Can Weather Forecasts Predict Norwegian Home Insurance Claims?

Mahsa Gorji and Sondre Lykke Rødal

Supervisors: Håkon Otneim, Sondre Nedreås Hølleland, and Etienne Dunn-Sigouin

Master thesis, MSc in Economics and Business Administration, Business Analytics profile

NORWEGIAN SCHOOL OF ECONOMICS

Abstract

Insurance companies are urgently seeking robust solutions to minimize the risks associated with property damage in response to more frequent extreme weather and increased regulation. This need is particularly significant in Norway, where extreme precipitation is common. Therefore, short-term prediction of future claims would be of great value for proactive decision-making and damage prevention. However, previous research focuses mainly on linking past insurance claims to past precipitation or on longer-term changes in insurance claims over the next few decades. Here, we develop a model that predicts the number of short-term Norwegian water damage property claims using precipitation forecasts from the European Center for Medium Range Weather Forecasting (ECMWF) from 2014-2021. The insurance dataset is unique in that it includes private claims data from a large insurance provider, Tryg Forsikring, where each claim is categorized as either a Natural Perils claim or not.The results show that precipitation forecasts can be used to predict the number of insurance claims within a few days. Although the signal is weak, with an AUC score of 0.65, it is better than chance alone. Specifically, the predictive skill is not very sensitive to details of the predictors derived from the weather forecast, and simpler logistic regression models perform as well as more complex machine learning models such as XGBoost and Neural Networks. The skill of the claims model can be largely explained by the skill of the precipitation forecast. Overall, our results suggest that using weather forecasts to predict insurance damage claims is possible, and demonstrate the potential to be used for operational decision making, such as in an early-warning system.

Acknowledgement

First and foremost, we would like to express our gratitude to our three supervisors: Etienne Dunn-Sigouin, a senior researcher at NORCE Norwegian Research Centre; Håkon Otneim, an associate professor at NHH; and Sondre Nedreas Hølleland, an assistant professor at NHH. They have provided invaluable support and guidance throughout this project. We also extend our thanks to NHH, Tryg Forsikring, NORCE, and the Climate Futures project, whose collaboration enabled us to complete this project.

Contents

List of Tables

1 Introduction

As observed globally, climate change is concomitant with escalating risks of weather-related damage, affecting virtually all sectors of the economy, ranging from fisheries and agriculture to power generation and tourism (Lyubchich & Gel, 2017). Due to the increased frequency and severity of extreme weather events, this phenomenon presents serious challenges for the insurance industry, affecting both the availability and affordability of coverage. This phenomenon increases the frequency and severity of weather events, and as the Norwegian Environment Agency, Miljødirektoratet, suggests in their 2015 report, it will lead to more severe storms, extreme precipitation, and floods, which in turn will lead to heightened risks for insurers and difficulties in risk assessment (Mills, 2005). In 1992, southeastern Florida experienced the devastating impact of Hurricane Andrew, classified as a Category 5 storm. The insured damage amounted to $22.3 billion in 2005 U.S. dollars (Swiss Re, 2006). As a result of these significant losses, nine insurance companies went bankrupt. Their underwriting practices had relied on past claim experiences, spanning approximately 30 years, which had not encountered a catastrophic event like Hurricane Andrew (Grossi & Kunreuther, 2005). Additionally, climate change can result in correlated losses, where a single event causes widespread damage, leading to issues such as the withdrawal of private insurers from flood coverage and the creation of the National Flood Insurance Program (Herweijer et al, 2009).

This paper seeks to address the challenges insurance companies face due to the escalating impacts of climate change. Specifically, our research focuses on predicting house insurance claims using forecasted precipitation data. We aim to determine whether precipitation forecasts can predict home insurance claims in Norway and to quantify the risk of observing these claims in the short-term future. This endeavor aligns with the EU's objective of promoting resilience to climate change and sustainable economic development. A fundamental pillar of the EU taxonomy is damage prevention, which highlights the importance of initiatives aimed at mitigating the adverse effects of climate-related risks. However, the insurance sector faces the challenge of adapting to the impacts of climate change, necessitating the development of new methods and tools. Predicting claims in advance would enable insurance companies to proactively mitigate both physical and economic damage by implementing timely interventions. For instance, an early warning system could guide households to take preventive measures, while efficiently allocating resources to enhance resilience to climate-related risks.

Weather forecasts have been effectively applied to predict various weather-related insurance claims and improve disaster preparedness. Röösli et al. (2021) develop an open-source impact forecasting system that integrates numerical ensemble weather predictions with exposure and vulnerability data to forecast building damage from winter windstorms in Switzerland, with a 2-day lead time. The system's forecasts were generally accurate for winter windstorms but showed higher rates of missed events and false alarms for thunderstorms and foehn storms. In another study, Pappenberger et al. (2015) demonstrates that integrating hydrological forecasts with socio-economic data significantly enhances the accuracy of flood impact predictions, thereby improving the effectiveness of early warning systems and decision-making processes during emergencies. Similarly, Komatsu et al. (2012) highlight that impact-based typhoon warnings in Japan, which include specific damage predictions, result in significantly better public preparedness and response compared to traditional hazard-only warnings.

However, while weather forecasts have been utilized in other domains to predict socially significant outcomes, the field of predicting precipitation-related insurance claims has seen limited development. Previous literature links observed precipitation with observed insurance claims, with Haug et al. (2011) among the pioneers in modeling the dynamics of weather-related residential claims. They employ generalized linear models (GLMs) incorporating exogenous weather variables. Scheel et al. (2013) develop a Bayesian hierarchical method for classifying hazard-prone areas. They adopt a leave-one-out prediction strategy to estimate the number of claims in 2001. However, a critical issue remains unresolved: predicting the dynamics of future weather-related claims. Cheng et al. (2012) also examine a monthly rainfall index derived from hourly downscaled data. They establish a crucial threshold for this index, identifying the point at which a high volume of insurance claims is triggered. The works of a Dutch research group (Spekkers et al, 2015; Spekkers et al, 2013) delve into damages related to rainfall on private property and content in the Netherlands. Their approach involves modeling both claim frequency and claim sizes, utilizing four categories of predictors (rainfall-related, socio-economic, building related, and topographic) through GLMs and decision trees. Similar to Cheng et al. (2012), Spekkers et al. (2013) select thresholds for precipitation intensity but refrain from making forecasts about future insurance risks. Also, Rohrbeck et al. (2018) apply mixture and extremal mixture modeling with generalized Pareto distributions, temporal clustering algorithms, and novel covariates to effectively capture local dynamics and spatial dependence in multiple Norwegian cities. Their model accurately predicts the number of insurance claims over clustered periods, handling both low and high claim counts. While previous literature extensively links observed precipitation with insurance claims, this cannot be used directly to prevent potential damage in this area. Therefore, we aim to investigate the potential of utilizing precipitation forecasts for predicting insurance claims to address this gap. Specifically, we apply state-of-the-art weather forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF) to predict house insurance claims.

There are a few research studies employing machine learning methods to link claims and precipitation, and most of the studies rely on traditional statistical techniques which might not be capable enough to capture the complex nonlinear relationships inherent in insurance claims data. In property and casualty insurance, claim distributions often skew to the right with a concentration at zero for cases where claims do not occur. Traditional GLMs based on Poisson or negative binomial distributions struggle with handling excess zeros in insurance claim data. To address this, the industry has turned to zero-inflated or hurdle models, such as zero-inflated Poisson or negative binomial regression, as evidenced by Yip and Yau (2005), Mouatassim and Ezzahid (2012), and Chen et al. (2019). Even though the GLM has found extensive use in actuarial research, as demonstrated by Ayuso et al. (2019) and Lemaire et al. (2016), its limitation lies in the rigid linearity of the logarithmic mean structure. This constraint may prove overly inflexible for specific applications, especially in cases where nonlinearity is present in claim data. Though GLMs and Generalized Additive Models (GAMs) are potent tools, they frequently encounter challenges when attempting to discern intricate interactions within a multitude of closely intertwined risk features (So, 2023).

Research thus turned to machine learning techniques. Among these few research endeavors, Lyubchich et al. (2019) employ two methods to determine thresholds for observed insurance claim counts concerning precipitation and wind speed: Classification and Regression Trees, and Alternating Conditional Expectations transformations. However, considering the capabilities of machine learning methods to study the non-linear relationship in this area, our study employs such advanced methods to capture the intricate nonlinear patterns present in insurance claims data. Furthermore, recent advancements in nonlinear attribution analysis, as demonstrated by Lyubchich and Gel (2017), introduces innovative methods to identify triggering thresholds for increased claims. By applying this novel approach to flood-related home insurance claims in Norway, researchers have gained deeper insights into the factors influencing claim occurrences, thus enhancing risk management strategies. Early warning systems can contribute to the prediction of insurance claims by providing valuable data and insights, potentially enhancing the application of simple statistical models. While statistical models serve as the engine that generates the actual predictions, early warning systems enhance their utility and can potentially prevent property damage. These systems include key elements: risk awareness, monitoring and warning, communication of warnings, and response capability. By integrating statistical model predictions, they enhance risk awareness and offer continuous monitoring to issue timely warnings. These warnings are then communicated to relevant stakeholders, ensuring effective information sharing and enabling proactive responses. This approach ensures that predictions not only inform stakeholders but also drive actions to mitigate risks. Recent research underscores the importance of these systems in managing extreme weather events and highlights their essential components (Coughlan de Perez et al., 2022). This integration makes the statistical models' output actionable, ultimately reducing the impact of potential claims through preemptive measures.

Our research seeks to answer two main questions: First, can precipitation forecasts predict insurance claims? Second, can more advanced machine learning methods predict insurance claims better than traditional methods?

