

Nikolai Skaue

An investigation into effects of weather and seasonal factors on response time for helicopter emergency services in Central Norway

Master's thesis in Master of Science (MSc) in Mathematical Sciences

Supervisor: Andreas Asheim

June 2022

Nikolai Skaue

An investigation into effects of weather and seasonal factors on response time for helicopter emergency services in Central Norway

Master's thesis in Master of Science (MSc) in Mathematical Sciences
Supervisor: Andreas Asheim
June 2022

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

Abstract

With a varying topography in Norway and weather that can change quickly, well structured emergency medical services (EMS) is important. By looking at completed primary missions by helicopter in Trøndelag and Møre & Romsdalen, we have looked at how much different factors influence the response time.

The missions are from the period 2014-2021, and are combined with hourly updated weather data from weather stations in the same region. Regression models are used to analyse the data, with distance as the response variable. Different regression models are used to solve the problem with heteroskedasticity errors.

Our final model shows an increase in distance for tailwind as time increase, while a lower ceiling and precipitation decreases the distance as time increase, especially during the winter. In addition, starting a mission in Ålesund gives a shorter distance compared to Rosten. With neutral weather conditions, the two helicopters at Rosten and Ålesund covers 91.5% of the population in Central Norway during a flying time of 35 minutes (response time of 45 minutes). With more difficult condition, the coverage was reduced to 86.4%.

These results indicates that taking weather data into consideration when dispatching emergency medical services can be useful.

Sammendrag

Med en varierende topografi i Norge og vær som kan endres kjapt, er et godt strukturert akuttmedisinsk tilbud viktig. Ved å se på fullførte primær oppdrag med helikopter i Trøndelag og Møre & Romsdalen, har vi sett på hvor mye ulike faktorer påvirker respons tiden.

Oppdragene er fra perioden 2014-2021, og er kombinert med timesbasert værdata fra værstasjoner i samme område. Regresjonsmodeller er brukt for å analysere dataen, med avstand som respons variabelen. Ulike regresjons modeller er brukt til å problemet med heteroskedastiske residualer.

Vår endelige modell viser en økning i distanse for medvind nå tiden øker, mens et lavere skydekke og nedbør reduserer distansen, spesielt om vinteren. I tillegg gir oppdrag som starter i Ålesund en kortere distanse sammenlignet med Rosten. Med nøyttale værforhold dekker de to helikoptrene på Rosten og Ålesund 91.5% av befolkningen i Midt-Norge med en flytid på 35 minutter (reponstid på 45 minutter). Mew vanskeligere forhold er dekingen redusert til 86.4%.

Disse resultatene indikerer at det kan være nyttig å ta værdata i betraktning ved utsendelse av akuttmedisinske tjenester.

Preface

This thesis is the final part of a Master of Science in the program Mathematical Sciences at the Norwegian University of Science and Technology (NTNU). The thesis is a 45 credit individual project worked on in fall 2021 and spring 2022.

I want to give a large thanks to my supervisor Andreas Asheim for excellent guidance and interesting discussions on the subject. In addition, I would like to thank people at Rosten for good input to the thesis and the opportunity to visit them during my work with the thesis.

Some of the people I would like to mention from Norsk Luftambulanse and St. Olav's University Hospital are:

Josten Dale, Medical doctor (MD)

Lars Eide Næss, PhD candidate Norwegian Air Ambulance Foundation (SNLA)

Andreas Kruger, MD PhD

Ole Erik Ulvin, MD PhD candidate SNLA

Jens Fjelnset, helicopter pilot

Oddvar Uleberg, MD PhD

Eivinn Årdal Skjærseth, MD medical director at Rosten

Definitions and abbreviations

Definitions

Dispatch:	Request from Emergency medical communication centre (EMCC) to helicopter about a mission. The pilot, with the crew will be able to make a decision to accept or reject the mission.
Mission:	A mission is a request from the Emergency medical communication centre (EMCC) which is accepted by the helicopter.
Primary mission:	Missions where the patient is located outside of the hospital
Secondary mission:	Missions where the patient is transported from one hospital to an other because of a medical reason.
Return mission:	Mission where a patient is transported to an other hospital to reduce the level of care, or to a hospital in the region that the patient belongs to.
Acute mission:	Patient is in a critical condition, and ambulance should be dispatched and doctor alarmed.
Urgent mission:	Patient is in a severe condition and in need av a assessment from a doctor or transportation to a hospital.
Activation time:	Time used from dispatch to ambulance start to drive/fly
Flying/driving time:	Time used from ambulance start to drive/fly to arrival on scene of incident.
Response time:	Total time used from alarm to arriving on scene. Activation time + flying/driving time.
Basic statistical units:	Municipalities divided into small, stable geographical units giving a flexible basis for presentation of regional statistics. Basic statistical units are geographically coherent and shall be as homogeneous as possible with respect to natural conditions, economic base, communication and building structure[1]. There are about 14 000 units in Norway, with about 400 persons per unit.

