Response to Reviewer #1

We thank the reviewer for the positive evaluation of the paper and the thorough review. Each comment is addressed separately and in detail below. The original comments of the reviewers are given in italics, followed by our response in normal letter type.

Referee's Report: Assessing effects of nature-based and other municipal adaptation measures on insured heavy rain damages

The study examines the effect of pluvial flood adaptation measures by analyzing insurance claims in two adjacent neighborhoods in the Netherlands before and after the interventions. The authors find that the adaptation measures in their totality cause a decline of 3700 euros per rainy day.

I agree with the authors that we still know too little about the effectiveness of mitigation measures, especially from an economic perspective. Studying the governmental mitigation projects adds to both the academic literature and has high social relevance. The use of insurance data provides a unique and objective measure to study the effectiveness of such measures.

While I think the set-up is interesting, multiple questions regarding the method and economic significance of the findings remain. Addressing these points should largely contribute to the paper, in my opinion. I recommend a revise and resubmit with major revisions.

Methods

I don't think that the visual interpretation around Figure 1 provides convincing evidence that the vital common pre-trend assumption is met. The study would largely benefit from a more formal test of the common pre-trend assumption. For example, through an event study.

We thank the reviewer for drawing our attention to this important point. Based on their comments, we plan to delete figure 1 from the main text (now in appendix 4). We would like to draw the reviewer's attention to the placebo test results, which is a formal test of whether the common trend assumption is met. The detailed results are shown in Appendix 2. The following sentences are part of the revised section 2.4:

'(...) a placebo test can be performed to check for the common trend assumption (Eggers et al., 2021). The placebo test checks the common trend assumption by creating "fake" treatment groups before and after the intervention. We select a different treatment timeframe and observe whether the effects are significant as well. If no effect is found in any of the placebo groups, it supports that the found treatment effect can be attributed to the treatment rather than pre-existing trends. Angrist and Pischke (2008) used lag and lead values of treatment status to show that no significant effects occurred in the placebo periods. In Appendix 2, we apply placebo tests by using one- and two-month leads and lags for the treatment variable. These placebo treatment variables resulted in non-significant outcomes, reinforcing the validity of the common trend assumption for causal inference.'

Related to this, I wonder if it is likely that the amount of rainfall varies between your treatment and control group? If purely coincidental, the treatment group was affected by a large shower while the control group was not this biases results.

The rain data we used from the Royal Dutch Meteorological Office is the same for both the treatment and control group. In both cases the nearest weather station is used. We expect only very minor differences in actual rainfall between the treatment and control group, since they are neighborhoods adjacent to one another. Therefore, they likely receive similar rainfall volumes and extremes. Pre-existing infrastructure is also similar, where only in the treatment area measures were taken to reduce rain damage (Amsterdam Weerproof, 2024).

Both your treatment and control groups are affected by the general information campaign. If the effect of this campaign is homogenous for both areas, this should not affect the estimates that you get for the specific projects. However, if this is not the case, your results might be biased. In this regard, I think it is interesting to think about why the government chose to implement mitigation measures in the Scheldebuurt and not (yet) in the Rijnbuurt? This is ideally random. However, I can imagine that in this case, the firmer was chosen because the initial risk was higher, and thus the information campaign might also be more effective here.

We do not have detailed information about why the Scheldebuurt (treatment) was chosen first. According to the municipality, the sewage- and rainwater system needed to be renewed. In both neighborhoods, large areas were once covered with concrete and tiles, thus reducing water drainage (Amsterdam Weerproof, 2024). The municipality started implementing FDM measures in Rijnbuurt (control) from 2025 (Amsterdam Weerproof, 2024). The information campaigns are the same across both neighborhoods, (Amsterdam Weerproof, 2025). We thus expect the effect of information provision to have a similar effect in both neighborhoods. We plan to incorporate a sentence in the paper that the treatment effects are attributed to the other FDM measures only. The impact of advice through information campaigns was expected to be low based on previous research (Osberghaus & Hinrichs, 2021), compared to tangible community focused measures like constructing green areas and water buffering zones.

Account for the insurance provider composition. The mitigation measures might be ex-ante observed by insurers who update their fees. Hence, the insurer composition before and after treatment might be altered. If certain insurers receive fewer claims in general (e.g., due to less coverage and user-friendliness of the claim portal), this could drive results.

