# Algorithmic Fairness in Geo-intelligence Workflows through Causality

Brian K. Masinde[1], Caroline M. Gevaert[1], Michael H. Nagenborg[2],
Marc van den Homberg[1] and Jaap A. Zevenbergen[1]

[1]*Faculty of Geo-information Science and Earth Observation, University of Twente, Enschede, The Netherlands*
[2]*Department of Philosophy, University of Twente, Enschede, The Netherlands*

### Abstract
In this paper, we investigate how causality (causal inference) can be used to detect bias and ensure fairness in geo-intelligence workflows . We investigate the usefulness of such a causality-based approach in the context of an early warning system that predicts building damage at municipality levels in The Philippines. We use directed acyclic graphs to reason about the causal relationships in the model case study and quantify the relationships using structural equation modelling. Mediation analysis is also used to validate the causal relationships between variables. We find cases of confounder bias and Simpsons paradox that could potentially bias the damage predictions. However we note that the objective and outcome variable in the early warning system needs to be defined in a manner that allows for more nuanced investigation on fairness (i.e., from damage assessment to impact assessment).

### Keywords
Biases, algorithmic fairness, geo-intelligence, disaster early warning systems

## 1. Introduction

Geo-intelligence (Artificial Intelligence (AI) and geodata) workflows are increasingly being used in disaster early warning early action systems to determine areas and communities at risk. Geo-intelligence workflows include collection, processing using AI, and dissemination of geodata such as satellite/drone images and other geo-tagged data (i.e., data with location references). An example of such an application is in trigger models for anticipatory action. Trigger models predict the location and impact of natural disasters enabling responders and communities to prepare resources (e.g., finances and personnel) [1]. Specific examples of use of AI include building detection in drone images and using historical data of natural hazards to predict impact of hazards.

As with other applications of AI for social problems, use of geo-intelligence workflows raise a number of concerns including fairness and transparency [2, 3]. Since, the context involves vulnerable communities, organizations deploying geo-intelligence workflows are obliged to be fair and transparent in how they decide to distribute aid resources. However, AI challenges

fairness and transparency as data biases and technical limitations can affect the reliability and trustworthiness of geo-intelligence workflows.

In this paper, we investigate how causality (causal inference) can be used to promote fairness by detecting bias in the data and design of geo-intelligence workflows. As a case/study we consider an early warning system that predicts building damage by tropical cyclones (TC) at municipality levels in The Philippines. Fairness is a pertinent issue in disaster response/management because of the limited resources and the lives and livelihoods at stake. We explore how causality can be used for algorithmic fairness since non-causal (associational) models have been shown to replicate, reinforce and propagate biases in observational data. Furthermore, some biases and statistical anomalies cannot be addressed using association/correlation language used in non-causal models (e.g., confounder bias [Figure 3] and the Simpson's paradox) [4].

We follow Friedman and Nissenbaum's [5] definition of a bias in computer systems. By their definition, a bias is a *systematic*, and *unfair* discrimination of individuals or groups of individuals. Friedman and Nissenbaum develop a framework of understanding bias in computer systems as either preexisting bias (e.g., data generated from biased society or institutions), technical bias (technical limitations of the algorithms) and or emergent bias (i.e., biases changing real world use such as changing societal values). In the context of our case study, the data does not contain the quintessential sensitive variables (e.g., gender, race/ethnicity). However, the case study uses building typologies as predictor variables which are often considered an indicator of socio-economic well-being. Therefore a biased damage assessment in this context would consequently cause unfair distribution of aid resources to municipalities with a high number of building typologies vulnerable to tropical cyclones.

Although there is literature on using causality for quantifying disaster risk (e.g., [6, 7]), to the best of our knowledge there are none so far on assessing biases and fairness using causality in this context. Though our case study is on aggregated data and does not contain the typical problematic sensitive variables (e.g., gender, ethnicity), it is still important to ensure that there are no biases in the early warning system and that it is fair. In this paper we present preliminary results on detecting biases through causality.