Abbreviations

AIC:	Akaike information criterion
AMIS:	Acute medical information system
API:	Application programming interface
BIC:	Bayesian information criterion
CV:	Cross validation
EMCC:	Emergency medical communication centre
EMCC-AA:	Emergency medical communication centre Air ambulance
EMS:	Emergency medical services
GAMLSS:	Generalized additive model for location, scale and shape
GEMS:	Ground emergency medical services
HEMIT:	Helse Midt-Norge IT
HEMS:	Helicopter emergency medical services, including rapid response car
HemsWX:	Helicopter Emergency Medical Service Weather
IDW:	Inverse distance weighting
LABAS:	Pasientadministrativt systemi luftambulansetjenesten
OLS:	Ordinary least squares
SAR:	Search and rescue
WLS:	Weighted least squares
RMSE:	Root mean square error
MAE:	Mean absolute error

Table of Contents

Abstract	i
Sammendrag	ii
Preface	iii
Definitions and abbreviations	iv
List of Figures	vii
List of Tables	viii
1 Introduction	1
1.1 Aim	2
1.2 Background information about the ambulance	3
2 Data	6
2.1 HEMIT data warehouse	6
2.2 Weather data	7
2.3 Population data	8
2.4 Data quality	8
2.5 Data approval	8
3 Theory	9
3.1 Classical linear model	9
3.2 General linear model	11
3.3 Heteroskedasticity	11
3.4 Model choice	14
3.5 Inverse Distance Weighting	15
3.6 Weather calculations	16
4 Method	17
4.1 Preparation of the data	17

4.2	Building a Model	19
4.3	Using the model	22
4.4	Software	22
5	Results	23
5.1	Data exploration	23
5.2	Models	29
5.3	Visualizations in different weather conditions	36
6	Discussion and Conclusion	42
6.1	Impact and further work	44
6.2	Conclusion	45
	Bibliography	46
	Appendix	48
A	Confidence interval	48
A.1	OLS model	48
A.2	GAMLSS model	49
B	R code	50
B.1	Extract weather data	50
B.2	Models	53
B.3	Visualizations	56

List of Figures

1	A map of the ambulance service in Central Norway, updated in 2020[5].	3
2	Illustration of prehospital time concept[11].	5
3	Illustration of the tailwind configuration used to calculate tailwind and crosswind [21].	16
4	Time vs distance for helicopter missions before and after some missions where removed.	25

5	Time vs distance for rapid response car missions before and after some missions where removed.	25
6	Comparison of response time vs. distance for helicopter and rapid response car. Response time is time from alarm until arriving at the patient. Figure b) is a zoomed in version of Figure a).	26
7	Figure including 3 plots showing heat maps of missions and location of weather stations used to retrieve weather data.	28
8	Standardized residuals against fitted values from a classical linear model with OLS.	30
9	Standardized residuals plotted against fitted values for the alternative models. . .	32
10	Predictions of distance plotted against time using different models including all variables/features. In plot b), c) and d), the black points are predictions (mean) and the purple points are predictions with added variance. Plot a) is the true observations, while b), c) and d) have been predicted on simulated data.	36
11	Map with circles showing the area the two different helicopters cover during a flying time of 35 minutes, along with the center of all basic statistical units in Central Norway (black dots).	37
12	Circles showing the ares that the helicopter at Rosten can cover during 15min, 25min and 35 min with different tailwind. The arrow indicates wind direction. .	38
13	Circles showing the ares that the helicopter at Rosten can cover during 15 min, 25 min and 35 min with low, medium or high ceiling (in season 1).	39
14	Circles showing the ares that the helicopter at Rosten can cover during 15 min, 25 min and 35 min with a precipitation amount of 0 mm, 1 mm and 3 mm (in season 1).	40
15	Circles showing area covered during a flying time of 25 minutes in two different sets of weather during season 1. Red circle is wind speed of 20 km/h (red arrow shows direction), 0 mm precipitation and a high ceiling. Blue circle is wind speed of 20 km/h (blue arrow shows direction), 1 mm precipitation and a low ceiling. .	41

List of Tables

1	Table with descriptive statistics of missions included in the final model. Mission are differentiated by helicopter of rapid response car, and again by flying/driving time of 25/15 minutes respectively.	23
2	Number of dispatches/mission each year grouped by base or helicopter/rapid response car.	24
3	Summary Statistics of the dispatch/mission status for all missions.	24
4	Number of completed mission each year grouped by base or helicopter/car. . . .	26

5	Summary Statistics for times and distance for completed missions with the following quantiles; 25%, 50% (median) and 75%. Air: helicopter, Ground: Rapid Response car. Activation time, flying time and response time are in minutes, and distance are in kilometers.	26
6	Summary Statistics of weather elements for the remaining completed primary missions.	27
7	Summary of the linear model with OLS predicting distance(km). x:y represents an interaction between x and y. df:degree of freedom.	30
8	Results from Breuch-Pagan test on different models to test heteroscedasity. OLS is a linear model with Ordinary least square. Log model is a model with log(distance) as response. WLS linear model with weights from the two-stage method, and WLS2 uses 1/time as weights.	31
9	Summary of the linear model with OLS and logarithm of distance as response.	33
10	Summary of the linear model with WLS where the weights is estimated by the two-stage method.	33
11	Summary of the linear model with WLS where 1/time is used as weights.	34
12	Summary of GAMLSS regression model, where both μ and σ are estimated.	34
13	RMSE and MAE values for the different models after 10-fold cross validation.	35
14	AIC and BIC values for the different models after 10-fold cross validation.	35
15	Confidence interval to estimated coefficients for the OLS model.	48
16	Confidence interval to estimated coefficients for the GAMLSS model.	49

1 Introduction

In a country like Norway, with large geographical distances, mountains and fjords, traveling from a place to another may be time consuming. With challenges like this, emergency medical services (EMS) need to be well structured to provide necessary help in time. Here, both the helicopter emergency medical services (HEMS) and the ground emergency services (GEMS) play a crucial part.

If an incident occurs, one of the first things you do is to call the EMS for help. When responding to a medical emergency, the EMS may use different resources, including ground ambulance and air ambulance. Several factors will affect the choice of resource, such as the severity of the injury and the patient's need, which medical service is available and location of the patient. Time to reach the patient and time to arrival at the hospital may in severe cases be decisive in the choice of resource.

