NTNU
Norwegian University of
Science and Technology

Utilizing spatial and temporal dependencies for prediction of avalanches

A case study of the road network in Sognefjorden.

Jostein Ballestad

Master of Science in Physics and Mathematics
Submission date: March 2016
Supervisor: Ingelin Steinsland, MATH
Co-supervisor: Eivind Juvik, Statens Vegvesen

Norwegian University of Science and Technology
Department of Mathematical Sciences

Abstract

Accurate prediction of avalanches is demanded to ensure safety along Norwegian roads during the winter season. On the initiative of Statens vegvesen (SVV), different statistical models has been tested, but have been found inadequate to properly explain the avalanches. Inspired by recent studies of the French Alps, we seek to find a better avalanche model by allowing separate stretches of road to share information.

In this case study, we used a data set comprising avalanche observations from the last 39 years in an area based around Sognefjorden in Norway. In addition, interpolations of the snow depth, based on the seNorge snow model, were used as explanatory variables for the different local avalanche sites. Different models were tested to evaluate the effect of utilizing spatial and temporal dependencies for prediction of avalanches, at a stretch of road-scale. The spatial and temporal dependencies were represented as explanatory variables in a Poisson regression model and as random effects in a latent Gaussian model (LGM) framework. Model inference and predictions were carried out using Poisson regression and integrated nested Laplace approximations (INLA) for the LGMs. The predictive performance of the various models was evaluated using different skill scores with cross validation for the Poisson regression models, and on a test set for the LGMs.

It was found that the inclusion of spatio-temporal explanatory variables improved the prediction accuracy of the Poisson regression model. However, the random effects in the more flexible latent Gaussian models were not found to strengthen the predictive performance. Overall, this implies that spatial and temporal dependencies can be exploited to improve avalanche models. Still, poor data quality was found to be a problem and more research is required to better understand the underlying spatio-temporal processes.

Sammendrag

Nøyaktige snøskredvarslinger er nødvendige for å kunne ivareta sikkerheten langs norske veier om vinteren. Ulike statistiske modeller har blitt testet på initiativ fra Statens vegvesen (SVV), men ingen av modellene har så langt vist seg å kunne forklare dataene på en tilfredstillende måte. Ny forskning i de franske alpene inspirerte oss til å forsøke å finne bedre modeller for snøskred ved å la separate veistrekninger dele informasjon.

I dette eksempelstudiet benyttet vi et datasett som bestod av snøsskredobservasjoner fra de siste 39 årene i et område rundt Sognefjorden i Norge. I tillegg ble interpolerte snødybder, basert på seNorges snømodell, brukt som forklaringsvariabler for ulike lokale skredområder. Forskjellige modeller ble testet for å undersøke effekten av å benytte romlige- og tidsavhengigheter til å predikere snøskred for gitte veistrekninger. Avhengighetene ble innkludert som forklaringsvariabler i en Poisson regresjonsmodell og som tilfeldige effekter i et latent Gaussisk modellrammeverk. Både inferens og prediksjoner ble utført ved hjelp av Poisson regresjon og ved bruk av integrerte nøstede Laplace approksimasjoner (INLA) for de latente Gaussiske modellene. Nøyaktigheten til prediksjonene ble vurdert ut i fra forskjellige nøyaktighetsmål, både ved kryssvalidering for Poisson regresjonsmodellene og for ett testsett for de latente Gaussiske modellene.

Resultatene viste at nøyaktigheten til prediksjonene basert på Poisson regresjonsmodellene ble bedre ved å innkludere romlige- og tidsavhengige forklaringsvariabler. Det ble derimot ikke funnet bevis for at de tilfeldige effektene, som ble benyttet i de mer fleksible latente Gaussiske modellene, forbedret nøyaktigheten til prediksjonene. Generelt antyder dette at romlige avhengigheter og tidsavhengigheter kan utnyttes til å forbedre snøskredvarsling. Allikevel ble det oppdaget at dårlig datakvalitet var et problem og det kreves mer forskning for å kunne forstå de underliggende romlige og temporale avhengighetene bedre.

Contents

Preface

This thesis concludes my Master of Science degree in Applied Physics and Mathematics with specialization in Industrial Mathematics. The work on this thesis has been carried out at the Norwegian University of Science and Technology (NTNU), in the months from October 2015 to March 2016. The thesis has been written in cooperation with Statens vegvesen (SVV), who provided me with the necessary data for the study.

I would like to thank my supervisor Ingelin Steinsland for her guidance and encouragement throughout my work with the thesis. Her ideas, good feedback and positive spirit have been invaluable. I would further like to thank my co-supervisor Eivind Juvik and his deputy Tore Humstad, at Statens vegvesen. I am grateful for the opportunity they have given me to do research in such an interesting field and for their help along the way. I would also like to thank Jens Tveit, Andreas Haslestad and Tuomo Saloranta for helpful feedback on my inquiries regarding the study area and the seNorge snow model. Recognition should also be given to Jabir Ali Ouassou, for providing the LaTeX template that was used in this thesis.

Finally, I would like to thank Elisa de la Fuente Sanz, as well as my family and friends for their encouragement and support during my work with the thesis.

Jostein Ballestad

March 2016

1 Introduction

Every winter, central parts of the Norwegian road network are closed due to avalanches, both when there is a significant risk of one and due to actual occurrences. Statens vegvesen (SVV) is an institution administering the main roads in Norway. Thereby, SVV has a responsibility of securing the road network against the hazards avalanches pose. With this objective, SVV launched a review of the k-nearest neighbors model (Juvik et al., 2015), which is a popular decision support tool due to its simplicity and interpretability by practitioners in operational use. The initiative motivated more comprehensive investigations seeking to explore alternative models, see Hennum (2015), as well as my own specialization project Ballestad (2015) were the problem of under-reporting was examined. The pervasive result from both studies pointed to a lack in the models' ability to properly explain the data. This was also a problem of poor data quality and a result of avalanches being rare events.

In the last few decades, the availability of spatio-temporal data has increased substantially as a result of technological and methodological developments. In addition, the development of Bayesian inference methods have made it feasible to work with spatio-temporal models on large datasets. In recent years, this approach has also received growing attention in relation to avalanche forecasting. A number of studies have shown promising results by modeling avalanches as spatio-temporal processes. Based on data from Savoie, in the French Alps, Eckert et al. (2007) showed spatial dependence of avalanche occurrence at a township scale, using annual count data in a discrete spatial Poisson model. This demonstrated that spatial models could provide extra insight. The research was later extended to also consider temporal fluctuations of avalanche occurrences (Eckert et al., 2010) and it was found that the annual effect accounted for 17% of the avalanche occurrence variability. This shows that also temporal dependencies are an intrinsic characteristic of the avalanche activity. Following a similar approach, Lavigne et al. (2011) showed that it was possible to bring information from registered to unregistered avalanche paths, thus substantiating the advantage of utilizing spatio-temporal dependencies. In the wake of this work, more sophisticated models were employed to include the effect of spatio-temporal clustering (Lavigne et al., 2012) and addition of expert contributions (Lavigne et al., 2015). Through these improvements, the evolution of avalanches over time in different regions of the French Alps were identified and related

to climate change.

Despite the growing attention, the research in this field is still limited. In all the mentioned previous studies, model inference was carried out in a hierarchical Bayesian framework using Markov Chain Monte Carlo (MCMC) simulations and the analysis was done at a township scale focusing on yearly avalanche frequencies in the French Alps. An alternative approach is presented in Pozdnoukhov et al. (2011), that suggests using Support Vector Machines (SVM), which is a non-parametric classification method, for spatio-temporal predictive modelling of avalanches. In the study, the method is used to explore the possibility of exploiting the increased data availability to develop a data-driven system that could assist in avalanche forecasting. The system produced avalanche danger maps, on a 10 meter resolution grid, for a region in Scotland. The danger maps were then used to make daily predictions for individual avalanche paths. Even though the method showed promising results, it might be hard to implement such a system for SVV, since avalanches are only registered once they interfere with the road network. Thus, providing little information about the terrain in general.

Returning to the situation in Norway, accurate daily avalanche forecasts are desired to attain a safe road network and ensure minimal unnecessary road closure. As mentioned before, the studies carried out in Norway so far showed limited accuracy due to poor data quality and lack in the models' ability to properly explain the data. It is therefore of interest to see whether information can be shared between different neighboring stretches of roads to improve local predictions and thereby overcoming the sparseness of local data. Even though the Bayesian hierarchical approach showed promising results for yearly predictions in the French Alps, it has still not been tested for daily predictions. This will further increase the sparsity of the data and it is important to select an area with a sufficiently large avalanche frequency. Moreover, to improve on the computational burden associated with MCMC, we will instead use a recently developed method called integrated nested Laplace approximations (INLA), see Rue et al. (2009).

In this study we will investigate the effect of using spatial and temporal dependencies to model avalanche occurrences in Norway. The spatio-temporal dependencies will be included both as explanatory variables in a frequentistic framework and as temporally structured random effects in the more flexible Bayesian hierarchical framework. For the frequentistic approach we will use Poisson regression. In the Bayesian approach, we also consider Poisson distributed responses, but specified as a latent Gaussian model (LGM).

Here we use integrated nested Laplace approximations (INLA) for the inference and predictions. Various models are tested for both approaches. The Poisson regression is used to assess different spatio-temporal explanatory variables that are used in addition to two basic weather covariates. Both first and second order random walks will be tested in different forms with the LGMs. The various models are assessed based on predictive performance. Hence, it can be investigated whether inclusion of spatio-temporal dependencies improve the accuracy of the avalanche predictions. In addition, it allows us to compare the simple Poisson regression with the more flexible latent Gaussian models. The increased flexibility of the LGM comes at the expense of increased computational costs and it is therefore of interest to see whether it significantly improves the predictions or not.

Since this study is a collaboration with the SVV, the focus will be on daily avalanche predictions for exposed parts of the road network. To make the work feasible, the investigation is restricted to a case study of an area based around Sognefjorden in Norway. For the various roads in the area, informative data was available. Other avalanche observations from the surrounding terrain were also available, however, with poor data quality due to inconsistent observations. It was therefore decided to treat the avalanche occurrences as an areal process, adopting the township scale methodology mentioned earlier, but instead using suitable stretches of road. Avalanche occurrences were then aggregated on a daily basis for each of the stretches.

In Chapter 2, the study area and data are described. In addition, some initial analysis of the data is carried out, exploring the spatial and temporal dependencies. This is followed by some theory on the Poisson regression, latent Gaussian models and INLA in Chapter 3. After having deliberated on the theoretical foundation, we continue with Chapter 4, where we describe the different Poisson regression models and latent Gaussian models that are used in this study. This chapter also provide details on the model validation procedure, as well as the skill scores and criteria used to compare the models. We go on to present the results of the model evaluation in Chapter 5. Finally, in chapter 6, we discuss the results and some data quality issues, before we present some suggestions for future work.

2 Study area, data and exploratory data analysis

2.1 Study area and avalanche data

In this case study it was chosen to use avalanche observations from an area based around central parts of Sognefjorden in Norway. The area is highlighted in Figure 2.1. As can be seen, it was located along the west coast of southern Norway. Climatically, this region was characterized by large amounts of precipitation caused by weather systems from the North Sea being forced up the coastal mountains. In the winter season much of the precipitation fall as snow, due to the lower temperatures in the mountainous regions. Moreover, since the topography is typically characterized by fjords and valleys with steep mountainsides, large parts of the road network is consequently exposed to avalanche-hazardous terrain where avalanches occur frequently. Additionally, the homogenous climate in the region together with the relative high frequency of avalanches, made it an area where spatial and temporal dependencies could provide extra insight and predictability.

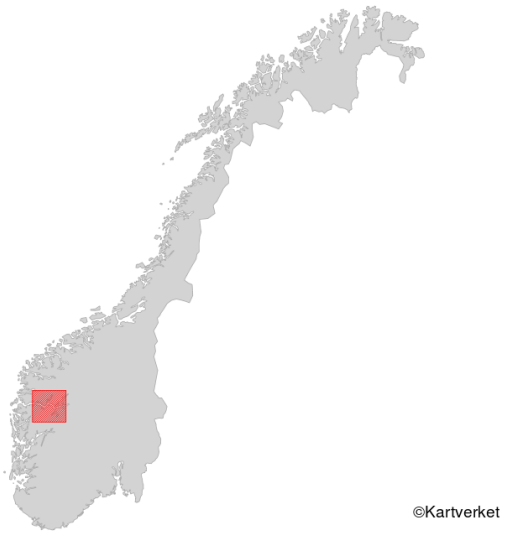


Figure 2.1: Location of the case study area.

The avalanche data was provided by the SVV database and comprised all registered avalanche observation, along nine different roads in the area, in the months January to April from 1977 until 2015. The avalanches were registered manually by local contractors. In Figure 2.2 the observed avalanches are plotted on a map of the area. The different colors indicate the various roads where the avalanches occurred. These roads were selected since they had the highest avalanche frequencies in the region. Additionally, it can be seen that the avalanches tended to cluster along shorter stretches of these roads. The avalanches along these stretches were presumably relatively similar in type and cause, thus easing the partition of the area into sub-areas which will be discussed in further detail in section 2.1.1. This was also the reason why avalanches along Fv13 were marked using both yellow and red dots.

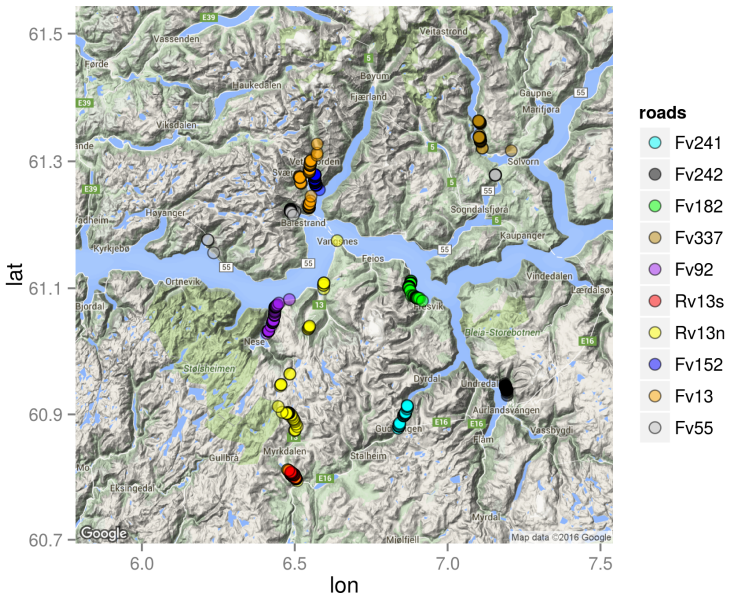
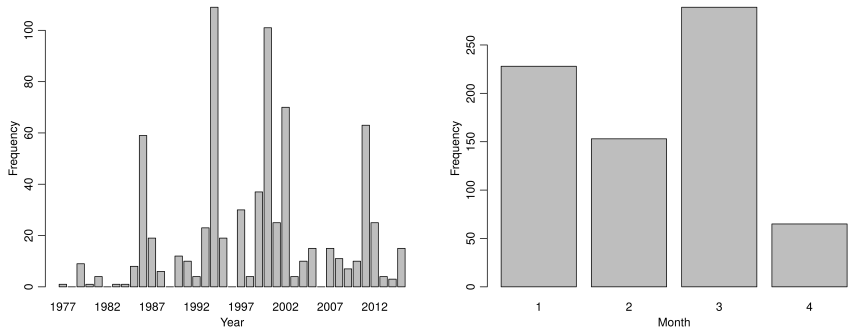


Figure 2.2: Sognefjorden with registered avalanches indicated by dots. The different colors represent different roads.

The first avalanche to be registered in this area took place along Fv55 in Esebotn on 4th of March 1977. Since then, a total of 735 avalanches have



(a) Yearly (b) Monthly

Figure 2.3: Avalanche frequency

been registered, in the given months and along the selected roads, up till 30th of April 2015. The figures 2.3a and 2.3b show the yearly and monthly avalanche frequencies respectively. It can be seen that in both of the years 1994 and 2000 more than 100 avalanches occurred in total in the region. However, there were also five years with no avalanches. On average there were approximately 20 avalanches per year with a standard deviation of 27, thus indicating a large variability in the yearly frequency. The monthly frequency plot show that the most avalanches occurred in March while the fewest occurred in April, 289 and 65, respectively. From the graph it can also be seen that there were fewer avalanches in February than in the neighboring months January and March.

2.1.1 Selection of stretches of roads

In order to investigate the spatial dependencies of avalanches the area was partitioned into sub-areas. A sub-area was restricted to be a stretch of road that could be considered homogeneous. The partition was therefore mainly based on the distance between the avalanche observations and the similarity in terrain. The roads Fv241, Fv242, Fv182, Fv337 and Fv 92 were relatively short and the avalanches along these roads were observed in proximity of one another, in similar terrain and isolated from observations along neighboring roads. Thus, each of these roads were selected to be a separate stretch of road. Rv13, on the other hand, was split into two separate stretches, a southern one and a northern one (red and yellow dots respectively), due to the length of the road and spread of avalanche observations. These groups

will be referred to as Rv13s and Rv13n.