To determine whether precipitation forecasts can predict insurance claims, we evaluate the model performance with the AUC (Area Under the Curve). The results show that for both Bergen and Oslo, the AUC is above 0.6 for precipitation forecasts, similar to using the observed precipitation as a feature. This exemplifies that the precipitation forecast is fairly good at predicting the correct precipitation levels, and a score above 0.6 indicates some ability to discriminate between 0, 1 and more than 1 claims. This indicates that companies can use precipitation forecasts to make proactive decisions. Furthermore, to evaluate whether more advanced machine learning models perform better than statistical models, we employ multinomial logistic regression as the baseline traditional method, and XGBoost and Neural Networks as more advanced models. However, we are unable to conclude that these advanced models perform better than the traditional logistic model. Additionally, we apply our findings to an early warning system designed for insurance companies to use in their day-to-day operations for risk mitigation and damage prevention. We believe such a model shows potential for further development and use in an early-warning system.

2 Data

Our data consists of home insurance claims, observed precipitation, and forecasted precipitation. The insurance claims dataset is provided by Tryg Forsikring, a Danish insurance group, which is the fourth-largest insurance company in Norway. The observed precipitation data is from ERA5, a comprehensive reanalysis dataset produced by ECMWF, the European Centre for Medium-Range Weather Forecasts (2020), while the precipitation forecast is obtained from ECMWF's operational archive (European Centre for Medium-Range Weather Forecasts, 2024). ECMWF is a research institute and operational service, possessing one of the largest supercomputers and weather data archives in the world, and produces state-of-the-art weather forecasts. All datasets cover the years 2014 through 2021 for both Bergen and Oslo areas. The insurance data is limited to the municipality borders, while both types of the precipitation data are extracted using latitude and longitude coordinates. These coordinates are specified to cover a geographic area with a resolution of 9 by 9 km when measured at the equator, centered over Bergen and Oslo municipalities in order to overlap with the insurance data. This approach ensures that we include precipitation that will land inside the municipalities and surrounding areas.

The home insurance claims data includes the daily aggregated number of house insurance claims. All claims are categorized into one of two categories: rain-associated claims or natural-perils claims. Rain-associated claims refer to those caused by water entering a property through leaks, cracks, or other unintended means, essentially covering all water-damage claims not related to a natural peril. Natural perils claims are those caused by extreme weather events, including weather variables other than precipitation. Natural perils claims encompass damages caused by extreme weather events and other weather-related variables beyond precipitation. The Norwegian Natural Perils Pool recognizes earthquakes, floods, landslides, storms, volcanic eruptions, and, starting in 2023, tsunamis and meteorite strikes as natural perils (Norsk Naturskadepool, 2024a). The insurance pool is a fund created to cover financial losses from these extreme weather-related home insurance claims. Each insurance company in Norway contributes to this fund based on their market share, measured by the number of customers. All companies that offer fire insurance in Norway must be a member of the insurance pool in Norway. The insurance companies collect a premium from their customers, which is then added to the pool. When a natural disaster occurs, these companies are reimbursed from the pool according to their market share, as long as there are sufficient funds to cover the losses. However, starting in 2025, insurance companies in Norway will only get coverage in the years where natural disaster compensation has exceeded contributions (Norsk Naturskadepool, 2024b). This mandatory collective arrangement was established by a law in 1980, and it has a coverage limit of NOK 16 billion - yet, the largest payout to date was NOK 1.3 billion, which occurred in the early 1990s (Norsk Naturskadepool, 2024c). Consequently, the financial impact of whether an insurance claim is categorized as a natural peril or not significantly affects insurance companies' bottom line, underscoring the importance of identifying the specific claim type when predicting insurance claims. For instance, if there is an increased risk of observing many claims in the short-term future, and most of them are rain-associated rather than natural-perils claims, it will result in a larger financial loss. This is because, as the insurance company will have to cover the full financial losses related to those claims themselves.

In Figure 2.1, we display the time series of the daily number of insurance claims, with both types of claims aggregated, in Bergen and Oslo municipalities from 2014 through 2021. Predicting and understanding this data is particularly challenging due to its complex trends and distribution, marked by extreme imbalance and a high proportion of zeros. In our dataset, over 75% of daily observations in both Oslo and Bergen recorded zero claims. Furthermore, both municipalities average roughly 3 claims per week, but as we can see in this figure, certain observations stand out immediately.

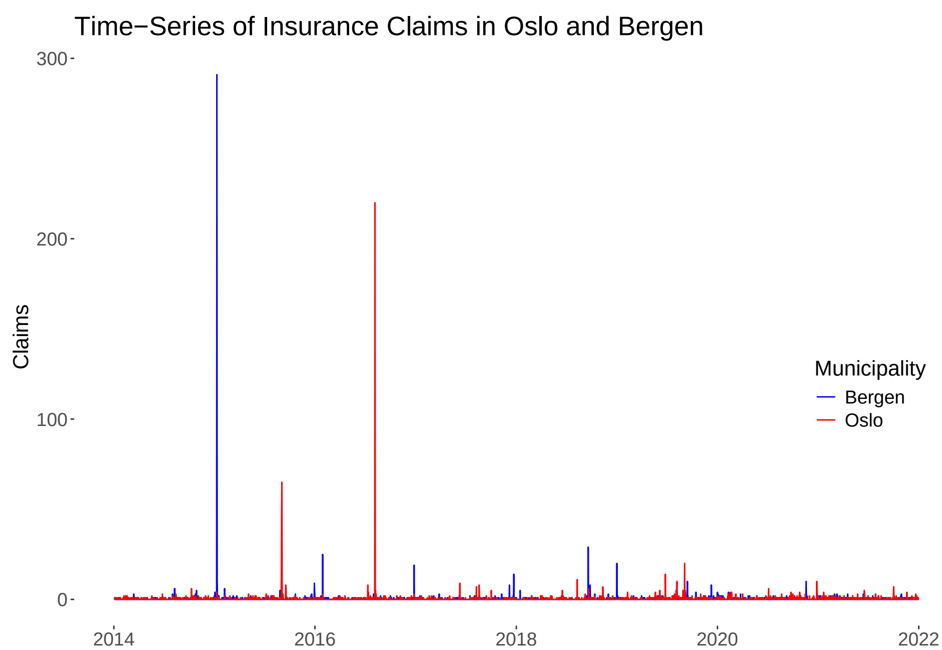

Figure 2.1: Time-series of daily aggregated insurance claims in Bergen and Oslo municipalities from 2014 through 2021

In Bergen, there are a total of 8 observations where the daily aggregate number of insurance claims reached 10 or more; all but one of these are below 30. Interestingly, the single outlier of 291 insurance claims occurred on January 10, 2015, coinciding with the peak of the extreme weather event ”Nina” (NRK, 2015). All of these claims were categorized as natural-perils claims. “Nina” caused massive material damages, as wind speeds reached over 160 km/h in Bergen according to TV 2 (2015). However, the precipitation amount was not particularly high, at 5 mm. The second highest number of insurance claims occurred on September 19, 2019, when the local newspaper in Bergen reported that a flood caused chaos in the city, which struggled with abnormal levels of water and concentrated claims (Bergens Tidende, 2016a). This event resulted in 29 claims, all but one categorized as natural-perils claims. Again, the precipitation amount was not extreme, at 9.5 mm. The hurricane “Tor” is another example of this, with a combination of 70 mm of precipitation and wind speeds of 55 m/s, which on January 29 resulted in 25 claims, all categorized as natural-perils claims (Bergens Tidende, 2016b). As a coastal city, Bergen is prone to the impact of strong winds that carry seawater, exacerbating potential damage. Also, it seems that a combination of various weather variables, specifically wind and precipitation at extreme levels, affects this area.

On the other hand, in Oslo there are also 8 observations of 10 or more claims. One notable instance involves 220 claims coinciding with extreme rainfall reported in Asker on August 6, 2016. According to the Norwegian University of Life Sciences (2017), NMBU, this event was classified as a ’200-year rain,’ with 101.9 mm of rainfall recorded within a 6-hour period in the morning. However, only 36 out of the 220 claims that day were categorized as natural perils claims. Additionally, during the three-day period from September 3 to 5, 2015, a total of 125 insurance claims were recorded. This month was marked by reported flooding in the southeastern areas of Norway. According to the Norwegian Water Resources and Energy Directorate (NVE), the heavy rainfall in September 2015 in southeastern Norway caused certain river basins to flood, exceeding the levels associated with a over 50-year recurrence interval (Norwegian Water Resources and Energy Directorate, 2016). Out of the 125 claims during that period, 12 were categorized as natural perils claims. Despite these two extreme events in Oslo, most of the recorded claims are not categorized as natural peril claims. Of all the observed insurance claims in the dataset, 55% of Bergen’s claims are categorized as natural perils claims, while for Oslo it is less than 14%. Yet, Oslo experiences more insurance claims on average.

This analysis reveals two main insights. First, during extreme weather events, most of the increased number of claims in Bergen are categorized as natural perils claims. In contrast, in Oslo, the majority of claims during such events are categorized as rain-associated. Second, in Bergen, on days with extreme weather, the main driver behind those insurance claims seems to be a combination of several extreme weather variables, typically high wind speeds and some precipitation, but not necessarily extreme precipitation.

To understand the relationship between insurance claims and observed precipitation, we visualize the monthly averages for both variables in Bergen and Oslo. Figure 2.2 shows that in Bergen, the monthly average precipitation begins to rise in June and continues to increase until January. Notably, January also marks a sudden spike in the average monthly observed insurance claims. In Oslo, the precipitation accelerates rapidly starting in April, followed by a sudden spike in average claims in August. The pattern in these two cities suggests that a period of increased precipitation leads to an increase in observed insurance claims. While this relationship has been well-documented in the existing literature, such as Haug et al. (2011) and Scheel et al. (2013), it remains valuable to verify its validity within the context of our data set.

