Supplementary Material: Quantifying Compounded Economic Impacts and Disease Burden of Flooding in Can Tho, Vietnam

Yamile Villafani[1,2], Heiko Apel[1], Laurens J.N. Oostwegel[1], Thi Thao Nguyen Huynh[3,4,5], Hong Quan Nguyen[5,6], Nigel K. Downes[7], Andrea Cominola[8,9], and Nivedita Sairam[1, *]

[1] GFZ German Research Centre for Geosciences, Hydrology, Potsdam, Germany
[2] Geography Department, Humboldt-Universität zu Berlin, Berlin, Germany
[3] IHE Delft Institute for Water Education, Delft, The Netherlands
[4] Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands
[5] Institute for Environment and Resources, Vietnam National University, Ho Chi Minh City, Vietnam
[6] Institute for Circular Economy Development, Vietnam National University, Ho Chi Minh City, Vietnam
[7] College of Environment and Natural Resources, Can Tho University, Can Tho 92000, Vietnam
[8] Chair of Digital Water Systems, Technische Universität Berlin, Straße des 17. Juni 135, Berlin, 10623, Germany
[9] Einstein Center Digital Future, Robert-Koch-Forum, Wilhelmstraße 67, Berlin, 10117, Germany

Correspondence to: Nivedita Sairam (nivedita@gfz.de)

S1. Data description

Table S1.1. Summary of flood survey data used for model development

Flood risks determinants	Value range [mean; median]	Units
Relative building loss	0.00–97.22 [10.90; 3.06]	%
Rotavirus A concentration	0.00–2.96 [0.41; 0.02]	gc/mL ($\times 10^6$)
E. coli concentration	0.00–8.99 [0.66; 0.00]	CFU/mL ($\times 10^3$)

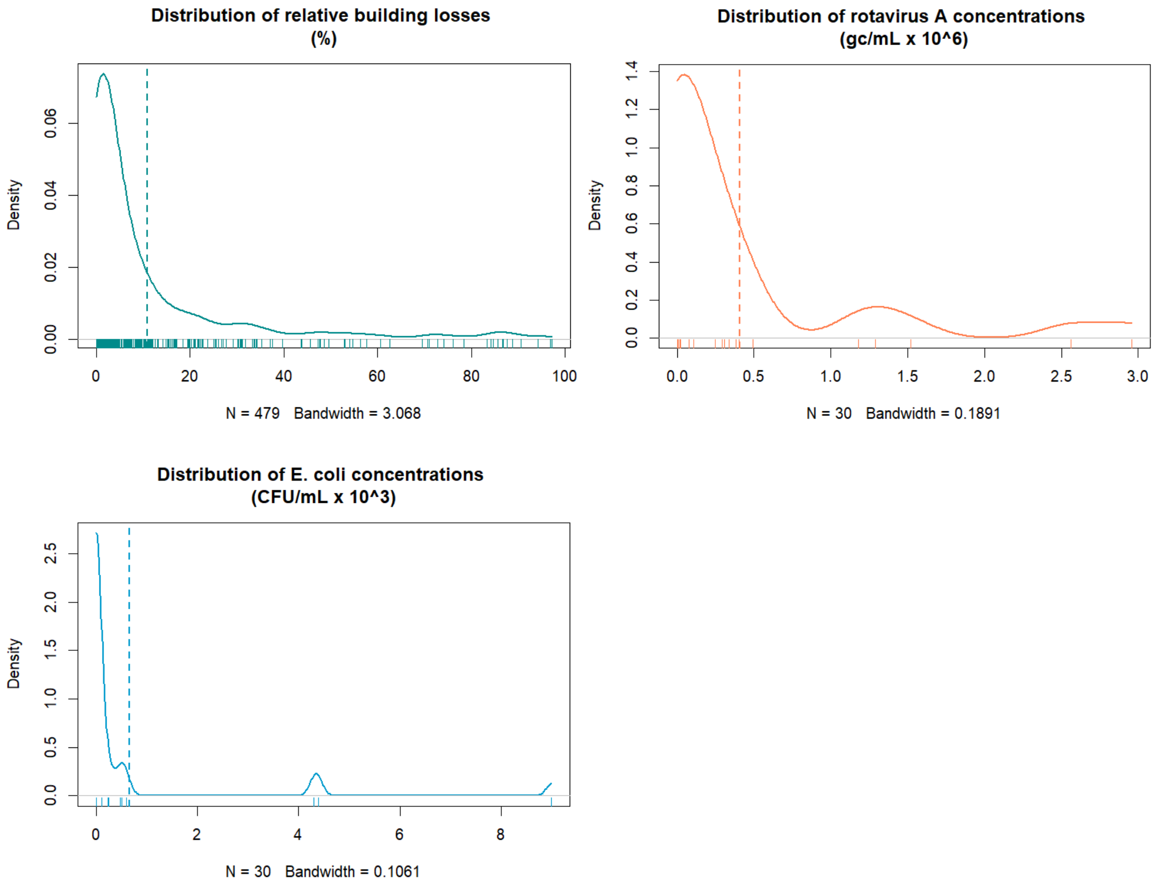

Figure S1.1. Density of survey data: relative building loss (*rbloss*; %), and concentration of rotavirus A and *E. coli* ($\times 10^6$ gc/mL, $\times 10^3$ CFU/mL, respectively).