The avalanches along the three remaining roads Fv152, Fv13 and Fv55 occurred close to one another and in similar terrain. A closer view of these roads is displayed in Figure 2.4. As can be seen, the avalanches along Fv13 (orange dots) can be roughly divided into four groups. A southern one, a middle one consisting of the avalanche on both sides of the little inlet called Sværafjorden, and a third group located north of these two. In addition there were three points located the furthest north, which we call the outlier group. The four groups varied in terrain and many of the avalanches were closer to avalanches along the neighboring roads than they were to the other groups. Still, the avalanches occurring along Fv152 would slide in a westward direction while most of the avalanches along Fv13 would slide in a eastward direction, hence signifying the difference in terrain. The avalanches occurring along Fv55 might, however, be more similar to the avalanches along Fv13 topographically. In Table 2.1 the correlations between the avalanche observations for the different groups and roads are displayed. Since the correlation between all combinations of the four Fv13 groups was stronger than any correlation between a group and a neighboring road, it was decided to adopt the natural road partition also in this region. I.e. the last three stretches were the roads Fv152, Fv55 and Fv13.

Table 2.1: Correlation between the various stretches of road in the Balestrand/Vetlefjorden area. Fv13 is split into Fv13o, Fv13n, Fv13m and Fv13s representing the outliers, northern, middle and southern group respectively.

	Fv152	Fv55	Fv13o	Fv13n	Fv13m	Fv13s
Fv152	1.00	0.07	0.17	0.14	0.14	0.17
Fv55	0.07	1.00	0.00	0.12	0.08	0.18
Fv13o	0.17	0.00	1.00	0.28	0.32	0.21
Fv13n	0.14	0.12	0.28	1.00	0.32	0.36
Fv13m	0.14	0.08	0.32	0.32	1.00	0.25
Fv13s	0.17	0.18	0.21	0.36	0.25	1.00

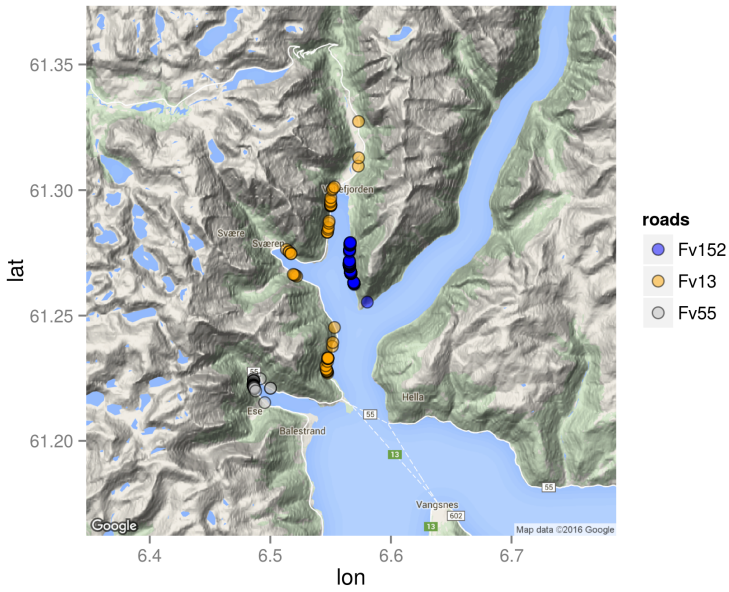


Figure 2.4: A closer view of the northwest area, Balestrand and Vetlefjorden. The registered avalanches are indicated by dots and the different colors represent various roads.

2.1.2 Omission of avalanche observations

The partition yielded ten different stretches following the color coding from Figure 2.2. In order to make the avalanche observations as homogeneous as possible, all observations that clearly stood out from the rest were omitted in further analysis. The omission was based on distance to the other avalanche observations and difference in terrain. Since Fv55 and Rv13n (grey and yellow dots) were the longest roads in the area, several clusters of avalanche observations could be identified along both. It was therefore necessary to decide which observations could be regarded as homogeneous and consequently kept for the prupose of this case study. In Figure 2.2, most of the avalanches along Fv55 (grey dots) occurs in the small inlet called Esebotn, visible in greater detail in Figure 2.4. The small spread made it a suitable stretch of road. The remaining avalanches along Fv55 were located far away and was omitted in further analysis. Most of the avalanches occurring along

Rv13n were observed in one valley with presumably homogenous conditions (yellow cluster furthest south). The remainder of the registered avalanches along Fv13n were omitted in further analysis, due to the large spread and since they were outside the valley, possibly making them different. Furthermore, the easternmost observation belonging to Fv337 was omitted in further analysis for similar reasons.

The resulting ten stretches and avalanche counts are displayed in Table 2.2. Almost half of the observed avalanches occurred along the two stretches A3 and A8, 106 and 219, respectively. A2, on the other hand, only had 26 registered avalanches. The other stretches vary between 40 to 66 registered avalanche observations. Note that the omission of data points, related to the selection of stretches, reduced the overall number of avalanches by 28 down to 707.

Table 2.2: Stretches of roads together with avalanche frequency.

Stretch	Road	Avalanches
A1	Fv241	55
A2	Fv242	26
A3	Fv182	106
A4	Fv337	42
A5	Fv92	63
A6	Rv13 s	62
A7	Rv13 n	40
A8	Fv152	219
A9	Fv13	57
A10	Fv55	37
Total	All	707

In Figure 2.5 the yearly avalanche frequency is plotted again, with color codes showing the distribution for the ten regions. In general the number of avalanches varies a lot for each of the stretches from year to year. Several avalanches were observed for A8 in most of the years. This was reasonable, considering that 219 out of the 707 avalanches occurred there. Moreover, it can be seen that for the stretches A6 and A7 (Rv13), avalanches were mostly registered from year 2000 and onwards. In addition, the frequency of avalanches in A3 appeared to decrease after 2002. After conversations with Tveit (2016) from SVV and Haslestad (2016) from Norges vassdrags- og energidirektorat (NVE) later in the study, it was found that these observa-

tions were related to poor data quality. This was, however, discovered after the analysis had been carried out and the issue will therefore be discussed as a source of error in Section 6.2.

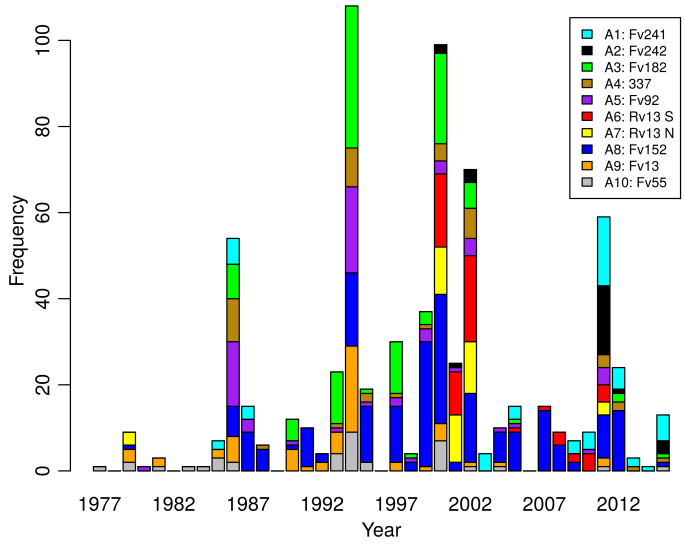


Figure 2.5: Yearly frequency of avalanches for each area.

2.2 Snow depth data from seNorge

In previous studies for SVV, Juvik et al. (2015) and Hennum (2015) found that snow depth observations from weather stations was a vital explanatory variable. However, since the weather stations measuring snow depths were often placed far away from the avalanche sites and at sea level it was difficult to get realistic observations. Moreover, the data quality was varying as measurements were sometimes lacking. In an attempt to improve this, it was decided to use interpolated snow depths based on the seNorge snow model. This model interpolated a snow depth, at 6:00 every morning, for every square kilometer in Norway based on precipitation and temperature observations, and a snow model. For more details, see Saloranta (2014).

The snow depth data were accessed through the web page (Norges vassdrags- og energidirektorat, 2012). Using this model, each stretch, Ai, was given a

different snow depth covariate, $\boldsymbol{s}_i$, from a selected grid point close to the avalanche observations. In this way it was attempted to account for local differences and to improve the accuracy of the covariates. However, several grid points were possible for many of the stretches, especially the stretches A1, A3, A4, A7 and A9, which were exposed to multiple avalanche paths. Furthermore, it was unknown at which height the avalanches were triggered. The correlation between snow depths from two potential grid points, of different altitude, was computed for each of these problematic stretches. The results in Table 2.3, show that the correlation was relatively strong for all the stretches. This indicated that the differences were small for grid points laying close to each other and closer inspection was not carried out. It was decided to use the grid points with altitudes that seemingly represented the surrounding mountains best. In Figure 2.6 the selected snow depth locations are displayed as blue squares, while the red squares represent the alternative locations.

Table 2.3: Correlation between snow depths from different grid points with different altitudes for problematic stretches.

Stretch	Correlation	Altitude difference
A1	0.68	638
A3	0.81	354
A4	0.95	150
A7	1.00	20
A9	0.75	160

Since avalanches often occurred during or after heavy snow falls it was decided to use the difference in snow depth as a second covariate, $\Delta\boldsymbol{s}_i$, in the avalanche model. This covariate was computed as the difference between the interpolated snow depth for a given day and the interpolated snow depth from the previous day. The correlation between the two covariates $\boldsymbol{s}_i$ and $\Delta\boldsymbol{s}_i$ was computed, for all the stretches combined, to make sure that $\Delta\boldsymbol{s}_i$ was not redundant. It was found to be 0.07, thus rejecting this concern.

2.3 Exploring data dependencies

2.3.1 Avalanches and snow depth

In order to investigate the relationship between avalanche occurrences and the snow depth, the yearly avalanche frequency was plotted together with

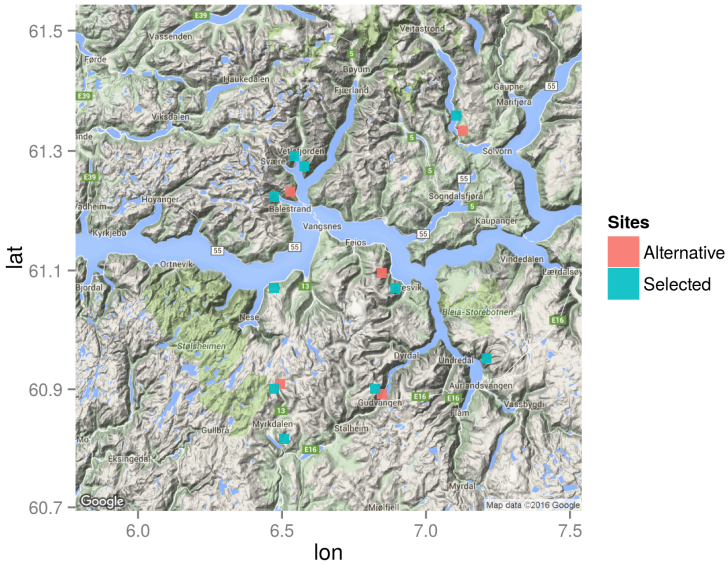


Figure 2.6: Map over area. Red squares show were the snow depth was computed, blue squares show alternative locations due to varied terrain and black dots show weather stations.

box plots of the snow depths, averaged over all the stretches for the corresponding years. The plots are displayed in Figure 2.7. It can be seen that the snow depths vary a lot from year to year, both in spread and size. There were no clear trends visible, although most years with high avalanche frequency also tended to have larger amounts of snow. This is better visualized in Figure 2.8, which displays the yearly snow depth median, averaged over all stretches, on top of the yearly avalanche frequency.

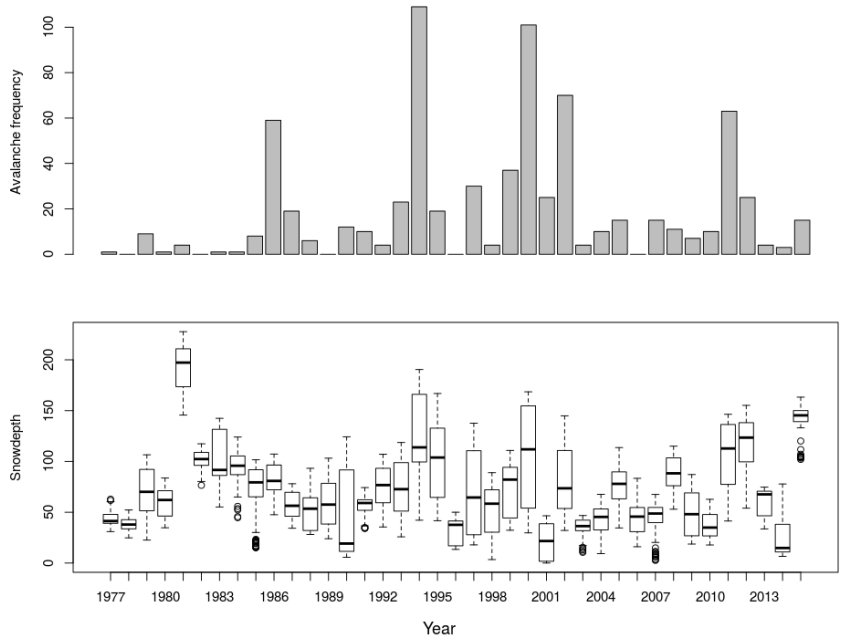


Figure 2.7: Yearly avalanche frequency and box plots of the snow depths, averaged over all stretches for each year. Compared with yearly avalanche frequency.

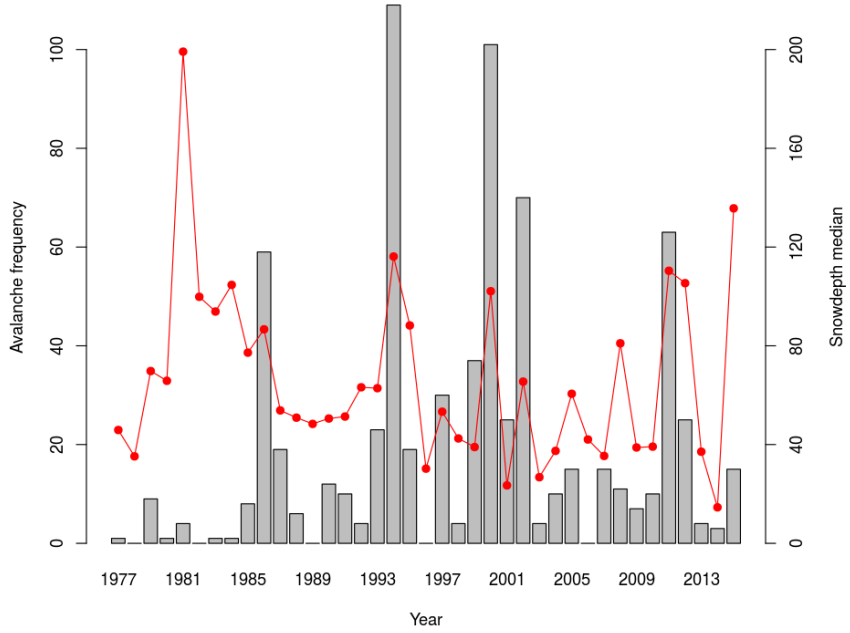


Figure 2.8: Median of snow depths per year, averaged over all stretches. Plotted on top of annual avalanche frequency.

2.3.2 Spatial and temporal dependencies

The spatial dependencies between the different stretches were studied by computing the correlation between the observed avalanches. The correlation was then plotted as a function of the distance separating each pair of the stretches. The results are displayed in Figure 2.9 and 2.10 in the panels to the left. A1 and A2 had the strongest correlation and were also located close to each other. The rest of the stretches had weaker correlation with at each other. However, no clear relation was found between the correlation and distance, for any of the stretches. To investigate the correlation pattern between the different stretches more closely a matrix was plotted, where all pairs of stretches with correlation above 0.18 had been marked with dark squares. The result is displayed in Figure 2.11. This plot revealed that there was a second group consisting of A3, A4, A5, A8, A9 and A10 that seemed to be better correlated than the rest, in addition to the isolated pair consisting of A1 and A2. This will be exploited later when validating the latent Gaussian models, see Section 4.6.