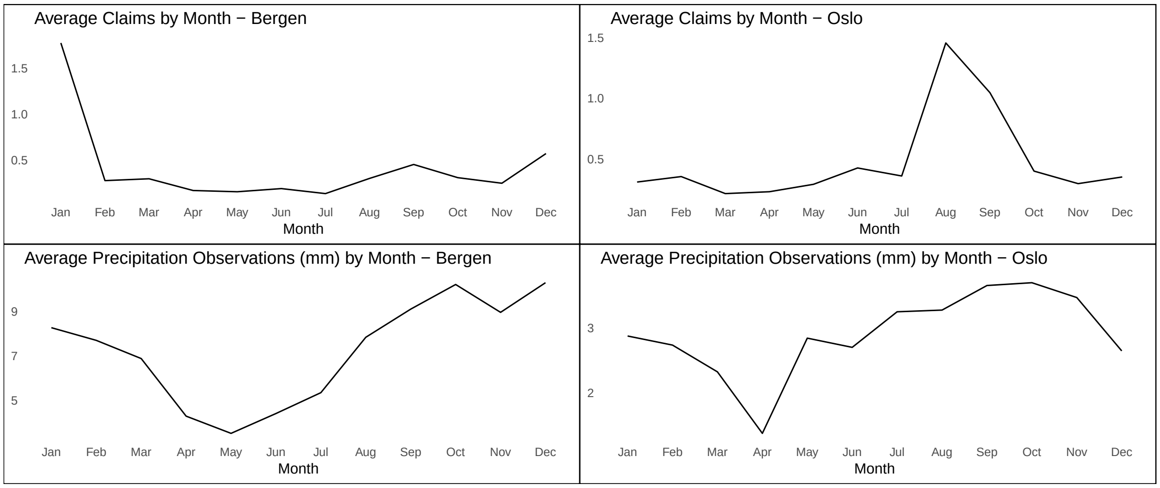

Figure 2.2: Monthly averages for claims and observed precipitation, for Bergen and Oslo (2014 - 2021)

Although there are clearly seasonal trends for both insurance claims and precipitation, the relationship between them is more complex than depicted in Figure 2.2. As detailed in the examples above and illustrated in Figure 2.3, extreme events with a large number of insurance claims do not necessarily correspond to high precipitation levels in Bergen. Precipitation can vary in intensity, duration, and type which affects insurance claims differently. This non-linear relationship motivates the use of more advanced machine learning models when utilizing precipitation forecasts to predict insurance claims. As Rohrbeck (2018) emphasizes, there are also regional differences in the relationship between insurance claims and precipitation. Figure 2.3 conveys this message; for instance, the extreme value of 220 claims in Oslo corresponds to a low precipitation value. However, as described above regarding this outlier in Oslo on August 6th, 2016, the extreme rainfall that day was observed in Asker, a nearby municipality. This discrepancy might be due to our precipitation data being sourced through latitude and longitude coordinates that cover a slightly larger area, which may not accurately reflect localized variations. However, it also emphasizes the complexity and dimensions that affect the relationship between insurance claims and precipitation.

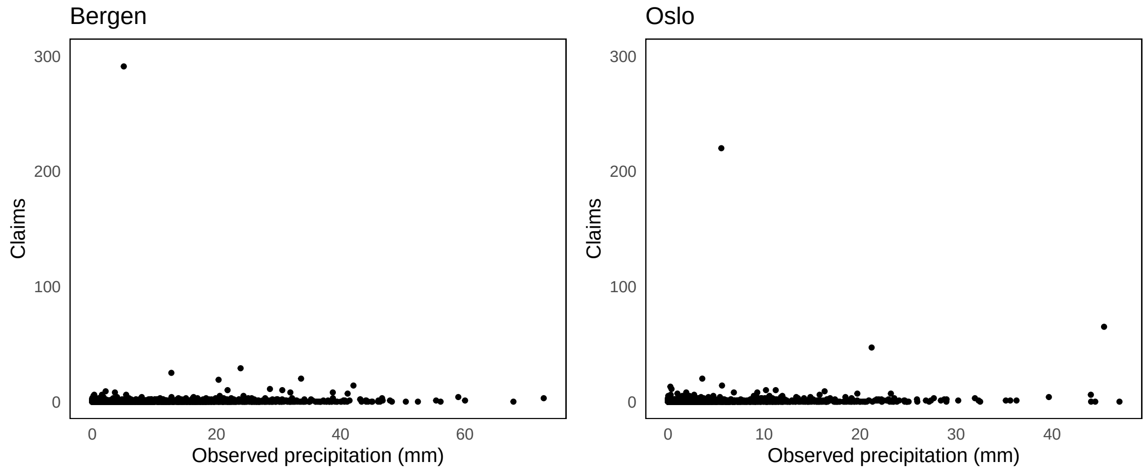

Figure 2.3: Scatterplot for claims and observed precipitation, for Bergen and Oslo (2014 - 2021)

While the relationship between observed precipitation and insurance claims is important to understand, in our project, we add another layer of complexity by using precipitation forecasts to predict claims. Thus, we must also understand the relationship between the forecasted precipitation and observed precipitaiton. In both of the precipitation datasets, the main variable precipitation refers to the daily accumulated precipitation in millimeters, which would also include melted snow and ice in addition to rain.

The forecast data includes three additional dimensions: initialization date, lead time, and ensemble member. The initialization date refers to the date on which the forecast was initialized, while lead time is the period between the initialization of the forecast and the actual forecasted date, ranging from 1 to 15 days ahead. ECMWF initializes a forecast every day. However, in our project, we use a dataset consisting of twice-weekly forecasts, specifically on Mondays and Thursdays. This approach is mainly due to the time cost associated with downloading the dataset from the ECMWF archive, given its 5-dimensional nature. Moreover, since not all forecasting centers provide daily forecasts, it is realistic for an insurance company to rely on forecasts issued twice a week. Even if more frequent forecasts were obtainable, an insurance company would still need to perform a cost-benefit analysis to determine if the added benefit of increased forecast frequency justifies the additional costs. Also we have ensemble members. Each forecast generates 51 unique simulations for the precipitation on a given day, and these simulations are referred to as ensemble members. Among these 51 ensemble members, initial conditions have been slightly adjusted to introduce variability and evaluate forecast sensitivity. This method addresses the inherent unpredictability in weather forecasting, often referred to as the "butterfly effect", where small changes in initial conditions can lead to vastly different outcomes later on (Lorenz, 1969). The variation among these ensemble members reflects the forecast's uncertainty, where a larger spread indicates greater uncertainty and a smaller spread suggests higher confidence. Figure 2.4 visualizes these three dimensions by displaying an example of a forecast made on January 16, 2020, for Bergen, showing all 15 lead times until January 31. The black solid line shows the observed precipitation, the red dashed line shows the average forecast, and all the blue solid lines show all 51 ensemble members. As shown in the plot below, weather forecasts tend to perform better in the first few days but become more uncertain over time. This increasing uncertainty is indicated by the widening spread of the blue lines and the decreasing correlation between the red line and the black line.

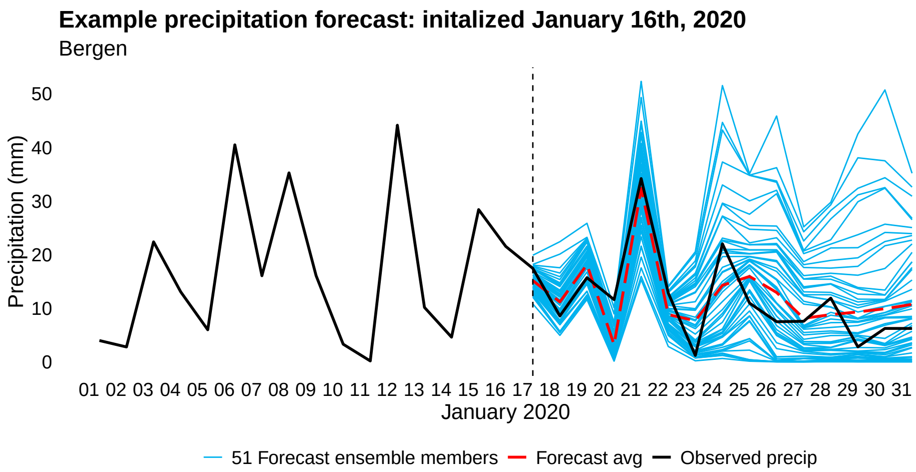

Figure 2.4: A precipitation forecast by ECMWF made on January 16th 2020 in Bergen. The black solid line shows the observed precipitation, all the blue solid lines show the 51 ensemble members, and the red dashed line shows the average of all those ensemble members. For lead time 1 through 15.

In Figure 2.5 we can see the correlation between forecasted and observed precipitation at various lead times for both Bergen and Oslo. Here we show the average of the 51 ensemble members. According to Figure 2.5, the correlation between observed precipitation and forecasted precipitation exhibits a notable trend, with large positive correlations at short lead times and smaller correlations at longer lead times. At shorter lead times, forecast models benefit from the relative proximity of their initial conditions to observed atmospheric states, allowing them to closely mirror observed precipitation trends. However, as lead times lengthen, the cumulative effects of small initial discrepancies magnify, leading to greater divergence between forecasted and observed precipitation which can indicate the butterfly effect. This is also evident in Figure 2.4, as the range between the maximum and minimum ensemble members increases as the time between initialization and prediction date increases, i.e., lead time. Our data clearly include complex relationships between both observed precipitation and claims, and forecasted and observed precipitation. In the next section we outline our methodology and feature engineering.

