High-resolution data, not model flexibility, drive accurate predictions of European winter storm damage

Preprint

D. van Ederen[1,2], W.J.W. Botzen[1], J.C.J.H. Aerts[1,3], P. Scussolini[1], E. Beutner[1], E.E. Koks[1]

1. Institute for Environmental Studies (IVM), Vrije Universiteit Amsterdam, Amsterdam, the Netherlands
2. Achmea, Zeist, the Netherlands
3. Deltares, Delft, the Netherlands

Contact Information

daan.van.ederen@achmea.nl

Predicting winter storm damage at the residential-building level

Expected damage =

Pr(damage occurs) × E(damage ratio | damage occurs)

× Reconstruction value

1. Chance-of-loss model predicts Pr(damage occurs) and is trained on residential building insurance portfolios with 9.4 million policy-storm observations from 9 Dutch winter storms.

2. Conditional-vulnerability model predicts E(damage ratio | damage occurs) and is trained on the 55,000 claims these storms generated, representing ~€136 million in insured damage.

Applied model classes:

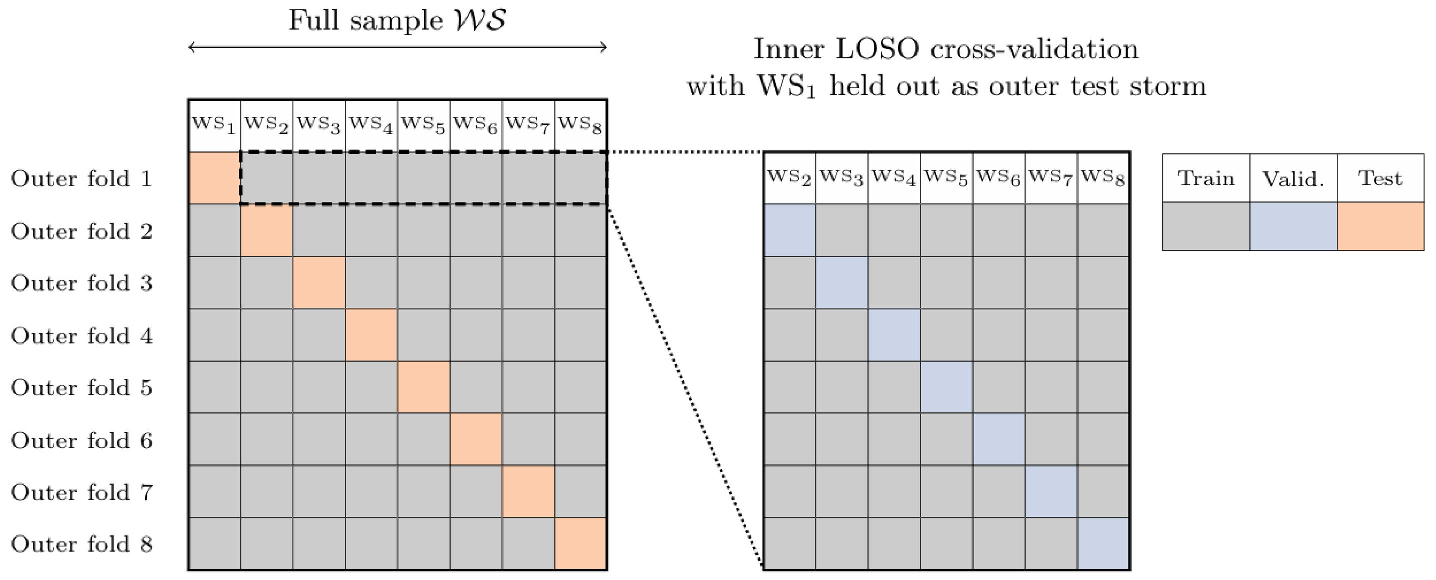
1. Intercept-only benchmark (IO)
Predicts sample average, used to evaluate gains from including features

2. Generalised linear models (GLM)
Least flexible functional form, but offers transparent interpretation

3. Generalised additive models (GAM)
More intricate non-linear effects through smooths, while interpretable

4. XGBoost (XGB)
Gradient-boosted tree method for learning highly non-linear relationships and complex interactions

We test predictions on truly unseen winter storms with nested leave-one-storm-out (LOSO) cross-validation



We measure prediction bias across multiple winter storms (MBE) and storm-level prediction errors (MAPE)

$$MBE = 100 \times \frac{\sum_s \widehat{Y}_s - \sum_s Y_s}{\sum_s Y_s} \qquad MAPE = \frac{100}{S} \sum_s \left| \frac{\widehat{Y}_s - Y_s}{Y_s} \right|$$

where $\widehat{Y}_s$ is the prediction for winter storm s, Y_s is the observed value, and S is the number of held-out winter storms.

