

# Estimating the Impact of a Flood Event on Property Value and Its Diminished Effect over Time

Nazia Ferdause Sodial ✉

City, University of London, UK

Oleksandr Galkin ✉

City, University of London, UK

Aidan Slingsby ✉

City, University of London, UK

---

## Abstract

With the increase in natural disasters, flood events have become more frequent and severe calling for mortgage industries to take immediate steps to mitigate the financial risk posed by floods. This study looked more closely at the underlying effects of flood disasters on historical house prices as part of a climatic stress test. The discount applied on house prices due to a flood event was achieved by leveraging a causal inference approach supported by machine learning algorithms on repeat sales property and historic flood data. While the Average Treatment Effect (ATE) was employed to estimate the effect of a flood event on house prices in an area, the Conditional Average Treatment Effect (CATE) aided in overcoming the heterogeneous nature of the data by calculating the flood effect on property prices of each postcode. LightGBM as a base estimator of the causal model worked as an advantage to capture the nonlinear relationship between the features and the outcome variable and further allowed us to interpret the contribution of each feature towards the decay of these discounts using SHAP values.

**2012 ACM Subject Classification** Computing methodologies → Machine learning approaches

**Keywords and phrases** Flood, Causal Inference, Machine Learning, Property Analytics

**Digital Object Identifier** 10.4230/LIPIcs.GIScience.2023.68

**Category** Short Paper

**Acknowledgements** This research was conducted under the initiative of MIAC Analytics LTD. The flood data was acquired by MIAC Analytics LTD from its data partners, WhenFresh.

## 1 Introduction

Buying houses has been a constant activity despite the surge in property value. Mortgage industries have profited from these purchases with the rise in house prices. However, with the drastic climate change, there has been a concern about these changes reflected in household insurance policies and house prices. There are two major climate change risks: physical and transitional. Physical risks are the direct risks posed by the physical impact of climate change like global warming, ocean circulation change, high flood levels, etc. These risks represent losses brought on by the more frequent and severe hazards or events related to the environment. This introduces transitional risks, or those brought on by market, technical, legal, and policy changes resulting from the transition to a low-carbon economy. With 1.9 million people in the UK exposed to the river, coastal, or surface water flooding on a regular basis, this danger is already of a high scale and is expected to grow further in the absence of higher degrees of flood risk mitigation [4]. A lot of mortgage organizations developed



© Nazia Ferdause Sodial, Oleksandr Galkin, and Aidan Slingsby;  
licensed under Creative Commons License CC-BY 4.0

12th International Conference on Geographic Information Science (GIScience 2023).

Editors: Roger Beecham, Jed A. Long, Dianna Smith, Qunshan Zhao, and Sarah Wise; Article No. 68; pp. 68:1–68:6

Leibniz International Proceedings in Informatics



LIPIC Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

## 68:2 Estimating the Impact of a Flood Event

an interest to participate in greening the financial system as a consequence of this drastic climate change as it could challenge the solvency of the companies. This has given rise to the demand for understanding the effect of flood events on house prices.

In order to better understand if severe climate change might have a significant impact on property values, it was attempted in this study to determine if historical flood events had an impact on house prices. The focus of this research was to identify a method that could handle the heterogeneity of the data and capture the non-linear relationship of the variables, which has been a challenge in this field of research. META Learner framework introduced in the `causalml` Python package was used to estimate the effect of the 2013 winter flood on Twickenham house prices and assess whether the effect eventually fades away with time.

### 2 Related work

Lamond *et al* [5] used coefficients of the regression to estimate the depression in the growth of house prices of part of the UK within a flood zone.

Beltrán *et al* [1] used a similar approach, modifying the repeat sales equation and using the coefficients to estimate the effect of flood events on different features of house prices. The findings indicate that for the majority of property types, the average post-flood price markdown of flood-affected properties is significant but generally transient.

