COST OF DAMAGE OF PLUVIAL FLOODING DUE TO CLIMATE CHANGE: A PRELIMINARY ASSESSMENT*

YOGEN P. SAMOO[1] AND TONY S. WIRJANTO[1,2]

[1]DEPARTMENT OF STATISTICS & ACTUARIAL SCIENCE
[2]SCHOOL OF ACCOUNTING & FINANCE
UNIVERSITY OF WATERLOO

This version: January 2023

Abstract

This paper adopts a deep learning methodology in assessing economic cost of damage arising from pluvial flooding which occurrence is increasingly exacerbated by global climate change. This investigation exploits a recently released, sizable claim database from the US National Flood Insurance Program and meteorological data from National Oceanic and Atmospheric Administration in training several competing models which are later used for prediction on test datasets which have not been used in the training process. Separate analyses are provided in this paper for claims on building and claims on content.

Keywords: Pluvial flooding, artificial neural network, multiple linear regression, NFIP, NOAA.

1 Introduction

With the advent of global climate change, both the frequency and intensity of extreme precipitation events have been rising unprecedentedly in recent times. However, assessing the timing and location of such events with any desired degree of accuracy has proven to be a daunting task for climate scientists in general. This is due primarily to the complexity of the climatic system (Kunkel, 2003, and Karl and Knight, 1998).

Flood is among the costliest natural disasters to this date. According to a report recently published by MunichRE (2020) losses due to flooding amounts to more than $1 trillion globally over the period of 1980-2019 and, moreover, only 12% of the losses is insured. The report published by MunichRE (2020) also posits that flood is perhaps among the most underestimated natural hazards and represents a staggering 40% of all loss-related natural catastrophes worldwide over the same time period. This report provides an important impetus for us to undertake research on assessing, or more precisely, predicting, economic costs of damage arising from pluvial flooding.

Insurance industry has a gargantuan challenge to reduce the insurability gap and also improve a societal natural-disaster resilience by offering affordable and innovative insurance products. In regard to the pluvial flooding, the insurance industry needs to ensure that future victims of flood events are able to receive adequate compensation to restore their properties and belongings through a sound flood-risk management. Achieving such an endeavor would

*This paper grows out of a mater research paper written by the first author under the supervision of the second (and corresponding) author in the summer of 2020. We thank Chengguo Weng for comments and suggestions on the early version of this paper. Tony Wirjanto is also grateful to Ernst Rauch, Chief Geo & Climate Scientist, Climate Change Solutions Department of MunichRE, and Tobias Ellenrieder, Senior Consultant Flood Risks of MunichRE for a series of conversations which have led to several revisions of this paper. The usual disclaimer applies.

necessitate a sound and reliable statistical methodology that is capable of assessing flood risk and also estimating the economic cost of extreme precipitation events at a reasonable degree of accuracy.

Most insured residents against floods in the US are part of the National Flood Insurance Program (NFIP), which was enacted in 1968 by the US Federal Emergency Management Agency (FEMA). In 2019, FEMA released a broad database consisting of more than 2 million historical claims for damages due to flooding.

This study provides a first-step analysis of the aforementioned database by exploiting a deep learning methodology in order to build predictive models to assess, or more precisely, forecast, the economic cost of damage due to pluvial flooding. Briefly pluvial flooding occurs when an extreme rainfall event causes flooding independent of an overflowing water body. The specific goal of this study is to use this newly released dataset to unearth any improvement that an artificial neural network (ANN) architecture from the deep learning domain could potentially offer in terms of improved accuracy in assessing, or more specifically predicting, the economic cost of flood damage in comparison to a multiple linear regression (MLR) framework which is more commonly used in this type of study.

The rest of the paper is structured as follows. Section 2 provides a selected review of the literature most relevant to our study and provides some background knowledge on the subject matter. Section 3 presents some descriptive statistics of the data and outlines the statistical methods and models that are used in this study. Section 4 provides a set of results on the preliminary analysis of the empirical research and proposes directions for further research, which is urgently needed. Section 5 summarizes the key findings of the study and concludes the paper.

2 Selected literature review

2.1 Climate change and extreme precipitation events

Global climate change poses a continuing threat to our socio-economic system as well as the terrestrial life on our planet. In essence climate change is a natural and cyclical occurrence caused by varying levels of greenhouse gases in the earth atmosphere and it happens over the span of a hundred of thousands of years or so. However, in the last few centuries alone, the level of anthropogenic greenhouse gases is reported to have risen substantially. In fact the world is already experiencing an increase in average global temperature of 1°C since 1901 (Storey et al., 2019). Additional evidence on the potential effects of the global climate change continues to accumulate in this field unabatedly (Manabe et al. (1967), Keeling et al. (1976), and Held et al. (2006)).

One major impact on climate is a potential increase in both the frequency and intensity in precipitation events. According to Karl and Knight (1998), the contiguous US has experienced a 10% increase in total annual precipitation since 1910. It is reported that about half of the increase is attributed to the increase in frequency of precipitation events and the other half to the increase in severity of heavy precipitation events. According to a widely cited study on trends in extreme precipitation in North America (Kunkel, 2003), a foreseeable increase in global temperature is likely to result in more frequent and severe precipitation events since atmospheric water vapor will be more readily available to the climatic system to generate precipitation events. As a result a potential increase in the frequency of pluvial flooding can not be ignored. This turn of events can pose a threat to, but also provide an opportunity for, the insurance industry, especially the property insurance sector, to engage more consistently in products related to this event.

Actuaries, far from being climate experts, are overtly expressing their growing concern about climate change and its precarious effects on the insurance industry. According to a global survey of more than 70 insurance analysts conducted by Ernst & Young in 2008, global climate change is considered to be the riskiest endeavor facing the industry (Mills, 2012). Insurers, being institutional investors, are already divesting carbon-intensive energy, such as underwriting coal projects (Storey et al., 2019). However a mitigation effort alone is unlikely to be realistically sufficient to ensure the continuingly vibrant existence of the industry. On the other hand having a thorough understanding and a proper perspective of climate change risks, and also being able to offer innovative and affordable insurance products in a timely manner can go a long way in enhancing the industry's ability in its efforts to improve the societal natural-disaster resilience.

In the case of the pluvial flooding, an exhaustive database which comprises well-specified flood metrics and a predictive model with the ability to predict accurately the economic cost of extreme precipitation events are undoubtedly twin essential to an equitable pricing of flood insurance and for closing the insurability gap. In 2018, the US Federal Emergency Management Agency (FEMA) released a wealth of data on its National Flood Insurance Program (NFIP), which acted as a game changer in research in this field. The following subsection presents a brief historical overview of NFIP claims.

2.2 A brief history of the national flood insurance program

In 1965 Louisiana was hit by a destructive hurricane known as Betsy, flooding 165,000 residences within a matter of just few hours and leaving behind thousands of victims with no means of restoring their possessions. In 1968 a continued lack of interest from the private sector and the provision of federal relief to flood victims, regarded as free insurance, eventually compelled the US federal government to enact the National Flood Insurance Program (NFIP). Under the NFIP, US citizens are eligible to purchase flood insurance policies including building coverage up to $250,000 and content coverage up to $100,000 (Michel-Kerjan, 2010). The absence of a private market for flood insurance can be explained by the nature of flood risk. Figure 1 depicts the NFIP annual aggregate claim amount paid in 2018 dollars and confirms that potential insurers bear a relatively high risk of heavy losses periodically. In other words a year teemed with a multitude of hurricanes and floods can result in catastrophic losses and put the potential insurers quickly out of business.

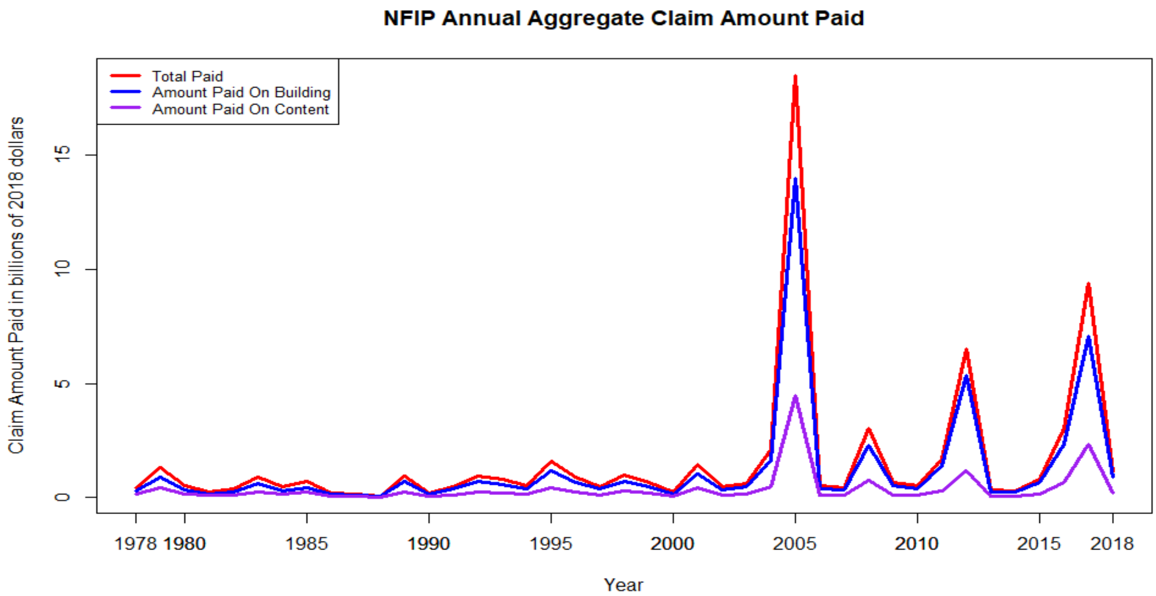

Figure 1. NFIP annual aggregate claim amount paid adjusted for inflation (in 2018 dollars)

In 2005 Hurricane Katrina hit New Orleans with such a devastating force that it costed the NFIP a staggering $18.5 billion (in 2018 USD) in claim payments and put the economic feasibility of the NFIP to test. Already operating on a deficit, the U.S Treasury had to lend $18.6 billion to the program (Michel-Kerjan, 2010). Hurricane Sandy in 2012 brought the total claim value paid by the NFIP to $6.5 billion (in 2018 USD). The total claim value paid by the NFIP in 2017 amounted to a hefty $9.4 billion (in 2018 USD) as the US was hit by Hurricanes Harvey, Irma and Maria. The economic viability of the NFIP is undermined due to a variety of reasons, such as inadequate premium charges, outdated flood maps, expensive administration fees charged by private insurers amongst others. Michel-Kerjan (2010) and Kousky et al. (2017) present an impressively exhaustive research on the operation of the NFIP and also propose a concrete set of measures to improve the effectiveness of the program.