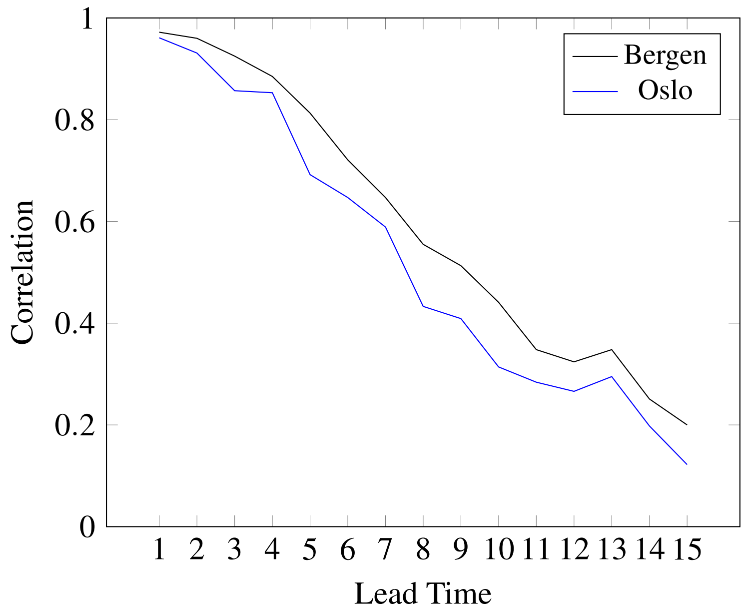

Figure 2.5: Correlation between observed and forecasted precipitation at various lead times, taking the average of all 51 ensemble members.

3 Methodology

To implement our model, we need to determine certain parameters and we begin with the outcome variable, insurance claims. A significant challenge in predicting insurance claims is the highly imbalanced and zero-inflated nature of the data, with over 75% of all observations, or days, showing 0 insurance claims, meaning the unconditional probability of observing zero claims is always the highest. Given this context, rather than focusing on predicting the exact number of insurance claims, our aim is to quantify the risk of observing insurance claims. To achieve this, we transform the insurance claims data into a categorical variable with three levels based on the number of observed claims including 0 claims, 1 claim, and more than 1 claim. Moreover, we must consider two types of claims classes: natural perils and rain-associated claims. In our modeling strategy, we examine the predictability of both types of claims separately and in aggregate.This is because, while we believe that the most critical information lies in predicting the aggregated number of claims, predicting the two types separately provides nuanced insights into the severity of claims. In terms of risk mitigation and damage prevention strategies, the category of claims decided in the aftermath does not matter to the household needing to take action now. However, it provides nuance into the financial impact on the insurance companies' bottom line, as claims losses related to natural perils will be covered by the natural perils pools.

Next, we need to determine the forecast time horizon for which we are interested in predicting insurance claims. Our general focus is on short-term predictions, where we have precipitation forecasts extending up to 15 days into the future. However, as illustrated by Figure 2.4 and 2.5, predictions beyond 7-8 days appear not to be relevant due to the low correlation between forecasted and observed precipitation. Given the twice-weekly forecasts on Mondays and Thursdays, we focus on the forecast with the shortest possible lead time for each claim. This approach ensures that we leverage the most reliable forecast intervals. This means the Monday forecast covers Tuesdays through Thursdays, and the Thursday forecast covers Friday through Monday. For instance, if we are interested in a prediction for Sunday, the forecast with the lowest lead time available would be from the preceding Thursday, a lead time of 3 days. Similarly, a prediction for Tuesday would occur with a lead time of 1 day, using the forecast initialized on Monday. We also make predictions for the aggregated lead times of 4 to 7 days. This involves summing claims and averaging the precipitation from days $t+1$ to $t+4$ and $t+1$ to $t+7$, respectively.

We apply three models to our data: a baseline model for comparison, which is the multinomial logistic regression, and two more advanced machine learning models, XGBoost and Neural Network, due to their capabilities to handle complex non-linear data. The logistic regression model estimates the probability of a particular outcome, such as the likelihood of an insurance claim falling into one of the three categories. The advantages of logistic regression include its simplicity, interpretability, and the fact that it requires fewer computational resources compared to more complex models. It also handles binary outcomes well and provides a clear coefficient for each predictor variable, indicating the impact of that variable on the outcome. However, logistic regression has its limitations. It assumes a linear relationship between the logit of the outcome and the predictor variables, which may not always hold true. Additionally, this model struggles with multicollinearity and is not suitable for complex and non-linear relationships, as evident in Figure 2.3. Consequently, this motivates us to apply more advanced methods as well as logistic regression.

XGBoost - or Extreme Gradient Boosting - is a machine learning algorithm that is often used for classification problems. It is a highly effective machine learning model due to its scalability, speed, and accuracy, making it particularly suitable for handling large and complex datasets. The XGBoost model structure, which combines weak learners to create a strong predictive model, has proven highly effective (Bentéjac et al., 2019). In Kaggle competitions, where data scientists and engineers compete to make the best predictive model, the model is often used. One of the notable features of XGBoost is its ability to provide variable importance measures, which can offer valuable insights into the most influential variables in the prediction model. This feature importance is calculated as the improvement in accuracy brought by a feature to the branches it is on, providing a clear indication of which variables are driving the model's predictions. We also use this as part of our feature engineering to find out which quantiles to extract from the precipitation forecast distribution. However, XGBoost models can be complex and lack the interpretability of simpler models like logistic regression. They can also be prone to overfitting, particularly if not properly tuned, and can be computationally intensive, requiring significant resources and time to be trained and optimized.

Neural networks also are a type of machine learning model inspired by the human brain. They consist of interconnected layers of nodes or ”neurons” that can learn to make predictions or decisions without being specifically programmed to perform the task. Neural networks are particularly effective at handling complex, non-linear relationships and large datasets. Shterev et al. (2022) demonstrate this capability, which indicates that neural networks are a powerful tool for predicting insurance claims. They can also handle a mix of numerical and categorical data and are capable of learning feature interactions. However, neural networks are often seen as a ”black box” due to their lack of interpretability. Understanding why a neural network has made a particular prediction can be challenging. They are also computationally intensive and require a significant amount of data to be trained effectively. Overfitting can also be a concern with neural networks if not properly regularized.

To evaluate our model, we use the AUC, which is a method for measuring how well a model can distinguish between different classification outcomes, in this case between the three classes of 0, 1, or more than 1 insurance claims. It represents the area under the Receiver Operating Characteristic

(ROC) curve, which plots the true positive rate against the false positive rate. This area quantifies the model's overall ability to correctly classify instances. It acts as a grading system for our model, where a score of 1 indicates perfect performance, while a score of 0.5 indicates performance no better than random guessing. Mandrekar (2010) finds that an AUC between 0.5 and 0.7 indicates some ability to predict with discriminating ability, meaning the model can separate between the classes. However, for the model to be considered excellent, the AUC must be above 0.8.

The AUC is particularly relevant when dealing with imbalanced data like insurance claims. This is because it is not misled by the overwhelming majority of one outcome, which is zero claims in our data as Bergen and Oslo have over 75% of daily observations with zero claims. Therefore, using accuracy as our performance metric would result in a 75% accuracy rate simply by always predicting zero claims. This would not be helpful for an insurance company, and the AUC solves the problem by offering a more standardized way of measuring the performance. While we use a classification model with three categorical outcomes, the AUC score is originally intended for binary classification problems. However, Hand et al (2001) describe a useful method to extend this measure to multi-class classifications. They propose treating each class as the positive binary class and all other classes as the negative binary outcome, then taking the average of those AUC scores. This approach effectively decomposes the multi-class problem into multiple binary classification problems, averaging the AUC scores across all classes to provide a comprehensive performance measure. In our modeling strategy, we evaluate various models by comparing their AUC scores. To achieve this, we utilize the standard binary AUC for the binary classification of insurance claims, distinguishing between claims greater than one or not. This approach allows us to test more models efficiently because calculating the multi-class AUC for each model requires more computational effort. We designate the "more than one claim" category as the positive class, as we are primarily interested in predicting that class. This method indicates which model and features may be the most useful in predicting insurance claims. However, in practice, one would use a multiclass classification model instead of the binary model, as we demonstrate in Section 5, possibly with even more classes than described here.

To enhance the reliability and robustness of the AUC score and address the imbalanced data, we employ cross-validation and resampling techniques. We use k-fold cross-validation with values ranging from 5 through 10, partitioning the dataset into k equally sized folds. This approach involves training the model on k-1 folds and validating it on the remaining fold, repeated k times to ensure each subset served as both training and validation data, thus providing a more reliable average AUC score. Additionally, we apply resampling methods such as Synthetic Minority Oversampling Technique and random sampling. These methods are intended to modify the dataset to ensure that minority and majority classes are more equally represented, which is necessary for our data due to its imbalance. However, neither of these techniques changed the AUC score significantly in either direction. This might imply that our data is prone to overfitting, or that we do not have a large enough dataset (Kuhn & Johnson, 2013).

The concept of feature engineering means different things to different data scientists. For some, it involves selecting the essential features required for building supervised models, such as those used to predict a specific response or outcome variable (Ozdemir, 2022). Therefore, in the context of predicting house insurance claims using precipitation forecasts, feature engineering refers to the process of selecting and creating the most relevant predictors from 51 ensemble forecasts to use them in machine learning models. This is one of the essential parts of the forecasting task to find relevant features which are informative for insurance claims. This process is complicated by the fact that the accuracy of the forecast can vary depending on how far into the future it predicts, lead time, and the time of year when the forecast is made. By extracting and analyzing these features we aim to improve the predictive power of our models. The process involves analyzing and extracting various summary statistics or quantiles from the precipitation forecast distribution, such as lagging and temporal features.

To extract the correct features we use two methods. First, due to the low computational power needed to run logistic models, we run 51 different logistic models, using each of the 51 quantiles of the precipitation forecast distribution, such that we can determine the impact on the AUC score from the various quintiles. Secondly, we fit an XGBoost model with all 51 ensemble members as 51 unique features, and then analyze the feature importance distribution, which essentially ranks each quantile according to its importance in predicting insurance claims. We further incorporate temporal and lagging features. Temporal features are characteristics of data that capture time-related aspects, such as patterns and trends over different time periods. Essentially, these are categorical variables. For instance, the temporal feature ”quarter” is a categorical variable with four levels representing the quarters of the year: 1 for January through March, 2 for April through June, 3 for July through September, and 4 for October through December. These features can help in identifying and modeling temporal dependencies and seasonality in the data, which appears to be present in our data. In our modeling, we incorporate temporal features by including week number, month, and quarter. On the other hand, lagging features are variables that use past values of a time series to predict future values. They capture temporal dependencies and consider persistence and delay effects in time series data. In our analysis, we incorporate lagging features for observed precipitation because rainfall from previous days can serve as a proxy for soil moisture levels. Wet soil has a reduced capacity to absorb additional moisture, making further rainfall likely to cause damage. We also include lagging features for insurance claims to account for the possibility that claims might not be recorded on the actual day of damage, despite efforts to ensure timely reporting. For both lagging variables, we utilize the observed values at $t-1$, meaning the day before the prediction day t. Additionally, we use three different time periods starting from $t-1$ and extending to either $t-2$,

$t-3$, or $t-4$. For the periods longer than a day, we used the aggregated sum of claims and the average precipitation, similar to our structure for future predictions of 4 or 7 days.

