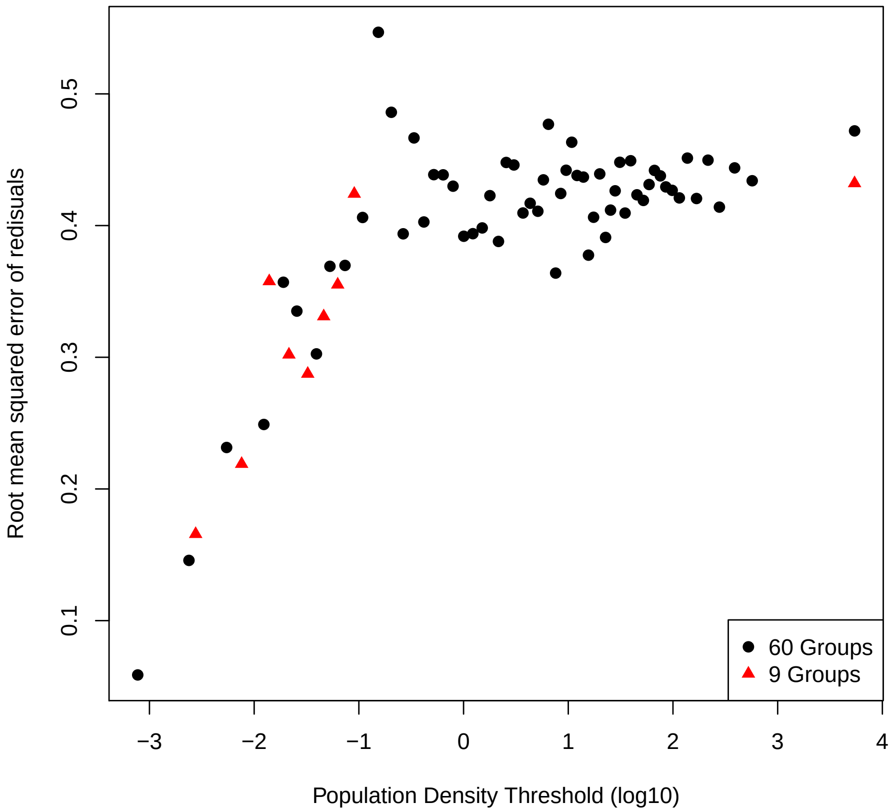

Figure S10. Comparison of the residual root mean squared errors when partitioning into 9 and 60 groups by mean watershed population density. The 60 group partitioning shows that the RMSE increases up to a threshold of -1, from where it remains flat. This allows us to use a simpler standard deviation structure with only 9 groups that preserves the increasing and then flat structure.