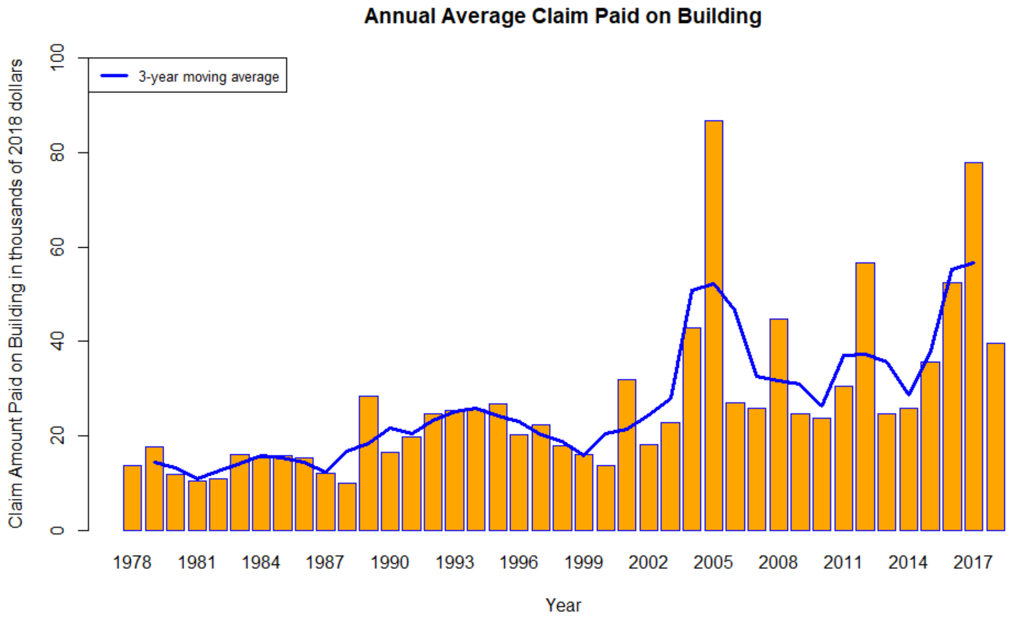

Figure 2. Annual average claim paid on building adjusted for inflation (in 2018 dollars).

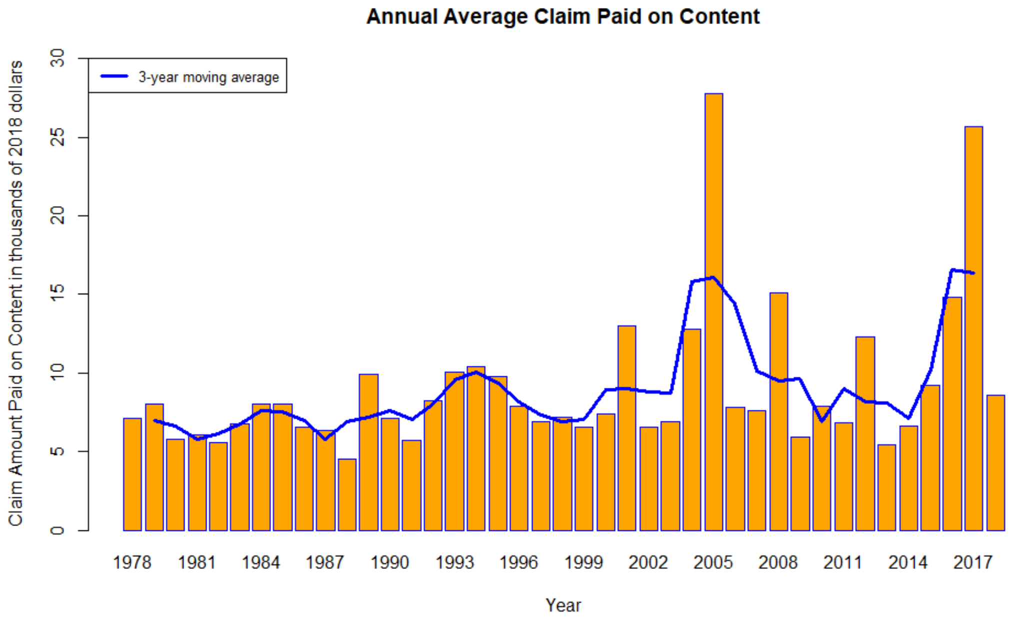

Figure 3. Annual average claim paid on content adjusted for inflation (in 2018 dollars).

Figures 2 and 3 present an annual average claim value paid in 2018 US dollars for building and content respectively. One stark observation from these figures is that the highest average claim values are on the latter part of the timeline for both types of claims, in particular, for the calendar years mentioned above. Another important point worth highlighting from these figures is that, even after adjusting for inflation, both of them still display a highly visible upward trend in the average claim value. As a note, a 3-year moving average line is overlaid on both figures in view to highlight a more distinctly discernible pattern.

2.3 Selected literature review on methodology used for predicting flood risk

With the advent of global climate change, insurers more than ever need to revisit their methodology used to model and predict the economic cost of damage caused by flooding more critically. With regards to insuring against floods, researchers in this field often have to concede that the lack of data availability, but most importantly the lack of data quality, may have seriously impaired studies on flood substantially (Spekkers et al., 2013). It appears to be the case that insurance claims databases cannot be accessed readily by investigators interested in studying this issue. Moreover, even when an access is granted to researchers, recorded measurements are often redacted or become obsolete and, therefore, are of little use for modelling complex events such as flooding.

Spekkers et al. (2015) pointed out that existing literature scarcely investigates the causes of flooding (i.e. drainage system in urban areas). This is because innovative data collection methods must often be devised first in order to capture information about the underlying reason and mechanism behind the damage caused by the pluvial flooding.

Cheng et al. (2012) presented a study of heavy rainfall-related damage claims for the province of Ontario in Canada. By averaging results from five downscaled global climate models, the study shows a substantial increase in potentially incurred losses solely due to the pluvial flooding. Spekker et al. (2013) used a logistic regression model to perform an analysis on how the probability of damage caused by the pluvial flooding is affected by rainfall intensity. But the model leaves a substantial portion of variance in damage on both building and content still largely unaccounted for.

More recently and also most related to our study Lyubchich et al. (2019) provided an extensive overview of the limited existing literature on machine learning techniques which are used to model natural hazard risk for property insurance. It concludes that as far as research using the machine learning techniques to predict flood risk is concerned, daily precipitation is found consistently to be the most studied meteorological variable and is also consistently regarded as the most important predictor in the models.

3 Data and research methodology

3.1.1 NFIP claim dataset

The National Flood Insurance Program (NFIP) claim dataset, which is publicly available from the URL: https://www.fema.gov/, is used in this study. This dataset is composed of 39 variables and 2,432,888 observations, which span the period from 1968 to 2019. The description of each variable originally contained in this dataset is provided in Appendix A. The dataset is truncated for the time range starting from January 1st, 1978, to December 31st, 2018, which is a time period equivalent to 41 years. The dataset contains both claims which resulted in payment and claims without payment. Since the study focuses on estimating the economic cost of the pluvial flooding, non-payment claims are filtered out, reducing the number of claims to 1,282,090. The monthly Consumer Price Index (CPI) for the same time period is retrieved from the URL: https://fred.stlouisfed.org/ which is used to construct annual inflation rates and adjust every claim to the 2018 US dollars.

3.1.2 NOAA daily precipitation dataset

The claim dataset contains the latitude and longitude of the location, where the damage caused by flooding occurred. The latitude and longitude for each claim is used to find the nearest meteorological station in a radius of 500 km. Due to our computational limitations, only meteorological stations, which are related to more than 500 claims, are retained in our study. Of this, 432 meteorological stations satisfied the aforementioned criteria and the number of claims in the dataset exceeds 900, 000 claims. Four meteorological variables, namely daily precipitation, 1-day lag precipitation, daily snowfall and 1-day lag snowfall for the period January 1^{st}, 1978, to December 31^{st}, 2018, are retrieved from these stations. A column containing daily precipitation data from the nearest meteorological station for the date, when water first entered the insured building for each claim is appended to the original claim dataset. Columns for the three other meteorological variables are constructed and appended to the original claim dataset in a similar fashion.