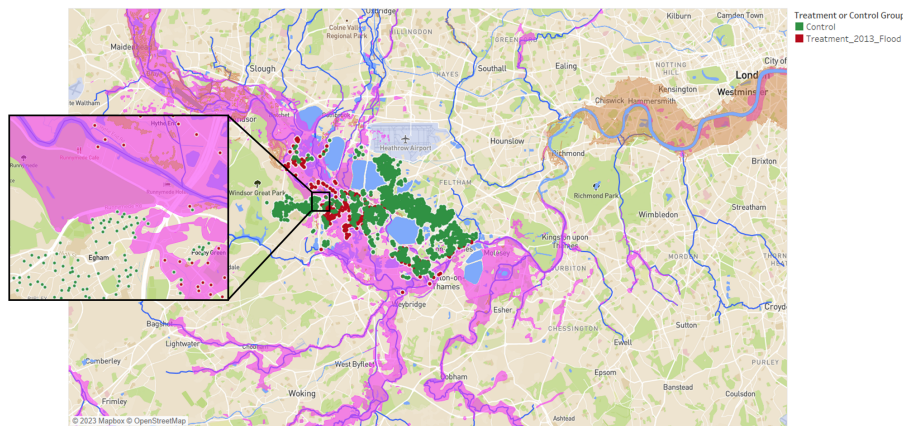
Again, N. Bui *et al* [2] used a similar approach by integrating “the hedonic property model in a difference-in-differences framework” and identified a discount of 9% was applied to house prices in some parts Ho Chi Minh City, Vietnam as a result of a disastrous flood event on 30 September 2017.

The aforementioned papers offer insightful and useful techniques into how flood occurrences affect home prices and the price decay that follows. In this study, causal models supported by machine learning algorithms were used in an effort to enhance the findings.

### 3 Methodology

#### 3.1 Data

Hedonic approaches need many characteristics of each sold unit. Case and Shiller [3] recommended a different strategy that uses the information on units sold repeatedly. Repeat sales technique proponents contend that because it is based on the actual housing units’ observed appreciation, it more correctly accounts for property characteristics [3]. In this research, the repeat sales/transactional data were provided by MIAC Analytics Ltd which included previous price, recent price, previous transaction date, recent transaction date, property type, and the geographic details of the properties from 1995 to 2019. The flood data provided by the data provider of MIAC Analytics Ltd, WhenFresh Ltd included the history of all the flood events like flood cause, flood count, flood source, flood start date, flood end date, and property level details from the year 1900 to 2020. The elevation of each postcode was gathered from Ordnance Survey, and Shapefiles of the historic flood map, river and coastal bodies of the UK, area benefiting from flood defense, and geography level of the UK were collected from DEFRA and EA. This study was conducted on the data from 2010 to 2020 to avoid the impact of the house price crash of 2008.



■ **Figure 1** Map of Twickenham. Blue lines: river bodies; orange shading: areas protected by flood defence; pink shading: historic flood shadings; red shading: overlap of historic flood shading and areas protected by flood defence; red dots: postcodes only impacted by 2013 flood; green dots: postcodes which did not experience flood events and are outside flood risk areas.

### 3.2 Assumptions

When a property's structural attributes remain constant between transactions, one or more of the following variables could be to blame for the price variation: inflation, significant local changes like the construction of new transportation infrastructure or a flood occurrence, and random variation [5]. It is vital to make the assumption that all properties are equally affected by changes in location variables other than flood occurrences when building repeat sales models. By selecting relatively small areas for investigation and by getting access to local knowledge about any significant events, this can be made more likely [5]. Due to the above assumptions, all the analysis on flood effects were conducted on smaller regions that are geographically adjacent to each other. The districts TW15, TW16, TW17, TW18, TW19, and TW20 of Twickenham were considered for this analysis as the areas were partially covered in historic flood, flood risk, and flood defence. Every other flood-affected region of the UK either had fewer data points or is well protected by flood defences. The 2013 winter flood had an influence on 1404 Twickenham transactions. Thus, Twickenham and the 2013 flood disaster were taken into account for the subsequent analysis.

### 3.3 Treatment Control Group

The winter flood of 2013 was considered as the treatment effect in this research. While a small group of districts in Twickenham was considered in this study to hold other effects on the house price constant, the area was further divided into treatment and control groups. The treatment group was created as a collection of postcodes that experienced the 2013 winter flood, with the flooding having no influence on transactions prior to treatment but having an impact on transactions after treatment. The control group had the collection of postcodes that were never impacted by the flood and are outside flood risk as designated by the Environmental Agency.