## 2. The Case Study and Data

The Philippines is prone to tropical cyclones (characterized by high speed winds and heavy rainfall) which cause loss of lives and infrastructure damage. While there are accurate models that predict the characteristics of tropical cyclones (e.g., wind speed and rainfall) it is still a challenge to predict and quantify the impact it would have on people. Because of socio-economic inequalities that influence choice of building materials, often this impact is measured by assessing building damage. Because of these challenges (high frequency of tropical cyclone events and cascading effects), geo-intelligence workflows are increasingly leveraged to guide Impact-based Forecasting (IbF) that in turn inform disaster anticipatory actions such as Forecast-based Financing (FbF) [8]. 510, an initiative of The Netherlands Red Cross developed an early warning system for The Philippines. Their model and data serves as our case study.

The observation data is from previous 39 tropical cyclone incidents in The Philippines. The data is aggregated at the municipality level (1486 municipalities), therefore the variables are an

aggregate representation of the hazard effects in each municipality. Hazard variables include wind speed and amount of rainfall and physical vulnerability variables are captured by house typology numbers (grouped by roof type and wall type). Geographical variables are also included (e.g., coastal length). The resulting early warning system under study is a regression model that predicts the damage percentage in municipalities identifying the most impacted (see Table 1 for metrics).

## 3. Methods: Causality

Causal inference is a methodology that aims at quantifying cause and effect from observational data. Causal inference has found applications in algorithmic fairness [e.g., 9, 10, 11] and in explainable AI [12]. There are different methods on causal inference such as Bayesian Networks and Structural Equation Models (SEMs). BNs and SEMs are popular because of the ease of representing causal relationships using graphs, specifically directed acyclic graphs (DAGs) which do not have feedback loops [4]. In this paper we use SEMs (Equation 1) because they can also be used in mediation analysis to estimate and validate causal chain graphs (Figure 2).

$$x_i = f_i(pa_i, u_i), \quad i = 1, \ldots, n \tag{1}$$

Where $f_i$ is a function model (e.g., linear regression models) with $pa_i$ as parent nodes of $x_i$ and $u_i$ a random error.

Mediation analysis aims to quantify both the direct effects and indirect effects of an exposure variable to an outcome. This is done in two steps, one regressing on the outcome variable on the exposure variable (Equation 2) and two regressing the outcome on both exposure and mediator variable (Equation 3). The difference of the exposure coefficients in the two model gives the effects of the exposure (E) interfered by mediator variable (M) (Equation 4) [13].

$$Y = \beta_0 + \beta_{dir} \times E + \epsilon \tag{2}$$

$$Y = \beta_0 + \beta_1 \times E + \beta_2 \times M + \epsilon \tag{3}$$

$$B_{indirect} = \beta_{dir} - \beta_1 \tag{4}$$

To detect potential biases through causality in the early warning system we first define a DAG informed by literature (e.g., [14]) of how tropical cyclones cause damage. This is however limited by the available variables in the data. We use correlation analysis to identify relationships, mediation analysis to confirm chain relationships (Figure 2), and linear regression models to identify cases of Simpsons paradox. Importantly, the DAG help identify potential confounder biases. Secondly, we conduct a path analysis, quantifying the effects of variables on the outcome. We use SEMs with the functions specified as linear regression models to quantify the direct effects among the variables. These steps are a precursor to conducting further causal fairness checks, for example path-specific counterfactual fairness as illustrated by Chiappa [15].

## 4. Preliminary results

Figure 1 shows our causal representation of the tropical cyclones early warning system. It is informed by literature of how tropical cyclones cause infrastructure damage [e.g., 14] and this damage varies because of the building typologies[1]. With many variables to consider, we focus on the ones that have a strong correlation with the outcome variable and among each other. Figure 5 shows the correlation between the variables. Furthermore, departure from the initial associational ML design to a causal model leads to removing variables that do not have a causal link (or interpretation) to the other variables. For example, the number of households registered for social protection can be associated with higher damage levels but there's no causal link/path between this variable and the outcome.

The DAG is based on the causal reasoning that natural hazards have a direct causal effect on damage of buildings. The hazard components - wind speed and total rainfall[2] - have a strong correlation (see Figure 5H) and statistical significance to the outcome (damage). Since every tropical cyclone event is different and there might be other characteristics not captured in the data, we consider the event itself as a parent node to wind speed and total rainfall. The event and municipalities are parent nodes of distance which causes the municipalities to experience different levels of wind speed and rainfall. For example, Table 6 shows that wind speed is indeed a significant mediator between distance and damage. Tables 4 and 3 show that the significance of rainfall as a predictor of damage changes depending on whether or not we account for the specific tropical cyclone events in the model. This is therefore a case of the Simpsons paradox. Furthermore, here we consider distance to express other hazard effects on damage and therefore it becomes a confounder variable. The hazard components may not be entirely representative of tropical cyclones since it does not also consider storm surges as a side effect of the wind speed.