3.1.3 Data cleansing and data partitioning

A substantial number of missing values for daily precipitation is observed in the dataset. Discarding the claims with missing values for daily precipitation brings the number of claims in the dataset down to 117,542, which is almost a nine-fold reduction in the size of the dataset. This is attributed to the fact that the temporal availability of meteorological data varies from station to station and do not always match the date when the damage occurred. Since the original NFIP dataset does not include the reason behind the flood related claims, a rather loose but logical assumption is made in order to proceed with the investigation reported in this study. The assumption entails that claims for damage occurred on rainy days is related to the pluvial flooding, thereby discarding all claims with daily precipitation data of value zero. This further reduces the effective number of claims in the dataset to 98,655.

The prima facie redundant variables (e.g. ID, census tract, etc.) are manually discarded from the analysis. Variables with more that 10 percent of missing values are also removed from the dataset with the exception of the two meteorological variables, namely, daily and 1-day lag snowfall which are imputed with zero. Variables with less than 10 percent of missing values are imputed with the measure of center, which is deemed to be more appropriate individually. A list of the final variables retained for our study can be viewed in Appendix B. Customary practices for pre-processing the data for machine learning algorithms are used before partitioning the data. Specifically the one hot-encoding method is used to create dummy variables for categorical variables and numerical variables are standardized prior to the analysis. Finally the dataset is split into two subsets: 80% of the dataset, which consists of 78,614 observations, is used to train the models (representing a training dataset) and the remaining 20%, which is equivalent to 20,041 observations, is used to assess the performance and robustness of the models (representing a test dataset). The partitioning of the dataset is carried out in this study is done on an ad-hoc basis and a more refined data-driven method can readily be used.

3.1.4 Exploratory data analysis

Random Forest is an ensemble learning algorithm in the machine learning domain. It averages the output of standard decision trees. As a by-product of this, it very usefully allows the ranking of variables by importance (as measured by a percentage increase in Mean Square Error), which is an informative way of learning about the variables in a given dataset and this particular feature of random forest is judiciously exploited in this subsection.

The analysis of claims on property and content damage is carried out separately in order to increase the visibility on variable importance in each case. Random Forest is performed to examine the predictive strength of each feature variable or predictor. The hyper-parameter *m-try* that controls the split-variable randomization feature of Random Forests is set to 29, which is a customary third of the number of predictors as generally prescribed for regression problems. The results are depicted in the figures below and brief descriptions of the most important variables are also presented below.

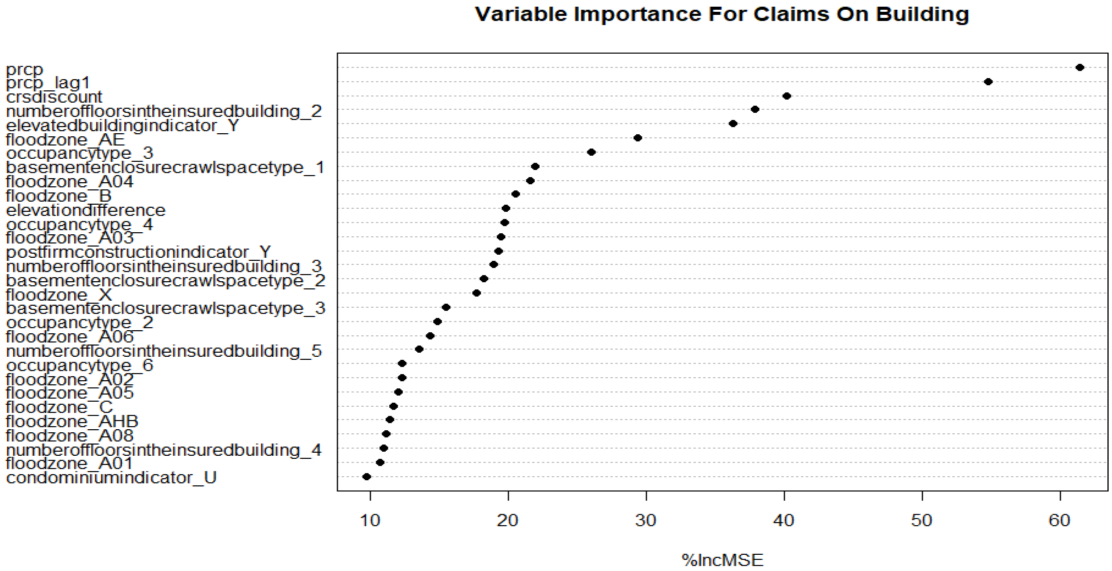

Figure 4. Variables importance for claims on building based on percentage increase in Mean Square Error.

Consistent with the existing literature daily precipitation (*prcp*) is the most important variable when it is used to predict damage caused by the pluvial flooding. It is ranked in the first position for claims on building damage and second for claims on content damage.

The 1-day lag precipitation (*prcp_lag1*) also appears in other related studies and is used to build rainfall intensity measures. Rainfall intensity measures dissolve some of the rigidity that temporal variables might carry. A1-day lag precipitation is ranked second for claims on building damage and first for claims on content damage. This indicates that content will more likely be damaged after a certain amount of rainfall.

The community rating system discount (*crsdiscount*) is a discount on premium for policy-holders in a community, who are collectively taking migration measures (i.e. installing drainage) to reduce flood risks (See Appendix A for further detail). Both the above figures confirm its importance.

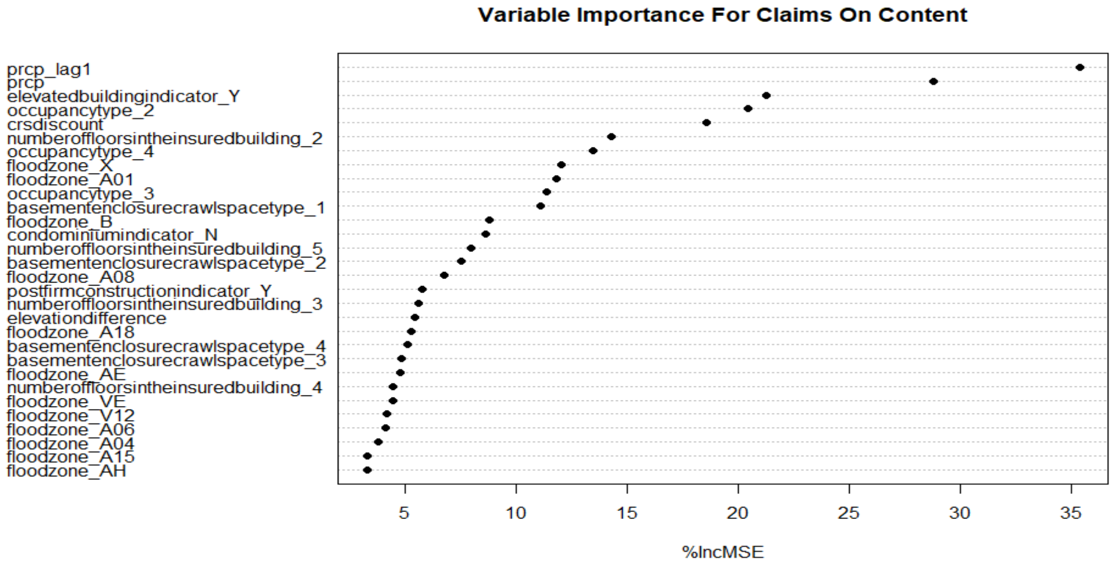

Figure 5. Variables importance for claims on content based on percentage increase in Mean Square Error.

A dummy variable, labelled a *numberoffloorintheinsuredbuilding_2*, indicates that the insured building consists of two floors. It is an interesting variable that could potentially affect the structure of buildings in an effort to mitigate flood risks.

A dummy variable, labelled as *elevatedbuildingindicator_Y*, indicates whether the insured building has no structure below the ground (i.e. the basement). Further detail of specifications to satisfy this criterion is provided in Appendix A. This variable is expected to be an important feature in predicting economic damage.

Flood Insurance Rate Map (FIRM), denoted as *floodzones*, comprises different zones which bear different risks. The zonification by flood risk facilitates pricing of insured buildings according to their peculiarities, although a great deal of criticisms about its obsolescence has been raised (Michel-Kerjan, 2010).

A dummy variable, labelled as *occupancytype_2*, refers to a residential building composed of two to four units. The description for the other occupancy type is also provided in Appendix A.

3.2.1 Multiple linear regression framework

Multiple linear regression (MLR) is a basic statistical technique which is used to model a linear (or approximately linear) relationship between a response variable, y_i, and a set of p feature variables or predictors, x_{ji} (with the first element being an intercept term)

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi} + \varepsilon_i \qquad (1)$$

or

$$y_i = X_i \beta + \varepsilon_i \qquad (2)$$

where $X_i = (1, x_{2i}, \dots, x_{pi})$, $\beta = (\beta_0, \beta_1, \dots, \beta_p)'$, with $\varepsilon_i \sim i.i.d(0, \sigma^2)$ and $\mathrm{Cov}(X_i, \varepsilon_i) = 0$ for $i = 1, \dots, n$. As noted above, it is usually assumed that ε_i is independent and identically distributed (i.i.d) with mean equal to 0 and variance equal to σ^2 MLR is used as a benchmark model in this study and its performance is assessed by a root mean square error (RMSE) criterion which is calculated on the test dataset which has not been used in the training of the MLR model.

3.2.2 Artificial Neural Network architecture

Artificial Neural Network (ANN) is a learning algorithm which is intended to mimic the functioning of neurons in a human brain. Essentially the neurons in the hidden layers are weighted sums of the input data and a predetermined activation function selects which sums will be used to predict the output.