Time and type of help are important factors when dispatching, and these may vary from different incidents. Some incidents need a fast response time, while a fast transportation to the hospital may be more important for others. Helicopters do as well have a medical doctor onboard, which the GEMS do not have. Therefore, if a medical doctor is required, HEMS may be preferred. Sometimes a combination of resources may be the best solution. As an example, GEMS can be dispatched first, with HEMS arriving to either have a faster transportation to the hospital or to transport a medical doctor to the location of incident.

A mission with helicopter engenders rest time for the crew. As a result, helicopter and crew may be unavailable to respond to future incidents. Therefore, it is important that a decision to fly is made with care. In addition, there are other variables such as landing condition, concurrency conflicts and weather that can influence whether HEMS may respond. The decision of which medical service to dispatch is important, both for the specific incident and for future incidents. This decision may need to be taken quickly.

The weather tends to change quickly in Norway. Therefore, knowing as much as possible of the weather is crucial. After an idea from a crew member in 2013, cameras financed by "Norsk Luftambulans" was set up, providing real-time picture of the weather along the most commonly used routes[2]. In the beginning of 2022, there was around 140 of these with new cameras already planned. By using real-time pictures, pilots (and the HEMS crew) are able to check if there is possible to fly a specific route or not. This can save much time. If they begin a mission toward a patient and discover that the planned route has difficult conditions, they risk flying a detour or aborting the mission, which results in the need of dispatching a new unit. The cameras take pictures every 15 minutes, and can be accessed via a service called Helicopter Emergency Medical Service Weather (HemsWX), both at the base and during a mission.

Earlier research have compared air ambulance with ground ambulance, and looked at when and where air ambulance can reduce prehospital time[3]. In this research they concluded that helicopter was faster at a distance of 12.5 km (7.7 miles) from the hospital, but with factors that could increase or decrease this distance. Under rush hours the helicopter was faster at 10.5 km (6.5 miles), and with bad weather and difficult flying condition, the distance increased to 27.5 km (17.1 miles). This proves that specific factors may influence the use of air ambulance compared to ground ambulance.

In 2018, Zakariassen et al. conducted a study to document accurate flying time to the scene and

the rate of acute primary mission in Norway in 2011[4]. Median flying time for all mission was 19 minutes (25th - 75th percentile: 13-28 min). For respective helicopter bases in Central Norway, the median flying time for Rosten was 19 minutes (15min - 32 min), Ålesund 15 minutes (11 min - 21 min) and Ørlandet 17 minutes (10 min - 27 min). By finding vast differences for both flying times and rate of missions, they concluded with a substantial difference in the accessibility of HEMS both within and between different geographical regions.

We have used data from dispatches of the air ambulance in Central Norway (Trøndelag and Møre & Romsdalen) from 2014 to 2021, including the helicopters and the rapid response cars. Combined with weather data, we use regression models to analyse the influence different factors have on the performance of HEMS.

1.1 Aim

In this thesis we focus on time used to reach the patient. Our aim is to investigate different factors that may affect the performance of HEMS. First by establishing different factors, then looking at how they affect a helicopter mission. In addition to looking at the effect of different factors, we will look at *how much* influence they have. With this, we try to highlight how the performance of helicopters can differ during missions in different condition, which again can provide support to EMS dispatchers.

1.2 Background information about the ambulance

Before looking at the data, we will give a short description about the emergency medical service (EMS) in Central Norway.

Structure

The medical service in Norway is divided into different regions. This thesis concerns missions in Central Norway, hence our focus is on "Helse Midt-Norge" which serves Central Norway (Trøndelag & Møre og Romsdal).

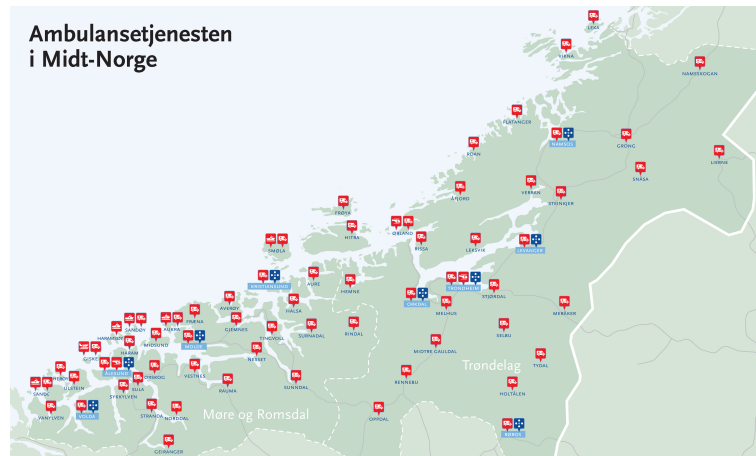


Figure 1: A map of the ambulance service in Central Norway, updated in 2020[5].

Figure 1 shows the ambulance service in Central Norway, updated in 2020. In 2019 Central Norway had 65 ambulance station, with 120 ambulance cars including back-up cars, where 101 of these ambulance cars were operative[5]. The Emergency medical communication centres (EMCC) answers emergency calls, and dispatches the most suited resource for each specific mission. There are three different EMCCs in Central Norway. The respective EMCC are located in "Helse Nord-Trøndelag HF", "St. Olavs hospital HF" and "Helse Møre og Romsdal HF".

Air ambulance

"Luftambulansetjenesten HF" have the responsibility of the air ambulance in Norway[6]. The air ambulance service is operative all day, covers all of Norway and serves approximately 20 000 patients per year. It consists of 14 helicopters on 13 bases and 9 ambulance planes on 7 bases. In addition, 7 Search and Rescue (SAR) helicopters are available. These may be used for ambulance missions and are operated by the Royal Norwegian Air Force.

