Response to Reviewer #2

We thank the reviewer for the positive evaluation of the paper and the feedback. We address each comment separately and in detail below. The original comments of the reviewers are given in italics, followed by our response in normal letter type.

General comments

General comments The manuscript evaluated the aggregated impact of a bundle of nature-based and other municipal adaptation measures on insured rain damages by comparing the insured data from two adjacent areas within the Rivierenbuurt neighbourhood, one with flood damage mitigation (FDM) measures and one without intervention. Using the statistical difference-in-difference method, the authors identified significant relations in some variables, highlighting the causal effect of FDM measures. The discussion and conclusions were valid and supported by data evidence.

The manuscript proposes a research initiative on a topic within the scope of Natural Hazards and Earth System Sciences (NHESS). I would recommend this manuscript for publication with the following suggestions, particularly regarding data processing, model interpretation, and widening the discussion.

Specific comments

Data heterogeneity and process

The listed adaptation measures in Table 1 may be applied to a specific area instead of the whole region. It is unclear how the local effects are translated to the larger treatment or local area. Moreover, given the spatial-temporal distribution of insured damage, is it possible to identify the relationship between adaptation measures and the observed reduced damages?

We account for this relationship by selecting the Difference-in-Difference method. This method controls time-invariant unobserved differences between neighborhoods (in this case the treatment and control neighborhood), such as historical infrastructure and socioeconomic factors, as well as time-specific shocks, like extreme weather events.

Furthermore, due to privacy restrictions we are not allowed to look at the damages at the address level. It is therefore difficult to estimate the local effect of a measure, more detailed than neighborhood level (PC4). We plan to add a footnote in section 2.2.1.

'Due to privacy restrictions on the claims data it is not possible to analyze the damages on address level.'

The study collected heterogeneous data (e.g., hydro-meteorological data and demographic characteristics) in space and time. It is unclear how authors address the heterogeneity and aggregate across the areas.

See response directly above. We discuss the characteristics of both neighborhoods in appendix 1.

In the data description section (table 2 and 3) we describe for each variable (1) the time period between which the data is recorded; and (2) the level of aggregation, e.g. average at postcode or neighbourhood level.

Data distribution

It is noted in Table 2, the standard deviation of insured rain damage was much higher than the mean. Is it driven by some extreme events that cause extensive insured rain damage? The same comment applies to "Rain data" in Table 3.

This is indeed driven by extreme events. Specifically, August 2010 was a month where extreme damages occurred. We performed a robustness test by deleting this month and rerunning the analysis again. However, this caused only minor changes to the results. We plan to add a footnote to describe this in section 4.1:

'In an additional analysis, we omitted the month August 2010, where large damages occurred in the treatment group and the control group. We observe minor changes to the results: the interaction coefficient is -704.461, compared to the -646.963 in the model with August 2010 included, and the relation is significant at the same level ($p < 0.01$).'

We also plan to add a sentence in section 4.1 linking the additional analysis, with the omission of August 2010, to the high standard deviation compared to the average damages (table 2) and rain data (table 3). We plan to add the following sentence:

'This also explains why the standard deviation is very high compared to the average of damage data (table 2) and rain data (table 3).'

Related to Figure 1, the highest peak occurred in 2010. Was it due to any specific event or insurance claim?

This was due to the damages in the month of August 2010. See response to the previous comment on how we addressed this.

Related to Tables 2 and 3, it will be nice, if possible, to visualise the data distribution through histograms.