### 3.4 META Learners

META Learner is a framework that uses multiple base learners/machine learning algorithms to build a model to estimate Average Treatment Effect (ATE) and Conditional Treatment Effect (CATE). ATE is the effect of a treatment on a population whereas CATE is the effect

of a treatment on a subgroup of the population based on a condition. `causalml` [6] is a Python package provides the tree models – Random Forest, LightGBM, and XGBoost – as the base learners for the META Learners. In this research, the four types of META Learners: S, T, X, and R Learners were used to estimate and validate CATE. To validate which base learner would perform the best in estimating the treatment effect, the preprocessed covariates and outcome variables were passed through all three algorithms, and random search hyperparameter tuning was applied to determine the best parameter. Once the best machine learning algorithm was selected as the base algorithm, the data were passed through all the META learners to calculate CATE. The average of CATE was used to calculate Average Treatment Effect (ATE) which indicated the flood effect on Twickenham.

### 3.5 Equations

Lamond *et al* [5] explain the derivation of the equations used to estimate the treatment effect. This approach was used to determine the growth effect by considering the market effect to be constant. For property  $i$ , the growth in price ( $P$ ) from time  $t$  to  $t + k$  is:

$$Y = \ln(P_{i(t+k)}) / \ln P_{it}$$

This term was used as the outcome variable to estimate the flood effect on the house prices of Twickenham. While each META Learner (S, T, X, and R) has different equations, in this section the equation of T learner is discussed as it proved to be the best learner. In the first stage, the T learner estimates the average outcome using machine learning algorithm[6]:

$$\mu_1(x) = E[Y(1)|X = x]$$

$$\mu_0(x) = E[Y(0)|X = x]$$

where  $\mu$  is the average outcome, 0 indicates the non-flooded (control) properties, 1 indicates the 2013 winter flood-impacted (treatment) properties,  $X$  values are the features, and  $Y$  is the outcome variable or growth effect. In the second stage, CATE ( $\hat{\tau}$ ) is estimated using the below equation[6]:

$$\hat{\tau}(x) = \hat{\mu}_1(x) - \hat{\mu}_0(x)$$

The ATE was further estimated by calculating the average of CATE.

## 4 Results and discussion

Although XGBoost outperformed all the models slightly, LightGBM was chosen as the base learner as it captured the contribution of most of the features which seemed more ideal to estimate the flood effect with an RMSE of 0.23. While ATE was efficient in capturing the impact of the 2013 winter flood on Twickenham districts, CATE was effective in capturing the impact of the flood event on each Twickenham postcode. With LightGBM as the base learner, the results of the META learners are below:

■ **Table 1** The flood effect estimated by all the META Learners of 2013 winter flood in Twickenham districts. The negative sign indicates that the flood effect resulted in a discount on house prices.

META Learners	S	T	X	R
ATE	-0.07	-0.08	-0.11	-0.07

## 4.1 Validation

The `causalml` package [6] was built using synthetic data. To validate the learners, the actual treatment effect was generated from the synthetic data and then the values of ITE and the actual treatment effect were used to validate the results and chose the best learner. However, in real-world applications, since the ground truth is not available, choosing the best learner was quite challenging. So, cumulative gain plots with a theoretical curve produced by the random model were used to validate the performance of the META learners. Once the theoretical random curve was in place, then the learners were compared to it as a benchmark. Every curve had the same beginning and end. Since the curve of the T learner deviated the most from the random line, it was considered to be the best learner amongst S, X, and R learners and the ATE value of the T learner was statistically significant. Meanwhile, from the permutation importance, it was observed that the T learner was capturing the effect of almost all the features. Hence, it was concluded that the T learner estimated the most authentic flood effect, and a discount of 8% was applied to the districts of Twickenham in 2019 as a result of the 2013 winter flood.