In Helse Midt-Norge there are two ambulance helicopters, one at Rosten (Trondheim) and one at Ålesund, an ambulance plane in Ålesund, as well as one SAR helicopter at Ørlandet. The three helicopters are responsible for the majority of helicopter missions in Central Norway, but helicopters from neighboring regions can be dispatched if necessary. Some reasons to dispatch

helicopter from different regions may be that other helicopters are occupied by other missions, not able to be dispatch due to rest time, issues on the helicopter or assumed to be the best option.

The Norwegian government requires that 90% of the population can be reached by air ambulance within 45 min after an emergency call. In 2015, a study concluded that 96.9% of the population and 91.8% of the incidents was reached within the 45 minute threshold with the location of air ambulance bases at the time [7].

At each helicopter base, the medical crew have a rapid response car available at their disposal in addition to the helicopter. The rapid response car is used by the same medical criteria as a helicopter, but in cases where the helicopter is unavailable due to bad weather conditions or nearby emergencies where the rapid response car may be more suited than a helicopter. The equipment in the rapid response car is the same as in the helicopter.

The helicopters are coordinated by four EMCCs, one in each region. Previously there was 11 of these, but they were reduced to four in 2018[8]. These centrals have control over the helicopter in their respective regions, and is referred to as the emergency medical communication center air ambulance (EMCC-AA) coordinator. The change from 11 to 4 centrals gives a more centralized, professionalized and uniform coordination of the helicopters. This is expected to give an even better distribution of tasks, more efficient management of resources and better continuity overview of the emergency situation and activity, especially in the border areas between the health regions.

Dispatch of a helicopter

When there is need for an air ambulance, EMCC alerts the EMCC-AA coordinator which immediately alerts the air ambulance and creates a conference call with the necessary people to make a decision. The coordination and choice of resource consists of two parts, a medical part and an operative part. The medical part consist of an assessment if the use of an air ambulance gives a medical gain compared to the use of other resources. The operative coordination involves, among others thing, knowledge about the resources and its status (crew and helicopter), as well as which resource will be fastest and most suited to the mission [9, 10].

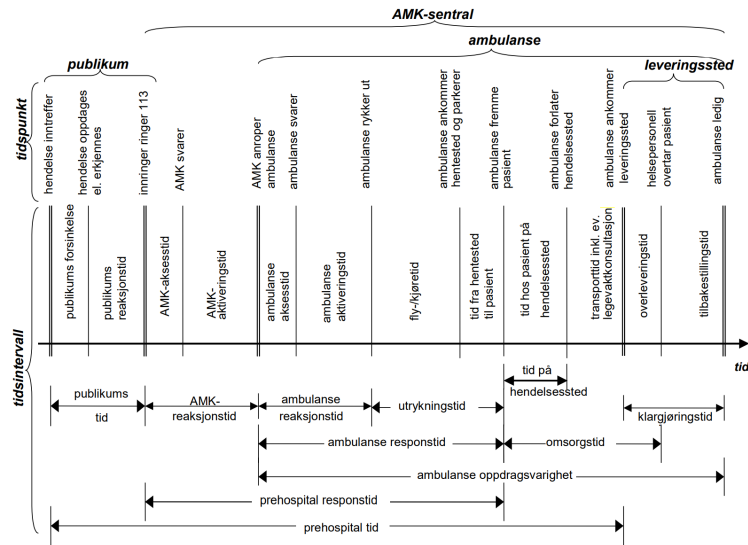


Figure 2: Illustration of prehospital time concept[11].

Rest time

Ambulance helicopters are considered to be commercial aviation, and therefore regulated by common European aviation rules[12]. The European Union Aviation Safety Agency (EASA) have regulations restricting both flying time and duty time[13]. Flight time is defined as the time in which the helicopter are in use, and duty time as the time the crew is actively working. Regarding flight time, the helicopter can have maximum of 8 hours flight time per 24 hours, 14 hours flight time per 48 hours, and 30 hours per week.

The restrictions regarding duty time is a maximum of 14 hours per 24 hours, and 30 hours per 72 hours. The on duty time count begins from the dispatch alarm and lasts until the mission is completed. For the helicopter there is a minimum of one hour added after landing. If there are less than two hours between landing and the next alarm, the duty time is counted together without a break. As well, if the dispatch alarm is triggered between 23:00 - 09:59, one hour of duty time is added even though the mission is not completed.

If a mission is starting before duty time limit is reached, the crew can decide to accept the mission if they considers themselves fit to fly. The crew can also at any time report themselves unfit to fly.

If the crew gets a minimum uninterrupted rest time of 8 hours between 23.00 - 09:59, the duty time is reset.

2 Data

Different data sources are used in this thesis. Data from St. Olav's University hospital including details of every dispatch/mission from EMCC, and weather data with information about different weather elements in the region during the time period. This section gives a description of what the different data sources includes.

2.1 HEMIT data warehouse

The ambulance data is managed by the regional IT-provider HEMIT, where data from different resources is gathered in a data warehouse. There are daily updated data on the population of individuals calling the emergency service within the area of the regional health trust, Helse Midt-Norge from January 1st, 2010. Relevant parts of the data are accessible on the internal computer system to St.Olav's University Hospital. These are data from the region's acute medical information system (AMIS) and the air ambulance system (LABAS). AMIS contains both air and ground ambulance, while LABAS contains just air ambulance including rapid response car.