The right panels in Figure 2.9 and 2.10, show the estimated autocorrelation functions (ACF) for the avalanche observations for each of the stretches. The ACF is clearly significant for the first few lags for both A3 and A5. Moreover, the ACF is weakly significant for many of the first 15 lags along the stretches A6, A7 and A8. Therefore, this indicates the presence of temporal dependencies, although weak, that could potentially be used to improve the accuracy of the avalanche predictions. It should also be noted that the ACF for both A1 and A2, had a pronounced value at lag 5. Closer examination, revealed that this was the result of the avalanche activity on the 11th and 16th of January in 2011. For A1, 6 and 5 avalanches were observed on these two days, respectively. Similarly 9 and 6 avalanches were observed along A2. Such high counts were rare. In fact, 9 avalanches was the highest count to be observed along a single stretch during one day. Therefore, since so many avalanches occurred with a 5 day spacing, it caused the ACF at lag 5 to be pronounced for the two stretches. This was also the reason why the two stretches were strongly correlated.

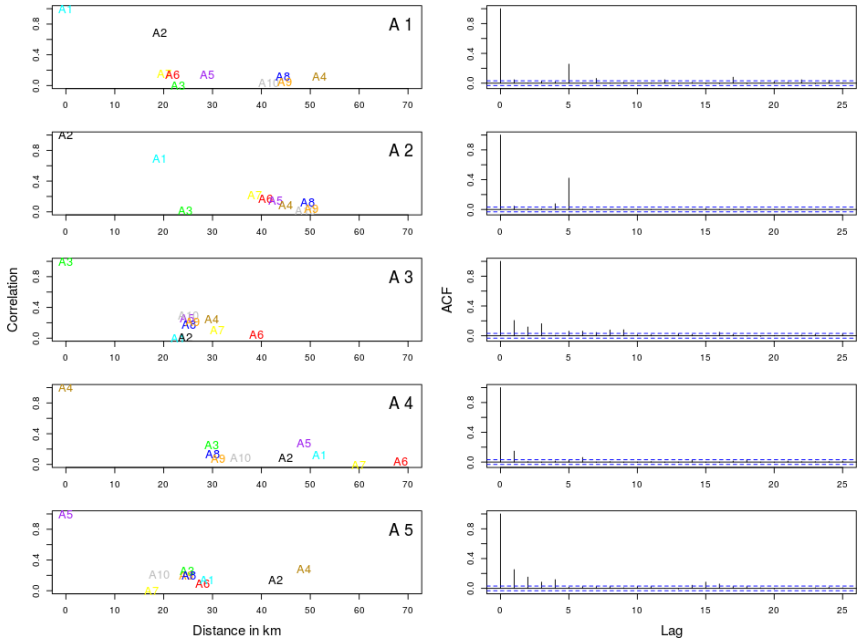


Figure 2.9: Correlation between avalanche occurences for the different stretches as a function of distance and the autocorrelation for regions A1-A5.

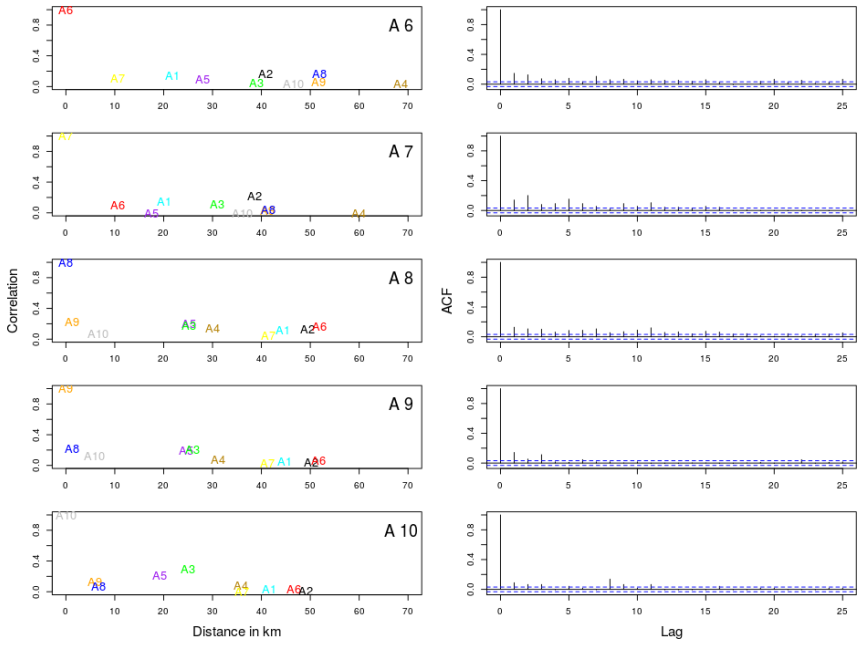


Figure 2.10: Correlation between avalanche occurences for the different stretches as a function of distance and the autocorrelation for regions A6-A10.

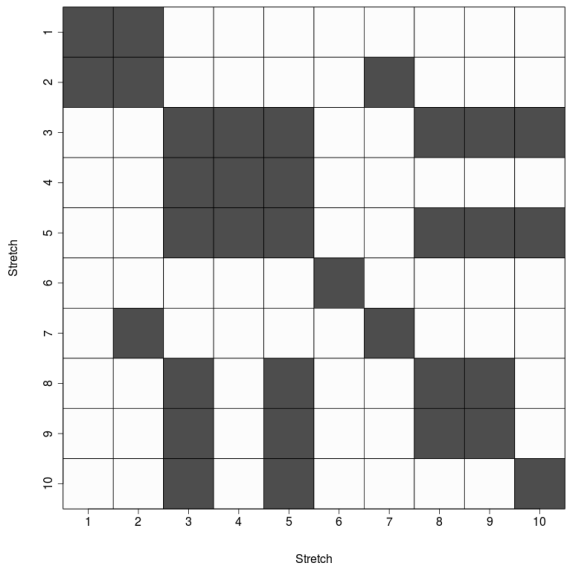


Figure 2.11: Stretches with avalanche occurrence correlation above 0.18 are illustrated with dark boxes.

3 Background

3.1 Notation

We first introduce some general notation that is useful to understand the following theory. The set of observations on a response is denoted by a column vector $\boldsymbol{y} = (y_1, ..., y_n)'$, where $'$ denote the transpose operation. Moreover, y_i is treated as a realization of the random variable Y_i. The set of covariates or explanatory variables is arranged as an $n \times (M+1)$ matrix $\boldsymbol{X} = (\boldsymbol{x}_0, \boldsymbol{x}_1, ..., \boldsymbol{x}_M)$. Each row of $\boldsymbol{X}$ refers to a different observation and each column to a different covariate. We let $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_M)'$ denote the set of coefficients associated with the covariates. Note that β_0 represents the intercept with covariate vector $(1, ..., 1)'$. Finally, we define the linear predictor as

$$\eta_i = \boldsymbol{x_i}\boldsymbol{\beta}. \tag{3.1}$$

3.2 General linear models: Poisson regression

Poisson regression is a model from the class of generalized linear models (GLM) introduced by Nelder and Wedderburn (1972). This is a class that extends the linear regression from concerning only responses from a Gaussian distribution to the case where it is sufficient that the distribution belong to the exponential family, see Casella and Berger (2002) for a definition. In other words the linear model

$$Y_i \sim \text{Normal}(\mu_i, \sigma^2), \tag{3.2}$$

$$\mu_i = \eta_i, \tag{3.3}$$

is generalized to include all distributions that belong to the exponential family. This is done by assuming instead that a one-to-one continuous differentiable transformation of the distribution mean follow the linear model, i.e. $g(\mu_i) = \eta_i$. The transformation $g(\cdot)$ is referred to as the link function and μ_i is the mean of an exponential distribution.

3.2.1 Poisson regression

The Poisson regression model is the typical statistical method for the analysis of the relationship between observed count data and a set of explanatory

variables. We describe the model here as formulated by McCullagh and Nelder (1989). The count data assume discrete values between 0 and ∞, i.e. $Y_i \in \{0, 1, 2, ...\}$. The parameter of interest is the average number of events, $\lambda_i = \mathrm{E}[Y_i]$, and the model assumes that λ_i has a linear relationship with the explanatory variables through a link function. For the Poisson regression the common link function is the logarithm, i.e

$$\eta_i = \log(\lambda_i) = \boldsymbol{x_i}\boldsymbol{\beta} \quad \text{and} \quad \lambda_i = \exp(\boldsymbol{x_i}\boldsymbol{\beta}), \tag{3.4}$$

where η_i is called the linear predictor. Equivalently the exponential function transforms the continuous values obtained by applying the linear predictor into the range of values of λ_i. The Poisson regression model can be written as follows:

$$Y_i \sim \text{Poisson}(\lambda_i) \tag{3.5}$$

$$\eta_i = \log(\lambda_i) = \boldsymbol{x_i}\boldsymbol{\beta}. \tag{3.6}$$

Note that the last equation can be expressed as

$$\eta_i = \beta_0 + \sum_{m=1}^{M} \beta_m x_{m,i}. \tag{3.7}$$

The coefficients are interpreted through the exponential function. Exponentiating the intercept yield the average count of events in an area or period when the predictors are at their reference category or at 0. Similarly, $\exp(\beta_m)$ for $m > 0$, give the multiplicative effect of the m-th predictor on λ_i when x_m changes by one unit.

3.2.2 Inference

The model parameters are estimated using the maximum likelihood, see McCullagh and Nelder (1989) for a detailed explanation. We present the method briefly here. The first and expected second derivatives of the logarithm of the likelihood is taken to obtain the score and information matrix. The estimates are then found using the iteratively re-weigthed last squares (IRLS) method. Based on a current estimate $\hat{\boldsymbol{\beta}}$, the linear predictor $\hat{\boldsymbol{\eta}} = \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}$ is computed. The fitted values can then be calculated as $\hat{\boldsymbol{\mu}} = \exp(\hat{\boldsymbol{\eta}})$. Using these values the working dependent variable $\mathbf{z}$, given for the Poisson regression as,

$$z_i = \hat{\eta}_i + \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i}, \tag{3.8}$$

is computed. An improved estimate for $\hat{\boldsymbol{\beta}}$ is then found by regressing $\boldsymbol{z}$ on the covariate, calculating the weighted least squares estimate

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}\boldsymbol{z}, \tag{3.9}$$

here $\mathbf{W}$ is a diagonal matrix of weights with entries $w_{ii} = \hat{\mu}_i$ and the large sample variance is

$$\text{Var}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}. \tag{3.10}$$

The final estimate is determined by repeating this procedure until convergence.

3.3 Latent Gaussian models

We continue by introducing the class of latent models, which extends the generalized linear model by allowing addition of random effects to the linear predictor. Following the description given in Blangiardo and Cameletti (2015):pp. 107-109, a latent model is typically defined within a three stage Bayesian hierarchical model framework consisting of responses, a latent process and hyperparameters for the priors that are assigned to the latent field. That is,

$$\text{Stage 1:} \quad \boldsymbol{Y} = (Y_1, ..., Y_n), \tag{3.11}$$

$$\text{Stage 2:} \quad \eta_i = \beta_0 + \sum_{m=1}^{M} \beta_m x_{mi} + \sum_{l=1}^{L} f_l(z_{li}), \quad \boldsymbol{\theta} = (\beta_0, \boldsymbol{\beta}, \boldsymbol{f}), \tag{3.12}$$

$$\text{Stage 3:} \quad \boldsymbol{\psi} = (\psi_1, ..., \psi_P), \tag{3.13}$$

where the collection of parameters in the linear predictor, $\boldsymbol{\theta}$, is the latent field and $\boldsymbol{\psi}$ is a vector of P hyperparameters corresponding to the assigned priors. The random effects are given by $f_l(z_{li})$, which are unknown functions defined in terms of a set of covariates $\boldsymbol{Z} = (\boldsymbol{z}_1, ..., \boldsymbol{z}_L)$. These can take many different forms such as smooth and nonlinear effects of covariates, time trends and seasonal effects, random intercept and slopes as well as temporal or spatial random effects. In this study the functions will be random walk processes of first and second order, see section 3.3.1 for more details. The responses in Stage 1 is typically assumed to belong to an exponential family where the mean, $\text{E}[Y_i] = \mu_i$, is linked to the structured additive predictor η_i through a link function $g(\cdot)$, so that $g(\mu_i) = \eta_i$, as for the GLM in Section 3.2.

A *latent Gaussian model* (LGM) is obtained by assigning Gaussian priors

to all elements of the latent field $\boldsymbol{\theta}$. The responses are assumed to be conditionally independent, given the latent field and the hyperparameters. Thus, we have for stage 1 that

$$p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{\psi}) = \prod_{i=1}^{n} p(y_i|\theta_i,\boldsymbol{\psi}), \tag{3.14}$$

where each data point y_i is connected to only one element θ_i in the latent field. In our case, since the responses are assumed to be Poisson distributed we will have that $\mathrm{E}[Y_i] = \lambda_i = \exp(\eta_i)$, as for the Poisson regression, and

$$p(y_i|\theta_i,\boldsymbol{\psi}) = \mathrm{Poisson}(y_i;\theta_i,\boldsymbol{\psi}) \tag{3.15}$$

The second stage is formed by the latent Gaussian field resulting from the multivariate Gaussian prior that is assumed on the latent field, with mean $\mathbf{0}$ and precision matrix $\boldsymbol{Q}(\boldsymbol{\psi})$. The density function thus becomes,

$$p(\boldsymbol{\theta}|\boldsymbol{\psi}) = (2\pi)^{-n/2}|\boldsymbol{Q}(\boldsymbol{\psi})|^{1/2}\exp\left(-\frac{1}{2}\boldsymbol{\theta}'\boldsymbol{Q}(\boldsymbol{\psi})\boldsymbol{\theta}\right), \tag{3.16}$$

where $|\cdot|$ is the determinant.

Finally, the third stage is formed by the hyperparameter prior distribution, i.e.

$$\boldsymbol{\psi} \sim p(\boldsymbol{\psi}). \tag{3.17}$$

In this study it was decided to use a standard non-informative gamma prior for the precision parameter associated with the random walk process. This will be specified further in Section 4.3.

3.3.1 Random walk model

The random effect, $f(z_i)$, that we included in the LGM was either a first-order random walk (rw1) or a second-order random walk (rw2). These are both improper GMRF's with rank deficiency one and two respectively (Rue and Held, 2005). The models can be described as follows. Given a time ordered vector $(z_1,...,z_T)$, a random walk is a model defined by an order r so that z_t only depends on the previous $t-r$ elements (Feller, 1968). The simplest RW model is defined when $r = 1$, then the conditional distribution of z_t given all the other elements of the vector is

$$z_t|z_{t-1} \sim \mathrm{Normal}(z_{t-1},\tau^{-1}),$$

so that there are independent increments

$$\Delta z_t = z_t - z_{t+1} \sim \text{Normal}(0, \tau^{-1}),$$

where τ is the precision. The density for $\boldsymbol{z}$ is derived from the $T-1$ increments as

$$\begin{aligned} p(\boldsymbol{z}|\tau) &\propto \tau^{(T-1)/2} \exp\left(-\frac{\tau}{2}\sum(\Delta z_t)^2\right) \\ &= \tau^{(T-1)/2} \exp\left(-\frac{1}{2}\boldsymbol{z}'\boldsymbol{Q}\boldsymbol{z}\right), \end{aligned} \tag{3.18}$$

where $\boldsymbol{Q} = \tau \boldsymbol{R}$ and $\boldsymbol{R}$ is the structure matrix reflecting the neighborhood structure of the model. Similarly for the second order random walk we have that

$$z_t | z_{t-1}, z_{t-2} \sim \text{Normal}(2z_{t-1} + z_{t-2}, \tau^{-1}),$$

with independent second-order increments

$$\Delta^2 z_t = z_t - 2z_{t+1} + z_{t+2} \sim \text{Normal}(0, \tau^{-1}).$$

The density for $\boldsymbol{z}$ thus becomes,

$$p(\boldsymbol{z}|\tau) \propto \tau^{(T-2)/2} \exp\left(-\frac{1}{2}\boldsymbol{z}'\boldsymbol{Q}\boldsymbol{z}\right), \tag{3.19}$$

where, again, $\boldsymbol{Q} = \tau \boldsymbol{R}$ and $\boldsymbol{R}$ represent the new neighborhood structure given by the second order increment $\Delta^2 z_t$.