> In practice, Dutch insurers do not update their fees based on mitigation measures (Kroes & Klok, 2024). Even though an insurance market changes and develops over time, rain damages were already widely insurable since the adoption of the so called Neerslagclausule (precipitation clause) in 2001 (Dutch Association of Insurers, 2018). For households and businesses, rain damage is insured by default (Dutch Association of Insurers, 2025). The data of households of all members of the Dutch Association of Insurers (over 95% of the Dutch market) is used (Dutch Association of Insurers, 2024). Also, we do not expect the insurer composition to change much over time. In the Netherlands, households tend to not switch insurance policies often. There is no data available for households changing insurer for non-life insurers, but for health insurers this is about 7.4% per year (Zorgverzekeraars Nederland, 2025). It is expected that this is much lower for non-life insurances, since health insurers actively campaign every year in december to change insurer. Also, for car insurance, households tend to stay for more than 12,5 years on average at the same insurer (Allianz Direct, 2025). Therefore, we expect that insurer composition does not change much across time.

Can spill-over effects bias the results? Many of the nature-based adaptation measures could also reduce risk in the control areas. For example, if the sewers are linked, increasing water storage capacity in the treatment area also reduces the chance that the water-bearing capacity in the control area is reached.

> We would argue that a possible spill-over effect, would be very minor. The measures are primarily focused on infiltration (e.g. green areas, water storage, green roofs) on a very local level, specifically focused on the 'Scheldebuurt' (treatment area). Large scale measures such as like dikes or large water barriers could influence groundwater flows across larger areas. However, these measures are not taken in the treatment area. Furthermore, most damages (about 60%) are roof leakages from rain, not coming from ground water (Amsterdam Weerproof, 2014). A green roof or similar measures could help prevent these damages. For these damages, no spill-over effect to another area occurs.

If I understand correctly, "post" is based on the date when the mitigation program started. However, realizing these projects does not happen instantly, and probably not all projects were implemented at the same time. Leaving out the intervention period *should provide a more reliable estimate of the actual benefit, as the current measure likely underestimates the effect.*

We agree with the reviewer that not including observations during the intervention period, would result in a cleaner analysis. We avoid potential bias from including the rollout period, when the policy's effect was only partial. Therefore, we ran the analysis without observations in the intervention period.

Table 1: Two-way fixed effects DiD regression on insured damage per day in case of maximum rain per hour exceeds 2mm per hour from 2007 with and without observations in the intervention period

Variables	(1) Model 1 (results with intervention period)	(2) Model 2 (results without intervention period)
Post × treatment (DiD)	-647.0*	-1,375**
	(392.5)	(558.2)
Sum of rain per day (in 0.1 mm)	6.267***	6.856***
	(1.949)	(2.308)
Sum of rain per day lag 1 (in 0.1 mm)	-2.158	-2.010
	(3.004)	(3.635)
Sum of rain per day lag 2 (in 0.1 mm)	0.434	-0.0834
	(3.423)	(4.274)
Maximum rain in an hour (in 0.1 mm)	-2.679	-3.053
	(5.075)	(6.315)
Maximum rain in an hour lag 1 (in 0.1 mm)	10.62	10.88
	(7.733)	(9.674)
Maximum rain in an hour lag 2 (in 0.1 mm)	-2.997	-0.665
	(9.572)	(12.79)
Constant	-61.99	-61.37
	(240.2)	(286.0)
Observations	1,766	1,416
R-squared	0.254	0.259
Adjusted R-squared	0.162	0.167

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

We performed multi-collinearity tests, which showed that the previous composition of variables resulted in multi-collinearity problems for the variables 'average number of people per household per address', 'percentage of real estate built before 1945' and 'address density' (all higher than 0.8). Therefore, we plan to remove the first two of these variables from the analysis and we plan to change 'address density' to a different variable called 'population density': the amount of people per area divided by the size of the area. The results are shown below. Model 1 shows the results with intervention period and model 2 without intervention period with the observations from 2007-2024. Model 3 (with intervention period) and model 4 (without intervention period) show the data from 2016-2024 with the extra variables 'population density' and 'value of property'.