For simplicity, we do not add paths between hazard nodes and build typologies nodes because this introduces a time component. That is over time communities adjust to impact tropical cyclones rebuilding damaged buildings with stronger materials. Mediation analysis however confirms that build typologies can be mediators of effect of hazard on damage levels in our data (see Table 5). Furthermore, the relationship between build typologies could be bidirectional but doing so in our model introduces cycles and structural causal models do not account for cyclic relationships [4].

## 5. Discussion

In this causal re-interpretation of the effect of tropical cyclones on damage levels in municipalities, based on our data, we find the problems of confounder bias (distance confounds both hazard components and the outcome damage levels). In addition, we find a case of Simpsons paradox with the direct effect of rainfall on damage. Following the correct causal path ways to establish effect of hazard on damage plays an important role in ensuring fairness. Path analysis is a precursor to carrying out further fairness tests (e.g., counterfactual analysis). In this case the weaker build typologies (e.g., light roof, light walls) are sometimes proxies for vulnerable

---

[1]Stronger building materials are more resilient to natural hazards
[2]We exclude rainfall measurements at 6 hours and 24 hours because of high correlation to total rainfall, see Figure 5
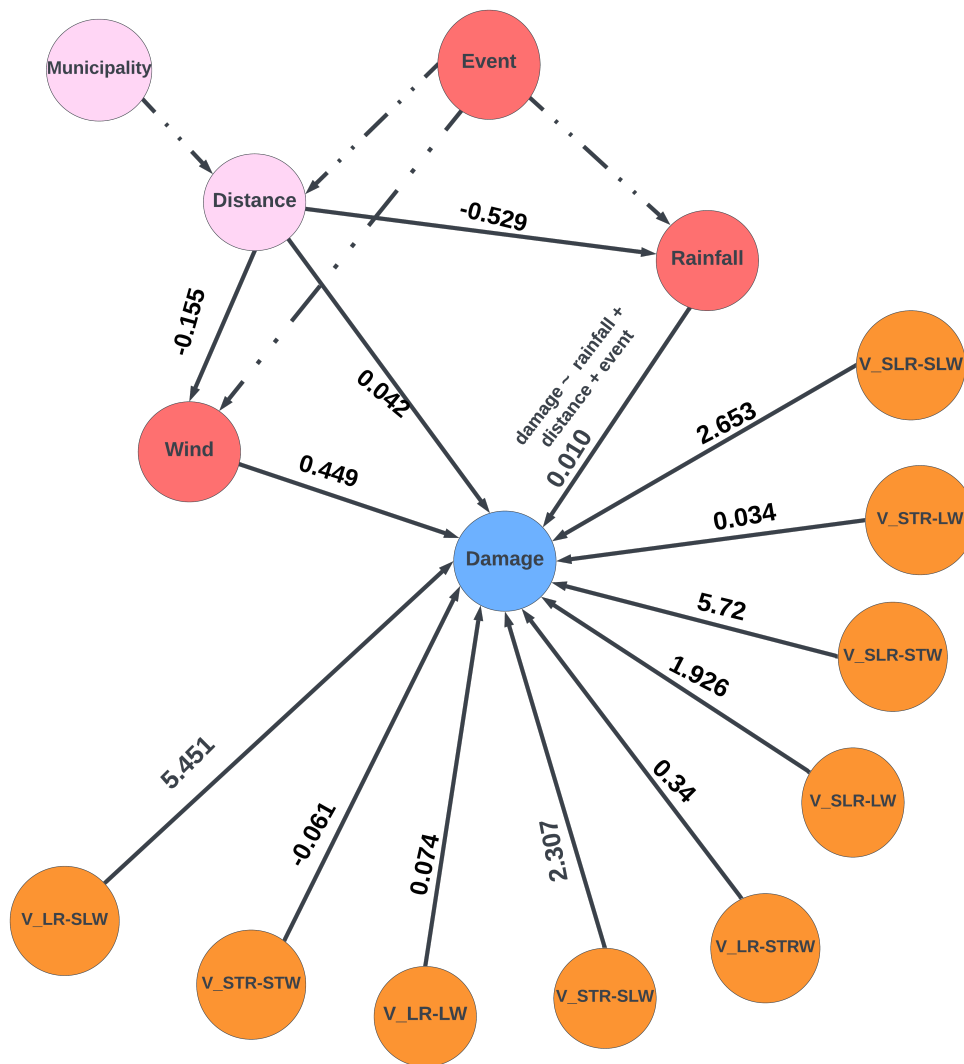
**Figure 1:** A DAG showing the causal pathways (with direct effects coefficients) of damage due to tropical cyclones. Dotted directed paths indicate event specific causal relationship. This allows to account for variation among different tropical cyclones events in the data.