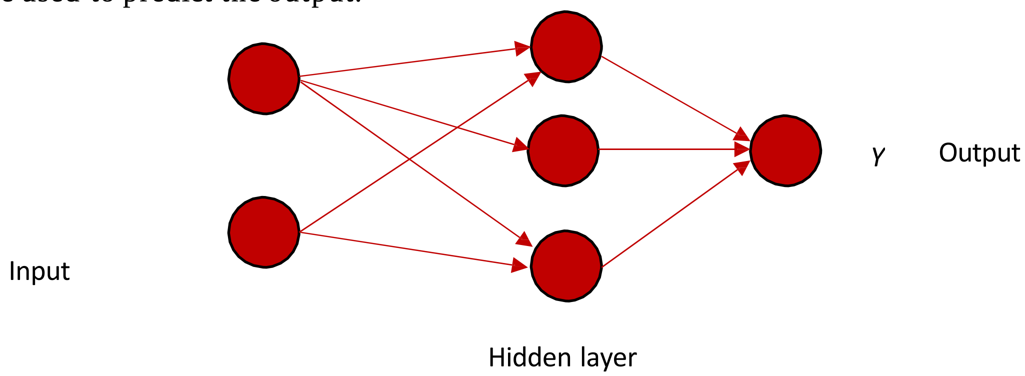

Figure 6. An ANN model with 2 input variables and 1 hidden layer with 3 neurons

In the above diagram only one hidden layer is shown to illustrate the rudimentary principle which governs a deep-learning technique. However, more complex neural networks are devised in real life especially when dealing with big datasets. This study seeks to exploit neural network's ability to capture nonlinear relationships between a set of relevant feature variables and response variables and to recognize underlying patterns in the data. For a coherent comparative analysis, the efficacy of this model is also assessed by the RMSE criterion which is again calculated on the test dataset as mentioned earlier.

4 Empirical analysis

4.1 Results

With the same rationale as in the previous section, results from damage on building and content are analyzed separately. Four models, which are a multiple linear regression model and three neural network architectures, are run for both types of claims. Detailed specifications of all of the models are provided in Appendix B. The visual aspect of plots is exploited by super-imposing the prediction of a trained model on the actual claim amount for the test dataset.

A quick observation from the diagram presented below is that some amount of variance is still left unexplained by all of the four models considered in this study after they have been trained on the training dataset. However the ANN architectures consistently outperform the benchmark MLR model, although none of the four models is able to predict high claim amounts with a high a degree of accuracy.

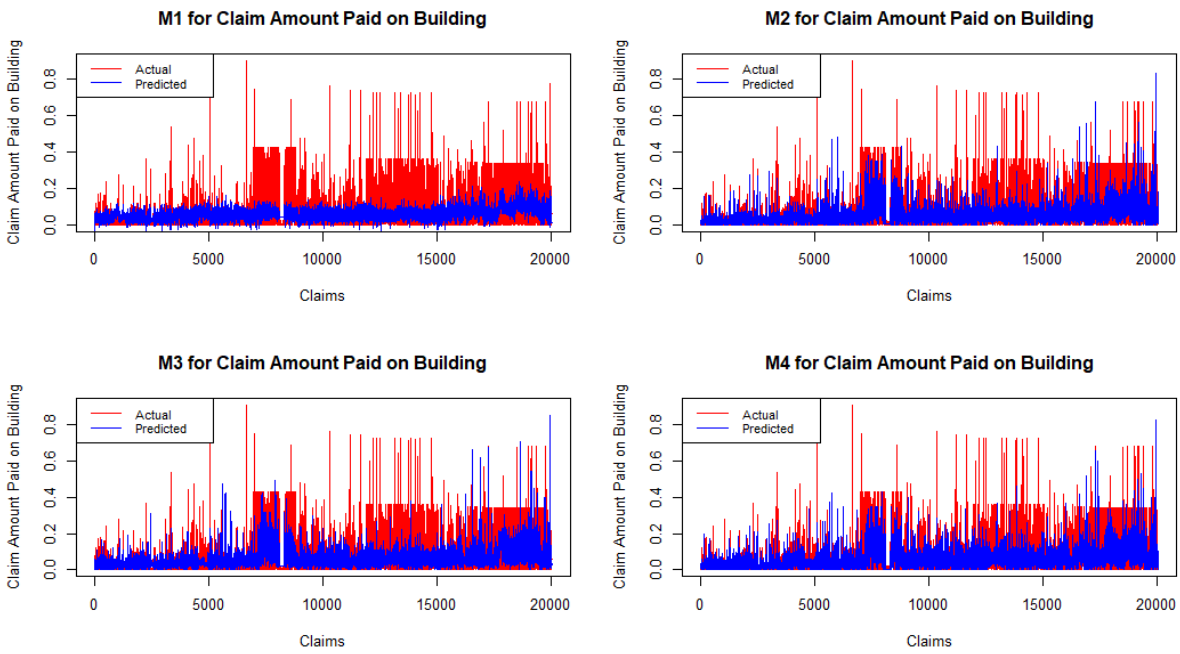

Figure 7. Comparison between prediction by each model and actual value of claim amount paid on building in test dataset.

The RMSE values, which are calculated on both the train and test datasets, are tabled below. The numerical value of the RMSE of a trained model to predict standardized claim amounts does not mean much standing alone. Moreover, the conversion back to 2018 dollars is deemed unnecessary for comparative purposes given the performance of the models.

Amount Paid on Building		
Model	TrainRMSE	TestRMSE
MLR (M1)	0.0781798	0.07788014
NN (M2)	0.0673668	0.0722917
NN (M3)	0.0676438	0.0718767
NN (M4)	0.0659166	0.0721947

Table 1. Train and test RMSE for each model for claim amount paid on building.

It is important to observe from Table 1 that the test RMSE values confirm that the ANN architectures bring about some degree of improvement over that of the MLR when they are used to model claim damage paid on building. The train RMSE for the neural network architectures is substantially smaller in magnitude than their respective test RMSE, indicating a potential for overfitting the models when the ANN architectures are trained.

For the claim amount paid on the content case, the figure below also shows that some amount of variance is still left unexplained by all of the four models considered in this study. The recorded performance of the models in comparison to claim amount paid on building appears to be relatively weaker. The ANNs outperform the benchmark MLR model to a lesser extent and, again, none of the four models predicts high claim amounts with a high degree of accuracy. The results reflect unerringly the complex nature of modelling claim amounts paid on content.

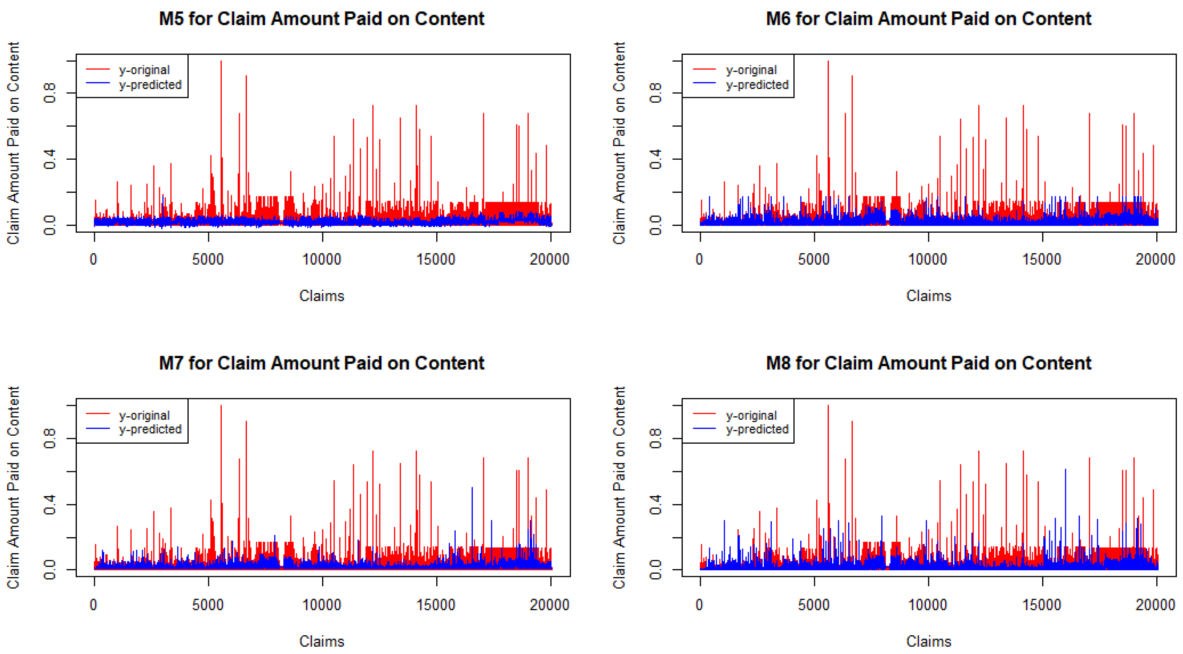

Figure 8. Comparison between prediction by each model and actual value of claim amount paid on content in test dataset.

Amount Paid on Content		
Model	Train RMSE	Test RMSE
MLR (M5)	0.03644489	0.03533696
NN (M6)	0.03390541	0.03482989
NN (M7)	0.03343500	0.03488170
NN (M8)	0.03323435	0.03615184

Table 2. Train and test RMSE for each model for claim amount paid on content.

The test RMSE values confirm that the predictions from the ANNs architectures, which are denoted as M6, M7 and M8 respectively, are generally more superior to those of the benchmark MLR model. However the train RMSE for the ANNs is uniformly smaller in magnitude than their respective test RMSE, providing again the potential for overfitting the models when the ANNs are trained.

4.2 Research extension

This study contributes to a limited amount of existing literatures on the U.S National Flood Insurance Program by showing that the deep learning method could bring about some potential improvement in predicting the economic cost of floods in comparison to the more conventional MLR framework, especially for claim amounts paid on building. However, it is also evident that all of the models presented in the previous subsection are not accurate enough to be used for real life estimation of economic cost for flood events. Numerous reasons for this failure can be readily put forward for the sources of the reported inaccuracy of the prediction results generated by these models. This section includes some of the most salient ones in the hope to encourage further investigation into this issue.