3.4 Bayesian inference and LGM

The integrated nested Laplace approximations (INLA) is a deterministic algorithm especially designed for latent models, that was introduced by Rue et al. (2009). We now present an overview of the INLA inference, for more details we refer the reader to Blangiardo and Cameletti (2015). In Bayesian inference the objectives are the marginal posterior distribution for each element of the parameter vector $\boldsymbol{\theta}$

$$p(\theta_i|\boldsymbol{y}) = \int p(\theta_i, \boldsymbol{\psi}|\boldsymbol{y}) d\boldsymbol{\psi} = \int p(\theta_i|\boldsymbol{\psi}, \boldsymbol{y}) p(\boldsymbol{\psi}|\boldsymbol{y}) d\boldsymbol{\psi}, \tag{3.20}$$

and for each element of the hyperparameter vector

$$p(\psi_k|\boldsymbol{y}) = \int p(\boldsymbol{\psi}|\boldsymbol{y}) d\boldsymbol{\psi}_{-k}, \tag{3.21}$$

where $\boldsymbol{\psi}_{-k}$ represent $\boldsymbol{\psi}$ without the kth element. It is therefore necessary to compute $p(\boldsymbol{\psi}|\boldsymbol{y})$ and $p(\theta_i|\boldsymbol{\psi},\boldsymbol{y})$ from which the all relevant marginals and marginal posteriors can be determined. The first entity can be approximated by

$$\begin{aligned} p(\boldsymbol{\psi}|\boldsymbol{y}) &= \frac{p(\boldsymbol{\theta},\boldsymbol{\psi}|\boldsymbol{y})}{p(\boldsymbol{\theta}|\boldsymbol{\psi},\boldsymbol{y})} && (3.22)\\ &\propto \frac{p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{\psi})p(\boldsymbol{\theta}|\boldsymbol{\psi})p(\boldsymbol{\psi})}{P(\boldsymbol{\theta}|\boldsymbol{\psi},\boldsymbol{y})} \\ &\approx \left.\frac{p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{\psi})p(\boldsymbol{\theta}|\boldsymbol{\psi})p(\boldsymbol{\psi})}{\tilde{p}(\boldsymbol{\theta}|\boldsymbol{\psi},\boldsymbol{y})}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*(\boldsymbol{\psi})} =: \tilde{p}(\boldsymbol{\psi}|\boldsymbol{y}), && (3.23)\end{aligned}$$

where $\tilde{p}(\boldsymbol{\theta}|\boldsymbol{\psi},\boldsymbol{y})$ is the Gaussian approximation of $p(\boldsymbol{\theta}|\boldsymbol{\psi},\boldsymbol{y})$, given by the Laplace method (Blangiardo and Cameletti, 2015:p. 105). The $\boldsymbol{\theta}^*(\boldsymbol{\psi})$ is the mode for a given $\boldsymbol{\psi}$. This approximation is accurate since $p(\boldsymbol{\theta}|\boldsymbol{\psi},\boldsymbol{y})$ appears to be almost Gaussian.

Several methods exists to compute the more complex second entity $p(\theta_i|\boldsymbol{\psi},\boldsymbol{y})$. We explain the standard option, called simplified Laplace approximation, which is both computationally efficient and usually sufficiently accurate. The parameter vector is rewritten as $\boldsymbol{\theta} = (\theta_i, \boldsymbol{\theta}_{-i})$ and the Laplace approximation is used to obtain

$$\begin{aligned} p(\theta_i|\boldsymbol{\psi},\boldsymbol{y}) &= \frac{p((\theta_i,\boldsymbol{\theta}_{-i})|\boldsymbol{\psi},\boldsymbol{y})}{p(\boldsymbol{\theta}_{-i}|\theta_i,\boldsymbol{\psi},\boldsymbol{y})} && (3.24)\\ &\propto \frac{p(\boldsymbol{\theta},\boldsymbol{\psi}|\boldsymbol{y})}{p(\boldsymbol{\theta}_{-i}|\theta_i,\boldsymbol{\psi},\boldsymbol{y})} \\ &\approx \left.\frac{p(\boldsymbol{\theta},\boldsymbol{\psi}|\boldsymbol{y})}{\tilde{p}(\boldsymbol{\theta}_{-i}|\theta_i,\boldsymbol{\psi},\boldsymbol{y})}\right|_{\boldsymbol{\theta}_{-i}=\boldsymbol{\theta}^*_{-i}(\theta_i,\boldsymbol{\psi})} =: \tilde{p}(\theta_i|\boldsymbol{\psi},\boldsymbol{y}), && (3.25)\end{aligned}$$

where $\tilde{p}(\boldsymbol{\theta}_{-i}|\theta_i,\boldsymbol{\psi},\boldsymbol{y})$ is the Laplace Gaussian approximation to $p(\boldsymbol{\theta}_{-i}|\theta_i,\boldsymbol{\psi},\boldsymbol{y})$ and $\boldsymbol{\theta}^*_{-i}(\theta_i,\boldsymbol{\psi})$ is its mode. Furthermore, a Taylor expansion of the Laplace approximation $\tilde{p}(\theta_i|\boldsymbol{\psi},\boldsymbol{y})$ that includes a correcting mixing term, is used in place of the Laplace approximation to reduce the computational cost.

For both the entities the computation of the joint posterior distribution for $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$, $p(\boldsymbol{\theta},\boldsymbol{\psi}|\boldsymbol{y})$ is required. It is given by the product of the likelihood (3.14), of the density (3.16) and of the hyperparameter prior distribution (3.17), that is

$$p(\boldsymbol{\theta},\boldsymbol{\psi}|\boldsymbol{y}) \propto\ p(\boldsymbol{\psi})\cdot|\boldsymbol{Q}(\boldsymbol{\psi})|^{1/2}\exp\left(-\frac{1}{2}\boldsymbol{\theta}'\boldsymbol{Q}(\boldsymbol{\psi})\boldsymbol{\theta}+\sum_{i=1}^{n}\log\left(p(y_i|\theta_i,\boldsymbol{\psi})\right)\right). \quad (3.26)$$

In the computations it is exploited that the matrix $\boldsymbol{Q}(\psi)$, is a sparse precision matrix. This is a consequence of the components of the Gaussian latent field $\boldsymbol{\theta}$ being conditionally independent. A specification that is known as Gaussian Markov random field (GMRF) (Rue and Held, 2005).

4 Models and evaluation

In this chapter we describe the different models that were used to investigate the effect of including spatial and temporal dependencies on the avalanche prediction accuracy. The spatial and temporal dependencies are considered in a GLM framework by using the number of avalanches in the previous days, both for a single stretch and for the area in general, as explanatory variables. In the LGM framework we introduce latent temporal variables that are shared between the stretches. The models are compared according to predictive performance and the potential improvements are assessed relative to a basic model that only includes weather covariates.

4.1 Notation

We let $i \in \{1,...,10\}$ denote the stretches of road. Furthermore, we let $j \in \{1,...,J\}$ denote the year and $t \in \{1,...,T\}$ the day. The number of years, J, will be described in section 4.6. The number of days in each year was $T \in \{120, 121\}$ depending on the year being a leap year or not. That is, y_{ijt} represented the number of avalanches observed along stretch i, in year j and on day t.

4.2 Poisson regression

The number of avalanches was assumed to be Poisson distributed, such that

$$y_{ijt} \sim \text{Poisson}(\lambda_{ijt}). \tag{4.1}$$

Recall, that the mean λ_{ijt} was linked to the explanatory variables through a logarithmic link function and we specify the linear predictor (3.7) as

$$\eta_{ijt} = \beta_{0,i} + \sum_{m=1}^{M} \beta_{m,i} x_{m,ijt}, \tag{4.2}$$

where $\beta_{m,i}$ were the coefficients, specified for each stretch i, and $x_{m,ijt}$ were the corresponding explanatory variables with daily observations. The stretch-specific coefficients allowed the different stretches to depend differently on the explanatory variables. This is reasonable, since each stretch has its own local topography and climate (Haslestad, 2016). For instance, each stretch had stretch-specific weather covariates, as discussed in Section 2.2.

Furthermore, also the intercept and spatio-temporal explanatory variables, which will be introduced shortly, were given stretch-specific coefficients, due to the different topography and since the average number of avalanches for each stretch varied.

An overview of the explanatory variables that were used is given in Table 4.1. The weather covariates snow depth and snow depth difference are described in Section 2.2. The four remaining explanatory variables were used to investigate whether avalanche occurrences significantly depended on avalanche activity in the previous days. This was done by taking the total number of avalanches, for all ten stretches in the area, the previous day and for the five previous days, denoted p_{jt}^1 and p_{jt}^5, respectively. Hence both the day-to-day dependence could be studied as well as the effect of a smoother 5 day interval. Similarly, the explanatory variables c_{ijt}^1 and c_{ijt}^5, were the number of avalanches, for stretch i, the previous day and the five previous days. Through comparison of these explanatory variables with p_{jt}^1 and p_{jt}^5, the dependency on avalanches along the neighboring stretches could be examined.

Table 4.1: Covariates used in model testing.

Covariate	Description
s_{ijt}	Snow depth
Δs_{ijt}	Snow depth difference
p_{jt}^1	Avalanches in area the previous day
p_{jt}^5	Avalanches in area the five previous days
c_{ijt}^1	Avalanches along stretch i the previous day
c_{ijt}^5	Avalanches along stretch i the five previous days

In total we tested three different models, within the GLM framework. The first model, *Model 1*, was a *basic model* and all other models, both in the GLM and the LGM framework, were extensions of this. Only the two weather covariates, s_{ijt} and Δs_{ijt}, were used in the basic model and it provided a reference when assessing the effect of including spatio-temporal dependencies in the rest of the models. We specify Model 1,

$$\text{Model 1:} \quad \eta_{ijt} = \beta_{0,i} + \beta_{1,i} s_{ijt} + \beta_{2,i} \Delta s_{ijt} =: \mathrm{B}_{ijt}, \tag{4.3}$$

where $\beta_{0,i}$ was the intercept for stretch i and $(\beta_{1,i}, \beta_{2,i})$ were the coefficients for the corresponding stretch. To improve readability we denote the linear predictor of Model 1 B_{ijt} and all the other models will be extensions of this.

The second model, *Model 2*, includes all the different spatio-temporal explanatory variables. In this way, the explanatory variables that were significant for the most stretches could be identified and in turn be used to make an updated Poisson regression model. Moreover, the model allowed us to compare the different spatio-temporal explanatory variables, so that the spatio-temporal dependencies could be better understood. We specify Model 2,

$$\text{Model 2:} \quad \eta_{ijt} = \mathrm{B}_{ijt} + \beta_{3,i} p^1_{jt} + \beta_{4,i} p^5_{jt} + \beta_{5,i} \varsigma^1_{ijt} + \beta_{6,i} \varsigma^5_{ijt}, \tag{4.4}$$

where B_{ijt} is represented in the basic model (4.3), while the remaining terms were the coefficients and corresponding spatio-temporal explanatory variables.

Model 3 was the final Poisson regression model and was the update of Model 2. The variables p^1_{jt} and ς^1_{ijt} were excluded, based on the findings in Section 5.1.1 which showed that they were less significant. The model was used both to investigate the effect of including spatio-temporal explanatory variables relative to the basic model, but also for comparison of the potential benefits when using the more flexible LGMs. We specify the Model 3,

$$\text{Model 3:} \quad \eta_{ijt} = \mathrm{B}_{ijt} + \beta_{3,i} p^5_{jt} + \beta_{4,i} \varsigma^5_{ijt}. \tag{4.5}$$

4.3 Latent Gaussian models

We first explain the generic structure of the tested LGMs, before we specify the individual models. The avalanche observations $\boldsymbol{y}$ were still assumed to follow a Poisson distribution and we had for the Stage 1 (3.14) of the LGM hierarchical structure that

$$y_{ijt} | \boldsymbol{\theta}, \boldsymbol{\psi} \sim \text{Poisson}(\lambda_{ijt}).$$

In Stage 2 (3.16), the latent field generally consisted of two main parts. The first part included the fixed effects given by the intercept and the weather covariates, that is B_{ijt}. The second part consisted of a temporally structured

effect $Temp$ which we will specify for each of the models. These two parts were combined to give the linear predictor

$$\eta_{ijt} = \mathrm{B}_{ijt} + Temp. \tag{4.6}$$

The temporal effect was controlled by the precision $\tau = 1/\sigma^2$, which was a Stage 3 hyperparameter (3.17) for all the LGMs. Since no information about the prior was available, a standard non-informative logGamma prior was assumed on the logarithm of the precision, i.e.

$$\tau \sim \mathrm{Gamma}(1, 5e-5). \tag{4.7}$$

The generic model structure was based on the belief that the different stretches shared a common time trend, much like the idea behind the explanatory variables p^1_{jt} and p^5_{jt}. Therefore, when an avalanche was observed along a stretch this would serve as a warning for the area in general that the present conditions might be avalanche hazardous. This seemed plausible, since topography, snowpack and weather conditions were thought to have some of the same characteristics throughout the area.

In practice we used three different versions of this model structure. For all three models both a) a first-order random walk and b) a second-order random walk was used as the temporal effect in $Temp$, see Section 3.3.1. In the first LGM, it was assumed that the temporal trend was common for all the years and stretches, such that no distinction was made between the different years and only one temporal trend was fitted to the days $\{1, ..., T\}$. This temporal trend therefore represented a seasonal effect. We specify *Model 4*,

$$\text{Model 4}: \quad \eta_{ijt} = \mathrm{B}_{ijt} + \epsilon^r(t), \tag{4.8}$$

where $\epsilon^r(t)$ represent the random walk process of order $r \in \{1, 2\}$, which was controlled by the precision parameter τ (4.7).

A second and more realistic GLM assumed that the time trend could be different for each year, but still sharing the same precision parameter τ. This seemed more likely considering the large variability in annual avalanche frequency and snow fall. Each year was therefore assumed to have a conditionally independent random walk, so that $\epsilon^r(t) = \epsilon^r_j(t)$, where j denotes the year. We specify *Model 5*

$$\text{Model 5}: \quad \eta_{ijt} = \mathrm{B}_{ijt} + \epsilon^r_j(t), \tag{4.9}$$

Finally, in the last version we assumed a seasonal effect, as in Model 4. However, the seasonal effect was allowed to affect each stretch differently, through a stretch-specific weight γ_i. This was also reasonable, considering that the average number of avalanches was different for the various stretches. Each stretch was considered to have the same time trend but scaled with the stretch-specific weight, γ_i. Note that the weights γ_i were also Stage 3 hyperparameters, so that $\psi = (\tau, \gamma)$. We specify *Model 6*

$$\text{Model 6} : \eta_{ijt} = \mathrm{B}_{ijt} + \gamma_i \epsilon^r(t). \tag{4.10}$$

Recall that, since $r \in \{1, 2\}$, this implies that each of the LGMs are tested in two different variants using first a rw1 and then rw2 process. The models will therefore be denoted using (a) for rw1 and (b) for rw2. So that Model 4a denotes Model 4 with a rw1 sub-model etc.

4.4 Model fit criteria

The deviance is a useful measure to compare the models in terms of fit, whether frequentist or Bayesian. Following the definition given in Blangiardo and Cameletti (2015):p. 169, the deviance is given by

$$\mathrm{D}(\theta) = -2\log p(\boldsymbol{y}|\theta), \tag{4.11}$$

where $p(\boldsymbol{y}|\theta)$ is the likelihood of θ given the responses $\boldsymbol{y}$. Since the deviance was general, it could be used both for the Poisson regression models as well as for the LGMs. We define the different criteria used for both the model classes. Note that the criteria could not be used to compare the Poisson regression models with the LGMs.

The relative quality of the different GLMs was compared using the Akaike information criterion (AIC) (Akaike, 1973). If we let $\hat{\boldsymbol{\beta}}$ be the maximum likelihood estimates of the parameters $\boldsymbol{\beta}$. Then the AIC is defined as

$$\mathrm{AIC} = 2d + \mathrm{D}(\hat{\boldsymbol{\beta}}), \tag{4.12}$$

where d is the number of estimated parameters and is used as a measure of the model complexity. The model with the lowest AIC value is in general believed to be better supported by the data.

Correspondingly, the deviance information criterion (DIC) (Spiegelhalter et al., 2002), which is a generalization of the AIC, was used to compare the LGMs. This is the most commonly used measure of model fit based on the deviance for Bayesian models. The DIC is defined as the sum of the estimated deviance, $\bar{\mathrm{D}}(\boldsymbol{\theta}) = \mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{y}}[\mathrm{D}(\boldsymbol{\theta})]$, and effective number of parameters given by

$$p_{\mathrm{D}} = \mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{y}}[\mathrm{D}(\boldsymbol{\theta})] - \mathrm{D}(\mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{y}}[\boldsymbol{\theta}]) = \bar{\mathrm{D}} - \mathrm{D}(\bar{\boldsymbol{\theta}}). \tag{4.13}$$

The DIC thus become,

$$\mathrm{DIC} = \bar{\mathrm{D}} + p_{\mathrm{D}}, \tag{4.14}$$

and the models with the smaller DIC were better supported by the data, as for the AIC.

In addition, the marginal likelihood, $p(\boldsymbol{y})$, was used to compare the LGMs. The quantity was approximated in the INLA framework by integrating the Laplace approximation (3.23) with respect to the hyperparameter $\boldsymbol{\psi}$ (Rue et al., 2009), that is

$$\tilde{p}(\boldsymbol{y}) = \int \tilde{p}(\boldsymbol{\psi}|\boldsymbol{y}) d\boldsymbol{\psi}. \tag{4.15}$$

The logarithm of the marginal likelihoods was used for the comparison and the model with the largest value is considered to fit the data better.