AMIS

AMIS, the acute medical information, includes all calls to the three EMCCs in the region of Helse-Midt. From AMIS the following data is used:

- All incidents(about 280 000 per year)
 - Time of calls to the Emergency Medical Communication (EMC) Centre
 - Coordinates of incident/location of the patient
- All Emergency Medical Service (EMS) responses (about 350 000 per year)
 - EMS identifier
 - Time stamps:
 - * EMS alerted
 - * Mission start
 - * Arrival at place of incident
 - * Leaving place of incident
 - * Arrival at destination
 - * EMS available after mission completed
- Patient info
 - Age
 - Gender
 - Date/year of birth

LABAS

LABAS includes data from the Air Ambulance service which primarily operates in the region Helse-Midt. This includes Rosten (Trondheim), Ålesund and Ørlandet.

From LABAS the following data is used:

- All dispatches from EMC involving air ambulance (about 5 000 per year)
 - Base of the air ambulance used
 - Time stamps:
 - * Crew alerted
 - * Mission start
 - * Arrival at place of incident
 - * Leaving place of incident
 - * Arrival at destination
 - * Mission completed
 - Deviation
 - * Type of deviation
 - * Reason for deviation
 - Patient info
 - * Age
 - * Gender
 - * Date/year of birth

The process of combining the data from AMIS and LABAS is described in Section 4.1. Some of the data, mainly the patient info, is primarily used in the preprocessing of the data. Data from LABAS comes directly from the air ambulance system, and timestamps from LABAS is therefore preferred.

We will only look at helicopter dispatches for primary acute missions from Rosten and Ålesund. The number of missions available from Ørlandet are small in our data, and therefore not used. In addition, the helicopter may be used to other typed of missions, such as secondary missions, return missions and urgent missions.

2.2 Weather data

Weather data is based on data from The Norwegian Meteorological institute, MET Norway[14]. More specifically, we use Frost API, which gives free access to MET Norway's archive of historical weather and climate data. From Frost API we can access quality controlled measurements of weather data on a daily, monthly and yearly basis. The following weather elements on a daily basis is retrieved:

-
- Air temperature
 - Dew point temperature
 - Wind speed
 - Direction of the wind
 - Precipitation amount

In Central Norway there are 228 different weather stations (Location of weather stations are shown in Figure 7b). The information we have about each station is longitude, latitude, ID and altitude. Some of the stations are active and measure the weather, while some are inactive and no longer in use. The reason for including the inactive stations is described in Section 4.1.

Different weather stations measure different weather elements. Hence, we may get different information about the weather from different stations. The API is used to retrieve as much data as possible from each station at an hourly basis. Missing values is registered as "NA", and will be estimated by interpolation.

Frost API measures air temperature and dew point temperature in degrees Celsius, precipitation in millimeters (mm), wind speed in meters per hour (m/s) and the wind direction in degrees. Wind direction is the direction from which the wind is coming from, where $0^\circ/360^\circ$ is north, 90° east, 180° south and 270° is west. Precipitation is the sum during each hour, while the other elements is measured as the mean during each hour[15].

2.3 Population data

We use population data from Statistics Norway (SSB) on the level of basic statistical units in Central Norway[16]. Each basic statistical unit have a recorded population per yer. We used the population of 2022 for this work.

2.4 Data quality

The overall quality of data is good. The data for each dispatch/mission contains reliable and complete information depending on if a mission is completed or not. The accuracy of time stamps is good, but can contain some uncertainty due to being registered by humans. Weather data is also reliable and accurate, but have weather stations with missing values, or weather stations that are not active at all.

2.5 Data approval

The use of data described in Section 2.1 has been approved, as a part of a quality assurance project, by the "Fagledernetverk for prehospitaltjenester i Midt-Norge". Individual patients cannot be directly identified. All of the analyses are done on St. Olav's University Hospital internal computer system, and the data always remain on this internal system.

3 Theory

This chapter presents statistical theory behind methods used in this thesis. In addition, formulas used for weather calculations will be described.

3.1 Classical linear model

Model definition

Linear regression plays a major role in statistics [17]. The aim in linear regression is to model the relation between a variable of interest $y_i, i \in \{1, 2, \dots, n\}$, called the response, and a set of explanatory variables $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ with $i \in \{1, 2, \dots, n\}$ observations and p distinct predictors. From this, the linear regression model is defined as

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i,$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)$, called the regression coefficients, describes the relation between y_i and \mathbf{x}_i . This relation is not exact and an error term, called the residual, are added to the model. It is assumed that the residuals are independent and identically distributed, $\varepsilon \sim N(0, \sigma^2)$. β_0 is called the intercept, and centers the model on the mean of the response.

Let $\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ and $\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$ be vectors for the response and residuals, and

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^t \\ \vdots \\ \mathbf{x}_n^t \end{pmatrix},$$

be the design matrix with all the observation of the predictors, with 1 representing the intercept β_0 . Then, the linear regression model can be written in a more compact form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

and is addressed as the *classical linear model*. In this model, uncorrelated and homoscedastic error are assumed, i.e. $\text{Cov}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$. Both \mathbf{y} and \mathbf{X} are observed and known, while the regression coefficients $\boldsymbol{\beta}$ are unknown and need to be estimated. Given a set of estimated coefficients $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)$, it is possible to make predictions by

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_p x_{ip}.$$

When fitting a model, the coefficients are estimated with the *least square method* by minimizing

the residual sum of squares (RSS)

$$\begin{aligned}RSS &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ &= \sum_{i=1}^n (y_i - \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_p x_{ip})^2 \\ &= \sum_{i=1}^n (y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})^2.\end{aligned}$$

The unique solution of the least square estimator is given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

and referred to as the ordinary least squares (OLS) estimator[17]. When the residuals are uncorrelated and homoscedastic, which implies that the error variance do not systematically vary across individuals, OLS is the most common method for estimating regression coefficients $\boldsymbol{\beta}$. First, from a mathematical point of view the principles of least squares is relatively simple. Secondly, estimators by the least square method have a number of desirable statistical properties[17].