S2. Bayesian Regression Models

S2.1. Model Definitions

We evaluated alternative models for predicting relative building losses (*rbloss*) and assessed the contribution of candidate predictors. We focus on water depth (*wd*), duration of the flood event (*dur*), frequency of events that hit the residential house (*f_events*), and building value (*bv*) as primary predictors. We contrasted the candidate multivariable model (m_1) to a bivariate model containing *wd* and *bv* (m_2), as well as with an an intercept model without predictors (m_3):

$$m_1 = rbloss \sim wd + bv + dur + f_{events}$$

$$m_2 = rbloss \sim wd + bv$$

$$m_3 = rbloss \sim 1$$

Model comparisons were performed using expected log predictive density (ELPD) via leave-one-out cross-validation. The ELPD comparison indicated no improvement in performance when including additional predictors beyond wd and bv (m_2), with an ELPD difference ≈ 0, while the simpler model (m_2) performed significantly better than the intercept-only model (m_3). Based on these results, we selected the model m_2 including only *wd* and *bv*.

We tested alternative distributions for modelling the concentration (*conc*) of rotavirus A (*rota*) and *E. coli* (*ecoli*), including a negative binomial, a hurdle lognormal model, and intercept-only models. We evaluated whether including additional predictors improved model performance. These are water depth (*wd*), median slope (*topo*), sewer diameter (*sewer*), and the inverse distance to a sewer (*invd*).

$$r_{nb1} = conc \sim 1$$
$$r_{nb2} = conc \sim wd + topo + sewer + invd, hu \sim wd + topo + sewer + invd$$

$$r_{hl1} = conc \sim 1$$
$$r_{hl2} = conc \sim wd + topo + sewer + invd, hu \sim wd + topo + sewer + invd$$

For both pathogens, the expected log-pointwise predictive density (ELPD) difference between the two intercept models (r_{nb1} and r_{hl1}) was greater than twice the standard error, indicating a meaningful improvement in predictive performance for the hurdle lognormal model. We tested the negative binomial model with additional predictors (r_{nb2}), which showed only a marginal improvement over r_{nb1}, confirming that the hurdle lognormal distribution provides superior predictive accuracy. Yet, adding the complementary predictors to the hurdle lognormal model (r_{hl2}) did not yield a meaningful improvement in predictive accuracy.

Consequently, we selected r_{hl1} as the final model to determine the concentration of rotavirus A and *E. coli*. The models for both pathogens remain essentially random and show no strong correlation with the tested covariates. Consequently, inclusion of predictors does not improve model performance.

S2.2. Metrics for predictive performance evaluation

Model performance is evaluated using a set of metrics that capture predictive accuracy and bias. We employ Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Bias Error (MBE), and Mean Continuous Ranked Probability Score (MEAN_CRPS). These are formulated as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \widehat{y_i}|$$

$$MBE = \frac{1}{n}\sum_{i=1}^{n} |(\widehat{y_i} - y_i)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} |(\widehat{y_i} - y_i)}$$

$$CRPS = \int_{-\infty}^{\infty} (F_i(x) - 1\{y_i \leq x\})^2 dx$$

$$Mean\ CRPS = \frac{1}{n}\sum_{i=1}^{n} CRPS_i$$

where $\widehat{y_i}$ is the predicted value, F_i is the cumulative distribution function of the predicted distribution for observation i, and $1\{y_i \leq x\}$ represents the indicator function. MAE measures the average magnitude of prediction errors, RMSE penalises larger deviations, MBE indicates systematic bias, and Mean CRPS corresponds to the average predictive accuracy across observations. Lower values indicate better model performance.

S2.3. Model parameters of the chosen candidate models

Table S2.3.1. Relative building losses model parameters (model m_2)

Parameter	**rbloss**
Model	$rbloss = f(wd, bv)$
Intercept	-1.59
Est. error	0.12
CI 95%	-1.82 – -1.35
zi_intercept	-1.53
Est. error	0.12
CI 95%	-1.76 – -1.31
wd	0.32
Est. error	0.28
CI 95%	-0.22 – 0.87
bv	-3.74
Est. error	0.92
CI 95%	-5.56 – -2.01

Table S2.3.2. Rotavirus A and *E. coli* model parameters (model r_{hl1})

Parameter	**rota**	**ecoli**
Model	$c_{rota} \sim 1$	$c_{ecoli} \sim 1$
Intercept	11.86	6.60
Est. error	0.56	0.61
CI 95%	10.71 – 12.95	5.34 – 7.81
sigma	2.37	1.74
Est. error	0.42	0.50
CI 95%	1.70 – 3.37	1.06 – 3.00
hu	0.34	0.69
Est. error	0.09	0.08
CI 95%	0.18 – 0.51	0.52 – 0.83