socio-economic groups of people and hence can be considered as sensitive variables. Because of the identified complexities (i.e., confounder bias and Simpsons paradox) further analysis is required to quantify the effect of these biases on the accuracy of damage prediction and how they would affect fair distribution of aid resources. It is however apparent that the outcome variable and the objectives also need to be designed or specified in a fair manner that ensures fairness. In this context we note that predicting impact instead of damage would allow for a deeper engagement on fairness.

## 6. Acknowledgments

# References

[1] S. Boeke, M. van den Homberg, A. Teklesadik, J. Fabila, D. Riquet, M. Alimardani, Towards predicting rice loss due to typhoons in the philippines, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 42 (2019) 63–70.

[2] M. J. Van den Homberg, C. M. Gevaert, Y. Georgiadou, The changing face of accountability in humanitarianism: Using artificial intelligence for anticipatory action, Politics and Governance 8 (2020) 456–467.

[3] C. M. Gevaert, M. Carman, B. Rosman, Y. Georgiadou, R. Soden, Fairness and accountability of ai in disaster risk management: Opportunities and challenges, Patterns 2 (2021).

[4] J. Pearl, Causality: Models, Reasoning, and Inference, 2nd ed., Cambridge Univeristy Press, 2009.

[5] B. Friedman, H. Nissenbaum, Bias in computer systems, ACM Transactions on information systems (TOIS) 14 (1996) 330–347.

[6] S. Xu, J. Dimasaka, D. J. Wald, H. Y. Noh, Seismic multi-hazard and impact estimation via causal inference from satellite imagery, Nature Communications 13 (2022) 7793.

[7] H. Burton, Causal inference on observational data: Opportunities and challenges in earthquake engineering, Earthquake Spectra 39 (2023) 54–76.

[8] A. Teklesadik, M. van den Homberg, Forecasting impacts of tropical cyclones with machine learning: A case study in the philippines, in: EGU General Assembly Conference Abstracts, 2022, pp. EGU22–12917.

[9] X. Wang, Y. Zhang, R. Zhu, A brief review on algorithmic fairness, Management System Engineering 1 (2022) 7.

[10] J. R. Loftus, C. Russell, M. J. Kusner, R. Silva, Causal reasoning for algorithmic fairness, arXiv preprint arXiv:1805.05859 (2018).

[11] C. Su, G. Yu, J. Wang, Z. Yan, L. Cui, A review of causality-based fairness machine learning (2022).

[12] Y.-L. Chou, C. Moreira, P. Bruza, C. Ouyang, J. Jorge, Counterfactuals and causability in explainable artificial intelligence: Theory, algorithms, and applications, Information Fusion 81 (2022) 59–83.

[13] S. J. Jung, Introduction to mediation analysis and examples of its application to real-world data, Journal of Preventive Medicine and Public Health 54 (2021) 166.

[14] U. M. K. Eidsvig, K. Kristensen, B. V. Vangelsten, Assessing the risk posed by natural hazards to infrastructures, Natural Hazards and Earth System Sciences 17 (2017) 481–504.

[15] S. Chiappa, Path-specific counterfactual fairness, in: Proceedings of the AAAI conference on artificial intelligence, volume 33, 2019, pp. 7801–7808.