4.5 Skill scores

We used the ranked probability score (RPS) (Epstein, 1969) to evaluate the accuracy of the predictions resulting from the different models. RPS is a categorical probabilistic forecast verification metric that was both strictly proper (Murphy, 1969) and sensitive to distance (Holstein, 1970). The error was measured through estimation of the mean squared error (MSE), which was, on the other hand, deterministic and continuous. In addition, it was decided to adopt the scores based on binary outcomes that were used in the SVV report, Juvik et al. (2015). We let $\boldsymbol{\lambda}^*$ denote the vector of n_{pred} predictions with elements $\{\lambda^*_{ijt}\}$, where i, j and t index the stretches, years and days for which avalanche activity is predicted. Using this notation the following definitions and specifications apply.

The RPS is defined as the sum of square differences between the cumulative distribution of the forecast and the observation (Murphy, 1971). If we let k denote the the possible forecasts, i.e. the number of avalanche along a

stretch on a given day, the RPS becomes

$$\mathrm{RPS} = \sum_{k=1}^{\infty} (F(k) - \mathbb{I}(y \leq k))^2, \tag{4.16}$$

where $F(k)$ is the cumulative forecast distribution and $\mathbb{I}(y \leq k)$ is the indicator function giving the cumulative observation probability of the observation y. The score is in the range $[0, \infty)$ with 0 being the optimal score.

In order to compute the RPS we needed to select an upper limit, K, representing the highest possible number of avalanches along stretch i on a given day $\{j, t\}$. The value of K was in general set to be 50 as this provided a sufficient upper limit for all forecasts smaller than 20, hence covering nearly all cases. For the rare cases where a larger number of avalanches was forecasted, K was selected such that the probability of forecasting $K+1$ avalanches was less than $10e-8$. The cumulative forecast probability was given by the cumulative Poisson distribution with parameter λ^*_{ijt}, i.e. $F(k) = F(k; \lambda^*_{ijt})$. This score was averaged over all predictions so that

$$\overline{\mathrm{RPS}} = \frac{1}{n_{pred}} \sum_{\lambda^*_{ijt} \in \boldsymbol{\lambda}^*} \sum_{k=1}^{K} (F(k; \lambda^*_{ijt}) - \mathbb{I}(y_{ijt} \leq k))^2. \tag{4.17}$$

The mean square error was estimated using

$$\widehat{\mathrm{MSE}} = \frac{1}{n_{pred}} \sum_{\lambda^*_{ijt} \in \boldsymbol{\lambda}^*} (\lambda^*_{ijt} - y_{ijt})^2, \tag{4.18}$$

i.e. the square sum of difference between the predicted average number of avalanches λ^*_{ijt} and the observation y_{ijt}.

4.5.1 Skill scores used by the SVV

The scores suggested by SVV in Juvik et al. (2015) was based on binary outcomes, were avalanche activity was either forecasted or not, depending on a preset threshold. If the prediction λ^*_{ijt} was above this threshold avalanche activity was forecasted and similarly no avalanche was forecasted if the prediction was below this threshold value. The selection of the threshold value will be discussed in more detail in Section 4.6. Based on this threshold there were four possible validation outcomes that are displayed in Table 4.2.

Table 4.2: Definition of validation outcomes.

Forecasted	**Observed**	
	Avalanche	¬ Avalanche
Avalanche	a (true positive)	b (false positive)
¬ Avalanche	c (false negative)	d (true negative)

The quantities a, b, c and d denote the number of outcomes in the respective categories during validation.

We define the skill scores as reported by Juvik et al. (2015). The *Hit Rate* (HR) is defined as the proportion of correct forecast for both days with and without avalanche activity,

$$\mathrm{HR} = \frac{a+d}{a+b+c+d}. \tag{4.19}$$

The value lies in the range $[0,1]$, with 1 being the optimal score. However, the HR is considered to be unreliable for rare events such as avalanches, since forecasting no avalanches will be beneficial.

Next, we define the *Probability of Detection* (POD) as the probability of forecasting avalanche activity before it occurs,

$$\mathrm{POD} = \frac{a}{a+c}. \tag{4.20}$$

The value lies in the range $[0,1]$, with 1 being the optimal score.

The *Unweighted Average Accuracy* (UAA) is defined as the average of the accuracy of the predictions for both days with and without avalanche activity,

$$\mathrm{UAA} = \frac{1}{2}\left(\frac{a}{a+c} + \frac{d}{b+d}\right). \tag{4.21}$$

The value lies in the range $[0,1]$, with 1 being the optimal score. This is considered to be a better skill score for rare events, since both events a and d are equally weighted.

The *Pierce Skill Score* (PSS) is a measure of the accuracy relative to a unbiased random reference forecast, i.e.

$$\mathrm{PSS} = \frac{a}{a+c} - \frac{d}{b+d}. \tag{4.22}$$

The value lies in the range $[-1, 1]$, with 1 indicating a perfect forecast. Moreover, 0 indicates that the model performs equivalently to a model forecasting the same every day.

The *Bias* is defined such that values above 1 indicate that an event is forecasted more often than it occurs, while values below 1 indicate the opposite,

$$\text{Bias} = \frac{a+b}{a+c}. \tag{4.23}$$

The value lies in the range $[0, \infty)$, with 1 being the optimal score.

4.6 Inference and validation schemes

The different GLMs were evaluated and compared, using k-fold cross validation. We let each fold correspond to a year, so that one year at a time was used as test set, while the remaining years were used to train the models. Since the focus was on predicting avalanches, all years without avalanche observations were excluded. This resulted in 34 years. Hence, the procedure was repeated 34 times for each model. The skill scores were computed for each repetition and the average was used as a general measure of the models predictive performance. In order to compute the different scores suggested by SVV, we needed to set a threshold value, deciding whether an avalanche was to be forecasted or not. The threshold was set such that the Bias (4.23) was approximately equal for all the models, to ease comparison.

The complexity of the latent models caused the computations related to the inference to be expensive. The cross validation procedure was therefore no longer feasible. It was instead decided to validate the different models using a single training set $\mathcal{D}_{train}$ to train the models and a single test set $\mathcal{D}_{test}$ to evaluate the predictions resulting from the trained model. To further reduce the computational costs only a subset of the complete data set $\mathcal{D} = \{\boldsymbol{y}, \boldsymbol{X}\}$ was selected for the validation procedure, i.e. $(\mathcal{D}_{train} \cup \mathcal{D}_{test}) \subset \mathcal{D}$. The two sets were selected so that they were mutually exclusive, i.e. $\mathcal{D}_{test} \cap \mathcal{D}_{train} = \emptyset$, to prevent overfitting. Note that this validation procedure also had to be carried out for the Poisson regression models, in order to compare the GLMs and LGMs.

We continue by describing how the test set and training set were selected. Firstly, we used the five stretches with strongest correlations in terms of avalanche occurrence. Hence, the spatio-temporal dependencies were more likely to have an effect. Recall, Figure 2.11, which display all pairs of

stretches that have a correlation above 0.18. As can be seen, the areas A3, A4 and A5 form a group of correlated stretches. Moreover both A3 and A6 were well correlated with the stretches A8, A9 and A10. However, A10 is less correlated with A8 and A9. Based on this the the five stretches A3, A4, A5, A8 and A9 were selected. Secondly, to further reduce the computational expenses it was decided to let the training set consist of 10 selected years: 1986, 1991, 1993, 1994, 1995, 1997, 1998, 2002, 2011 and 2012. These years were chosen to include varying numbers of avalanches. However, a minimum of 4 avalanches per year was required, in total for the 5 stretches, as the focus was on predicting avalanche *occurences*. Similarly the test set was constructed of the three years: 1990, 1999 and 2000 and used to evaluate the different models' prediction accuracy. For the three years it occurred 12 avalanches (few), 37 avalanches (moderate) and 62 avalanches (many), respectively. Thereby allowing assessment of the models' performance in the different cases.

The spatio-temporal dependency from the random effects in the LGMs must be estimated every day, based on data from the previous days the same year. The model therefore needs to be run again for each day in the test set. The associated computational cost is high, since the whole test set comprises 361 days. This would not, however, be a problem for SVV operationally. The mode of the precision parameter, τ (4.7), from the trained model was kept fixed, mainly to reduce computation time. Thus, for each day a prediction λ^*_{ijt} was made for each of the five stretches, based on the corresponding weather covariates and the previously trained model. The data set was then updated with the observed values y_{ijt}. and the procedure was repeated. In this way it was prevented that future information falsely enhanced the prediction estimates. The same routine was not required for predictions for the GLMs, since no information exchange with the assumed future observations would occur. These predictions could therefore be computed as normal.

The iterative prediction and validation procedure for the LGMs is illustrated in Algorithm 1. Here l was the index iterating through the days of the test set. Thus, each value of l corresponded to a unique combination of the index $\{j, t\}$ identifying the year and day for which the prediction, $\boldsymbol{\lambda}^*_l$, was made. Note that $\boldsymbol{\lambda}^*_l = (\lambda^*_{3jt}, \lambda^*_{4jt}, \lambda^*_{5jt}, \lambda^*_{8jt}, \lambda^*_{9jt})$, corresponding to the predictions for each of the selected stretches, on the given day. The predictions $\boldsymbol{\lambda}^*_l$ were taken to be the exponentiated mode of the linear predictors that was estimated using INLA. In addition, the threshold required to compute the scores for binary outcomes was selected such that the Bias, see equation (4.23), was approximately equal for the different models, as for the cross

validation earlier.

Algorithm 1 LGM validation

1: **procedure** VALIDATION PROCEDURE($\mathcal{D}_{train}$, $\mathcal{D}_{test}$)
2: **for** each day l **do**
3: make prediction $\boldsymbol{\lambda}_{*l}$ based on day $1 : (l-1)$
4: update $\mathcal{D}_{train}$ with observation y_l
5: compute skill scores for predicted values $\boldsymbol{\lambda}^*$

4.6.1 Software

All analysis, data handling, as well as all computations related to the model were done using R (R Development Core Team, 2008). The GLM inference and prediction were carried out using the *glm()* and *predict.glm()* functions which was available in the R environment. Similarly the *inla()* function, available from the R-INLA package, see Rue et al. (2009) and Martins et al. (2012), was used for the inference and also prediction related to the LGMs. In addition the *rgdal* package (Bivand et al., 2015) was useful in the treatment of the spatial data. At last, the *ggmap* package (Kahle and Wickham, 2013) deserves a mention for the usefulness when plotting coordinates and maps.

5 Results

This chapter presents the results for the models investigated in this study. We start with the coefficients, AIC and results from the cross validation for the Poisson regression models. Thereafter, we show examples of the temporal random effects that were found using the LGMs. We also present the DIC and marginal log-likelihood, as well as the estimated random walk precision, τ, for the different LGMs. Finally, we present the results of the predictive performance for all the models of interest, both for each of the years; 1990 (few avalanches), 1999 (moderate) and 2000 (many) individually, as well as for the joint test set.

5.1 Poisson regression

5.1.1 Selection of coefficients

The estimated regression coefficients for Model 1, 2 and 3 are found in Tables 5.1, 5.2 and 5.3, respectively. The gray entries indicate coefficients that were not significant at a 0.05 level. For more details, consult Table A.1, A.2 and A.3 in Appendix A, which display the coefficients and corresponding standard error.

In the basic model (Model 1) both the explanatory variables were significant for most of the stretches. The snow depth difference was only found to be nonsignificant for the single stretch A4, while the snow depth was nonsignificant for the three stretches A1, A2 and A10. Both weather covariates therefore seemed to be informative for the modelling of avalanches.

In Model 2, all the available explanatory variables where tested simultaneously. That is, both the weather covariates as well as the four spatio-temporal explanatory variables. Note that no normalization of the variables was used. From Table 5.2 we see that the snow depth difference is the explanatory variable that is significant for the most stretches. Moreover, both the variables p_{jt}^5 and c_{ijt}^5, were found significant for seven of the stretches each. These variables were the number of avalanches the 5 previous days both for the area and single stretches. The variables p_{jt}^1 and c_{ijt}^1, one the other hand were significant for two and six stretches, respectively. These represented the number of avalanches on the previous day for the area and single stretches. It therefore seemed that the day-to-day dependence might be less informative

than the smoother 5 day interval. It should also be noted that the inclusion of the additional explanatory variables caused the weather coefficients to be smaller and of less significance.

Based on the results for Model 2, we excluded the explanatory variables for the day-to-day dependence, p_{jt}^1 and ς_{ijt}^1, in the updated spatio-temporal Poisson regression model. For Model 3, all coefficients were significant for at least seven of the stretches except for the snow depth, which was only significant for three of the stretches. This is displayed in Table 5.3.

Table 5.1: Coefficients for Model 1, which acted as the basic model depending only on weather covariates, fitted to the full data set. The coefficients that were not significant at a 0.05 level are marked gray.

	Stretch									
	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
$\mathbf{1}$	-4.14	-5.44	-4.23	-5.39	-5.22	-4.69	-5.80	-3.42	-5.27	-5.00
$\boldsymbol{s}_i$	-0.25	0.36	1.97	1.23	2.40	0.60	0.77	0.50	0.48	0.10
$\Delta\boldsymbol{s}_i$	0.08	0.15	0.07	0.03	0.13	0.09	0.05	0.11	0.12	0.10

Table 5.2: Coefficients for Model 2, where all the explanatory variables were included to investigate which were more significant, fitted to the full data set. The coefficients that were not significant at a 0.05 level are marked gray.

	Stretch									
	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
$\mathbf{1}$	-4.44	-6.02	-4.27	-5.60	-5.27	-4.86	-5.90	-3.52	-5.70	-5.02
$\boldsymbol{s}_i$	-0.17	0.52	1.15	1.13	1.87	0.36	0.50	0.27	0.33	-0.19
$\Delta\boldsymbol{s}_i$	0.08	0.15	0.04	0.02	0.10	0.08	0.07	0.10	0.11	0.10
$\boldsymbol{p}^1$	0.25	0.59	-0.02	-0.12	0.01	-0.18	0.09	-0.01	0.02	-0.08
$\boldsymbol{p}^5$	-0.16	-0.24	0.12	0.15	0.08	0.09	0.04	0.02	0.21	0.12
ς_i^1	-0.96	-2.09	0.31	1.74	0.50	0.50	-0.11	0.06	0.10	0.38
ς_i^5	1.25	1.25	-0.05	-0.93	0.05	0.42	0.79	0.39	-0.49	0.16

5.1.2 Model fit and predictive performance

The AIC for the three models are listed in Table 5.4. Both Model 2 and 3 have a smaller value than Model 1, thus suggesting that the additional explanatory variables improve the model fit. The AIC score for Model 2 and Model 3 was relatively similar, with Model 2 scoring slightly lower. However, since the extra explanatory variables in Model 2 were nonsignificant for many of the stretches, it is reasonable to assume that Model 3 fitted the data better.

Table 5.3: Coefficients for Model 3, which was the updated Poisson regression model, fitted to the full data set. The coefficients that were not significant at a 0.05 level are marked gray.

	Stretch									
	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
$\mathbf{1}$	-4.41	-5.91	-4.25	-5.46	-5.21	-4.84	-5.86	-3.52	-5.66	-5.03
s_i	-0.18	0.39	1.15	1.04	1.80	0.33	0.46	0.27	0.28	-0.17
Δs_i	0.08	0.15	0.05	0.01	0.11	0.08	0.08	0.10	0.11	0.10
p^5	-0.09	0.04	0.11	0.09	0.07	0.06	0.07	0.01	0.22	0.10
ς_i^5	1.01	0.54	0.03	0.11	0.24	0.50	0.75	0.40	-0.48	0.20

Table 5.4: AIC for the three models.

Model	AIC
1	6888.9
2	6264.3
3	6297.9

Table 5.5: Skill scores from cross validation for the three different GLM models.

Model	RPS	MSE	HR	POD	UAA	PSS	Bias
1	**0.02**	**4.15e-2**	**0.88**	0.35	0.62	0.24	30.4
2	15.44	3.13e6	**0.88**	**0.39**	**0.64**	**0.28**	30.4
3	5.63	4.44e5	**0.88**	**0.39**	**0.64**	**0.28**	30.4

The results from the cross validation are displayed in Table 5.5. The best result is marked with bold for each skill score. As can be seen, Model 1 had the best score in terms of RPS and MSE. In fact, the resulting values for these two scores was extremely large and therefore weak for Model 2 and 3, in comparison. This indicated that it was a problem with the cross validation. The cause of this will be explained shortly. Moreover, both Model 2 and 3 had slightly higher values for the POD, UAA and PSS score, which suggested that the models detected more avalanches in general. The large MSE, therefore, indicated that too many avalanches were being forecasted on some of the days. However, this was not an operational problem as the general interest was on determining whether avalanche activity would occur or not.