4 Results

Our predictive modeling yields three main results. The first major finding is that precipitation forecasts can predict insurance claims, but with a weak signal. This is indicated by the AUC score of 0.6, which Mandrekar (2010) describes as showing discriminating ability. In Table 4.1, we can see the AUC scores for our best model for both Bergen and Oslo, which is the logistic model for both municipalities. We also show the AUC score for the observed precipitation for comparison.The models have one feature, either a quantile from the forecast distribution or the actual observed precipitation value.

The AUC scores of all our logisitc models with various features, ranked by AUC score, are in Appendix B. When using forecasted precipitation as the feature to predict insurance claims, using the maximum ensemble forecast, or the 100th percentile, results in the highest AUC score for Bergen. In contrast, using the 90th percentile results in the highest AUC score for Oslo. Using these quantiles to compare against the model with observed precipitation resulted in a difference of only 0.01 in the AUC score for both municipalities. In fact, for Bergen, the AUC score is slightly higher using the forecasted precipitation compared to the observed precipitation (Table 4.1). It is also interesting to note that using forecasted precipitation instead of observed precipitation results in similar AUC scores, demonstrating that the precipitation forecast is reliable.

Bergen

Model	Feature(s)	AUC
Logistic	Observed precipitation	0.64
Logistic	Forecasted precipitation (100th percentile)	0.65

Oslo

Model	Feature(s)	AUC
Logistic	Observed precipitation	0.67
Logistic	Forecasted precipitation (90th percentile)	0.66

Table 4.1: AUC scores of the best model for Oslo and Bergen, compared to using observed precipitation. Predicting the aggreageted insurance claims observed at $t+1$.

Our second finding is that neither advanced machine learning models nor feature engineering appear to enhance the model's performance compared to using the traditional logistic model. In Table 4.2 below, we show the best models using XGBoost and Neural Networks for Bergen and Oslo. An interesting point to note is that using all 51 ensemble members as 51 unique features to predict insurance claims in the Neural Network for Bergen gives the same AUC score (Table 4.2) as when using the observed precipitation in the logistic model (Table 4.1).

Regarding our feature engineering, the XGBoost feature importance analysis clearly indicates that the extreme quantiles, both bottom and top, of the precipitation forecast distribution have a more substantial impact on the model's results compared to the middle quantiles (Appendix A). Utilizing specific quantiles such as the 6th percentile and 96th percentile results in the best XGBoost model for Bergen. Conversely, the best Neural Network model for Oslo includes the use of all bottom-10 and top-10 quantiles (Table 4.2). However, these models still achieved approximately the same AUC score as the logistic model using only one feature from the forecast distribution. Furthermore, the temporal and lagging features did not significantly impact the AUC score either, only by an increase of less than 0.02 when adding the quarter feature for Bergen. Detailed results from our models, the XGBoost feature importance and the AUC of 51 logistic models are presented in Appendix A and B.

Bergen

Model	Feature(s)	AUC
Neural Network	All 51 ensemble members	0.64
XGBoost	6th percentile and 96th percentile	0.63

Oslo

Model	Feature(s)	AUC
Neural Network	All 10 bottom quantiles, and all 10 top quantiles	0.65
XGBoost	All 51 ensemble members	0.63

Table 4.2: AUC scores of the best XGBoost and Neural network models for Oslo and Bergen, predicting the aggreageted insurance claims observed at $t+1$.

The third main results (Table 4.3) show that for Bergen, the best logistic model (Table 4.1) performs significantly better at predicting natural perils claims. These weather events are likely quite predictable using weather forecasts. Bergen experiences significant precipitation and extreme weather, characterized by strong seasonal patterns. Bergen also experiences a higher number of observed natural perils claims compared to Oslo, as over half of Bergen's claims are categorized as natural perils. While predicting both types of insurance claims together is crucial for risk mitigation, distinguishing between natural perils claims and rain-associated claims provides valuable information. This distinction allows the insurance company to predict potential financial losses more accurately, as only losses associated with natural perils claims are covered.

Claims type	**Bergen (AUC)**	**Oslo (AUC)**
Natural perils claims	0.79	0.63
Water induced	0.59	0.66
Both	0.65	0.66

Table 4.3: The different AUC scores when predicting either the natural perils claims, the water-induced claims, or both. Using the best logistic model with the forecasted precipitation as the feature, predicting the aggreageted insurance claims observed on $t+1$.

Lastly, we want to show the difference in AUC scores when using the best model (Table 4.1) to predict insurance claims over the three different time horizons (Table 4.4). There is not a clear difference in AUC between next day predictions and next 4-day predictions, but this is likely due to the fact that we only include two precipitation forecast initializations per week. As a result, we use a mix of lead times for the next day predictions depending on what day of the week it is. For instance, since the precipitation forecast is initialized each Monday and Thursday, the predictions for Tuesdays will use a lead time of 1 while the predictions for Wednesdays will use a lead time of 2. Including more frequent initialization forecasts will likely improve this, but it is unclear by how much.

Forecast horizon	**Bergen (AUC)**	**Oslo (AUC)**
$t+1$	0.65	0.67
$t+4$	0.66	0.63
$t+7$	0.58	0.62

Table 4.4: AUC scores using the best logistic model to predict either of the three time-horizons, for both Bergen and Oslo.

5 Operational example

In this section, we demonstrate how an insurance company might incorporate precipitation forecasts to predict insurance claims in their day-to-day operations and determine appropriate actions based on the models' outcomes. Specifically, we aim to show how the model could enhance risk mitigation and damage prevention strategies. We begin by describing the unconditional probabilities and their implications.

The unconditional probabilities, or baseline probabilities, reflect the chances of an outcome occurring if we had to make a prediction without any additional data, relying solely on the observed values of the outcome variable. We are interested in understanding the changes and their magnitudes when comparing the unconditional probabilities to the probabilities predicted by our model. We begin by defining the unconditional probabilities, which are simply the observed frequencies of each class divided by the total number of observations. Table 5.1 below summarizes these probabilities.

	Bergen		*Oslo*	
	Observations	%	Observations	%
0 claims	2351	80.5	2226	76.2
1 claim	449	15.4	520	17.8
>1 claim	120	4.1	174	6.0
Total	2920	100	2920	100

Table 5.1: The observed frequencies of each of the three classes of claims in both Bergen and Oslo from 2014 through 2021. The "%" column refers to the unconditional probabilities of observing that class of claim.

When we use the precipitation forecast into the predictive model, the output provides new probabilities for each class of claims: 0, 1, or more than 1 claim. A reasonable next step would be to classify the prediction into the class with the highest probability. However, this approach presents a problem with insurance claims data because it is highly imbalanced and zero-inflated. This means that the unconditional probability of observing 0 claims will always be the highest. Even when conditioned on the precipitation forecast, the model might still assign the highest probability to the 0-claims class, despite an increased risk of observing insurance claims. The intention of the predictive model is to provide more nuanced insights than the baseline probabilities. By using the precipitation forecast, the model aims to refine the probability estimates for each class of claims. Ideally, the output probabilities from the predictive model will be more accurate than relying solely on the baseline probabilities. We will now illustrate this with two specific examples.

Pretend today is Thursday, September 30, 2021, and the insurance company has just received the weather forecasts from ECMWF. Currently, we are interested in daily predictions, but since we only receive the weather forecasts twice a week—on Mondays and Thursdays—we can predict the outcome of claims for Friday, October 1, as well as the following three days until we receive the next forecast on Monday. Ideally, if we received weather forecasts every day, we would update the next-day claims predictions with the new data, as it would include a forecast with a shorter lead time, which is a more reliable precipitation forecast. This is a cost-benefit analysis that the specific business should undertake to determine if sourcing more frequent forecasts is valuable. In Table 5.2 below, we show how the model would be applied in practice by utilizing the best model for Oslo (Table 4.1). After receiving the precipitation forecast initialized today, we extract the 90th percentile from the ensemble member distribution for each of the lead times 1 through 4 and place them according to the corresponding date of prediction, essentially running 4 different logistic models. The 90th percentile was the quantile resulting in the best model for Oslo using logistic regression. These models provide the updated distributed probabilities for each of the three classes of claims. Our interest lies in the change in probabilities from the unconditional probabilities to these new probabilities, which is shown as a percentage change in the model output below.

	Fri 10/1	**Sat 10/2**	**Sun 10/3**	**Mon 10/4**
Lead time:	**1**	**2**	**3**	**4**
Model Input on Thu 9/30:				
Precipitation forecast 90th percentile (mm)	24	36	29	68
Model Output:				
0 claims prob change (%)	-19	-39	-26	-83
1 claim prob change (%)	+43	+73	+55	+65
>1 claims prob change (%)	+113	+281	+169	+863
Outcome:				
Total claims (both types)	3	7	1	1
Natural perils claims	1	0	0	0

Table 5.2: Overview of the inputs, outputs and outcomes when applying the logisitc predictive model on September 30th 2021 in Oslo for 1-day predictions. Here we use the 90th percentile from the precipitation forecast distribution.