We agree with the reviewer that more detailed information would be an addition to the paper. However, since it is damage data that is guided by extremes, the dataset contains large outliers. Distributions shown via histograms are then not too helpful, because these are far apart and low damage observations are highly clustered. We instead plan to show how the data is distributed via tables showing the damage percentiles. We plan to include the following tables in the paper in appendix 4:

Table 1: Distribution insured rain damage data full dataset (from 2007)

	1%	5%	10%	25%	50%	75%	90%	95%	99%	Largest
All observations (n=12568)	€0.00	€0.00	€0.00	€0.00	€0.00	€0.00	€169.00	€1000.00	€3761.00	€169305.00
Only damages (n=1360)	€1.00	€119.13	€202.57	€498.23	€956.75	€1727.25	€3470.00	€5325.44	€17703.44	€169305.00

Table 2: **Detailed description insured rain damage data full dataset (from 2007)**

Variable	**Mean (standard deviation if non-binary in parentheses)**		**Median**		**Range**	
From 2007	*From 2007*	*From 2016*	*From 2007*	*From 2016*	*From 2007*	*From 2016*
Insured rain damage	€202.12 (€1928.92)	€242.32 (€3029.46)	€0.00	€0 .00	€0 – €169305.00	€0 – €169305.00
Insured rain damage (damages only)	€1867.82 (€5594.01)	€2191.74 (€8883.30)	€956.73	€1000.00	€1 – €169305.00	€1 – €169305.00

Table 3: **Distribution of rain data (from 2007)**

	1%	5%	10%	25%	50%	75%	90%	95%	99%	Largest
Sum of rain per day (in 0.1 mm, n=12568)	0	0	0	0	1	26	76	115	202	672
Maximum rain in an hour (in 0.1 mm, n=12568)	0	0	0	0	1	11	26	39	89	281

Model development

Based on the results in Table 4, some variables which were conceived to be significant turned out to be not so significant (e.g., p-value >0.1) statistically in both models. Why the authors choose to keep these variables?

We believe it is important to hold certain control variables constant when analyzing the intervention effect. We describe in section 2.2.2 the addition of the lag variables to account for the fact that the claims data consists of observations where the claim was filed by the insurer. This is often the same day, but can also be one or two days later. We account for this by adding the lags of one and two days in the analysis of the variables ‘maximum rain in an hour’ and ‘um of rain per day of one and two days in the analysis’.

If we only include the significant variables in model 1, we obtain the results below. We plan to add this result as a robustness test to appendix 5.

Table 4: Two-way fixed effects DiD regression on insured damage per day in case of maximum rain per hour exceeds 2mm per hour from 2016 with only significant variables

Variables	(1) Model 1 with intervention period (2007-2024)	(2) Model 2 with intervention period (2016 – 2024)
Post × treatment (DiD)	-646.6*	-4,188**
	(392.3)	(1,626)
Sum of rain per day (in 0.1 mm)	5.562***	
	(1.424)	
Population density		-6.961**
		(2.798)
Constant	-35.75	98,843**
	(196.0)	(39,437)
Observations	1,766	886
R-squared	0.253	0.266

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Some variables were correlated to some extent, e.g., three variables related to the sum of rain per day. Can this statistical modelling handle correlated variables? It will be nice to clarify the statistical assumptions.

We tested for multicollinearity through the use of the Variance Inflation Factor (VIF) method. VIF is commonly used in econometric analyses (Woolridge, 2016). If a VIF-value is between 1 and 5, it is deemed as acceptable collinearity. If a VIF-value is above 10, we can conclude that multicollinearity is problematic for that variable (Woolridge, 2016). We performed a VIF-test for the models in the manuscript: model 1 and model 2. For model 1 we observe no issues (VIF-scores below 5). However, for model 2 we observe high VIF-values (>10) for the variables ‘percentage of real estate built before 1945’, ‘Address density’ and ‘Average number of people per household per address’. We therefore plan to delete these variables from the previous model 2. The first reviewer commented that it would be cleaner to do the analysis with omitting the intervention period. We plan to follow this comment. The results of model 2, which we include in two models, with and without the intervention period, are now as follows:

Table 5: Two-way fixed effects DiD regression on insured damage per day in case of maximum rain per hour exceeds 2mm per hour from 2016 with and without observations in the intervention period