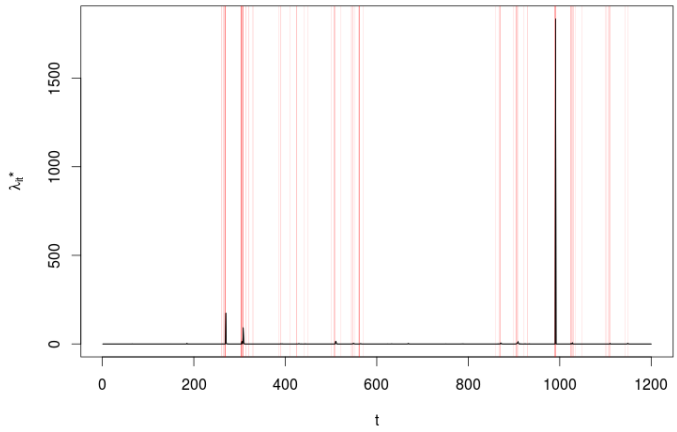


Figure 5.1: Predictions λ_{ijt}^* based on model 2 for the year 1994. Red vertical lines indicate avalanches.

After closer examination, it was found that the poor results for Model 2 and 3 was caused by the two years 1994 and 2011. Both years included several avalanche observations that were observed either on the same or on consecutive days. This was discovered when inspecting the RPS and MSE scores for the individual iterations of the cross validation algorithm, see Table B.1 and B.2 in Appendix B for more details. Therefore, when either of these years were used as the test set, and hence excluded from the training set, in the cross validation algorithm, it resulted in a large coefficient for the spatio-temporal explanatory variables. This caused the predictions to spike for the relevant consecutive days in the test set. The problem is illustrated in Figure 5.1, were the predictions based on Model 2, for 1994, are plotted. For one of the days it was predicted close to 2000 avalanches, even though eight actually occurred. Still, eight avalanches on the same day was unusual and since the training data did not include enough similar incidents, the estimated model coefficients for the temporal explanatory variables were too high.

To reduce the impact of unusually many avalanches we tested the model using the square root of the spatio-temporal explanatory variables instead, i.e. $\sqrt{p_{jt}^1}$, $\sqrt{p_{jt}^5}$, $\sqrt{c_{ijt}^1}$ and $\sqrt{c_{ijt}^5}$. This lowered the maximal values of these

variables, so that that overall the range of possible values would be smaller. Hence, decreasing the chance of the extreme predictions. The results from the cross validation with the modified explanatory variables are displayed in Table 5.6. The updated models are denoted 2.1 and 3.1 respectively. The modification improved the scores and all three models were now found to have identical RPS. The MSE was still lowest for Model 1 and the HR had decreased slightly for both the updated models. However, for the remaining scores for binary outcomes, the updated models performed clearly better than Model 1. Furthermore, Model 3.1 appeared to be slightly better than 2.1, as was expected. It should be noted that this was not a problem when evaluating the LGMs in Section 5.2.2, since the problematic years 1994 and 2011 was included in the training set $\mathcal{D}_{train}$.

Table 5.6: Skill scores from cross validation for the updated GLMs: 2.1 and 3.1.

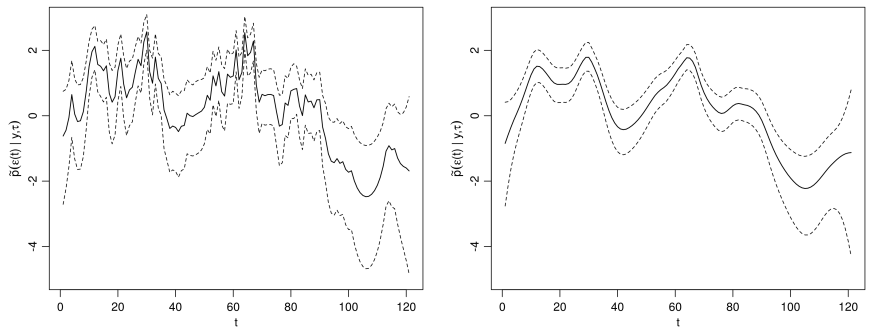
Model	RPS	MSE	HR	POD	UAA	PSS	Bias
1	**0.02**	**0.04**	**0.88**	0.35	0.62	0.24	30.4
2.1	**0.02**	0.09	0.87	0.45	0.66	0.32	30.4
3.1	**0.02**	0.05	0.86	**0.47**	**0.67**	**0.34**	30.4

To conclude, Model 3.1 was found to provide the most accurate predictions, in spite of the MSE and HR being better for Model 1. This suggested that the spatio-temporal explanatory variables could improve binary avalanche forecasts.

5.2 Latent Gaussian models

5.2.1 Temporal effects

The approximated temporally structured effects fitted to Model 4a and 4b are displayed in Figure 5.2a and 5.2b, respectively. In Model 4 all the years were assumed to have a common time trend that was shared between all the stretches. Recall, that 4a was the version using rw1 and 4b was the version using rw2. The rw1 process show distinctively more fluctuations than the smoother rw2 process. Both the effects increase overall during the first 30 days, corresponding to January, before revealing a drop in February. The effects then increase again, into March, before gradually wearing off in late March and April. This corresponded well with the reported monthly avalanche frequencies, see Figure 2.3b in Section 2.1.

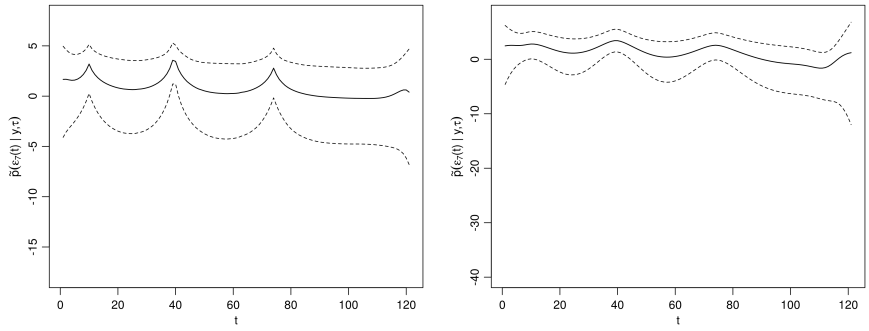


(a) Model 4a, random effect (rw1) (b) Model 4b, random effect (rw2)

Figure 5.2: Approximated posterior mode with 0.025 and 0.975 quantiles for the temporally structured effect fitted to Model 4a (left) and 4b (right).

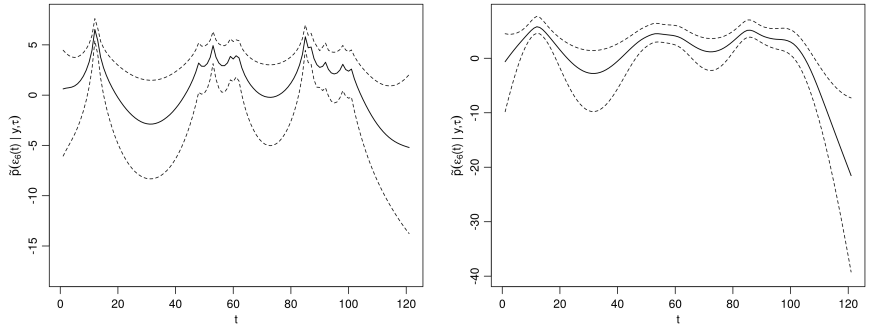
In Model 5a and 5b each year was considered to be a replicate of the same process, sharing a common precision parameter. This resulted in ten random effects with rw1 structure for Model 5a and equivalently ten random effects with rw2 structure for Model 5b. In Figures 5.3a and 5.3b the rw1 and rw2 for a year with few avalanches (1998) are displayed. We see that the effect is relatively flat with small fluctuations revealing the few avalanche incidents. Correspondingly, the Figures 5.4a and 5.4b show the random walk processes for a year with a moderate number of avalanches (1997). It is observed that the fluctuations increase in size and number due to the increased avalanche activity. Finally, Figure 5.5a and 5.5b show the random walk processes for a year with many avalanches (1994). The fluctuations increased again for the rw1 process. The rw2 process, however, is smoother with fewer fluctuations. This might be due to choice of prior.

Since Model 6 assumes that the stretches A3, A4, A5, A8 and A9 have the same random effect with a different scale parameter γ_i, the approximated random effects for Model 6a and 6b were similar to the effects fitted to Model 4a and 4b, respectively. The resulting scales are listed in Table 5.7. It was found that the scales were approximately equal in the two models. Furthermore, it can be seen that the scaling for both the stretches A4 and A5 were approximately 1, thus suggesting that the temporal effect was equally influential as for A3. This matched the ACF findings in Section 2.3.2, see Figure 2.9. The scale for A8, on the other hand, was approximately 0.5.



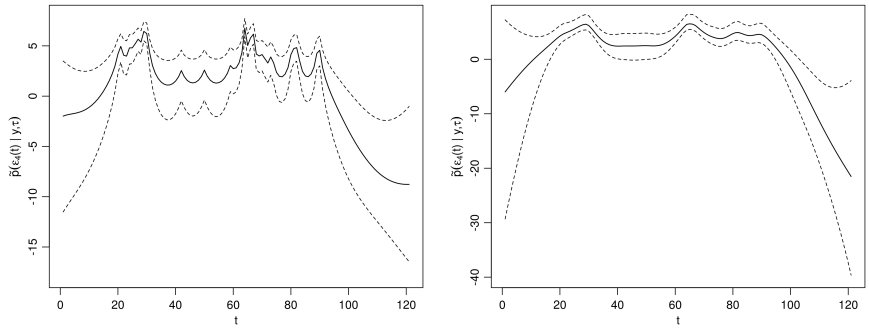
(a) Model 5a, random effect (rw1) (b) Model 5b, random effect (rw2)

Figure 5.3: Approximated posterior mode with 0.025 and 0.975 quantiles for the temporally structured effect for a year with few avalanche observations (1998) fitted to Model 5a (left) and 5b (right).



(a) Model 5a, random effect (rw1) (b) Model 5b, random effect (rw2)

Figure 5.4: Approximated posterior mode with 0.025 and 0.975 quantiles for the temporally structured effect for a year with moderate avalanche observations (1997) fitted to Model 5a (left) and 5b (right).



(a) Model 5a, random effect (rw1) (b) Model 5b, random effect (rw2)

Figure 5.5: Approximated posterior mode with 0.025 and 0.975 quantiles for the temporally structured effect for a year with many avalanche observations (1994) fitted to Model 5a (left) and 5b (right).

Hence, the temporal effect would have less impact on the predictions, in spite of this being the stretch with the most avalanche occurrences. The scale for A9 was the largest at approximately 1.4, indicating that the temporal effect was more influential for this stretch. Still the differences were small, which was reasonable considering that these were the stretches that were strongest correlated.

Table 5.7: Approximated scale parameters, γ_i, for Model 6a and 6b.

Model	Scale	Mean	0.025	0.975	Mode
6a	γ_4	1.03	0.66	1.40	1.03
	γ_5	1.04	0.70	1.39	1.03
	γ_8	0.49	0.29	0.71	0.49
	γ_9	1.36	0.97	1.76	1.35
6b	γ_4	0.96	0.57	1.36	0.96
	γ_5	1.03	0.67	1.40	1.03
	γ_8	0.41	0.21	0.62	0.40
	γ_9	1.35	0.94	1.76	1.34

5.2.2 Model selection

In Table 5.8 the DIC and marginal log-likelihood for each of the latent models are displayed. In general, the DIC and log-likelihood was better for the models with the rw1 random effect compared to their counterparts, where the rw2 was used instead. I.e. 4a, 5a and 6a had better scores than 4b, 5b and 6b respectively. This suggested that the temporal structure of the avalanche observations had more resemblance to a first-order random walk than a second-order. The best DIC and log-likelihood value was obtained for Model 5a, thus indicating that the rw1 random effect with yearly replicates might be the best model. Model 5b, which was the same model, but with the rw2 replacing the rw1, had a DIC score that was a lot worse than for the rest of the models. Still, the log-likelihood value was the second best. The poor DIC value was thought to be related to numerical instability. Therefore, it was decided to discard Model 5b from the prediction based model validation.

The precision parameter τ can be seen to be generally smaller and more accurately determined for the models with rw1. This implied that the variance was in general larger for these temporal random effects.

Table 5.8: DIC and approximated hyperparameter for the different LGM's

	Model fit		**Precision τ**			
Model	DIC	log-likelihood	Mean	0.025	0.975	mode
4a	2072.93	-1147.32	2.30	1.09	4.39	1.87
4b	2138.80	-1168.50	105.65	27.78	63.91	63.91
5a	**1567.24**	**-1015.17**	0.71	0.48	1.03	0.67
5b	5.59e12	-1105.84	20.79	8.27	42.5	16.33
6a	2051.14	-1136.76	1.87	0.85	3.63	1.52
6b	2101.86	-1157.03	33.46	6.53	96.75	16.86

5.3 Evaluation of predictive performance

In this section we present the skill scores resulting from the years 1990, 1999 and 2000 for Model 1, 3, 4a, 4b, 5a, 6b and 6b. Hence, the basic model can be compared to the other models that include various forms of spatial and temporal dependencies. In addition, it was of interest to compare Model 3, where the temporal dependencies were included using explanatory variables, with the LGMs to see whether the flexibility of the LGMs provided any gain on the accuracy. Finally, it was of interest to compare the LGMs against each

other.

The various resulting skill scores for 1990 are found in Table 5.9. Overall Model 5a, seemed to yield the best predictions, although the basic model (Model 1) yielded equally good values for the binary scores. The probability of detection was as high as 0.75 for Model 1 and Model 5a thus 8 out of 12 avalanches where correctly forecasted. Still, the Bias of 4 indicate that in general too many avalanches were predicted. Figure 5.6 illustrate the forecasts for Model 1, 3, 4b and 5a for stretch A9 this year. The red vertical lines indicate the actual avalanche observations with width proportional to the number of avalanches on a given day. The gray horizontal line indicate the threshold value used for the scores based on binary outcomes. Model 5a appears to fit the avalanche observations well for this year, explaining the low RPS and MSE scores. There was also an avalanche that non of the models could detect, hence illustrating the models' weakness.

Table 5.9: Skill scores for predictions made for the year with few avalanches (1990).

Model	RPS	MSE	HR	POD	UAA	PSS	Bias
1	0.026	0.046	**0.930**	**0.750**	**0.842**	**0.684**	4.000
3	0.023	0.039	0.920	0.500	0.714	0.429	4.000
4a	0.022	0.036	0.917	0.417	0.672	0.344	4.000
4b	0.021	0.035	0.923	0.583	0.757	0.514	4.000
5a	**0.018**	**0.030**	**0.930**	**0.750**	**0.842**	**0.684**	4.000
6a	0.022	0.038	0.927	0.667	0.799	0.599	4.000
6b	0.022	0.037	0.923	0.583	0.757	0.514	4.000

In Table 5.10 the resulting skill scores for 1999, a year with few avalanche observations, is listed. In general, the differences between the models were much smaller than for 1990. It can be seen that Model 3 and 4b had a slightly better RPS and MSE value. Moreover it can be seen that Model 1 and 3 had identical and best values for the binary scores. Among the GLMs Model 6a seemed to perform slightly better for the binary scores. The forecasts for stretch A8 is plotted in Figure 5.7. Almost all the avalanches occurred along this stretch this year. The plots show that the forecasts given by Model 1 and Model 3 were identical, explaining the identical scores. Also Model 4b had a similar forecast, however, the predictions for 5a are observed to be less accurate.

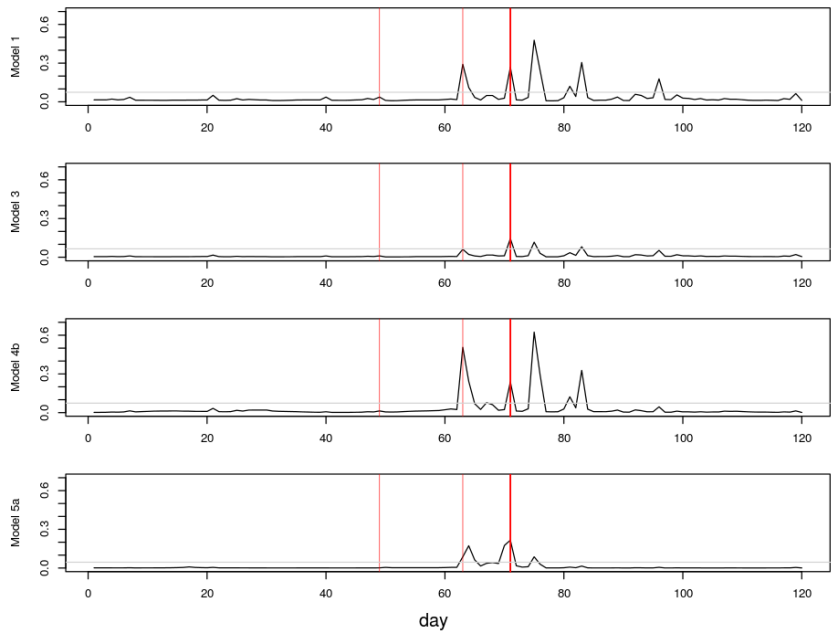


Figure 5.6: Predictions λ^*_{ijt} for the year with few avalanches (1990) for stretch A9. The observed avalanches are marked with red vertical lines. The grey horizontal line indicates the threshold value.