Table 2: 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

The reviewer argues that the model without observations in the interaction period would result in a cleaner analysis. Also, the fit of one model improves compared to the analysis with observations in the intervention period (model 1). Therefore, we plan to use the models without observations in the intervention period for the analysis. The models with observations in the intervention period we plan to use as robustness tests. Therefore, we plan to show the following models without intervention period in the results chapter:

Table 3: Two-way fixed effects DiD regression on insured damage per day in case of maximum rain per hour exceeds 2mm per hour from 2007 and 2016 without observations in the intervention period

Variables	(1) Model 1	(2) Model 2
Post × treatment (DiD)	-1,375**	-5,648**
	(558.2)	(2,512)
Sum of rain per day (in 0.1 mm)	6.856***	7.100
	(2.308)	(5.375)

Sum of rain per day lag 1 (in 0.1 mm)	-2.010	-0.986
	(3.635)	(9.271)
Sum of rain per day lag 2 (in 0.1 mm)	-0.0834	0.624
	(4.274)	(11.80)
Maximum rain in an hour (in 0.1 mm)	-3.053	-13.90
	(6.315)	(16.89)
Maximum rain in an hour lag 1 (in 0.1 mm)	10.88	14.21
	(9.674)	(27.48)
Maximum rain in an hour lag 2 (in 0.1 mm)	-0.665	2.483
	(12.79)	(37.70)
Population density (per km²)		-6.391
		(5.845)
Value of property (in euros)		48.00
		(56.33)
Constant	-61.37	66,643
	(286.0)	(98,878)
Observations	1,416	536
R-squared	0.259	0.271
Adjusted R-squared	0.167	0.173

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

Time-specific neighborhood-level shocks are controlled for through fixed effects for each month. This addresses general temporal variation but not temporal variation at the neighborhood level. This would require 'month by neighborhood' FE.

We believe that time fixed effects per neighborhood may absorb or obscure some of the intervention effects. General time fixed effects control for common shocks or trends over time (this is why we opt to include these effects in our analysis). Neighborhood-by-time fixed effects are likely to control for time-specific changes within each neighborhood. Therefore, we plan to retain the models presented under the above comment.

Does the control percentage of real estate built before 1945 add new info? Unless a lot of new buildings were added during the study period, all this information should already be absorbed by the PC4 fixed effects.

See the previous comment about the models. We decided to remove the variable 'real estate built before 1945' due to multi-correlation problems with this variable.

Also, it is unclear how the "address density" and the "value of property" controls are measured. Are those PC4 averages? If not, how can you use singular-level controls in combination with an aggregated dependent variable?

See our answer related to multi-collinearity above. Specifically, the previous composition of variables resulted in multi-collinearity problems for the variables 'average number of people per household per address', 'percentage of real estate built before 1945' and 'address density'. Therefore, we removed the first two variables from the analysis and changed 'address density' to a different variable called 'population density': the amount of people per area divided by the size of the area. The inclusion of these two control variables does not result in multi-collinearity.

We agree with the reviewer that the description of 'address density' and 'value of property' could be improved.

On the variable we plan to add 'Population density:
'The amount of people within each neighborhood per km^2.'

On 'Value of property':
'Average price per real estate asset per neighborhood based on the Valuation of Immovable Property Act (WOZ) (€x1000)."

Economic significance
The authors state that their outcomes can be used to make a cost-benefit analysis. I would encourage them to do this already. If costs are difficult to obtain, at least provide some intuition on the benefits. What is the economic significance of saving 3696 euros per rainy day? Is that a lot or a little? How many rainy days are there?

We agree with the reviewer that interpreting a result like 'saving an x-amount of euros per rainy day' can be difficult. Therefore, we plan to add more information that helps understanding the results more clearly. We plan to look at days that resulted in extreme damages and estimate the average total number of extreme rain days that may result in damage on an annual basis. In this way, the result can be expressed in damage prevented annually in the treatment area.

Rather than controlling for rainfall, can you not incorporate this in your measure? E.g., a triple diff with post x treatment x sum of rain? Likely, the benefits from mitigation are not equal for different levels of rain. E.g., the measures only work to a certain limit, or they become even more effective beyond a certain amount of rain. It would be interesting to disentangle this.

We tested for this by including a three-way interaction: post x treatment x sum of rain. The results are different from the previous results. This might be the case because ''sum of rain'' only takes the day into account on which damage was registered. However, in the models we deliberately use lags of one day and two days, to account for late registrations of damages that actually occurred one or two days earlier.