Variables	(3) Model 3 (results from 2016 with intervention period)	(4) Model 4 (results from 2016 without intervention period)
Post × treatment (DiD)	-5,017***	-5,648**
	(1,863)	(2,512)
Sum of rain per day (in 0.1 mm)	5.405	7.100
	(3.650)	(5.375)
Sum of rain per day lag 1 (in 0.1 mm)	-1.136	-0.986
	(5.774)	(9.271)
Sum of rain per day lag 2 (in 0.1 mm)	2.537	0.624
	(6.711)	(11.80)
Maximum rain in an hour (in 0.1 mm)	-8.205	-13.90
	(9.669)	(16.89)
Maximum rain in an hour lag 1 (in 0.1 mm)	11.49	14.21
	(15.41)	(27.48)
Maximum rain in an hour lag 2 (in 0.1 mm)	-8.074	2.483
	(18.22)	(37.70)
Population density (per km²)	-7.325***	-6.391
	(2.827)	(5.845)
Value of property (in euros)	24.67	48.00
	(27.79)	(56.33)
Constant	91,656**	66,643
	(40,384)	(98,878)
Observations	886	536
R-squared	0.269	0.271
Adjusted R-squared	0.174	0.173

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

With changing the variables, we also changed the variable description table of the control variables. We plan to add the description below.

Table 6: **Added control variable**

Variable	**Variable description**	**Data source**	**Mean and standard deviation if non-binary**
Area characteristics (per day from 2016)			
Population density	The amount of people per km²	CBS	13897.77 (744.84)

There are other factors related to disaster management and capacity building (e.g., social vulnerability and infrastructure interruption) which is not considered in this study.

If these factors were included, what would be the potential impact on the developed model and its implications?

We understand the comment of the reviewer. To examine social vulnerability, one would probably need to collect and couple survey data on individual socio-economic characteristics and combine it with damage and rainfall data. This is practically difficult, given the level of aggregation of the damage (pc4) and rainfall data. For a future study, it could be interesting to understand social vulnerability as well. Given that social vulnerability might influence insurance uptake. In the discussion, under section 4.3 'Limitations and research implications' we plan to add the following sentence:

'Additionally, it would be insightful to look at social vulnerability, since that could influence insurance uptake.'

However, this should not impact the results, since external developments are the same in both neighborhoods. Socio-economic characteristics are likely similar across both neighborhoods. The DiD two-way fixed effects approach controls for socioeconomic differences. We test for the common trend assumption using a placebo test. We show the results of these tests in appendix 2. We explain the placebo tests in section 2.4. We see no influence of the factors mentioned by the reviewer, because the treatment and control groups follow a similar trend before the interventions (see placebo tests in appendix 2).

Results

In the DiD model, the time and unit fixed effect is not well discussed. It would be nice to understand its implications.

We agree with the reviewer that we could expand on the time and unit fixed effects explanation. The main implication is that fixed effects in a DiD are used to give a more robust causal estimate, controlling for time-invariant unobserved differences like socioeconomic factors (on unit) or extreme weather events (on time). We plan to alter the text in the manuscript as follows:

'In this study, we use a DiD two-way fixed effects model to estimate the impact of municipal adaptation measures on rainfall damage in Amsterdam. One can compare a situation before and after an intervention. We compare two adjacent areas within the Rivierenbuurt neighborhood: one where flood damage mitigation (FDM) measures have been implemented (Scheldebuurt) and another where no interventions have been implemented (Rijnbuurt). The DiD approach allows us to compare changes in outcomes over time between these areas, while controlling for unobserved factors and broader trends (Card & Krueger, 1993; Wooldridge, 2014). By leveraging insurance claims data, we can isolate the causal impact of these measures under the assumption that both areas would have followed similar trends in the absence of interventions. We test this assumption in the next section.