# A. Basic DAGs

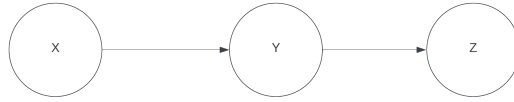Chain graph conditional independence.
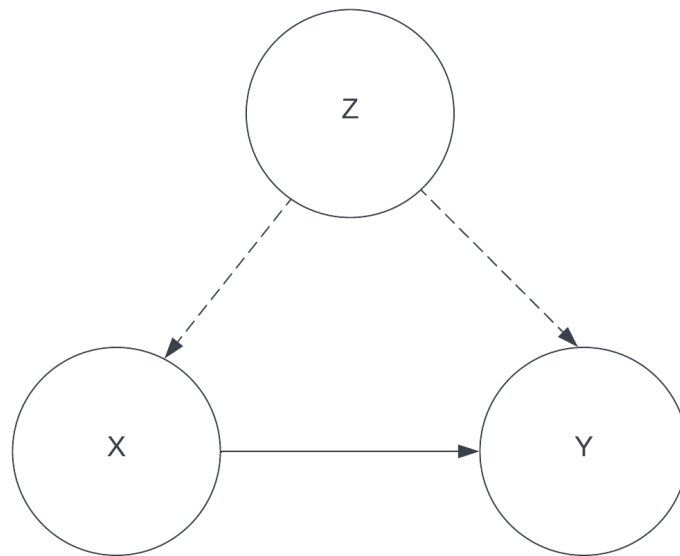


**Figure 2:** Chain DAG

$$Z \perp X | Y \qquad (5)$$



**Figure 3:** An example DAG of a confounder ($Z$) which influences both the explanatory variable $X$ and the outcome $Y$
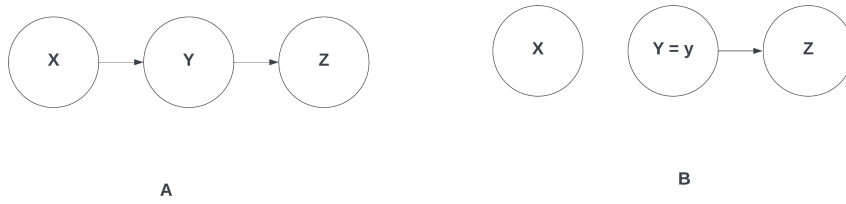
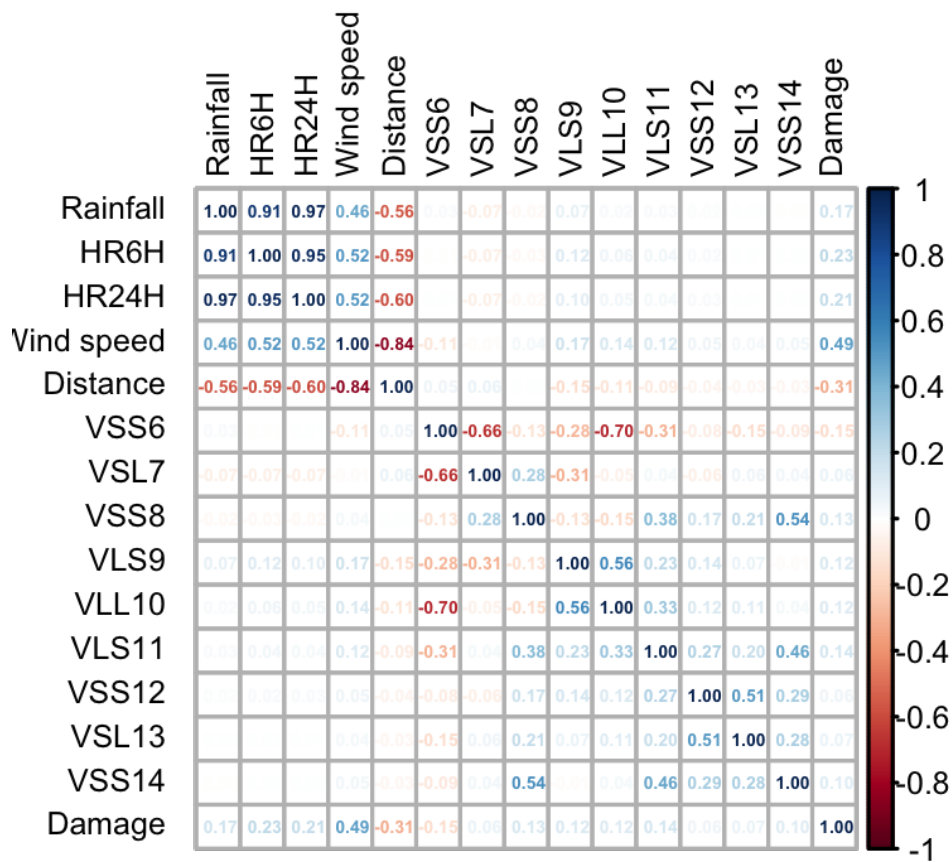**Figure 4:** Example DAGs before (A) and after intervention (B) on variable Y