In Table 5.11 the skill scores for 2000, a year with a moderate number of avalanches, are listed. Model three scored clearly better than the rest of the models this year. Also Model 1 seemed to be competitive to the LGMs. Model 4b was again the LGM with the best RPS and MSE score. However, Model 5a yielded the best binary scores. The predictions for A3 are plotted in Figure 5.8. As for 1999, Model 1 and 3 yielded fairly similar predictions, but the temporal explanatory variables appeared to make a positive contribution. Especially for detecting several avalanches at the end of the time series. Model 5a also detected these avalanches and the predictions peak for these few days.

The skill scores resulting from all three years combined are listed in Table 5.12. Model 3 was clearly better than all the other models according to the skill scores, while Model 4b had the best RPS and MSE value of the GLMs. Model 5a was the second best model in terms of the skill scores for

Table 5.10: Skill scores for predictions made for the year with a moderate number of avalanches (1999).

Model	RPS	MSE	HR	POD	UAA	PSS	Bias
1	0.057	**0.090**	**0.788**	**0.784**	**0.786**	**0.572**	4.000
3	**0.056**	**0.090**	**0.788**	**0.784**	**0.786**	**0.572**	4.000
4a	0.058	0.092	0.775	0.676	0.729	0.457	4.000
4b	**0.056**	**0.090**	0.782	0.730	0.757	0.515	4.000
5a	0.057	0.091	0.782	0.730	0.757	0.515	4.000
6a	0.057	0.091	0.785	0.757	0.772	0.544	4.000
6b	0.057	0.091	0.782	0.730	0.757	0.515	4.000

Table 5.11: Skill scores for predictions made for the year with many avalanches (2000).

Model	RPS	MSE	HR	POD	UAA	PSS	Bias
1	0.095	0.182	0.630	0.694	0.658	0.316	4.000
3	**0.093**	**0.176**	**0.666**	**0.871**	**0.757**	**0.514**	4.000
4a	0.102	0.199	0.620	0.645	0.631	0.262	4.000
4b	0.096	0.186	0.613	0.613	0.613	0.226	4.000
5a	0.106	0.216	0.636	0.726	0.676	0.352	4.000
6a	0.099	0.192	0.613	0.613	0.613	0.226	4.000
6b	0.099	0.200	0.617	0.629	0.622	0.244	4.000

binary outcomes, but the worst in terms of RPS and MSE. Moreover, the basic model proved competitive having the third best results both in terms of RPS and MSE, as well as for the scores for the binary outcomes.

Table 5.12: Skill scores for predictions for the whole test set.

Model	RPS	MSE	HR	POD	UAA	PSS	Bias
1	0.059	0.106	0.778	0.694	0.739	0.477	4.000
3	**0.058**	**0.102**	**0.792**	**0.811**	**0.801**	**0.602**	4.000
4a	0.061	0.109	0.768	0.613	0.695	0.391	4.000
4b	**0.058**	0.104	0.775	0.667	0.724	0.448	4.000
5a	0.061	0.113	0.782	0.730	0.758	0.515	4.000
6a	0.060	0.107	0.772	0.649	0.715	0.429	4.000
6b	0.059	0.109	0.773	0.658	0.719	0.439	4.000

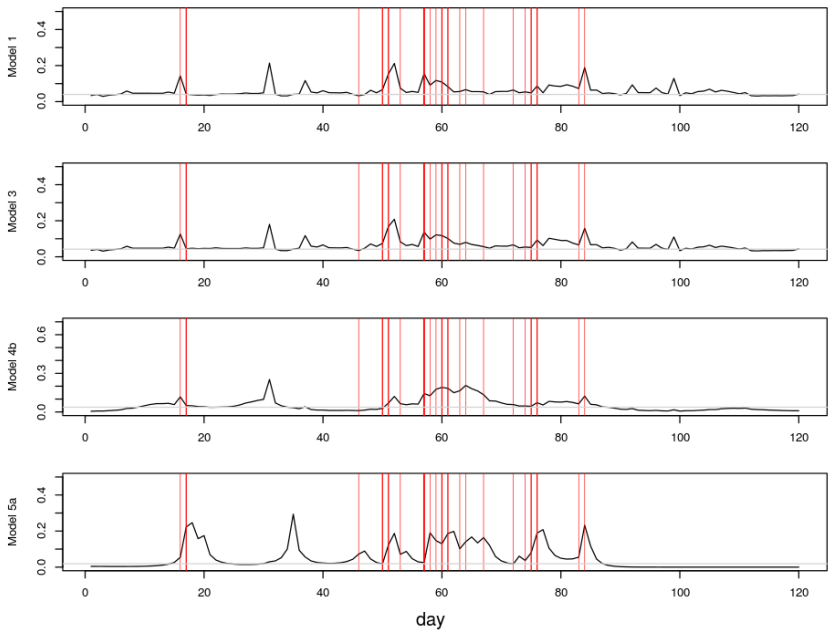


Figure 5.7: Predictions λ^*_{ijt} for the year with a moderate number of avalanches (1999), for stretch A8. The observed avalanches are marked with red vertical lines. The grey horizontal line indicates the threshold value.

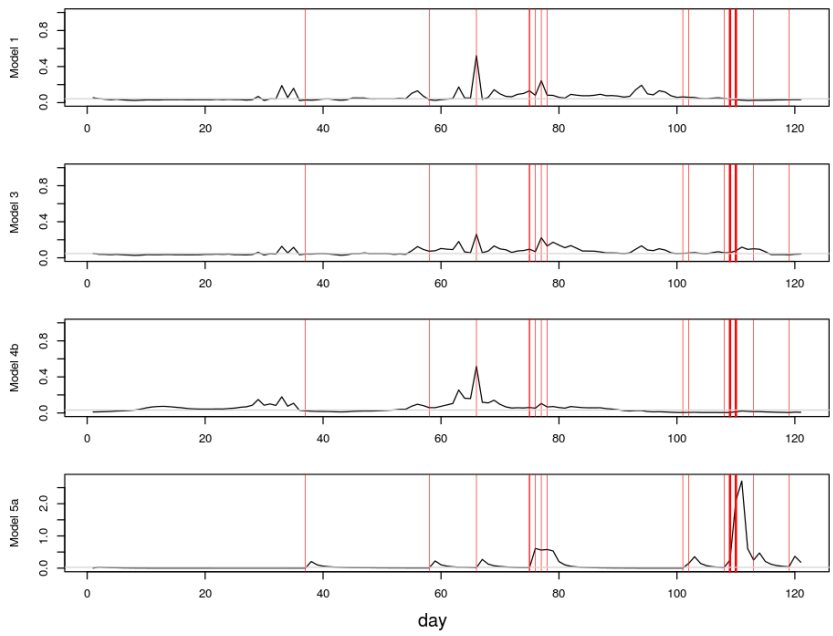


Figure 5.8: Predictions λ^*_{ijt} for the year with a moderate number of avalanches (2000), for stretch A3. The observed avalanches are marked with red vertical lines. The grey horizontal line indicates the threshold value. Note that the y-axis is different for Model 5a, to fully display the predictions for the last days.

6 Concluding remarks

6.1 Discussion and main results

In this study we investigated whether spatial and temporal dependencies could be utilized to improve the accuracy of daily probabilistic avalanche predictions at a stretch of road scale. Spatio-temporal dependencies were included as explanatory variables in a generalized linear model framework. In addition, temporal random effects were included in the extended latent Gaussian model framework. The temporal trend was estimated based on the best correlated stretches, thus exploiting spatial dependencies. This was done to examine whether the increased flexibility of these models could further improve the predictions. Our findings suggested that the inclusion of spatio-temporal dependencies through explanatory variables improved the avalanche prediction accuracy. However, the inclusion of random effects was not found to further improve the predictions. In fact the best latent Gaussian model performance was comparable to the basic model, that was only based on two weather covariates.

Our first finding, that spatio-temporal dependencies improved the avalanche predictions was supported by our evaluation of the predictive performance of the various models. Model 3, which was the preferred model that included spatio-temporal explanatory variables, performed better than the basic model (model 1) in terms of the selected skill scores. Since Model 3 also performed better than all the latent Gaussian models, this demonstrated that the random effects did not further improve the predictions. However, the differences were small for most of the skill scores, indicating that the improvements were limited. The cross validation, used to compare the generalized linear models, gave less evidence of improvements as the basic model scored slightly better both in terms of MSE and HR. Still, HR is considered an unreliable score for rare events and both Model 2 and 3 yielded better results for the POD, UAA and PSS, indicating that the binary forecasts were better. This was in general more important, since the focus would be on whether avalanches occurred or not, rather than how many in operational use.

Our study partly extends the previous findings in Eckert et al. (2010) and Lavigne et al. (2011), showing that spatio-temporal dependencies might also be intrinsic for daily predictions. However, these results were obtained in

the GLM framework and could not analogously be shown for the Bayesian hierarchical models. This is therefore inconsistent with the previous findings, where Bayesian hierarchical models were used. However, there were several factors distinguishing the models and data in this study from the previous ones. Thus, making direct comparison challenging. Most importantly, the daily temporal effects might be more difficult to identify than the annual fluctuations that was in focus in the previous works.

There could be several other reasons as to why the LGM's did not improve the prediction accuracy. We point out the most important ones here. Firstly, the models tested in this study were different from the ones in previous studies (Eckert et al., 2010; Lavigne et al., 2011). For instance no distinct spatial random effects were included, since the spatial correlation between the stretches was in general weak. Secondly, in comparison to the previous studies which were based on data from the French Alps, this study was carried out in Norway, thus differing both topographically and climatically. Thirdly, the model assumptions were also slightly different, focusing on stretches of roads rather than townships. In addition, it could be a problem that the selected random walk process did not sufficiently resemble the underlying temporal process and other models should be tested. As a final point, the combination of stretch-specific weights and conditionally independent yearly temporal effects might have improved the results. However, this was not tested in this study as the increased model complexity did not fit the INLA framework.

In conclusion, this study demonstrates that spatio-temporal dependencies can be included as explanatory variables to improve prediction accuracy. This shows that spatio-temporal dependencies provide extra insight, that can be useful to secure the Norwegian road network against the hazards posed by avalanches. Moreover, this is a simple model that can easily be applied, provided that the stretches of road of interest are sufficiently similar in terms of topography and weather and that sufficient data are available.

6.2 Data quality

As previously mentioned in Section 2.1.2, conversations with Tveit (2016) from SVV and Haslestad (2016) from NVE revealed that poor data quality was an issue. However, this was first fully discovered after the results were ready. This was consequently an important source of error, that should be considered in future studies to give more credible results. For this reason, we give a detailed account of the problem.

The avalanche data used in this study were observations from a relatively large area and spanning almost 40 years. Hence, different personnel have been involved in the observation procedure. This has resulted in local variations, and strongly varying data quality is known to be a widespread problem due to under-reporting according to Tveit (2016) and Haslestad (2016). Furthermore, under-reporting is especially known to occur when several avalanches happen on the same day, along the same stretch. This inflicts a problem with the assumption of Poisson distributed observations.

According to the same sources, it does not exist any record of when avalanche registration was initiated along the different stretches. Indeed, there is no available collective overview of when the different roads was built. Consequently there is no certain way to decide how far back avalanche observations can be reliably used. In Table 6.1 the year of the first registered avalanche observation for each stretch is listed. The overall earliest registration was in 1977. However, for many of the other stretches the first registration was several years later. For example, the first registered avalanche observation in domain A2 was from 2000. According to Tveit (2016), this was probably related to change of personnel that took place the same year. Hence, the lack of avalanches before this year was probably due to under-reporting. This shows that large scale under-reporting was in fact a problem and more careful investigation is required to determine which data that can be used. The problem of under-reporting is also supported by Figure 2.8 in Section 2.3.1. As can be seen, it appeared to be more snow than in any other year in 1981, still only a few avalanches were observed. Again suggesting that under-reporting occurred.

Another problem arises, as the avalanche observations span several years. For roads where avalanches pose a big threat, it is likely that avalanche protection measures could be installed. Hence, causing the data basis to change. An example of this was found for stretch A3, where a tunnel and an avalanche gallery had been built in 2002. The observed avalanche occurrences, were seen to drop significantly after this year, see Figure 2.5 in Section 2.1.2. This clearly demonstrates how the premisses changes. Moreover, since it is difficult to get an overview of these modifications for historic data, this represents another issue regarding the data quality.

Also the snow depth weather covariate is prone to uncertainty, being interpolated values on a square kilometer grid. According to Saloranta (2016), it had been found that the present version of the seNorge snow depth model

Table 6.1: The year of the first registered avalanche observation for the different stretches of road.

Stretch	First Observation
A1	1985
A2	2000
A3	1986
A4	1986
A5	1980
A6	2000
A7	1979
A8	1979
A9	1979
A10	1977

often predicted too high temperatures at the mountains near the fjords along the western coast of Norway, resulting in too modest snow depths. This will also affect the models ability to properly explain the data.

In general the under-reporting and avalanche protection measures yielded a unreliable data set, with fewer avalanche observations than what really occurred. In addition, since the snow model probably estimated too little snow, this is likely overall weaken the models' predictive performance. Therefore, by taking these issues into account in future studies better and more realistic results are expected.

6.3 Future work

In this study data quality has been pointed out as an issue that affected the credibility of the results. We will present here some recent developments that might be used in future studies to improve on this issue. Furthermore, we will give some suggestions of alternative models that should also be explored.

6.3.1 Improving data quality

More realistic snow depth data might be obtained using the new version of the seNorge snow depth model that was released at the end of 2015 according to (Saloranta, 2016). In the new version updated Bayesian interpolation

methods are used in interpolation of precipitation and air temperature to improve the estimates of the snow depth. Furthermore, research exploring the possibility of utilizing electronic systems for surveying avalanche activity has been growing in the last few years. For example Eckerstorfer and Malnes (2015) has shown good ability to detect avalanches using satellite-borne high resolution synthetic aperture radars. These radars are insensitive to clouds and bad weather, which is typical for periods where avalanche intensity is large. Moreover, it allows avalanches to also be detected during the night. Such a system could potentially replace the contractors role as avalanche observers and provide more consistent and accurate observations over larger areas. In this way, avalanches that does not interfere with the road network additionally be detected.

6.3.2 Alternative models

As of today, avalanche forecasts are provided at a regional basis in Norway, as a danger level scale ranging from 1-5. The appropriate danger level for a specific region is made by experts based on field observations and weather conditions (Norges vassdrags- og energidirektorat, 2013). In order to increase a statistical models ability to explain the avalanche occurrences it should be considered to include the forecasted danger levels as a covariate. This was unfortunately not possible at the time when this thesis was written, since the forecast service had only been operational for 2015 for our region.

In future works the models should also be tested for other regions to investigate whether comparable results can be obtained. Alternative temporal models, such as the autoregressive (AR) model should be explored in the LGM framework and closer examinations of the underlying spatial structures should be attempted, for instance by including spatial random effects in the LGM. Other hierarchical structures were also possible, e.g. non-linear effects for the intercept, the explanatory variables or both. New area partitions should also be assessed, as the assumption of homogeneous behavior of all the avalanches paths along a stretch is probably not true for longer stretches, that run along several mountains. Moreover, given sufficient availability of data, the SVM approach (Pozdnoukhov et al., 2011), mentioned in the introduction, should be tested as an alternative, for comparison.

A Extended results on significance of coefficients

Table A.1: Regression coefficients and standard errors for Model 1.

Covariate	Stretch	Estimate	Std. Error
$\mathbf{1}$	1	-4.14	0.29
	2	-5.44	0.34
	3	-4.23	0.14
	4	-5.39	0.27
	5	-5.22	0.22
	6	-4.69	0.22
	7	-5.80	0.40
	8	-3.42	0.13
	9	-5.27	0.31
	10	-5.00	0.31
$\boldsymbol{s}_i$	1	-0.25	0.20
	2	0.36	0.50
	3	1.97	0.28
	4	1.23	0.26
	5	2.40	0.30
	6	0.60	0.25
	7	0.77	0.23
	8	0.50	0.14
	9	0.48	0.21
	10	0.10	0.32
$\Delta\boldsymbol{s}_i$	1	0.08	0.01
	2	0.15	0.03
	3	0.07	0.02
	4	0.03	0.03
	5	0.13	0.02
	6	0.09	0.02
	7	0.05	0.02
	8	0.11	0.01
	9	0.12	0.01
	10	0.10	0.02

Table A.2: Regression coefficients and standard errors for Model 2.