The authors claim that studying multiple interventions in a single study is a strength. I would say that knowing about individual intervention contributions adds more value, as it allows for more optimal adaptation given budget constraints. Within the study set-up, it *might be difficult to disentangle the benefits of individual adaptation measures. However, I would like to see some discussion on this. Do all measures equally contribute, are some more important than others, or is it exactly the joint effect that makes these measures so effective? Disentangling these added values would help with more efficient adaptation.*

We agree with the reviewer that it would be interesting to look at individual measures. We point this out in the limitations section, where we state the following:

'Lastly, this study shows the impact of all adaptation measures combined. In a future study, it might be of value to understand the impact of these measures separately.'

There are papers that look into the effects of single measures, 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. Municipal adaptation measures tend to focus on a package of measures in reality, not on one measure only. Given the information we have about the measures from the municipality, 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. This would give information on which measures policymakers could prioritize.'

It is unclear whether the study utilizes all insurance claims or only those accepted. The latter is probably a cleaner measure. However, it can mask real damages and thus underestimate the net effect or bias results entirely if certain insurers reject more claims than others, and the composition of insurers changes during the study period. Related to this, I would like some reflection on how well insurance claims as a whole measure the effectiveness of mitigation measures. How large is the share of people not making claims despite damages, and what does this imply for your results?

We look at the claims that are registered by the insurer. We plan to add the following line to section 2.1.1:

' The Dutch Association of Insurers registers claims of households filed by insurance companies that are member of the association. Since rain damage is covered by default (Dutch Association of Insurers, 2025), we expect that the vast majority of the claims are accepted.'

We do not know the percentage of the people who are insured but do not claim their damages, since there are no public numbers available on this topic. We argue that most people claim their damage when they are insured, and that when they are insured they will get compensated. Rain coverage is by default part of property and contents insurance products in the Netherlands (Dutch Association of Insurers, 2025). Very minor damages (e.g. of a few euros) may not be claimed, but we expect these damages to not alter the main findings of our study. On the other hand, bad maintenance or negligence can be a ground to not accept a claim. Nevertheless, according to anecdotal evidence from the insurance industry, this does not occur often.

In the discussion, under section 4.3 'Limitations and research implications' we plan to add the following sentences:

'It would be of value to look into uninsured damages (e.g. public infrastructure) and claims of businesses as well. Insured damage of households is only a part of total damage of extreme rain, but can still give valuable insights into the effectiveness of FDM measures.'

References

Allianz Direct. Overstappen van autoverzekering: https://www.allianzdirect.nl/autoverzekering/overstappen-autoverzekering/, last access: 15 August 2025.

Amsterdam Weerproof. Rivierenbuurt buurtpagina: https://weerproof.nl/rivierenbuurt/, last access: 15 August 2025.

Amsterdam Weerproof. Bewonersbijeenkomst Rijnbuurt Oost: Bewonersbijeenkomst Rijnbuurt Oost - Weerproof, last access 15 August 2025.

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

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.

Dutch Association of Insurers: Ledenlijst / Lid worden: https://www.verzekeraars.nl/over-het-verbond/lid-worden, last access: 15 August 2025.

Dutch Association of Insurers: Dutch Insurance Industry in Figures 2016: verzekerd-van-cijfers-2016-eng.pdf, last access 15 August 2025.

Dutch Association of Insurers, Overstroming en droogte: schade en verzekeringen: https://www.verzekeraars.nl/verzekeringsthemas/klimaatbestendig-verzekeren/overstroming-en-droogte, last access: 15 August 2025.

Eggers, A. C., Tuñón, G., and Dafoe, A.: Placebo Tests for Causal Inference, American Journal of Political Science, 68, 1106–1121, https://doi.org/10.1111/ajps.12818, 2024.

Kroes, S., Klok, L., and van de Kerkhof, A.: Enablers and Barriers of Nature-based Solutions, n.d.

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.

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.

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.

Zorgverzekeraars Nederland. Definitief overstapcijfer bekend: Definitief overstapcijfer bekend: 7,4% wisselt van zorgverzekeraar - Zorgverzekeraars Nederland, last access 15 August 2025.