## B. Figures



**Figure 5:** Correlations among variables in the Tropical Cyclone early warning system.

## C. Tables

|  | XGBoost | Random Forest | Simple Linear Regression |
|---|---|---|---|
| Mean Absolute Error (MAE) | 2.37 | 2.48 | 3.89 |
| Root Mean Square Error (RMSE) | 7.96 | 7.73 | 8.39 |

**Table 1**
The Philippines Typhoon early warning system metrics as analyzed by Teklesadik et.al.[8]

| Variable name | Description |
|---|---|
| Damage | Percentage of damage buildings observed in the previous typhoons |
| Wind | Hazard - wind speed ($m/s$) |
| Rainfall | Hazard - total rainfall ($mm$) during the tropical cyclone |
| Distance | Geography - Distance from the epicentre of the tropical cyclone |
| V_STR-STW | Percentage of strong roof and strong wall build types |
| V_STR-SLW | Percentage of strong roof and salvage wall build types |
| V_STR-LW | Percentage of strong roof and light wall build types |
| V_SLR-STW | Percentage of salvage roof and strong wall build types |
| V_SLR-SLW | Percentage of salvage roof and salvage wall build types |
| V_SLR-LW | Percentage of salvage roof and light wall build type |
| V_LR-STW | Percentage of light roof and strong wall build type |
| V_LR-SLW | Percentage of light roof and salvage wall build type |
| V_LR-LW | Percentage of light roof and light wall build types |

**Table 2**
Causal graph variable names and their descriptions.

| Coefficients | Estimate | P-value |
|---|---|---|
| distance | 0.041 | <2e-16 *** |
| wind speed | 0.439 | <2e-16 *** |
| rainfall | 0.004 | 0.000120 *** |
| str_stw | 0.069 | 0.560 |
| str_lw | 0.070 | 0.555 |
| str_slw | 1.041 | 1.05e-05 *** |
| lr_stw | 0.127 | 0.300 |
| lr_lw | 0.069 | 0.561 |
| lr_slw | 0.500 | 0.246 |
| slr_stw | -1.451 | 0.158 |
| slr_lw | 0.693 | 0.024 |
| slr_slw | 0.444 | 0.149 |

**Table 3**
Linear regression model with tropical cyclone events accounted for.

| Coefficients | Estimate | P-value |
|---|---|---|
| distance | 2.902e-02 | <2e-16 *** |
| wind speed | 3.618e-01 | <2e-16 *** |
| rainfall | 5.090e-04 | 0.516 |
| str_stw | 1.056e-01 | 0.402 |
| str_lw | 1.265e-01 | 0.316 |
| str_slw | 1.648e+00 | 3.73e-11 *** |
| lr_stw | 2.097e-01 | 0.106 |
| lr_lw | 1.335e-01 | 0.289 |
| lr_slw | 5.532e-01 | 0.227 |
| slr_stw | -6.973e-01 | 0.524 |
| slr_lw | 6.367e-01 | 0.051 |
| slr_slw | 5.978e-01 | 0.068 |

**Table 4**
Linear regression model with tropical cyclone events un-accounted for.

| | Estimate | P-value |
|---|---|---|
| Average Causal Mediation Effects(ACME) | 0.005 | <2e-16 *** |
| Average Direct Effects (ADE) | 0.235 | <2e-16 *** |
| Total Effect | 0.240 | <2e-16 *** |
| Prop. Mediated | 0.021 | <2e-16 *** |

**Table 5**
Mediation analysis: str_stw as mediator of the effect between wind speed and damage

| | Estimate | P-value |
|---|---|---|
| Average Causal Mediation Effects(ACME) | -0.058 | <2e-16 *** |
| Average Direct Effects (ADE) | 0.030 | <2e-16 *** |
| Total Effect | -0.028 | <2e-16 *** |
| Prop. Mediated | 2.065 | <2e-16 *** |

**Table 6**
Mediation analysis: wind speed as mediator of the effect between distance and damage