This is an example of an abnormally high number of claims within a short period. These periods are of particular interest to insurance companies in terms of risk mitigation. The model clearly signals an increased risk for all four days, but it does not clearly distinguish between the occurrence of one claim and more than one claim. For instance, on Saturday, October 2nd, the probability for one claim increased the most out of the four days, even though seven claims were observed that day. Similarly, the probability for more than one claim increased the most on Monday, October 4th with an 863% increase, with an outcome of 1 claim, while it increased only 281% 2 days prior, where we have 7 claims.

In the next example we are specifically interested in predicting natural perils claims. As shown in Table 4.3, the AUC score indicates that when predicting only natural perils claims in Bergen using the best logistic model, the model performance improves compared to predicting both types or only the rain-associated claims. We want to examine the model's performance by using the day with the highest increase in the probability of observing multiple natural peril claims. In 2021, this day was October 29th, which showed a 1054% increase, as detailed in Table 5.3 below. Here, the probability changes for each of the three claims classes refers to only natural perils claims. In this example, we train the model exclusively on natural perils claims instead of both types of claims. As such, the unconditional probabilities, compared to a model predicting both types of claims (Table 5.1), will be different. For natural perils claims in Bergen, the unconditional probabilities for 0, 1, and more than 1 claim are 94%, 4%, and 2%, respectively. This is why the magnitude of the change in probabilities is different from the first example in Table 5.2, and should thus only be compared among the same type of model.

	Fri 10/29	Sat 10/30	Sun 10/31	Mon 11/1
Lead time:	**1**	**2**	**3**	**4**
Model Input on Thu 10/28:				
Precipitation forecast 100th percentile (mm)	70	43	47	28
Model Output (natural perils claims probabilities):				
0 claims prob change (%)	-26	-7	-9	-2
1 claim prob change (%)	+145	+64	+75	+24
>1 claims prob change (%)	+1054	+233	+303	+63
Outcome:				
Total claims (both types)	0	3	0	0
Natural perils claims	0	1	0	0

Table 5.3: Overview of the inputs, outputs and outcomes when applying the predictive model for natural perils claims on October 28th 2021 in Bergen. Here we use the max quantile from the precipitation forecast distribution.

What is interesting in the table above is that the model gives very strong signals for more than one natural perils claim on the 29th, but only the next day actually observed a natural perils claim. This might indicate that even though the model is not perfect for predicting the exact day, it might provide information on observed claims for the near future. Next, we show the model probabilities for the full year of 2021, when trained on 2014 through 2020 data.

In Figure 5.1, we show the predicted probabilities from the logistic model when predicting claims for $t+1$ throughout 2021 for both Bergen and Oslo, using the best features found in Table 4.1. We use two Y-axes. The left Y-axis displays in colors both the unconditional probabilities represented by the constant stippled lines, and the predicted probabilities in solid lines, where the green lines represent the probabilities for 0 claims, the blue lines show the probabilties for 1 claim, and the red for more than 1 claims. The right Y-axis shows the observed claims by black dots, for the 3 different classes of claims. We filter out all the 0 claim observations in the figures, because it would dominate the bottom of the plot if they are included.

A perfect model would indicate increased observed claims when the red line increases and the green line decreases, which seems to be the case sometimes, but not always. Oslo also seems to give more extreme values, two instances where the predicted probabilities for 0 claims is less than that for either 1 or more claims; but neither of those days resulted in more than 1 claim. Ideally, the insurance company would have such a model for each chosen geographical area which in our dataset is limited to the Bergen and Oslo municipalities. If each municipality were divided into smaller locations, the model could identify local variations in risk.

The main goal of our examples above, is to show how insurance companies can implement the model to minimize potential damage resulting from insurance claims. To achieve this, the insurance company should create a system detailing when to inform households, through which mediums, and what information should include. The primary information in the warning system should guide households on how to divert water away from properties by channeling it safely or allowing it to pass through designated safe routes. Specific actions will depend on the predicted risk levels. Additionally, it is essential to develop and communicate clear evacuation plans to residents, ensuring they understand the safest routes and procedures during severe flooding. This approach can serve as a critical warning mechanism for insurance companies, helping them mitigate potential losses.

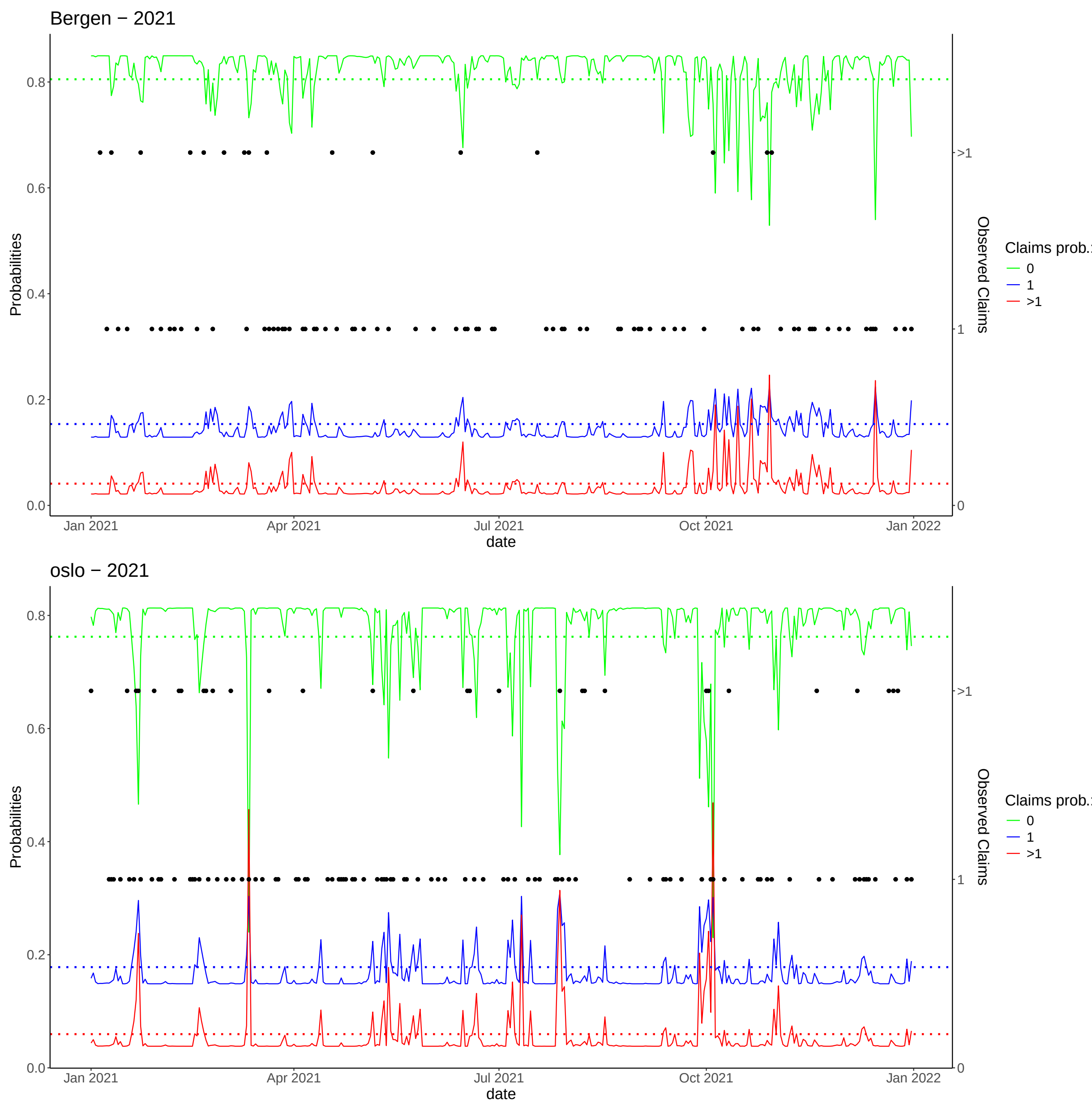

Figure 5.1: On the left Y-axis both the unconditional probabilities in stippled lines, and predicted probabilities in solid lines. The observed claims are on the right Y axis with black dots. The model used for each municipality is the best logistic model.

6 Conclusion and discussion

Our paper sought to answer two main questions. First, can precipitation forecasts predict insurance claims? And second, can more advanced machine learning methods predict insurance claims better than traditional methods?

Our results show that predicting insurance claims using precipitation forecasts within a few days is possible and more accurate than random guessing. We believe this is a promising initial step for the insurance industry in terms of increasing its ability to identify short-term risks related to insurance claims. Although the signal is weak, even incremental increases in the ability to quantify risks should be of interest to the insurance industry, and our research shows that incorporating the use of precipitation forecasts can achieve this. One contributing factor resulting in the weak signal could be the quality of the claims data, which has coarse spatial resolution (county level) and spans only a short term period (2014-2021). Additionally, since we used precipitation forecasts initialized only twice a week, this scheduling pattern introduces variability in lead times and results in slightly less accurate predictions compared to having forecasts initialized seven days a week. This is because short term lead times have higher correlation with actual precipitation events, which would improve the performance of the model.

It is worth noting that our model performs quite well for predicting natural perils claims over bergen, with an AUC of almost 0.8, a level considered excellent by Mandrekar (2010). One reason might be that weather forecasts are better at predicting extreme weather associated with damages in the natural perils pool, such as mid-latitude storms. Therefore, these extreme weather events might provide a window of opportunity for companies as during which the model demonstrates increased performance. For an insurance company, these periods of increased observed insurance claims would be of particular interest. Another reason why Bergen had a better performance in predicting natural claims compared to Oslo might be because Bergen experiences more extreme weather and has more natural peril claims than Oslo. This allows the model to receive more training on these types of claims in Bergen.