In this study meteorological data is retrieved from the nearest meteorological station in a radius of 500 km in order to preserve the maximum amount of claim data. Moreover, within the meteorological stations selected by the aforementioned criteria, only those with more than 500 claim payments are retained for analysis due to our computational limitations as well as for consistency. These concessions are necessary when dealing with conventional meteorological data. Unfortunately this also renders the meteorological data to have a substandard quality especially in the localized scale context. A better alternative would be satellite meteorological data, known as Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) which can be obtained from https://chrsdata.eng.uci.edu/.

The latitude and longitude values provided in the NFIP datasets for each claim are only up to two decimal places. FEMA rightly truncates the coordinates in order to protect the privacy concerns of policyholders. However a low level of precision of a grid location inevitably leads to to an inaccurate meteorological data which, in turn, has major ramifications on the accuracy of the model, especially when the most highly important variable is inaccurately measured for all observations.

Based on the results reported in this study we reach a tentative conclusion that the complexity of modelling flood events is unlikely to be addressed adequately by using the precipitation data alone. An improvement in the quality of feature variables or predictors is needed in order to build models with higher degree of accuracy in prediction. For instance, whether a damage is caused specifically by a rainfall event for a claim is not reported in the NFIP claim dataset. Moreover new data collection methods need to be carefully devised to replace existing records. For example the maximum water depth in the insured building on the date of loss is a very informative feature variable. Installation of innovative devices that can collect high quality data during flood events can in principle be mandated as underwriting criteria.

Lastly a reliable statistical model for assessing the economic cost of flood events accurately would not be complete unless the data also includes damage from non-insured property and content damages as predictors Models with a high degree of accuracy in prediction, which are trained on both insured and non-insured data, can ideally be used in combination with a projection of rainfall and other flood metrics data by the use of extreme value theory in statistics, to predict the total economic cost of prospective flood events.

5 Conclusion

In this paper the performance of the ANN models in predicting the economic cost of flood damage is compared to that from the standard MLR framework, which is more commonly used in this type of study. Claim data from the U.S National Flood Insurance Program (NFIP) is used to build a number of predictive models. The analysis for claims on building and claims on content are carried out separately. The grid location for each claim is used to identify the nearest meteorological station to where the damage occurred. Subsequently the meteorological data is retrieved for each station from the National Oceanic and Atmospheric Administration (NOAA). The meteorological variables of interest are selected specifically for the date when water first entered the insured building. The data is then preprocessed before they are utilized to train the models. The feature of Random Forest in determining the important variables is judiciously exploited to present a useful preview of the candidate variables needed for assessing, or more precisely predicting, the economic cost of damage due to the pluvial flooding.

ANN models are shown to perform uniformly better than the conventional MLR models for claims on building and to a lesser extent also for claims on content. The results endorse the fact that claims on content have a more complex nature and are not easy to model statistically. Another theme pointed by our results is the amount of variance of claim value left unaccounted for by all of the four models. This theme is recurrent in the existing literature and often justified by the fact that the current data collection methods and flood maps are seemingly obsolete. In other words improvement in a model's prediction accuracy is limited by default. Encouraging the installation of innovative devices that collects informative points of data could be the next game changer we are looking for in order to solve this aspect of the challenge. In conclusion ANN has demonstrated its ability to capture more informative elements than the conventional MLR framework in assessing economic damage caused by the pluvial flooding. Further research on this topic is urgently called for as the occurrence of pluvial flooding is expected become a less rare-event worldwide compared to even the most recent past due to the on-going global climate change.

References

Cheng, C., Li, Q., Li, G., and Auld, H. (2012). Climate Change and Heavy Rainfall-Related Water Damage Insurance Claims and Losses in Ontario, Canada. Journal of Water Resource and Protection. 04. 49-62.10.4236.

Held, I. M., and B. J. Soden, (2006). Robust Responses of the Hydrological Cycle to Global Warming. Journal Climate, 19, 5686–5699.

Horn, Diane. (2019). Selected Issues for National Flood Insurance Program (NFIP) Reauthorization and Reform. Washington, DC: Congressional Research Service.

Karl, T. R., and R. W. Knight, (1998). Secular Trends of Precipitation Amount, Frequency, and Intensity in the United States. Bulletin of the American Meteorological Society, 79, 231–242.

Keeling, C.D., Bacastow, R.B., Bainbridge, A.E., Ekdahl, C.A., Jr., Guenther, P.R., Waterman, L.S. and Chin, J.F.S. (1976), Atmospheric carbon dioxide variations at Mauna Loa Observatory, Hawaii. Tellus, 28: 538-551.

Kousky, C. and Michel-Kerjan, E. (2017), Examining Flood Insurance Claims in the United States: Six Key Findings. Journal Risk and Insurance, 84: 819-850. doi:10.1111/jori.12106.

Kunkel, K.E. (2003). North American Trends in Extreme Precipitation. Natural Hazards 29, 291–305.

Lyubchich, V, Newlands, NK, Ghahari, A, Mahdi, T, Gel, YR. (2019). Insurance risk assessment in the face of climate change: Integrating data science and statistics. WIREs ComputStat; 11: e1462. https://doi.org/10.1002/wics.1462/

Manabe, S., and R. T. Wetherald, (1967). Thermal Equilibrium of the Atmosphere with a Given Distribution of Relative Humidity. Journal of Atmospheric Sciences, 24, 241–259.

Merz, B., Heidi, K., Lall, U. (2013). Multi-variate flood damage assessment: A tree-based data-mining approach. Nat. Hazards Earth Syst. Sci. 13. 53-64. 10.5194/nhess-13-53-2013.

Michel-Kerjan, E., (2010). Catastrophe Economics: The National Flood Insurance Program. Journal of Economic Perspectives. 24. 165-86. 10.2307/20799178.

Mills, E., (2009). A Global Review of Insurance Industry Responses to Climate Change. The Geneva Papers on Risk and Insurance - Issues and Practice. 34. 323-359. 10.1057/gpp.2009.14.

Munich Re's NatCatSERVICE. https://www.munichre.com/en/risks/natural-disasters-losses-are-trending-upwards/floods-and-flash-floods-underestimated-natural-hazards.html. Accessed on July 24th, 2020.

Spekkers, M. H., Kok, M., Clemens, F. H. L. R., & ten Veldhuis, J. A. E. (2013). A statistical analysis of insurance damage claims related to rainfall extremes. Hydrology and Earth System Sciences,17(3), 913–922.

Spekkers, M. H., Kok, M., Clemens, F. H. L. R., & ten Veldhuis, J. A. E. (2014). Decision-tree analysis of factors influencing rainfall-related building structure and content damage. Natural Hazards and Earth System Science. 14. 10.5194/nhess-14-2531-2014.

Spekkers, M. H., Clemens, F. H. L. R., & ten Veldhuis, J. A. E. (2015). On the occurrence of rainstorm damage based on home insurance and weather data. Natural Hazards and Earth System Sciences Discussions. 2. 10.5194/nhessd-2-5287-2015.

Storey C, MacFarlane A, Spira J, Davangere M, Thulliez M, Bagree N, Hughes R, and Watt S. (2019). Climate Change for Actuaries: An Introduction. London: IFoA. https://www.actuaries.org.uk/documents/climate-change-actuaries-introduction.

Weber, A. (2020). FEMA Puts New Data on the Map for Policymakers. Cityscape, 22(1), 273-278. doi:10.2307/26915497.

Appendix A. NFIP variables description

Field Name	Description
agricultureStructureIndicator	Yes (Y) or No (N) indicator of whether or not a building is reported as being an agricultural structure in the policy application.
asOfDate	The effective date of the data in the file.
baseFloodElevation	Base Flood Elevation (BFE) is the elevation at which there is a 1% chance per year of flooding in feet from the elevation certificate.
basementEnclosureCrawlspaceType	Basement is defined for purposes of the NFIP as any level or story which has its floor subgrade on all sides. Basement structure values are as follows: 0 – none. 1 – Finished Basement/Enclosure. 2 – Unfinished Basement/Enclosure. 3 - Crawlspace. 4 – Subgrade Crawlspace.
reportedCity	This is the city of the insured property as reported to us by our Write Your Own (WYO) partners.
condominium Indicator	This is an indicator of what type of condominium property is being insured. Condominium Code - 1 character: - Not a condominium (N). - An individual condominium unit owned by a unit owner, or by a condominium association (U). The entire condominium building owned by the association insuring building common elements as well as building elements (additions and alterations) within all units in the building, not eligible under Condominium Master Policy (A). The entire residential condominium building owned by the association eligible under Condominium Master Policy, insuring the entire condominium building common elements as well as building elements (additions and alterations)within all units in the building, (H) for High-Rise or (L) for Low-Rise. Townhouses (T).

policyCount	Insured units in an active status. A policy contract ceases to be in an active status as of the cancellation date or the expiration date. Residential Condominium Building Association Policy (RCBAP) contracts are stored as a single policy contract but insure multiple units and therefore represent multiple policies.
countyCode	FIPS code uniquely identifying the primary county (e.g., 011 represents Broward County) associated with the project. Note, the County Code field may not reflect the individual county the property is located as projects can be associated with more than one county.
communityRatingSystemDiscount	The Community Rating System (CRS). Classification Credit Percentage used to rate the policy. The insurance premium credit is based on whether a property is in or out of the Special Flood Hazard Area (SFHA) as shown on the community's Flood Insurance Rate Map. The premium credit for properties in the SFHA increases according to a community's CRS class. 1 - SFHA 45% ** Non SFHA 10% ** 2 - SFHA 40% ** Non SFHA 10% ** 3 - SFHA 35% ** Non SFHA 10% ** 4 - SFHA 30% ** Non SFHA 10% ** 5 - SFHA 25% ** Non SFHA 10% ** 6 - SFHA 20% ** Non SFHA10% ** 7 - SFHA 15% ** Non SFHA 5% ** 8 - SFHA 10% ** Non SFHA 5% ** 9 - SFHA 5% Non SFHA 5% 10 - SFHA 0% Non SFHA 0% *For the purpose of determining CRS Premium Discounts, all AR and A99 zones are treated as non-SFHAs. **These percentages are subject to change. Always refer to the Flood Insurance Manual for the latest information.
dateOfLoss	Date on which water first entered the insured building.
elevatedBuildingIndicator	Yes (Y) or No (N) indicator of whether or not a building meets the NFIP definition of an elevated building. An elevated building is a no-basement building that was constructed so as to meet the following criteria: 1. The top of the elevated floor (all A zones) or the bottom of the lowest horizontal structural member of the lowest floor (all V zones) is above ground level. 2. The building is adequately anchored.