Hypothesis testing

With hypothesis tests, we are able to say something about the estimation of unknown regression coefficients $\boldsymbol{\beta}$. Both one by one, and for all together. By using these test we assume independently and identically normal distributed residuals, $\varepsilon_i \sim N(0, \sigma^2)$.

One aim of linear regression is to determine whether there is a relationship between the response and the predictors. To test the significance of the regression the following test can be used:

$$\begin{aligned}H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0, & \quad \text{against} \\ H_a : \beta_j \neq 0 \text{ for at least one } j \in \{1, \dots, p\}.\end{aligned} \tag{1}$$

To perform the test we compute the F-statistic

$$F = \frac{n-p}{p-1} \frac{R^2}{1-R^2} \sim F_{p-1, n-p},$$

where n is the number of observations, p is the number of predictors and R^2 is the coefficient of determination, described in more detail later.

If H_0 is rejected, the alternative hypothesis indicates that at least one of the j predictors have an influential relation with the response variable.

Further, we can test the significance of each of the coefficients with the following hypothesis

$$H_0 : \beta_j = 0 \quad \text{against} \quad H_a : \beta_j \neq 0. \tag{2}$$

This hypothesis is performed by computing the t-statistic

$$t_j = \frac{\hat{\beta}_j}{se_j} \sim t_{n-p},$$

where se_j is the estimated standard errors of the coefficient calculated by

$$se_j = \widehat{\text{Var}(\hat{\beta}_j)}^{1/2}.$$

H_0 in Equation 1 is rejected if $F > F_{p-1, n-p}(1 - \alpha)$ and H_0 is rejected in Equation 2 if $|t| > t_{n-p}(1 - \alpha)$. The most common is to use a significance level of $\alpha = 0.05$.

3.2 General linear model

If the assumptions of uncorrelated and homoscedastic errors are not satisfied, we can extend the classical linear model to the *general linear model*. In this model,

$$\text{Cov}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$$

is replaced by a more general assumption

$$\text{Cov}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{W}^{-1},$$

where \mathbf{W} is a positive definite matrix. The variance to the heteroscedastic errors are then given by $\text{Var}(\varepsilon_i) = \sigma_i^2 = \frac{\sigma^2}{w_i}$. For the case of uncorrelated and heteroscedastic errors, w_i is defined to be values on the diagonal of \mathbf{W} . To improve the estimator with the general linear model, ordinary least squares is generalized to a weighted least squares estimator (WLS). WLS is defined as

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^T \mathbf{X}^T \mathbf{W} \mathbf{y}.$$

which minimize the weighted residual sum of squares

$$WRSS(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \sum_{i=1}^n w_i (y_i - x_i^T \boldsymbol{\beta})^2.$$

The idea behind is that observations with a small variance receive a larger weight than observations with a larger variance. As a result, the observations with smaller variances have greater influence in the estimations of $\boldsymbol{\beta}$.

3.3 Heteroskedasticity

Detecting heteroskedasticity

With homoscedastic errors as an assumption in the classical linear model, it is important to check whether we have homoscedasticity or not. An approach to detect heteroscedastic errors is to analyse the residuals of the estimated (classical linear) model. Residuals are plotted against

the predicted values of \hat{y}_i . If the standardized residuals have a random fluctuation around zero with a constant variance, we have homoscedastic errors. Standardized residuals are given by

$$r_i = \frac{\hat{\varepsilon}_i}{\hat{\sigma}\sqrt{1-h_{ii}}},$$

where $\hat{\varepsilon}_i = y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}$ and h_{ii} is the diagonal elements of the hat matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$.

When this is not the case, and the variance is not constant, we have an indication that errors are heteroscedastic. Given that the model assumptions to the classical linear model are correct, the standardized residuals are homoscedastic[17].

Performing tests is a different approach to detect heteroscedastic errors, where the Breusch–Pagan test is one of these. The test assumes a multiplicative model for the error variance

$$\sigma_i^2 = \sigma^2 \cdot h(\alpha_0 + \alpha_1 z_{i1} + \dots + \alpha_q z_{iq}),$$

where z_1, \dots, z_q are covariates that may influence the variance, and h a function not dependent on the unit index i . The Breusch–Pagan test is formulated with the following hypothesis:

$$\begin{aligned} H_0 : \alpha_1 = \dots = \alpha_p = 0 \text{ against} \\ H_a : \alpha_j \neq 0 \text{ for at least one } p \in \{1, \dots, q\}, \end{aligned}$$

where H_0 indicates homoscedasticity and H_1 heteroskedasticity.

The test statistic is given by

$$T = \frac{1}{2} \sum_{i=1}^n (\hat{g}_i - \bar{g})^2,$$

where

$$\hat{g}_i = \frac{\hat{\varepsilon}_i^2}{\hat{\sigma}_{ML}^2}$$

is an auxiliary regression between the response variable and explanatory variables. The residuals $\hat{\varepsilon}_i$ for $i \in \{1, 2, \dots, n\}$, and $\hat{\sigma}_{ML}^2$ are the maximum likelihood estimator of σ^2 for a linear model $y = X^T \boldsymbol{\beta} + \varepsilon$. The mean, \bar{g} , is given by $\bar{g} = \frac{1}{n} \sum_{i=1}^n \hat{g}_i$.

Absent heteroscedasticity, $\hat{g}_i \approx \bar{g}$ gives a small value of T . A larger difference between \hat{g}_i and \bar{g} gives more evidence of heteroscedasticity and that the covariates influences the variance. T is asymptotically χ^2 -distributed under H_0 with q degrees of freedom, $T \stackrel{a}{\sim} \chi_q^2$.