Covariate	Stretch	Estimate	Std. Error
$\mathbf{1}$	1	-4.44	0.30
	2	-6.02	0.43
	3	-4.27	0.15
	4	-5.60	0.29
	5	-5.27	0.22
	6	-4.86	0.24
	7	-5.90	0.42
	8	-3.52	0.13
	9	-5.70	0.38
	10	-5.02	0.32
$\boldsymbol{s}_i$	1	-0.17	0.19
	2	0.52	0.59
	3	1.15	0.35
	4	1.13	0.28
	5	1.87	0.35
	6	0.36	0.28
	7	0.50	0.24
	8	0.27	0.16
	9	0.33	0.25
	10	-0.19	0.35
$\Delta\boldsymbol{s}_i$	1	0.08	0.01
	2	0.15	0.03
	3	0.04	0.02
	4	0.02	0.03
	5	0.10	0.02
	6	0.08	0.02
	7	0.07	0.02
	8	0.10	0.01
	9	0.11	0.01
	10	0.10	0.02
$\boldsymbol{p}^1$	1	0.25	0.12
	2	0.59	0.17
	3	-0.02	0.05
	4	-0.12	0.10
	5	0.01	0.05
	6	-0.18	0.11
	7	0.09	0.07

	8	-0.01	0.05
	9	0.02	0.03
	10	-0.08	0.08
p^5	1	-0.16	0.07
	2	-0.24	0.15
	3	0.12	0.02
	4	0.15	0.02
	5	0.08	0.02
	6	0.09	0.02
	7	0.04	0.04
	8	0.02	0.02
	9	0.21	0.02
	10	0.12	0.02
ς_i^1	1	-0.96	0.44
	2	-2.09	0.53
	3	0.31	0.12
	4	1.74	0.48
	5	0.50	0.20
	6	0.50	0.24
	7	-0.11	0.26
	8	0.06	0.14
	9	0.10	0.19
	10	0.38	0.50
ς_i^5	1	1.25	0.24
	2	1.25	0.34
	3	-0.05	0.06
	4	-0.93	0.38
	5	0.05	0.12
	6	0.42	0.09
	7	0.79	0.11
	8	0.39	0.06
	9	-0.49	0.15
	10	0.16	0.27

Table A.3: Regression coefficients and standard errors for Model 3.

Covariate	Stretch	Estimate	Std. Error
$\mathbf{1}$	1	-4.41	0.30
	2	-5.91	0.42
	3	-4.25	0.15
	4	-5.46	0.28
	5	-5.21	0.22
	6	-4.84	0.24
	7	-5.86	0.41
	8	-3.52	0.13
	9	-5.66	0.37
	10	-5.03	0.33
$\boldsymbol{s}_i$	1	-0.18	0.19
	2	0.39	0.60
	3	1.15	0.35
	4	1.04	0.29
	5	1.80	0.35
	6	0.33	0.28
	7	0.46	0.24
	8	0.27	0.16
	9	0.28	0.25
	10	-0.17	0.35
$\Delta\boldsymbol{s}_i$	1	0.08	0.01
	2	0.15	0.03
	3	0.05	0.02
	4	0.01	0.03
	5	0.11	0.02
	6	0.08	0.02
	7	0.08	0.02
	8	0.10	0.01
	9	0.11	0.01
	10	0.10	0.02
$\boldsymbol{p}^5$	1	-0.09	0.06
	2	0.04	0.05
	3	0.11	0.02
	4	0.09	0.02
	5	0.07	0.02
	6	0.06	0.02
	7	0.07	0.03

	8	0.01	0.01
	9	0.22	0.02
	10	0.10	0.02
ς_i^5	1	1.01	0.21
	2	0.54	0.12
	3	0.03	0.05
	4	0.11	0.13
	5	0.24	0.09
	6	0.50	0.07
	7	0.75	0.09
	8	0.40	0.05
	9	-0.48	0.10
	10	0.20	0.22

B Extended results from cross validation

In Table B.1 and B.2 the RPS and MSE for each iteration in the cross validation of the GLM's is listed. It was found that the years 1994 and 2011, $j = 15$ and $j = 30$ respectively, yielded poor predictions for model 2 and 3. These models included temporal explanatory variables, that were found to be the problem cause. The years are marked with bold font in both tables and it can be seen that both the RPS and MSE is extremely large for these two years, compared to the rest. This indicated a weakness with the model and the temporal explanatory variables were modified so that instead the square root was used. The results for the updated models are given in the same Tables, denoted as model 2.1 and 2.3 and it can be seen that this roughly solves the problem.

Table B.1: Estimated RPS for each iteration in the cross validation, for the three generalized linear models. Model 2.1 and 3.1 are the updated versions of model 2 and 3, were the square root of the temporal explanatory variables is used instead. The years were the predictions were poor is marked with bold font.

Test set		**Model**			
j	1	2	3	2.1	3.1
1	1.1e-03	1.0e-03	1.0e-03	9.6e-04	9.6e-04
2	7.7e-03	7.6e-03	7.5e-03	8.1e-03	7.5e-03
3	1.2e-03	1.0e-03	1.0e-03	9.7e-04	9.6e-04
4	2.0e-02	6.1e-03	6.1e-03	3.7e-03	3.6e-03
5	2.8e-03	1.5e-03	1.6e-03	1.2e-03	1.3e-03
6	1.5e-03	1.2e-03	1.2e-03	1.0e-03	1.0e-03
7	6.5e-03	6.1e-03	6.1e-03	6.0e-03	6.0e-03
8	4.4e-02	4.2e-02	4.2e-02	4.3e-02	4.2e-02
9	1.2e-02	1.3e-02	1.3e-02	1.3e-02	1.3e-02
10	5.0e-03	5.0e-03	5.0e-03	5.0e-03	5.0e-03
11	1.1e-02	1.1e-02	1.1e-02	1.1e-02	1.1e-02
12	8.2e-03	8.2e-03	8.2e-03	8.3e-03	8.3e-03
13	4.1e-03	3.8e-03	3.8e-03	3.7e-03	3.7e-03
14	1.9e-02	1.9e-02	1.9e-02	1.8e-02	1.8e-02
15	**8.7e-02**	**1.9e+00**	**6.1e-01**	**1.6e-01**	**1.2e-01**
16	1.5e-02	1.5e-02	1.5e-02	1.6e-02	1.6e-02
17	2.9e-02	2.9e-02	3.0e-02	2.7e-02	3.0e-02
18	3.6e-03	3.5e-03	3.5e-03	3.5e-03	3.5e-03
19	3.0e-02	2.8e-02	2.8e-02	2.9e-02	2.8e-02
20	8.1e-02	8.1e-02	8.1e-02	7.7e-02	7.8e-02
21	2.1e-02	2.1e-02	2.1e-02	2.1e-02	2.1e-02
22	5.7e-02	5.9e-02	5.7e-02	5.9e-02	5.7e-02
23	3.5e-03	3.8e-03	3.6e-03	3.5e-03	3.5e-03
24	8.0e-03	8.1e-03	8.1e-03	8.2e-03	8.2e-03
25	1.2e-02	1.2e-02	1.2e-02	1.2e-02	1.2e-02
26	1.2e-02	1.2e-02	1.2e-02	1.1e-02	1.1e-02
27	7.5e-03	7.4e-03	7.3e-03	7.4e-03	7.4e-03
28	6.0e-03	6.0e-03	5.9e-03	6.0e-03	6.0e-03
29	8.1e-03	8.1e-03	7.7e-03	8.1e-03	7.6e-03
30	**4.9e-02**	**5.2e+02**	**1.9e+02**	**6.9e-02**	**7.1e-02**
31	1.9e-02	1.9e-02	1.9e-02	1.8e-02	1.9e-02
32	2.7e-03	2.6e-03	2.6e-03	2.6e-03	2.6e-03
33	5.8e-03	3.0e-03	2.9e-03	1.7e-03	1.7e-03
34	1.2e-02	1.1e-02	1.1e-02	1.2e-02	1.1e-02

Table B.2: Estimated MSE for each iteration in the cross validation, for the three generalized linear models. Model 2.1 and 3.1 are the updated versions of model 2 and 3, were the square root of the temporal explanatory variables is used instead. The years were the predictions were poor is marked with bold font.

Test set		**Model**			
j	1	2	3	2.1	3.1
1	1.1e-03	1.0e-03	1.0e-03	9.6e-04	9.7e-04
2	1.1e-02	1.1e-02	1.1e-02	1.2e-02	1.1e-02
3	1.2e-03	1.0e-03	1.0e-03	9.7e-04	9.7e-04
4	3.1e-02	8.4e-03	8.4e-03	5.6e-03	5.3e-03
5	4.4e-03	1.8e-03	1.9e-03	1.3e-03	1.4e-03
6	1.6e-03	1.2e-03	1.2e-03	1.0e-03	1.0e-03
7	8.3e-03	7.8e-03	7.8e-03	7.6e-03	7.6e-03
8	1.1e-01	1.1e-01	1.0e-01	1.1e-01	1.1e-01
9	2.4e-02	2.4e-02	2.4e-02	2.4e-02	2.4e-02
10	6.4e-03	6.4e-03	6.4e-03	6.4e-03	6.4e-03
11	1.8e-02	1.8e-02	1.8e-02	1.8e-02	1.8e-02
12	1.6e-02	1.6e-02	1.6e-02	1.6e-02	1.6e-02
13	4.3e-03	3.9e-03	3.9e-03	3.9e-03	3.9e-03
14	2.9e-02	2.9e-02	2.9e-02	2.8e-02	2.8e-02
15	**2.8e-01**	**2.8e+03**	**1.5e+02**	**1.8e+00**	**5.4e-01**
16	2.2e-02	2.2e-02	2.2e-02	2.2e-02	2.2e-02
17	9.8e-02	9.9e-02	1.1e-01	6.8e-02	1.1e-01
18	3.6e-03	3.6e-03	3.6e-03	3.6e-03	3.6e-03
19	4.7e-02	4.5e-02	4.4e-02	4.6e-02	4.4e-02
20	1.6e-01	1.6e-01	1.6e-01	1.5e-01	1.5e-01
21	5.4e-02	5.5e-02	5.4e-02	5.4e-02	5.4e-02
22	1.2e-01	1.4e-01	1.2e-01	1.4e-01	1.2e-01
23	5.2e-03	5.7e-03	5.2e-03	5.2e-03	5.1e-03
24	1.1e-02	1.1e-02	1.1e-02	1.1e-02	1.1e-02
25	1.3e-02	1.3e-02	1.3e-02	1.3e-02	1.3e-02
26	2.0e-02	1.9e-02	1.9e-02	1.9e-02	1.9e-02
27	1.2e-02	1.2e-02	1.2e-02	1.2e-02	1.2e-02
28	9.3e-03	9.2e-03	9.2e-03	9.3e-03	9.3e-03
29	1.5e-02	1.5e-02	1.4e-02	1.5e-02	1.4e-02
30	**2.0e-01**	**1.1e+08**	**1.5e+07**	**2.9e-01**	**3.2e-01**
31	2.4e-02	2.4e-02	2.4e-02	2.3e-02	2.3e-02
32	2.7e-03	2.6e-03	2.6e-03	2.6e-03	2.6e-03
33	2.4e-02	7.6e-03	7.1e-03	2.5e-03	2.6e-03
34	1.5e-02	1.3e-02	1.3e-02	1.7e-02	1.3e-02

Bibliography

Akaike, H. (1973). "Information theory and an extension of the maximum likelihood principle." In: *B. N. Petrov and F. Csáki (eds) Second international sympoisum on information theory. Akademia Kiado, Budapest*. Pp. 632–641 (cit. on p. 33).

Ballestad, J. (2015). "Effect of under-reporting on probabilistic avalanche predictions based on logistic regression model." (cit. on p. 1).

Bivand, R., T. Keitt, and B. Rowlingson (2015). *rgdal: Bindings for the Geospatial Data Abstraction Library*. R package version 1.0-7 (cit. on p. 39).

Blangiardo, M. and M. Cameletti (2015). *Spatial and Spati-temporal Bayesian Models with R-INLA*. 1st ed. John Wiley & Sons (cit. on pp. 23, 25, 26, 33).

Casella, G. and R. L. Berger (2002). *Statistical Inference*. 2nd ed. Wadsworth (cit. on p. 21).

Eckerstorfer, M. and E. Malnes (2015). "Manual detection of snow avalanche debris using high-resolution Radarsat-2 SAR images". In: *Cold Regions Science and Technology* 120, pp. 205–218 (cit. on p. 59).

Eckert, N., E. Parent, L. Belánger, and S. Garcia (2007). "Hierarchical Bayesian modelling for spatial analysis of the number of avalanche occurrences at the scale of the township." In: *Cold Regions Sciences and Technology* 50, pp. 97–112 (cit. on p. 1).

Eckert, N., E. Parent, R. Kies, and H. Baya (2010). "A spatio-temporal framework for assessing the fluctuations of avalanche occurrence resulting from climate change to 60 years of data in the northern French Alps." In: *Climatic Change* 101, pp. 515–553 (cit. on pp. 1, 55, 56).

Epstein, E. (1969). "A scoring system for probability forecasts of ranked categories." In: *Journal of Applied Meteorology* 8 (6), pp. 985–987 (cit. on p. 34).

Feller, W. (1968). *An Introduction to Probability Theory and its Applications*. 3rd ed. Vol. 1. John Wiley & Sons (cit. on p. 24).

Haslestad, A. (2016). personal communication (cit. on pp. 10, 29, 56, 57).

Hennum, A. A. (2015). "Statistical analysis of avalanche activity and weather at Senja, Norway" (cit. on pp. 1, 11).

Holstein, S. (1970).
"A family of strictly proper scoring rules which are sensitive to distance."
In: *Journal of Applied Meteorology* 9 (6), pp. 751–758 (cit. on p. 34).
Juvik, Eivind, Katharina Kahrs, and Tore Humstad (2015).
Snøskredvarsling med nærnabometoden. Test av den canadiske nærnabomodellen på skreddata fra Senja. (Cit. on pp. 1, 11, 34–36).
Kahle, D. and H. Wickham (2013).
"ggmap: Spatial Visualization with ggplot2".
In: *The R Journal* 5(1), pp. 144–161 (cit. on p. 39).
Lavigne, A., L. Bel, E. Parent, and N. Eckert (2011).
"Spatio-temporal modeling of avalanche frequencies in the French Alps."
In: *Procedia Environmental Sciences* 7, pp. 311–316 (cit. on pp. 1, 55, 56).
Lavigne, A., L. Bel, E. Parent, and N. Eckert (2012).
"A model for spatio-temporal clustering using multinomial probit regression: application to avalanche counts".
In: *Environmetrics* 23, pp. 522–534 (cit. on p. 1).
Lavigne, A., N. Eckert, L. Bel, and E. Parent (2015).
"Adding expert contributions to the spatiotemporal modelling of avalanche activity under different climatic influences."
In: *Journal of the Royal Statistical Society* 64, pp. 651–671 (cit. on p. 1).
Martins, T. G., D. Simpson, F. Lindgren, and H. Rue (2012).
"Bayesian computing with INLA: new features."
In: *Computational Statistics & Data Analysis* 67, pp. 68–83 (cit. on p. 39).
McCullagh, P. and J. A. Nelder (1989). *Generalized Linear Models*. 2nd ed.
White and Hall (cit. on p. 22).
Murphy, A. H. (1969). "On the "ranked probability score"."
In: *Journal of Applied Meteorology* 8 (6), pp. 988–989 (cit. on p. 34).
Murphy, A. H. (1971). "A Note on the Ranked Probability Score."
In: *Journal of Applied Meteorology* 10 (1), pp. 155–156 (cit. on p. 34).
Nelder, J. A. and R. W. M. Wedderburn (1972).
"Generalized Linear Models". In: *Journal of the Royal Statistical Society, Series A (General)* 135(3), pp. 370–384 (cit. on p. 21).
Norges vassdrags- og energidirektorat (2012).
Fellesportal for overvåknings- og varslingsdata. URL: www.xgeo.no
(cit. on p. 11).
Norges vassdrags- og energidirektorat (2013).
Varslingstjeneste for naturfarer i Norge. URL: www.varsom.no
(cit. on p. 59).
Pozdnoukhov, A., G. Matasci, M. Kanevski, and R. S. Purves (2011).
"Spatio-temporal avalanche forecasting with Support Vector Machines."

In: *Nautral Hazards and Earth System Sciences* 11 (2), pp. 367–382 (cit. on pp. 2, 59).

R Development Core Team (2008).
R: A Language and Environment for Statistical Computing.
R Foundation for Statistical Computing. Vienna, Austria (cit. on p. 39).

Rue, H. and L. Held (2005).
Gaussian Markov Random Fields: Theory and Applications. Vol. 104.
Chapman and Hall: London (cit. on pp. 24, 27).

Rue, H., S. Martino, and N. Chopin (2009).
"Approximate Bayesian inference for latent Gaussian model by using integrated nested Laplace approximations."
In: *Journal of Royal Statistical Society, Series B* 71 (2), pp. 319–392 (cit. on pp. 2, 25, 34, 39).

Saloranta, T. (2014).
New version (v.1.1.1) of the seNorge snow model and snow maps for Norway.
(Cit. on p. 11).

Saloranta, T. (2016). personal communication (cit. on pp. 57, 58).

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde (2002).
In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64(4), pp. 583–639 (cit. on p. 34).

Tveit, J. (2016). personal communication (cit. on pp. 10, 56, 57).