We expand upon a traditional DiD by employing a two-way fixed effects (TWFE) model (Callaway & Sant'Anna, 2021). Using fixed effects in a DiD gives a more robust causal estimate. This approach controls for time-invariant unobserved differences between neighborhoods, such as historical infrastructure and socioeconomic factors, as well as time-specific shocks, like extreme weather events. By accounting for both unit (neighborhood) and time (month) fixed effects, the TWFE model ensures that our estimated treatment effect reflects the impact of adaptation measures rather than underlying trends or external influences. This strengthens the causal interpretation of the DiD analysis.'

DiD with time and unit fixed effects is an econometric method that is used to estimate causal effects. You compare a situation before and after a policy intervention. Fixed effects help to control for unobserved time-invariant heterogeneity. Unit fixed effects control for time-invariant characteristics of a unit, for instance geography of a location or individual differences.

Time fixed effects control for shocks that equally affect the units. For instance a recession, policy changes, or, in this case, extreme weather events for the whole group. Using fixed effects in a DiD gives a more robust causal estimate.

Though the variables were discussed regarding their significance levels, the model performance (e.g., how well it fit the dependent variables) was not shown to visualise the goodness of fit and uncertainty.

We show the Adjusted R^2 in table 4. We highlight this in the text by the following sentence in chapter 3, which we plan to change with the updated results:

'According to the adjusted R^2, Model 1 explains 16.7% of the variation in insured damage and model 2 explains 17.3% of the variation.'

It is important to extend the discussion on the discussion, such as what other factors should be considered.

We agree with the reviewer that this could be done more extensively. In the discussion, under section 4.3 'Limitations and research implications' we plan to add the following sentences:

'It would be of value to analyze uninsured damages and claims of businesses as well, to present a more complete picture of the effectiveness of FDM.'

'Furthermore, it would be valuable to understand how much separate measures contribute to damage reduction.'

Also, related to social vulnerability, we plan to add the following:

'Additionally, it would be insightful to look at social vulnerability, since that could influence insurance uptake.'

The placebo results in Appendix B it is hard to understand for a reader without a statistical background. A short description to guide readers through the results is necessary.

We agree it is useful to clarify these test results. We plan to add the following sentences to appendix 2 to guide the reader through the results shown in the tables of appendix 2:

The goal of this placebo test is to identify whether the treatment and control groups were experiencing similar trends before the treatment. This can be done by creating 'fake' treatments that indicate treatment before it actually occurred (Angrist & Pischke, 2009). These placebo treatments should have no effect if the common trend assumption holds. If they do show significant effects, this suggests a violation of the assumption, as it indicates that treated and control groups were already on diverging paths prior to the intervention.

We apply placebo tests by using one- and two-month leads and lags for the treatment variable. The lead and lagged placebo treatments do not show any significant outcomes, which provides evidence in favour of the common trend assumption.

It is interesting to see the different trend patterns during the implementation of the adaptation measures. However, it is unclear whether it is due to the implementation temporarily reducing the overall adaptation effect or other reasons.

We account for this in our method. The difference-in-difference (DiD) approach controls for time-invariant unobserved differences between neighborhoods, such as historical infrastructure and socioeconomic factors, as well as time-specific shocks, like extreme weather events. Homogenous shocks are experienced in both neighborhoods. The use of the DID approach makes it possible to only estimate the effect of the intervention.

Discussion

This study evaluated the effectiveness of these measures combined in a municipality, only measured in terms of insured rain damage. However, it is worth providing insights regarding the long-term climate-adaptive benefit as well as the non-monetary impact.

We agree with the reviewer that there are other benefits of these FDM measures next to rain damage reduction. Measures like water storage can be used against drought, green roofs and greener areas can mitigate heat and these areas can also be used for recreational purposes. These measures can limit long-term impacts of climate change. We touch upon this in section 4.2 by the following sentences:

‘Local governments can use nature based and other adaptation measures (e.g. through green lanes, water storage facilities, green roofs, and greener gardens) as means to decrease rain damage in urban areas and increase livability and biodiversity in these areas (Skrydstrup et al., 2022). These nature based measures often come with co-benefits like mental and physical benefits (Tzoulas et al., 2007).’