Treating heteroskedasticity

To avoid incorrect conclusions from ignoring heteroskedasticity, there exist methods to treat heteroskedasticity. A simple method is *variable transformation*. With a multiplicative model, an example of variable transformation can a classical linear model with a logarithmic transformation of the response. This gives the following model

$$\log(y_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i.$$

An other approach is to add weights and use a general linear model instead of a classical linear model. With heteroscedastic variances the weights w_i for $i \in \{1, 2, \dots, n\}$ are unknown in addition

to $\boldsymbol{\beta}$ and σ^2 . Therefore, a joint estimation of all the unknown parameters $w_i, \boldsymbol{\beta}$ and σ^2 is a reasonable approach. In the general linear model we have $E(\varepsilon_i) = 0$ and $E(\varepsilon_i^2) = \text{Var}(\varepsilon_i) = \sigma_i^2$. From this, ε_i^2 can be represented as

$$\varepsilon_i^2 = \sigma_i^2 + v_i,$$

where v_i is the deviation of the squared errors from their expectation. Using the fact that σ_i^2 depends on one or more covariates, we assume that σ_i^2 can be written as

$$\sigma_i^2 = \alpha_0 + \alpha_1 z_{i1} + \dots + \alpha_p z_{ip} = \mathbf{z}_i^T \boldsymbol{\alpha},$$

where \mathbf{z} is a vector with all covariates that influence the variance. With this defined, we obtain the following method for the *two-staged least square* method[17]:

1. Obtain preliminary estimates of $\hat{\boldsymbol{\beta}}$ from an unweighted regression between y and \mathbf{x} , and compute the residuals $\hat{\varepsilon}_i$.
2. Obtain estimates of $\hat{\boldsymbol{\alpha}}$ from an unweighted regression between the squared residuals $\hat{\varepsilon}_i^2$ and the variance explanatory variables \mathbf{z}_i . Then fit a general linear model using the weights $\hat{w}_i = \frac{1}{\mathbf{z}_i^T \hat{\boldsymbol{\alpha}}}$.

However, to guarantee that $\mathbf{z}_i^T \hat{\boldsymbol{\alpha}}$ is greater than zero and avoiding negative weights \hat{w}_i , we assume the model

$$\sigma_i^2 = \exp(\mathbf{z}_i^T \hat{\boldsymbol{\alpha}}).$$

Then, $\boldsymbol{\alpha}$ is estimated by

$$\log(\hat{\varepsilon}_i^2) = \mathbf{z}_i^T \hat{\boldsymbol{\alpha}} + v_i,$$

which ensure that the estimated variances is positive. As a result, the weights is given by

$$\hat{w}_i = \frac{1}{\exp(\mathbf{z}_i^T \hat{\boldsymbol{\alpha}})}.$$

A similar approach to the two-stage methods is a *simultaneous estimation* of $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\sigma}^2$ [17]. This allows for a maximum likelihood estimation of (Gaussian) regression models with heteroscedastic variances. To do this, the linear regression model is extended such that both the mean and variance is dependent on the covariates, i.e. $y_i = \mathcal{N}(\mu_i, \sigma_i^2)$. By assuming that the mean and variance is dependent on the same covariates, we can write the mean as

$$\mu_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} = \mathbf{x}_i^T \boldsymbol{\beta}$$

and the variance as

$$\sigma_i = \exp(\alpha_0 + \alpha_1 x_{i1} + \dots + \alpha_k x_{ik}) = \exp(\alpha_0) \exp(\alpha_1 x_{i1}) \dots \exp(\alpha_k x_{ik}) = \exp(\mathbf{x}_i^T \boldsymbol{\alpha}).$$

The variance is written as $\exp(\mathbf{x}_i^T \boldsymbol{\alpha})$ to ensure a positive variance. Different covariates may be used in the estimation of mean and variance. Therefore, it is possible to drop the assumption that mean and variance are dependent on the same set of covariates.

3.4 Model choice

The coefficient of determination, R^2

The coefficient of determination, called R^2 (R-squared), is a statistic that measure the fit of the model. R^2 defines the proportion of total variance that is explained by the regression model, given by the formula

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^n \hat{\varepsilon}_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

where $0 \leq R^2 \leq 1$, \bar{y} the mean of the model, $\hat{\varepsilon}_i$ the residuals and \hat{y}_i the predictions.

The interpretation of R^2 is that a value closer to 1 suggest a model with a good fit to the data, with $R^2 = 1$ suggesting a perfect fit. Therefore, using R^2 is a useful metric when analysing the performance of the model. However, when comparing different models with a different number of predictors, the use of R^2 is limited. The reason is that for every predictor added to the model, the value of R^2 will never decrease. The solution to this is to introduce an adjusted version of R^2 which gives a penalty for each new predictor included in the model. The adjusted coefficient of determination is defined as

$$R_{adj}^2 = 1 - \frac{n-1}{n-p}(1 - R^2),$$

where n is the number of observations and p the number of distinct predictors. R^2 is also based on the underlying assumptions that the model assumptions in the model is met.

Akaike information criterion, AIC

Akaike information criterion (AIC) defined as

$$AIC = -2 \cdot l(\hat{\beta}, \hat{\sigma}^2) + 2(|M| + 1),$$

is one of the most common criteria on model choice. With the maximum likelihood estimators $\hat{\beta}$ and $\hat{\sigma}^2$, $l(\hat{\beta}, \hat{\sigma}^2)$ gives the maximum value of the log-likelihood. When assuming a normal distribution, the log-likelihood is given by

$$l(\beta, \sigma^2) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta)$$

with n number of observations. Since σ^2 is counted as a parameter, $|M|$ represent the total number of parameters included in the model. A smaller value of AIC indicates a better model fit.