Feature selection affects prediction accuracy more than model class

Model	IO	W	W+B	W+B+L	W+B+L+P24	W+B+L+P24+TSm	W+B+L+P24+TSd
MBE							
IO	-0.42%						
GLM		-1.33%	1.71%	1.51%	4.27%	3.62%	58.07%
GAM		-2.90%	1.52%	1.22%	-2.86%	8.35%	-14.35%
XGB		-2.67%	0.60%	**0.23%**	-2.88%	36.18%	-4.81%
MAPE							
IO	62.70%						
GLM		40.11%	41.09%	**37.32%**	43.05%	59.63%	76.48%
GAM		40.61%	43.39%	37.79%	37.84%	61.90%	61.42%
XGB		40.16%	40.85%	37.71%	38.64%	154.91%	42.63%

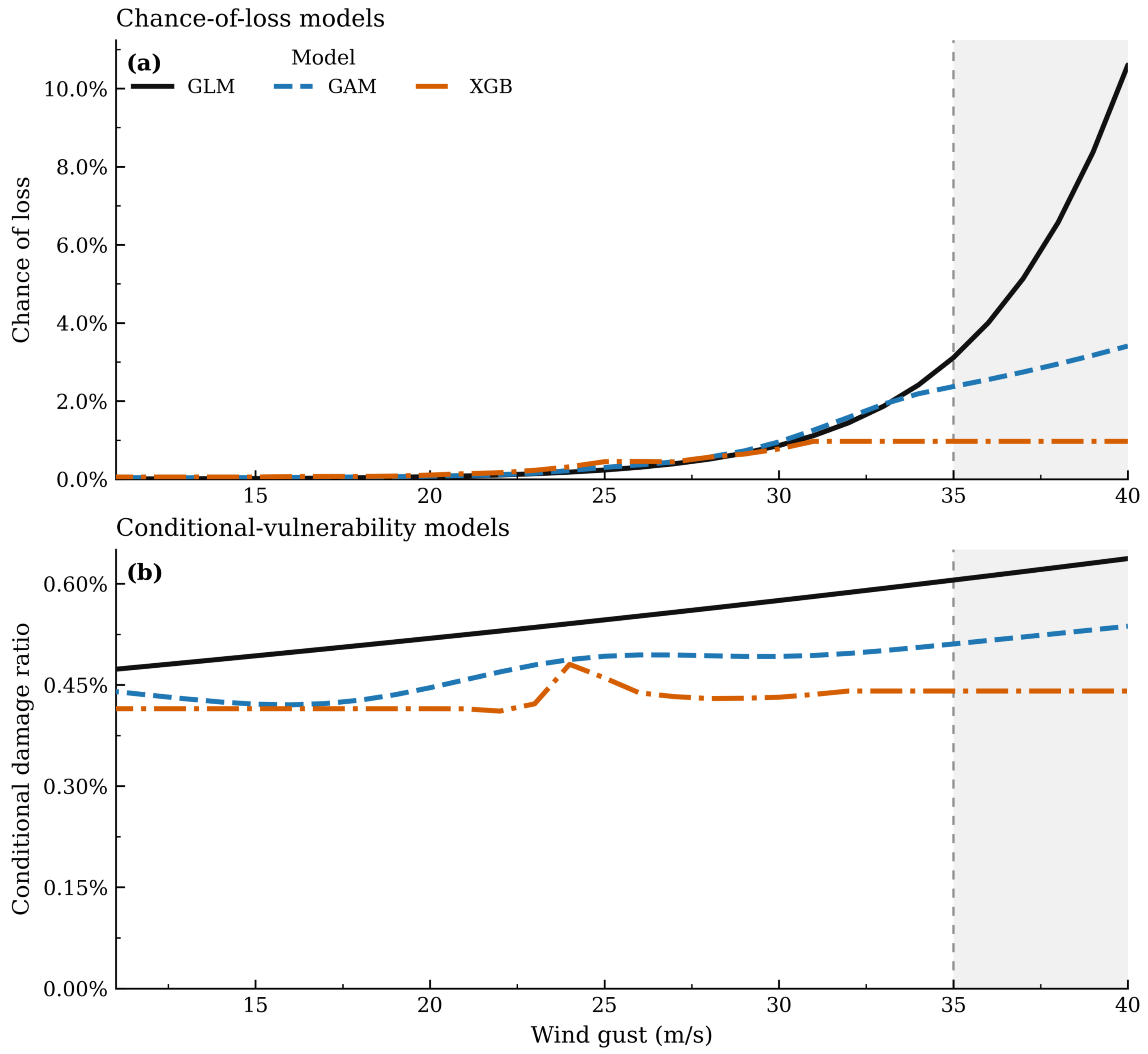
Features for model specifications:
- **IO** = Intercept-only
- **W** = Wind gust
- **B** = Building characteristics
- **L** = Random location effects
- **P24** = 24-hour cumulative precipitation
- **TSm** = mean of thermodynamic state variables
- **TSd** = changes in thermodynamic state variables