## 4.2 Diminishing of flood effect over time

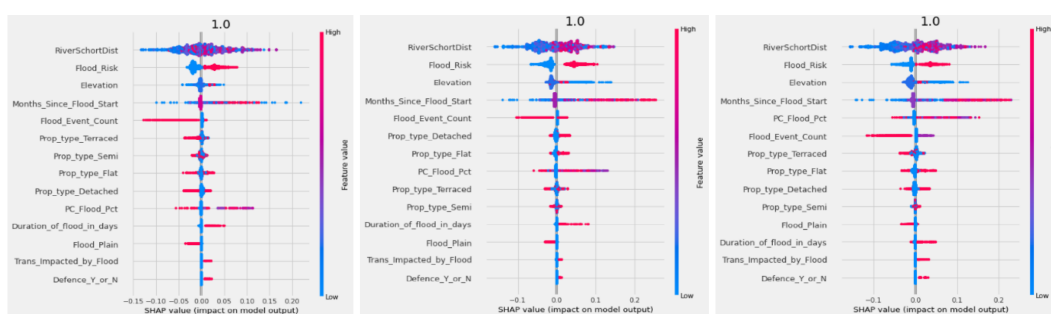
SHAP (Shapley Additive Explanations) values were used to interpret the Meta Learners. Each feature is given an importance value by SHAP for a specific prediction. These values were used to understand the Meta Learners better. As years pass by, customers tend to forget about the flood event, and in the absence of any other issues with the property, house prices tend to continue to increase. A slight decay in discount after 4 years of the flood event can be observed in Table 2. Fig. 2 shows that as the months between the transaction and the flood increase, there is growth in house prices. Whilst for semi-detached houses the discount remained the same from 2015 to 2019, for flats, terraced, and detached houses the discount decays and data points tend to contribute towards the growth of house prices by 2019. The SHAP plots also indicate that lesser values of the features: (a) distance from a river body, (b) elevation from mean sea level, (c) months since the flood event happened to contribute to higher flood discount. While properties within flood plains that experienced a higher number of flood events contribute to flood discounts, the areas protected by flood defence contribute to the growth of house prices.

■ **Table 2** The flood effect estimated by T learners with CI of Twickenham districts over the years.

Years after 2013 flood	Lower Bound	ATE	Upper Bound
2 years	-0.16	-0.09	-0.02
4 years	-0.16	-0.09	-0.02
6 years	-0.16	-0.08	-0.02

## 5 Conclusion

If flood effects do not reflect on the house prices, it raises the concern that a sudden risk re-pricing could be financially unstable if this risk is not represented in house prices. The two major papers that dealt with a similar goal [5, 1] used the "repeat sales method" along with linear or generalized regression method to determine the discount. This study contributes to the ongoing research in the field of climate change and its impact on transitional risk. The use of the causal inference algorithm backed by machine learning, Meta learners, presented



■ **Figure 2** The SHAP plot of T learner over the years. *Left*: after 2 years of the flood event (2015). *Middle*: after 4 years of the flood event (2017); *Right*: after 6 years of the flood event (2019).

a significant answer to the problems with the previous research efforts. First, by using CATE to estimate the flood effect on each postcode, it is possible to capture heterogeneity. Secondly, capturing the non-linear correlations between the data using LightGBM. Third, the algorithms' ability to be understood by using SHAP values.

While Beltrán *et al* [1] stated that “the discount is short-lived and the discount is no longer statistically significant for properties affected by inland flooding after 5 years, which falls to just 4 years for properties affected by coastal flooding”, we found that the discount begins to diminish after 4 years following the 2013 flood event. However, with a longer timeline, it could have been more interesting to capture the decay in the flood discount. One of the shortcomings of this project would be the fact that the treatment's random assignment is not assured as a lot of factors could contribute to the flood occurrences. Beltrán *et al* [1] agreed, as flooding occurs mostly in areas that are exposed to flood risk/hazard. It could be argued that the property or real estate market in such areas might already possess some special characteristics and attract households with distinctive preferences. The research was unable to calculate the treatment effect independently for the four property types due to insufficient data as a result of concentrating on a narrower area to hold other effects on the house prices constant. Although the results of the treatment effects cannot be generalized due to the assumptions associated with the repeat sales data, this methodology can be used to estimate the effect of flood events on house prices for any area and any timeline.

## References

- 1 Allan Beltrán, David Maddison, and Robert Elliott. The impact of flooding on property prices: A repeat-sales approach. *Journal of Environmental Economics and Management*, 95:62–86, 2019.
- 2 Nam Bui, Le Wen, and Basil Sharp. House prices and flood risk exposure: An integration of hedonic property model and spatial econometric analysis. *The Journal of Real Estate Finance and Economics*, pages 1–32, 2022.
- 3 Karl E Case and Robert J Shiller. Prices of single family homes since 1970: New indexes for four cities, 1987.
- 4 CCRA3. Housing briefing. <https://www.ukclimaterisk.org/wp-content/uploads/2021/06/CCRA3-Briefing-Housing.pdf>, 2021. [Online; accessed 17-April-2023].
- 5 Jessica Lamond, David Proverbs, and Adarkwah Antwi. Measuring the impact of flooding on uk house prices: A new framework for small sample problems. *Property Management*, 2007.
- 6 Uber. Meta-learner algorithms. <https://causalml.readthedocs.io/en/latest/methodology.html>, 2022. [Online; accessed 06-Nov-2022].