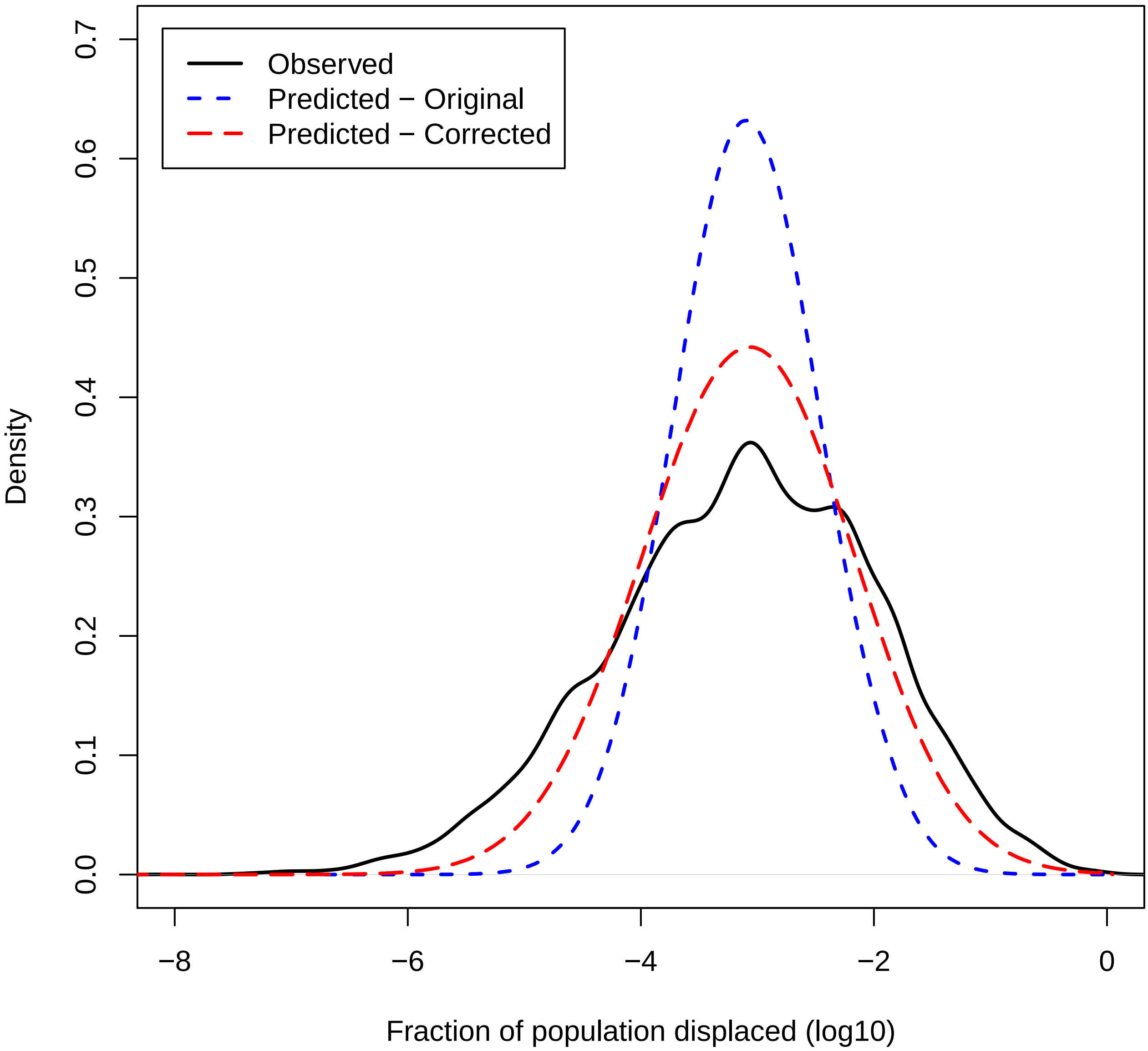

Figure S11. Comparison of observed impact distribution (log10 scale) with those simulated for the flood event catalogue. Observed impacts are solid black, whereas the original and corrected impact predictions are blue (short dash) and red (long dash), respectively.

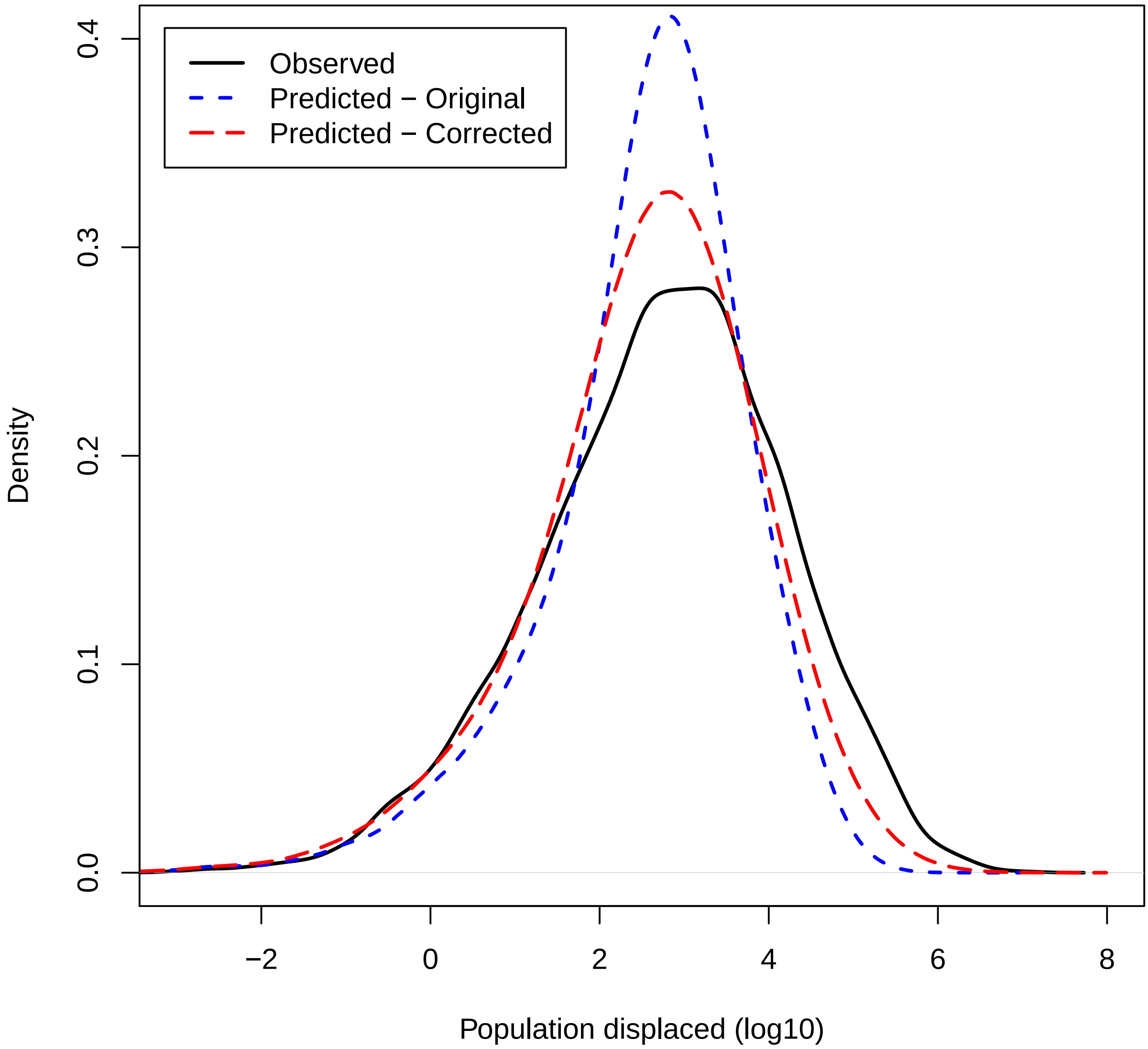

Figure S12. Comparison of observed population displaced distribution (log10 scale) with those simulated for the flood event catalogue. Observed population displaced is solid black, whereas the original and corrected impact predictions are blue (short dash) and red (long dash), respectively. Values are truncated at -3, or 0.001 people displaced. Population displaced of less than one can arise when the population is low and a low fraction of population affected is predicted. For example, a watershed with a population of 100 and and impact of 0.001 will result in 0.1 people displaced.