Thermodynamic state variables:
- Air pressure
- Air temperature
- Relative humidity

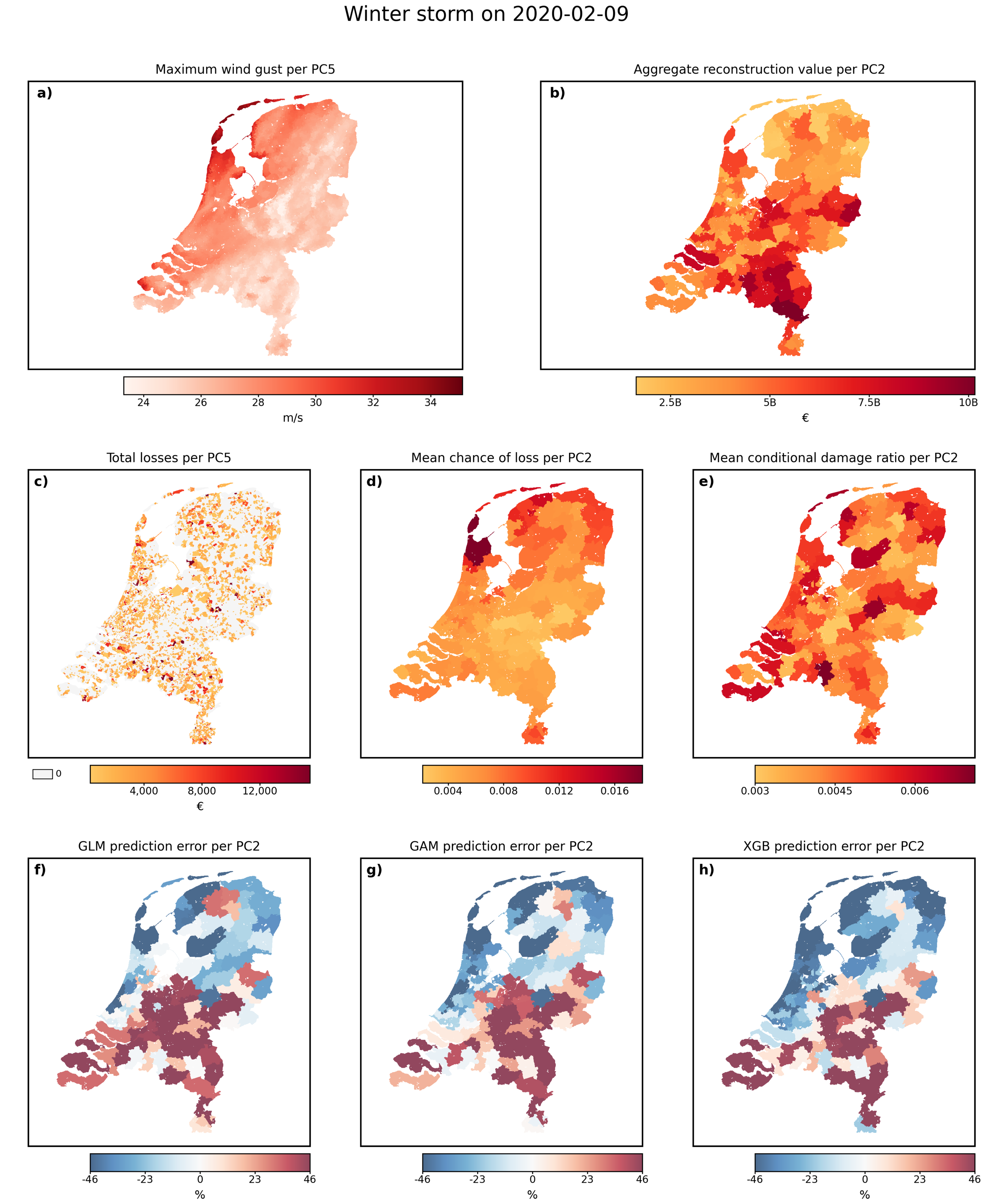
Building characteristics:
- Reconstruction values
- Building type (terraced, corner, semi-detached, and detached houses, and apartments)
- Building age at event
- Number of floors

Gridded meteorological data from the Dutch national weather service (KNMI) at:
2.5 km horiz. res. (W/TSm/TSd)
1 km horiz. res. (P24)

The winter storm damage process is characterised by numerous non-linear relationships



The spatial distribution of damage predictions is highly similar across model classes



Interactions act as secondary damage modifiers and give XGBoost no significant advantage over GLM and GAM

SHAP decomposition of the wind-gust contribution in the chance-of-loss model