To answer our second research question, the more advanced machine learning models XGBoost and Neural Networks do not seem to improve model performance over the traditional logistic model, despite their ability to model complex non-linear relationships, such as the ones that link forecasted precipitation and insurance claims. However, their performance might be limited by the short record of claims data. We also cannot exclude that further parameter tuning would improve those models, as we have only tried a limited number of parameter configurations. Incorporating further tuning of the layers and neurons in the neural network, or the number of trees or other boosting parameters in XGBoost, might increase model performance. Moreover, we observed that during extreme weather events in Bergen, high wind speeds were often present. As a coastal city, high wind speed can bring in water from sea, resulting in additional damage. Thus, incorporating a variable for wind and further feature engineering might improve the model performance of the more advanced machine. Despite these possibilities for improving the performance of the model, predicting claims accurately based on weather forecasts remains a significant challenge. This is mainly because the relationship between weather and damage is complex and influenced by numerous factors.

How could our results be relevant to insurance companies? We believe our research can contribute to building an early warning system for the insurance industry. However, it is important to determine the appropriate skill and performance of the model before it is ready to be used. Several factors must be considered to do so. While the ability of the model to correctly identify increased risk is the obvious metric for performance, one must also consider the false positives, where the system would indicate risks when there are none. As seen in our operational example, even a model with an AUC score close to 0.8 may not always correctly identify the right signals at the right times. Frequent false alarms can cause customer backlash and diminishing trust, as highlighted by LeClerc and Joslyn (2015). This effect is referred to as the "cry wolf" (Breznitz, 1984), in which people are less likely to respond to future warnings due to previous false alarms. Therefore, it is essential for each individual company to evaluate the system's feasibility before implementation, as the required model performance might vary between companies. The analysis should aim to determine whether the potential cost of losing customers is worth the added benefits of reducing losses related to insurance claims. Insurance companies might consider initially implementing a beta version and informing their customers that the system is in its early stages. This approach could help mitigate some of the negative effects of issuing false risk warnings. These areas warrant further study in future research.

Further research should concentrate on exploring the potential of an early warning system for insurance companies to more effectively mitigate the short-term risks associated with insurance claims. Achieving this goal requires better-quality insurance claims data, specifically more comprehensive data in terms of the geographical dimension. For instance, the data could be expanded beyond municipal borders to a gridded map with multiple, equally sized grids. The exact resolution or number of total households within each grid should be further researched to determine the optimal level of resolution. Nevertheless, increasing the resolution too much, thereby creating smaller spatial units, may lead to a more imbalanced dataset with a higher proportion of zero-claim observations, particularly in less densely populated areas. With such a map, an early warning system could be implemented as a traffic light system, similar to Google's Flood Hub (2022), which indicates increased risk of flooding using green, orange and red colors on a map to identify different levels of risk.This system can alert households to imminent risks by identifying and categorizing them based on their location and severity. Such a system could be used to warn customers of impending risk of damage to their homes, or used internally by an insurance company to make proactive decisions about impending damages, such as allocating resources at the right time. Similarly to our operational example in Section 5, such a model would be implemented for each of the grids included, which could provide deeper insights into regional risk factors.

In conclusion, our study shows the potential of using weather forecasts to predict home insurance damage claims. It is crucial to highlight the innovative nature of this effort while also acknowledging the need for considerable improvements in data quality and model variables to achieve practical applicability. This study represents a first step towards developing an ex ante forecast model for precipitation-related insurance claims. Our findings indicate that it is feasible to use weather forecasts to predict insurance damage claims. However, further research is required to enhance the accuracy of these predictions and integrate them into practical applications, such as early-warning systems. Building on these results, future research should concentrate on improving model variables and data quality to increase the predictive models' performance and resilience. By doing so, we might be able to unlock the full potential of using precipitation forecasts and machine learning in mitigating risks for insurance companies.

Declaration on the use of AI tools in the work on this master's thesis

Name (and version) of the AI tools: ChatGPT-4, ChatGPT-4o, and Sikt KI-chat (GPT-4)

Purpose of the tools: Data Handling, Internet Browsing, Grammar and Writing Assistance, Coding (R, Python and LaTex)

We are aware that we are responsible for all content of this master's thesis, including the parts where AI tools are used. We are responsible for ensuring that the thesis complies with ethical rules for privacy and publication.

7 References

Ayuso, M., Guillen, M., & Nielsen, J. P. (2019). Improving automobile insurance ratemaking using telematics: incorporating mileage and driver behaviour data. *Transportation*, 46(5), 735–752.

Bergens Tidende. (2016 a, January 29). Ekstremværet kan treffe Hordaland hardest. *Bergens Tidene*. https://www.bt.no/nyheter/lokalt/i/e5kA4/ekstremvaeret-kan-treffe-hordaland-hardest

Bergens Tidende. (2016 b, August 26). Regnskur og lyn skapte kaos flere steder i Bergen fredag ettermiddag. *Bergens Tidene*. https://www.bt.no/nyheter/lokalt/i/bKv6V3/regnskur-og-lyn-skapte-kaos-flere-steder-i-bergen-fredag-ettermiddag

Bentéjac, C., Csörgo, A., & Martínez-Muñoz, G. (2019). A comparative analysis of XGBoost. *College of Science and Technology, University of Bordeaux*.

Breznitz, S. (1984). Cry wolf: The psychology of false alarms (1st ed.). Lawrence Erlbaum Associates.

Chen, K., Huang, R., Chan, N. H., & Yau, C. Y. (2019). Subgroup analysis of zero-inflated poisson regression model with applications to insurance data. *Insurance: Mathematics and Economics*, 86, 8–18.

Cheng, C. S., Li, Q., Li, G., & Auld, H. (2012). Climate change and heavy rainfall-related water damage insurance claims and losses in Ontario, Canada. *Journal of Water Resource and Protection*, 4, 49–62.

Coughlan de Perez, E., Harrison, L. C., Berse, K., Easton-Calabria, E., Marunye, J., Marake, M., Binte Murshed, S., Shampa, & Zauisomue, E.-H. (2022). Adapting to climate change through anticipatory action: The potential use of weather-based early warnings. *Weather and Climate Extremes*, 38, 100508.

European Centre for Medium-Range Weather Forecasts. (2020). ERA5: A comprehensive reanalysis dataset. *Quarterly Journal of the Royal Meteorological Society*. https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.3803

European Centre for Medium-Range Weather Forecasts. (n.d.). Operational archive. *ECMWF*. Retrieved May 24, 2024, from https://www.ecmwf.int/en/forecasts/dataset/operational-archive

Google. (2022). Google Flood Hub. Retrieved from https://floodhub.google.com/

Grossi, P., & Kunreuther, H. (eds). (2005). Catastrophe Modelling: A New Approach to Managing Risk. Springer: New York.

Hand, D. J., & Till, R. J. (2001). A simple generalization of the area under the ROC curve for multiple class classification problems. *Machine Learning*, 45(2), 171–186. https://doi.org/10.1023/A:1010920819831

Haug, O., Dimakos, X. K., Vardal, J. F., Aldrin, M., & Meze-Hausken, E. (2011). Future building water loss projections posed by climate change. *Scandinavian Actuarial Journal*, 2011(1), 1–20.

Herweijer, C., Ranger, N., & Ward, R. E. (2009). Adaptation to climate change: threats and opportunities for the insurance industry. *The Geneva Papers on Risk and Insurance-Issues and Practice*, 34, 360-380.

Komatsu, R., Maeda, Y., & Tachikawa, Y. (2012). Impact-based typhoon warnings and public response in Japan. *Environmental Science Policy*, 23, 27-35.

Kuhn, M., & Johnson, K. (2013). Applied Predictive Modeling. Springer.

LeClerc, J., & Joslyn, S. (2015). The Cry Wolf Effect and Weather-Related Decision Making. Risk Analysis, 35(3), 385-395.

Lemaire, J., Park, S. C., & Wang, K. C. (2016). The use of annual mileage as a rating variable. *ASTIN Bulletin: The Journal of the IAA*, 46(1), 39–69.

Lorenz, E. N. (1969). The predictability of a flow which possesses many scales of motion. *Journal of the Atmospheric Sciences*, 26(6), 636-646.

Lyubchich, V., Gel, Y. R. (2017). Can we weather proof our insurance?. *Environmetrics*, 28(2), e2433. https://doi.org/10.1002/env.2433

Lyubchich, V., Kilbourne, K. H., Gel, Y. R., & Richardson, T. X. (2019). Where home insurance meets climate change: making sense of climate risk, data uncertainty, and projections. *Variance*, 12, 278-292.

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology*, 5(9), 1315-1316.

Miljødirektoratet. (2015). Klima i Norge 2100. *Miljødirektoratet*. https://www.miljodirektoratet.no/publikasjoner/2015/september-2015/klima-i-norge-2100/

Mills, E. (2005, August 12). Insurance in a climate of change. *Science*, 309(5737), 1040-1044.

Mouatassim, Y., & Ezzahid, E. H. (2012). Poisson regression and zero-inflated poisson regression: application to private health insurance data. *European Actuarial Journal*, 2(2), 187–204.

Norwegian University of Life Sciences. (2017). Rain in Asker. *NMBU Brage*. https://nmbu.brage.unit.no/nmbu-xmlui/handle/11250/2465135

Norwegian Water Resources and Energy Directorate. (2016). Rapport $2016_4 4.https://publikasjoner.nve.no/rapport/2016/rapport2016_4 4.pdf$

Norsk Naturskadepool. (2024a). Types of damages. https://www.naturskade.no/en/types-of-damages/

Norsk Naturskadepool. (2024b). Nye regler om naturskadeforsikring trer i kraft 1. januar 2025. *Norsk Naturskadepool*. https://www.naturskade.no/nyheter/2024/nye-regler-om-naturskadeforsikring-trer-i-kraft-1.-januar-2025/

Norsk Naturskadepool. (2024c). Om naturskadepoolen: Spørsmål og svar. *Norsk Naturskadepool*. https://www.naturskade.no/om-naturskadepoolen/sporsmal-og-svar/

NRK. (2015, January 10). Slik herjer stormen 'Nina' i Norge. *NRK*. https://www.nrk.no/vestland/slik-herjar-stormen-$_nina\_i-norge-1.12142845$

Ozdemir, S. (2022). Feature engineering bookcamp. Simon and Schuster.