However, an example of a long term benefit is less taken into consideration. Therefore, we plan to add the following sentence:

‘(...), which can have a long term impact on health as well by incentivizing people to exercise for instance.’

In the current text we do not yet point out that the municipal climate adaptation measures not only limit rain damage, but also limit impacts of other natural hazards like drought and heat. We plan to add the following sentence to section 4.2:

‘Rain damage is the focus of this study. The measures the municipality applied can also limit impacts of other natural hazards, like drought and heat. These measures can limit long-term impacts of climate change in the area.’

A broad range of adaptation measures was studied as a whole. Can the data show the single contribution of respective measures to the overall climate-adaptation effect? Can the authors identify the most effective adaptation measures among all the considered measures?

The reviewer is correct that it would be interesting to look at individual measures. We now point this out in the limitations section, where we plan to add the following text:

‘Lastly, this study shows the impact of all adaptation measures combined. Because of privacy regulations, it was not possible to localize claims on a more detailed level than PC4-level. This makes it difficult to attach effects of local measures to single damage claims. In a future study, it might be of value to understand the impact of these measures separately.’

There are papers that assesses the risk reduction of a single measure, for instance an old stormwater system (Sörensen & Emilsson, 2019), blue-green roofs (Busker et al., 2021) or awareness campaigns (Osberghaus & Hinrichs, 2020). A unique characteristic of this paper is that we look at all the measures combined. Given how the measures were implemented, we cannot research the impact of a measure individually. Therefore, we plan to add the following sentence to the limitations section:

'Further, it would be valuable to understand how much separate measures contribute to damage reduction.'

References

Busker, T., de Moel, H., Haer, T., Schmeits, M., van den Hurk, B., Myers, K., Cirkel, D. G., and Aerts, J.: Blue-green roofs with forecast-based operation to reduce the impact of weather extremes, Journal of Environmental Management, 301, 1–12, https://doi.org/10.1016/j.jenvman.2021.113750, 2022.

Angrist, J. and Pischke, J.-S.: *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, Princeton, NJ, 392 pp., 2009.

Callaway, B. and Sant'Anna, P. H. C.: Difference-in-Differences with multiple time periods, Journal of Econometrics, 225, 200–230, https://doi.org/10.1016/j.jeconom.2020.12.001, 2021.

Card, D. and Krueger, A. B.: Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Reply, The American Economic Review, 90, 1397–1420, 2000.

Osberghaus, D. and Hinrichs, H.: The Effectiveness of a Large-Scale Flood Risk Awareness Campaign: Evidence from Two Panel Data Sets, Risk Analysis, 41, 944–957, https://doi.org/10.1111/risa.13601, 2021.

Sörensen, J. and Emilsson, T.: Evaluating Flood Risk Reduction by Urban Blue-Green Infrastructure Using Insurance Data, Journal of Water Resources Planning and Management, 145, https://doi.org/10.1061/(ASCE)WR.1943-5452.0001037, 2019.

Skrydstrup, J., Löwe, R., Gregersen, I. B., Koetse, M., Aerts, J. C. J. H., de Ruiter, M., and Arnbjerg-Nielsen, K.: Assessing the recreational value of small-scale nature-based solutions when planning urban flood adaptation, Journal of Environmental Management, 320, 115724, https://doi.org/10.1016/j.jenvman.2022.115724, 2022.

Tzoulas, K., Korpela, K., Venn, S., Yli-Pelkonen, V., Kaźmierczak, A., Niemela, J., and James, P.: Promoting ecosystem and human health in urban areas using Green Infrastructure: A literature review, Landscape and Urban Planning, 81, 167–178, https://doi.org/10.1016/j.landurbplan.2007.02.001, 2007.

Wooldridge, J. M.: Introduction to Econometrics (Europe, Middle East & Africa Edition), Cengage Learning EMEA, Andover, United Kingdom, 912 pp., 2014.