	3. The method of elevation is pilings, columns (posts and piers), shear walls (not in V zones), or solid foundation perimeter walls (not in V zones).
elevationCertificateIndicator	Indicates if a policy has been rated with elevation certificate 1 - No Elevation Certificate, original effective date prior to October 1, 1982, with no break in insurance coverage or change in insurable interest. Policies will be rated using “No Base Flood Elevation” +2 to +4 feet rates. 2 - No Elevation Certificate, original effective date on or after October 1, 1982, with no break in insurance coverage or change in insurable interest. Policies will be rated using “No Elevation Certificate” rates. 3 - Elevation Certificate with BFE. Policies will be rated using “With Base Flood Elevation” rates. 4 - Elevation Certificate without BFE. Policies will be rated using “No Base Flood Elevation” rates.
elevationDifference	Difference in feet between the elevation of the lowest floor used for rating or the flood proofed elevation and the base flood elevation (BFE), or base flood depth, as appropriate from the elevation certificate.
censusTract	US Census Bureau defined census Tracts; statistical subdivisions of a county or equivalent entity that are updated prior to each decennial census. The NFIP relies on our geocoding service to assign census tract code. 11 digit code defining census tract.
floodZone	Flood zone derived from the Flood Insurance Rate Map (FIRM) used to rate the insured property. A - Special Flood with no Base Flood Elevation on FIRM. AE, A1-A30 - Special Flood with Base Flood Elevation on FIRM. A99 - Special Flood with Protection Zone AH, AHB* - Special Flood with Shallow Ponding. AO, AOB* - Special Flood with Sheet Flow X, B - Moderate Flood from primary water source. Pockets of areas subject to drainage problems.

	X, C - Minimal Flood from primary water source. Pockets of areas subject to drainage problems. D - Possible Flood. V - Velocity Flood with no Base Flood Elevation on FIRM. VE, V1-V30 - Velocity Flood with Base Flood Elevation on FIRM. AE, VE, X - New zone designations used on new maps starting January 1, 1986, in lieu of A1-A30, V1-V30, and B and C. AR - A Special Flood Hazard Area that results from the decertification of a previously accredited flood protection system that is determined to be in the process of being restored to provide base flood protection AR Dual Zones–(AR/AE, AR/A1-A30, AR/AH, AR/AO, AR/A) Areas subject to flooding from failure of the flood protection system (Zone AR) which also overlap an existing Special Flood Hazard Area as a dual zone. *AHB, AOB, ARE, ARH, ARO, and ARA are not risk zones shown on a map, but are acceptable values for rating purposes.
houseWorship	Yes (Y) or No (N) indicator of whether or not a building is reported as being a house of worship in the policy application.
Latitude	Approximate latitude of the insured building (to 1 decimal place). This represents the approximate location of the insured property. The precision has been lessened to ensure individual privacy. This may result in a point location that exists in an incorrect county or state. Use the state and county fields for record aggregation for these dimensions.
locationOfContents	Code that indicates where within the structure the contents are located. 1 - Basement/Enclosure/Crawlspace/Subgrade Crawlspace only. 2 - Basement/Enclosure/Crawlspace/Subgrade Crawlspace and above. 3 - Lowest floor only above ground level (No basement/ enclosure/crawlspace/subgrade crawlspace). 4 - Lowest floor above ground level and higher floors (No basement/enclosure/crawlspace/subgrade crawlspace). 5 - Above ground level more than one full floor. 6 - Manufactured (mobile) home or travel trailer on foundation.

Longitude	Approximate longitude of the insured building (to 1 decimal place). This represents the approximate location of the insured property. The precision has been lessened to ensure individual privacy. This may result in a point location that exists in an incorrect county or state. Use the state and county fields for record aggregation for these dimensions.
lowestAdjacentGrade	Lowest natural grade adjacent to the insured structure prior to excavating or filling. The difference in feet of the lowest natural grade adjacent to the building from the reference level of the building.
lowestFloorElevation	A building's lowest floor is the floor or level (including basement/enclosure/crawlspace/subgrade crawlspace) that is used as the point of reference when rating a building. This includes the level to which a building is flood proofed*. The elevation in feet of the reference level of the building from the elevation certificate.
numberOfFloorsInTheInsuredBuilding	Code that indicates the number of floors in the insured building. 1 = One floor. 2 = Two floors. 3 = Three or more floors. 4 = Split-level. 5 = Manufactured (mobile) home or travel trailer on foundation. 6 = Townhouse/Rowhouse with three or more floors (RCBAP Low-rise only).
nonProfitIndicator	Yes (Y) or No (N) indicator of whether or not a building is reported as being a non-profit in the policy application.
obstructionType	Code that gives the type of obstruction (if any) in the enclosure (if any). With obstruction: enclosure/ crawlspace with proper openings not used for rating (not applicable in V zones) – 15. With obstruction: less than 300 sq. ft. with breakaway walls, but no machinery or equipment attached to building below lowest elevated floor, or elevation of machinery/ equipment is at or above Base Flood. Elevation - 20 With obstruction: less than 300 sq. ft. with breakaway walls or finished enclosure and with machinery or equipment attached to building below lowest elevated floor, and elevation of machinery/equipment is below. Base Flood Elevation - 24 With obstruction: 300 sq. ft. or more with breakaway walls, but no machinery or equipment attached to building below the Base Flood Elevation – 30. With obstruction: 300 sq. ft. or more with breakaway walls or finished enclosure and with machinery or equipment attached to building below the Base Flood Elevation - 34 With obstruction: no walls, but the

	the building is located below the BFE. – 95. With Obstruction: Elevated buildings with elevator below the BFE in V zones. Breakaway wall obstruction is finished or is used for other than parking, building access, or storage. - 96 With Obstruction: Elevated buildings with elevator below the BFE in V zones. No other obstruction, but has M&E servicing the building located below the BFE. – 97. With Obstruction: Elevated buildings with elevator below the BFE in V zones. Breakaway walls obstruction and M&E servicing the building are located below the BFE. – 98.
occupancyType	Code indicating the use and occupancy type of the insured structure. One digit code: 1=single family residence. 2 = 2 to 4 unit residential building. 3 = residential building with more than 4 units. 4 = Non-residential building.
originalConstructionDate	The original date of the construction of the building.
originalNBDate	The original date of the flood policy.
amountPaidOnBuildingClaim	Dollar amount paid on the building claim. In some instances, a negative amount may appear which occurs when a check issued to a policy holder isn't cashed and has to be re-issued.
amountPaidOnContentsClaim	Dollar amount paid on the contents claim. In some instances, a negative amount may appear, which occurs when a check issued to a policy holder isn't cashed and has to be re-issued.
amountPaidOnIncreasedCostOfComplianceClaim	Dollar amount paid on the Increased Cost of Compliance (ICC) claim. Increased Cost of Compliance (ICC) coverage is one of several flood insurances resources for policyholders who need additional help rebuilding after a flood. It provides up to $30,000 to help cover the cost of mitigation measures that will reduce the flood risk.
postFIRMConstructionIndicator	Yes or No Indicator on whether construction was started before or after publication of the FIRM. For insurance rating purposes, buildings for which the start of construction or substantial improvement was after December 31, 1974, or on or after the effective date of the initial FIRM for the community, whichever is later, are considered Post-FIRM construction.

rateMethod	Indicates policy rating method: 1 – Manual. 2 – Specific. 3 – Alternative. 4 - V-Zone Risk Factor Rating Form. 5 - Underinsured Condominium Master Policy. 6 – Provisional. 7 - Preferred Risk Policy (PRPs issued for eligible properties located within a non-Special Flood Hazard Area [non-SFHA]). 8 – Tentative. 9 - MPPP Policy. A - Optional Post-1981 V Zone. B – Pre-FIRM policies with elevation rating - Flood Insurance Manual rate tables. E – FEMA Pre-FIRM Special Rates. F – Leased Federal Property. G – Group Flood Insurance Policy (GFIP) P – Preferred Risk Policy (A PRP renewal issued in the first year following a map revision for an eligible property that was newly mapped into the SFHA by the map revision, or new business written for an eligible property that was newly mapped into the SFHA by a map revision effective on or after October 1, 2008 – PRP Eligibility Extension). Q – Preferred Risk Policy (subsequent PRP renewals where the previous policy year was reported as a ‘P’ or ‘Q’). S – FEMA Special Rates. T – Severe Repetitive Loss Properties (formerly Target Group Full Risk). Effective October 1, 2013, code will no longer be valid. W – Pre-FIRM policies with elevation rating – Submit-for-Rate procedures.
smallBusinessIndicatorBuilding	Yes (Y) or No (N) indicator of whether or not the insured represents a small business. Small business is defined as a business with fewer than 100 employees in the policy application.
State	The two-character alpha abbreviation of the state in which the insured property is located.
totalBuildingInsuranceCoverage	Total Insurance Amount in dollars on the Building.
totalContentsInsuranceCoverage	Total Insurance Amount in dollars on the Contents.
yearofLoss	Year of Loss = Year in which the flood loss occurred.
reportedZipCode	5 digit Postal Zip Code of the insured property.

primaryResidence	Yes (Y) or No (N) indicator of whether or not a building is a primary residence. A primary residence is a single family building, condominium unit, apartment unit, or unit within a cooperative building that will be lived in by the policyholder or the policyholder's spouse for: More than 50% of the 365 calendar days following the current policy effective date; or 50% or less of the 365 calendar days following the current policy effective date if the policyholder has only one residence and does not lease that residence to another party or use it as rental or income property at any time during the policy term. A policyholder and the policyholder's spouse may not collectively have more than one primary residence.
Id	Unique ID assigned to the record.