Pappenberger, F., Cloke, H. L., Parker, D. J., Wetterhall, F., Richardson, D. S., & Thielen, J. (2015). The monetary benefit of early flood warnings in Europe. *Journal of Hydrology*, 541, 105-116.

Rohrbeck, C., Eastoe, E. F., Frigessi, A., & Tawn, J. A. (2018). Extreme value modelling of water-related insurance claims. *The Annals of Applied Statistics*, 12(1), 246-282.

Röösli, T., Appenzeller, C., & Bresch, D. N. (2021). Towards operational impact forecasting of building damage from winter windstorms in Switzerland. *Meteorological Applications*, 28(1), Article e2035. https://doi.org/10.1002/met.2035

Scheel, I., Ferkingstad, E., Frigessi, A., Haug, O., Hinnerichsen, M., Meze-Hausken, E. (2013). A Bayesian hierarchical model with spatial variable selection: The effect of weather on insurance claims. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 62, 85–100.

Shterev, V. A., Metchkarski, N. S., & Koparanov, K. A. (2022). Time series prediction with neural networks: A review. In *2022 57th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST)*. IEEE. https://ieeexplore.ieee.org/document/9828735

So, B. (2023). CatBoost Versus XGBoost and LightGBM: Developing Enhanced Predictive Models for Zero-Inflated Insurance Claim Data. *arXiv preprint arXiv:2307.07771*.

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, 15(2), 261–272.

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.

Swiss Re. (2006). Natural catastrophes and man-made disasters 2005: high earthquake casualties, new dimension in windstorm losses. *Swiss Re Sigma Series*, 2/2006.

TV2. (2015, January 11). Nina blåste med over 160 km/t. *TV2*. https://www.tv2.no/storm/nina-blaste-med-over-160-km/t/6431808/

Yip, K. C., & Yau, K. K. (2005). On modeling claim frequency data in general insurance with extra zeros. *Insurance: Mathematics and Economics*, 36(2), 153–163.

A Appendice: XGBoost Feature Importance

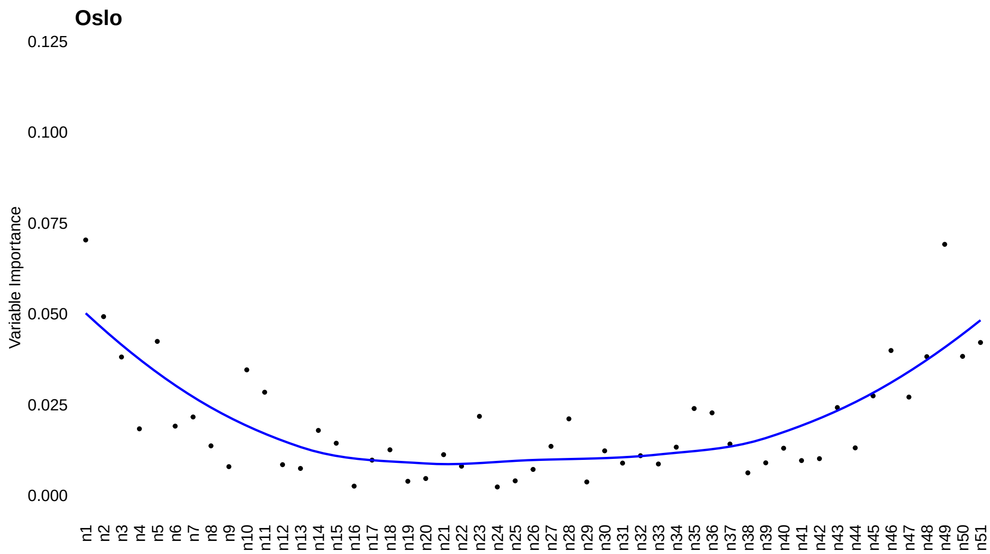

Figure A.1: XBgboost feature importance of all forecast ensemble members for Oslo municipality. "n1" represents the minimum quantile, and "n51" represents the max quantile.

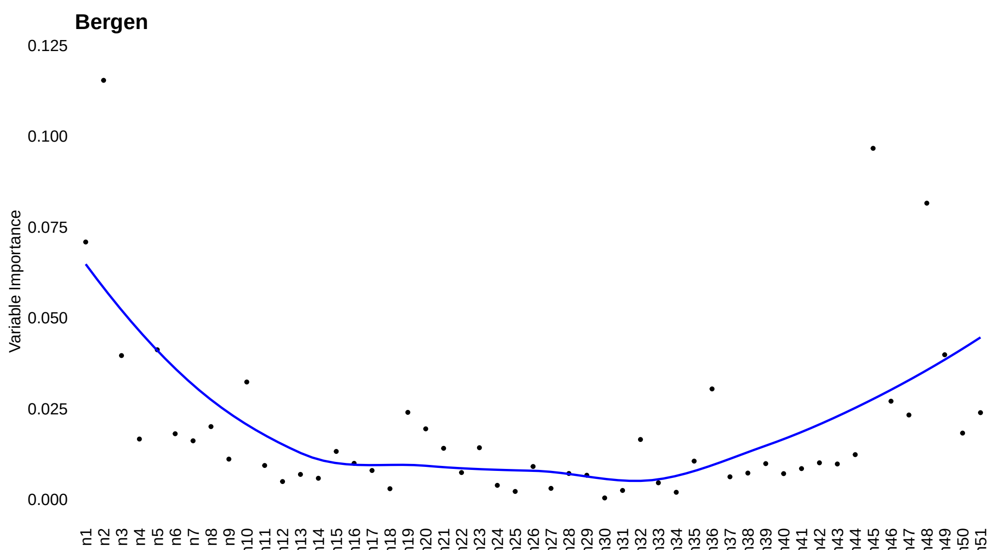

Figure A.2: XBgboost feature importance of all forecast ensemble members for Bergen municipality. "n1" represents the minimum quantile, and "n51" represents the max quantile.

B Appendice: Logistic model AUC plot

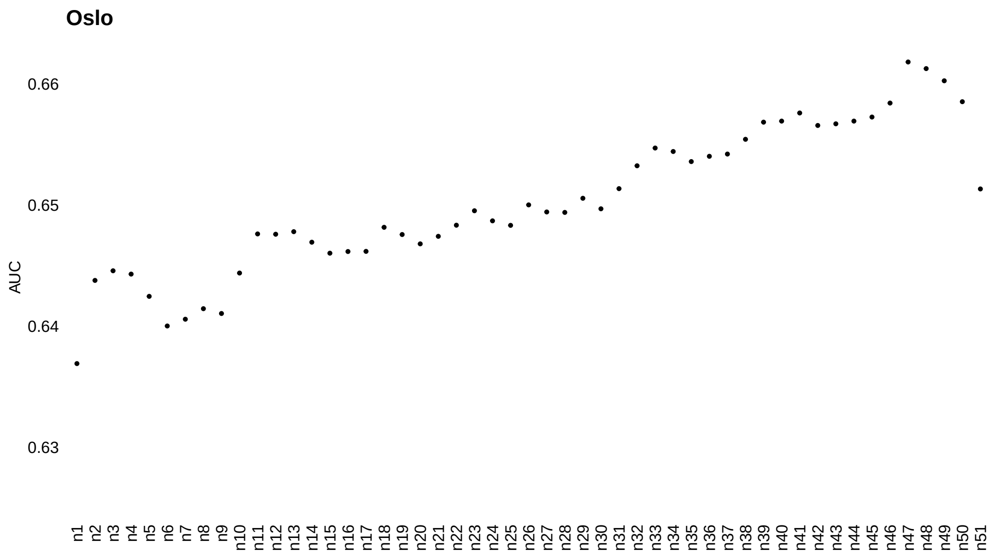

Figure B.1: AUC score for 51 logistic models using each 51 forecast ensemble members as features for Oslo municipality. ”n1” represents the minimum quantile, and ”n51” represents the max quantile.

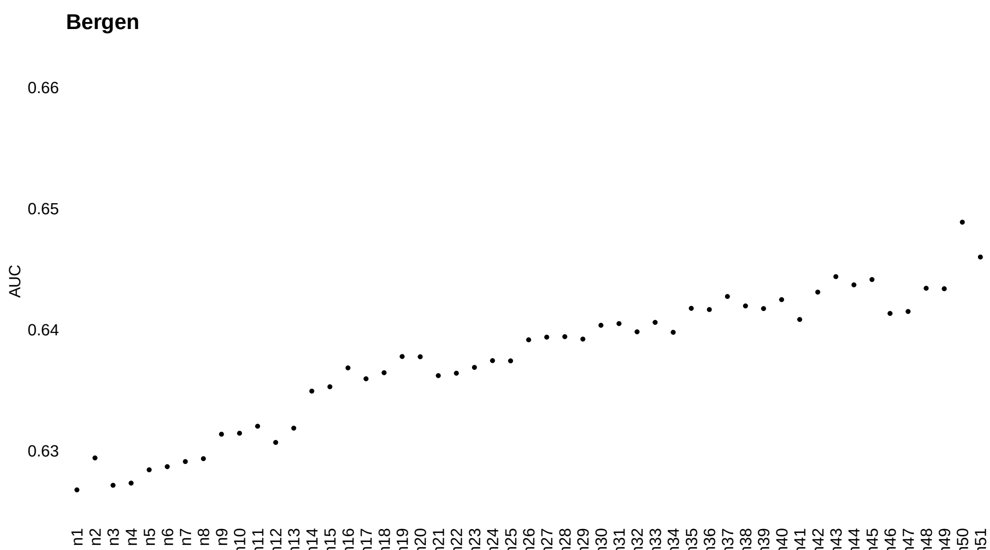

Figure B.2: AUC score for 51 logistic models using each 51 forecast ensemble members as features for Bergen municipality. ”n1” represents the minimum quantile, and ”n51” represents the max quantile.