Appendix B. List of variables used for regression

- Daily precipitation.
- 1-day lag precipitation.
- Daily snowfall.
- 1-day lag snowfall.
- Community rating system discount.
- Elevation difference.
- Basement enclosure crawl space type.
- Condominium indicator.
- Elevated building indicator.
- Flood zone.
- Number of floors in insured building.
- Occupancy type.
- Post FIRM construction indicator.

Appendix C. Summary of Results

M1 (R-Output)							
Call:							
lm(formula = amountpaidonbuildingclaim ~ ., data = train1)							
Residuals:							
Min	IQ	Median	3Q	Max			
-0.22351	-0.04566	-0.0211	0.02151	0.92086			
Coefficients							
				Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)				-0.034	0.023689	-1.435	0.151213
crsdiscount				0.129048	0.004622	27.921	< 2e-16 ***
Elevationdifference				-0.00612	0.001447	-4.228	2.36e-05 ***
prcp				0.032865	0.001636	20.093	< 2e-16 ***
prcp_lag1				0.090638	0.002392	37.887	< 2e-16 ***

snow	-0.06522	0.016092	-4.053	5.06e-05 ***
snow_lag1	-0.0821	0.030598	-2.683	0.007296 **
basementenclosurecrawlspacetype_1	-0.02651	0.001221	-21.709	< 2e-16 ***
basementenclosurecrawlspacetype_2	-0.01609	0.001177	-13.673	< 2e-16 ***
basementenclosurecrawlspacetype_3	-0.0196	0.001578	-12.423	< 2e-16 ***
basementenclosurecrawlspacetype_4	-0.00154	0.002783	-0.555	0.578812
Condominiumindicator H	0.080108	0.024777	3.233	0.001225 **
condominiumindicator_L	0.129439	0.026245	4.932	8.16e-07 ***
condominiumindicator_N	0.051127	0.02362	2.165	0.030427 *
condominiumindicator_U	-0.00441	0.023852	-0.185	0.853435
elevatedbuildingindicator_Y	-0.01191	0.000916	-13.009	< 2e-16 ***
floodzone_A00	-0.00753	0.01102	-0.683	0.494693
floodzone_A01	0.009895	0.001999	4.950	7.45e-07 ***
floodzone_A02	0.031224	0.002987	10.453	< 2e-16 ***
floodzone_A03	0.058869	0.002472	23.811	< 2e-16 ***
floodzone_A04	0.042971	0.002282	18.833	< 2e-16 ***
floodzone_A05	0.039338	0.002186	17.997	< 2e-16 ***
floodzone_A06	0.031118	0.002027	15.352	< 2e-16 ***
floodzone_A07	0.017234	0.002579	6.681	2.38e-11 ***
floodzone_A08	0.024736	0.002205	11.219	< 2e-16 ***
floodzone_A09	0.014334	0.003061	4.683	2.83e-06 ***
floodzone_A0B	-0.02765	0.014614	-1.892	0.058454
floodzone_A10	0.031467	0.002742	11.475	< 2e-16 ***
floodzone_A11	0.03275	0.003592	9.117	< 2e-16 ***
floodzone_A12	0.006728	0.005133	1.311	0.18995
floodzone_A13	0.01231	0.003337	3.689	0.000225 ***
floodzone_A14	0.021158	0.00347	6.098	1.08e-09 ***
floodzone_A15	0.015962	0.003732	4.277	1.90e-05 ***
floodzone_A16	0.01919	0.005891	3.258	0.001124 **
floodzone_A17	0.016822	0.005991	2.808	0.004984 **
floodzone_A18	-0.0081	0.00691	-1.172	0.241313
floodzone_A19	-0.00701	0.020948	-0.335	0.737774
floodzone_A20	0.011005	0.009059	1.215	0.224447
floodzone_A21	0.014191	0.013893	1.021	0.307047
floodzone_A22	0.033322	0.019598	1.700	0.089076
floodzone_A23	0.024166	0.045189	0.535	0.592812
floodzone_A25	0.019158	0.045217	0.424	0.671799
floodzone_A30	0.012473	0.039136	0.319	0.749953
floodzone_A4	-0.00646	0.078247	-0.083	0.934223
floodzone_A99	-0.00029	0.004764	-0.062	0.950784
floodzone_AA	-0.0159	0.078233	-0.203	0.838968
floodzone_AE	0.021032	0.001319	15.952	< 2e-16 ***
floodzone_AH	0.018583	0.006893	2.696	0.007019 **
floodzone_AHB	-0.02652	0.002782	-9.534	< 2e-16 ***
floodzone_AO	0.024725	0.004017	6.155	7.55e-10 ***

floodzone_AOB	0.031497	0.005411	5.821	5.86e-09 ***
floodzone_AR	0.004629	0.078243	0.059	0.952824
floodzone_B	0.037847	0.001655	22.87	< 2e-16 ***
floodzone_C	0.009416	0.001621	5.809	6.29e-09 ***
floodzone_D	0.01041	0.009493	1.097	0.272829
floodzone_V	0.001808	0.016358	0.111	0.911967
floodzone_V02	-0.0058	0.023623	-0.246	0.805975
floodzone_V03	-0.01082	0.031963	-0.339	0.734875
floodzone_V04	0.003262	0.021732	0.150	0.880699
floodzone_V05	0.01843	0.011363	1.622	0.1048
floodzone_V06	0.03164	0.011367	2.783	0.005379 **
floodzone_V07	0.009931	0.023622	0.420	0.674175
floodzone_V08	-0.00236	0.009187	-0.257	0.797174
floodzone_V09	-0.00166	0.010937	-0.152	0.879299
floodzone_V10	0.020988	0.009238	2.272	0.023101 *
floodzone_V11	-0.01767	0.011729	-1.506	0.131979
floodzone_V12	-0.0026	0.007129	-0.365	0.715358
floodzone_V13	-0.00236	0.008831	-0.267	0.789723
floodzone_V14	-0.00778	0.011024	-0.705	0.480605
floodzone_V15	0.043444	0.013477	3.223	0.001267 **
floodzone_V16	0.00737	0.010927	0.675	0.499981
floodzone_V17	0.015382	0.023619	0.651	0.514888
floodzone_V18	0.109887	0.031966	3.438	0.000587 ***
floodzone_V19	0.076048	0.013692	5.554	2.79e-08 ***
floodzone_V20	-0.00037	0.018493	-0.02	0.984156
floodzone_V21	0.020063	0.017126	1.171	0.241407
floodzone_VE	0.017032	0.00349	4.881	1.06e-06 ***
floodzone_X	0.037113	0.001367	27.16	< 2e-16 ***
numberoffloorsintheinsuredbuilding_2	0.017652	0.00072	24.51	< 2e-16 ***
numberoffloorsintheinsuredbuilding_3	0.012957	0.001095	11.838	< 2e-16 ***
numberoffloorsintheinsuredbuilding_4	0.010037	0.00231	4.346	1.39e-05 ***
numberoffloorsintheinsuredbuilding_5	-0.02498	0.002863	-8.724	< 2e-16 ***
numberoffloorsintheinsuredbuilding_6	-0.02054	0.02179	-0.942	0.345972
occupancytype_2	-0.00798	0.001376	-5.799	6.70e-09 ***
occupancytype_3	0.038724	0.002911	13.304	< 2e-16 ***
occupancytype_4	0.017833	0.001373	12.988	< 2e-16 ***
occupancytype_6	0.078634	0.003489	22.536	< 2e-16 ***
postfirmconstructionindicator_Y	0.012385	0.000778	15.924	< 2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1				
Residual standard error: 0.07822 on 78526 degrees of freedom				
Multiple R-squared: 0.122, Adjusted R-squared: 0.121				
F-statistic: 125.4 on 87 and 7852 DF, p-value: < 2.2e-16				

M2 (Python-Output)		
Model: "M2"		
Layer (type)	Output Shape	Param #
dense_562 (Dense)	(None, 87)	7656
dense_563 (Dense)	(None, 87)	7656
dense_564 (Dense)	(None, 16)	1408
dense_565 (Dense)	(None, 1)	17
Total params	16,737	
Trainable params	16,737	
Non-trainable params	0	

M3 (Python-Output)		
Model: "M3"		
Layer (type)	Output Shape	Param #
dense_558 (Dense)	(None, 58)	5104
dense_559 (Dense)	(None, 58)	3422
dense_560 (Dense)	(None, 16)	944
dense_561 (Dense)	(None, 1)	17
Total params	9,487	
Trainable params	9,487	
Non-trainable params	0	

M4 (Python-Output)		
Model: "M4"		
Layer (type)	Output Shape	Param #
dense_498 (Dense)	(None, 58)	5104
dense_499 (Dense)	(None, 512)	30208
dense_500 (Dense)	(None, 8)	4104
dense_501 (Dense)	(None, 1)	9
Total params	39,425	
Trainable params	39,425	
Non-trainable params	0	

M5 (R-Output)					
Call:					
lm(formula = amountpaidoncontentsclaim ~ ., data = train2)					
Residuals:					
Min	1Q	Median	3Q	Max	
-0.16819	-0.01652	-0.00792	0.00563	0.9654	

Coefficients				
	Estimate	Std. Error	t-value	Pr(>|t|)
(Intercept)	-5.62E-03	1.10E-02	-0.509	0.610711
crsdiscount	1.98E-02	2.16E-03	9.167	< 2e-16 ***
Elevationdifference	-2.22E-03	6.75E-04	-3.286	0.001017 **
prcp	1.39E-02	7.63E-04	18.276	< 2e-16 ***
prcp_lag1	3.14E-02	1.12E-03	28.128	< 2e-16 ***
snow	-1.48E-02	7.50E-03	-1.969	0.048957 *
snow_lag1	-1.26E-02	1.43E-02	-0.884	0.376865
basementenclosurecrawlspacetype_1	-9.33E-03	5.69E-04	-16.384	< 2e-16 ***
basementenclosurecrawlspacetype_2	-7.06E-03	5.49E-04	-12.866	< 2e-16 ***
basementenclosurecrawlspacetype_3	-3.63E-03	7.35E-04	-4.938	7.91e-07***
basementenclosurecrawlspacetype_4	-4.80E-03	1.30E-03	-3.700	0.000216 ***
condominiumindicator_H	1.56E-02	1.16E-02	1.354	0.17575
condominiumindicator_L	2.28E-02	1.22E-02	1.867	0.061943
condominiumindicator_N	1.56E-02	1.10E-02	1.419	0.155848
condominiumindicator_U	1.22E-02	1.11E-02	1.093	0.274397
elevatedbuildingindicator_Y	-7.43E-03	4.27E-04	-17.393	< 2e-16 ***
floodzone_A00	-6.26E-03	5.14E-03	-1.218	0.223229
floodzone_A01	5.73E-04	9.32E-04	0.615	0.5388
floodzone_A02	5.84E-03	1.39E-03	4.190	2.79e-05 ***
floodzone_A03	1.24E-02	1.15E-03	10.781	< 2e-16 ***
floodzone_A04	7.92E-03	1.06E-03	7.447	9.67e-14 ***
floodzone_A05	3.82E-03	1.02E-03	3.748	0.000178 ***
floodzone_A06	2.91E-03	9.45E-04	3.075	0.002105 **
floodzone_A07	6.77E-04	1.20E-03	0.563	0.573491
floodzone_A08	3.10E-03	1.03E-03	3.013	0.002590 **
floodzone_A09	4.98E-03	1.43E-03	3.489	0.000485 ***
floodzone_A0B	-2.13E-02	6.81E-03	-3.119	0.001815 **
floodzone_A10	8.16E-03	1.28E-03	6.383	1.74e-10 ***
floodzone_A11	6.47E-03	1.68E-03	3.866	0.000111 ***
floodzone_A12	1.31E-03	2.39E-03	0.545	0.585412
floodzone_A13	4.20E-03	1.56E-03	2.703	0.006878 **
floodzone_A14	4.26E-03	1.62E-03	2.635	0.008404 **
floodzone_A15	6.91E-03	1.74E-03	3.971	7.16e-05 ***
floodzone_A16	7.91E-03	2.75E-03	2.881	0.003971 **
floodzone_A17	1.72E-03	2.79E-03	0.615	0.538641
floodzone_A18	-5.07E-03	3.22E-03	-1.575	0.115208
floodzone_A19	-1.59E-02	9.77E-03	-1.624	0.104391
floodzone_A20	-3.94E-03	4.22E-03	-0.933	0.351052
floodzone_A21	2.62E-04	6.48E-03	0.040	0.96776
floodzone_A22	5.76E-02	9.14E-03	6.305	2.90e-10 ***
floodzone_A23	1.31E-01	2.11E-02	6.226	4.82e-10 ***
floodzone_A25	-2.94E-05	2.11E-02	-0.001	0.998888
floodzone_A30	-3.73E-03	1.82E-02	-0.204	0.838137

floodzone_A4	-1.02E-02	3.65E-02	-0.28	0.779641
floodzone_A99	-6.80E-03	2.22E-03	-3.063	0.002192 **
floodzone_AA	-1.05E-02	3.65E-02	-0.289	0.772935
floodzone_AE	4.15E-03	6.15E-04	6.757	1.42e-11 ***
floodzone_AH	1.61E-03	3.21E-03	0.501	0.616121
floodzone_AHB	-5.84E-03	1.30E-03	-4.502	6.74e-06 ***
floodzone_AO	3.17E-03	1.87E-03	1.693	0.090445
floodzone_AOB	7.74E-03	2.52E-03	3.067	0.002161 **
floodzone_AR	-1.94E-02	3.65E-02	-0.531	0.59517
floodzone_B	1.15E-02	7.72E-04	14.935	< 2e-16 ***
floodzone_C	1.02E-03	7.56E-04	1.353	0.175923
floodzone_D	-2.02E-04	4.43E-03	-0.046	0.963639
floodzone_V	-4.98E-03	7.63E-03	-0.653	0.513937
floodzone_V02	-3.41E-03	1.10E-02	-0.309	0.757159
floodzone_V03	-8.86E-03	1.49E-02	-0.594	0.552296
floodzone_V04	-1.15E-02	1.01E-02	-1.132	0.25751
floodzone_V05	3.25E-03	5.30E-03	0.614	0.539213
floodzone_V06	-4.36E-03	5.30E-03	-0.823	0.410612
floodzone_V07	-2.42E-03	1.10E-02	-0.22	0.825966
floodzone_V08	-3.83E-03	4.28E-03	-0.894	0.371462
floodzone_V09	-7.83E-03	5.10E-03	-1.535	0.124708
floodzone_V10	-9.45E-05	4.31E-03	-0.022	0.982492
floodzone_V11	-1.05E-02	5.47E-03	-1.927	0.053996
floodzone_V12	-4.22E-03	3.32E-03	-1.270	0.204189
floodzone_V13	-2.90E-03	4.12E-03	-0.704	0.481267
floodzone_V14	-1.27E-02	5.14E-03	-2.472	0.013439 *
floodzone_V15	4.10E-03	6.28E-03	0.653	0.514076
floodzone_V16	-1.19E-04	5.09E-03	-0.023	0.981421
floodzone_V17	-2.10E-04	1.10E-02	-0.019	0.984814
floodzone_V18	3.41E-02	1.49E-02	2.290	0.022029 *
floodzone_V19	1.64E-02	6.38E-03	2.564	0.010360 *
floodzone_V20	1.71E-03	8.62E-03	0.199	0.842519
floodzone_V21	4.09E-03	7.98E-03	0.512	0.608892
floodzone_VE	-1.53E-03	1.63E-03	-0.939	0.347672
floodzone_X	7.06E-03	6.37E-04	11.075	< 2e-16 ***
numberoffloorsintheinsuredbuilding_2	2.47E-03	3.36E-04	7.349	2.01e-13 ***
numberoffloorsintheinsuredbuilding_3	-4.39E-04	5.10E-04	-0.860	0.389911
numberoffloorsintheinsuredbuilding_4	1.61E-03	1.08E-03	1.497	0.134349
numberoffloorsintheinsuredbuilding_5	-8.98E-03	1.34E-03	-6.725	1.77e-11 ***
numberoffloorsintheinsuredbuilding_6	-4.62E-03	1.02E-02	-0.455	0.64927
occupancytype_2	-6.84E-03	6.41E-04	-10.67	< 2e-16 ***
occupancytype_3	-1.15E-02	1.36E-03	-8.445	< 2e-16 ***
occupancytype_4	2.53E-02	6.40E-04	39.533	< 2e-16 ***
occupancytype_6	3.39E-02	1.63E-03	20.809	< 2e-16 ***
postfirmconstructionindicator_Y	4.98E-03	3.63E-04	13.742	< 2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 0.03647 on 78526 degrees of freedom
Multiple R-squared: 0.08632, Adjusted R-squared: 0.0853
F-statistic: 85.27 on 87 and 78526 DF, p-value: < 2.2e-16

M6 (Python-Output)		
Model: "M6"		
Layer (type)	Output Shape	Param #
dense_566 (Dense)	(None, 87)	7656
dense_567 (Dense)	(None, 87)	7656
dense_568 (Dense)	(None, 8)	704
dense_569 (Dense)	(None, 1)	9
Total params	16,025	
Trainable params	16,025	
Non-trainable params	0	

M7 (Python-Output)		
Model: "M7"		
Layer (type)	Output Shape	Param #
dense_597 (Dense)	(None, 87)	7656
dense_598 (Dense)	(None, 512)	45056
dense_599 (Dense)	(None, 64)	32832
dense_600 (Dense)	(None, 4)	260
dense_601 (Dense)	(None, 1)	5
Total params	85,809	
Trainable params	85,809	
Non-trainable params	0	

M8 (Python-Output)		
Model: "M8"		
Layer (type)	Output Shape	Param #
dense_702 (Dense)	(None, 87)	7656
dense_703 (Dense)	(None, 128)	11264
dense_704 (Dense)	(None, 128)	16512
dense_705 (Dense)	(None, 8)	1032
dense_706 (Dense)	(None, 1)	9
Total params	36,473	
Trainable params	36,473	
Non-trainable params	0	