Bayesian information criterion, BIC

Similar to AIC, the Bayesian information criterion (BIC) is used in model choice. BIC is defined as

$$BIC = -2 \cdot l(\hat{\beta}, \hat{\sigma}^2) + \log(n) \cdot 2(|M| + 1),$$

where n is the number of observations. The main difference compared to AIC is that BIC penalizes complex models to a greater extend than AIC. However, similar to AIC, a smaller BIC value indicates a better model fit.

Cross-validation

Cross validation(CV) is a statistical method used to assess the predictive quality of a model and/or tune hyperparamters in a model[18]. By dividing the data into k subsets of similar size, we can test a model with multiple training-test splits instead of just one. This approach is called k -fold cross validation, where $k - 1$ folds is used to fit a model and the last subset is used for validation. The aim is to minimize the error between the estimated value and the observed value. The error is computed by a loss function, where mean squared error (MSE) are one possible loss functions. In cross-validation the MSE is calculated for each fold, which results in a k -fold cross validation estimate by computing the average of these MSE values,

$$CV_k = \frac{1}{k} \sum_{i=1}^k MSE_i,$$

and

$$MSE_i = \frac{1}{n} (y_i - \hat{y}_i)^2$$

where y_i is the observed value and \hat{y}_i is the estimated value.

Leave-one-out cross-validation (LOOCV) is a special case of CV when $k = n$, the number of observations. Here every observation except for one is used when estimating the value of the observation left out. This can be more time consuming than k -fold CV, but it will reduce the bias to the test MSE compared to k -fold CV with $k < n$.

3.5 Inverse Distance Weighting

Inverse distance weighting (IDW) is a deterministic model, and can be used to interpolate spatial data. It is a straightforward and non computationally intensive method, regarded as one of the more standard procedures when interpolating spatial data, developed by the U.S. National Weather Service[19]. The method is based on the idea that it is possible to estimate a value of an unsampled point by using a weighted average of known values from point in a neighbourhood around the unsampled point. By using this method it is possible to assign values to unknown points by values from a set of known points. In this paper the method is used to interpolate weather measurements from weather stations.

IDW is given by the formula:

$$\hat{z}_p = \frac{\sum_{i=1}^n w_i z_i}{\sum_{i=1}^n w_i},$$

where \hat{z}_p a point with an unknown value we want to estimate, z_i is the points with a known value, n is the number of n closest neighbour points used and w_i is the weight to a specific point, and is given by the formula

$$w_i = \frac{1}{d(\mathbf{x}, \mathbf{x}_i)^\beta},$$

where $d(\mathbf{x}, \mathbf{x}_i)$ is the Euclidean distance from the points with the unknown value to the station with a known value[20]. The inverse distance power β is used to control the degree of how much the nearer points are preferred over the more distant points. Both n and β can be tuned by using cross-validation.

3.6 Weather calculations

Tailwind and crosswind calculation

Wind arriving from different directions causes headwind/tailwind and crosswind. Therefore, it is useful to decompose wind speed and wind direction into tailwind (headwind) and crosswind. Headwind is defined as having a movement in the opposite direction as the wind, while tailwind is moving in the same direction as the wind. Crosswind is defined as wind arriving from the side, either left or right. Since it is rare to have complete tailwind/headwind or crosswind, the formula to calculate these components are given by

$$\text{Tailwind} = \text{Wind speed} \cdot \cos(\alpha)$$

and

$$\text{Crosswind} = \text{Wind speed} \cdot \sin(\alpha),$$

where α is the angle between movement of direction and the direction which the wind is coming from[21]. This is illustrated in Figure 3.

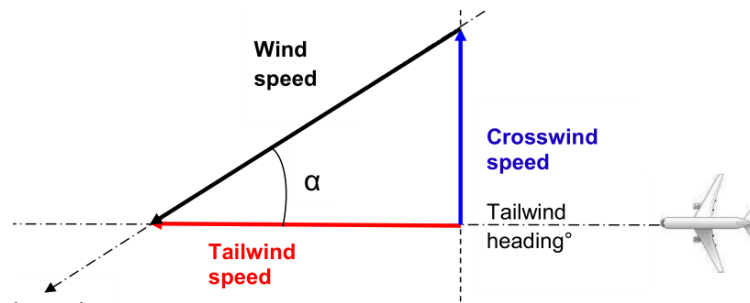


Figure 3: Illustration of the tailwind configuration used to calculate tailwind and crosswind [21].

Ceiling

By the International Civil Aviation Organization (ICAO), *ceiling* is defined as the height above the ground/water, in which the lowest layer of clouds cover more than half of the sky [22]. This indicates an approximate of the highest altitude an helicopter can operate in, but is it is not an exact conclusion of this. In some cases, helicopter are able to operate in higher altitudes. A result of a lower ceiling may be a reduced ability to fly in a straight line compared to a high ceiling.

Ceiling is a relation between air temperature and dew point temperature, and can be calculated by:

$$\text{Ceiling} = \frac{\text{Air temperature} - \text{Dew point temperature}}{2.5} \cdot 1000.$$

Dew point temperature indicates the temperature air need to be cooled to under constant pressure for water vapor to condense to liquid water[23]. This happens when the relative humidity is 100%.

4 Method

4.1 Preparation of the data

As described in Section 2, we use both ambulance data and weather data. In this part we describes different parts in the process from raw data to complete data used for modelling.

Merging Ambulance data

Both LABAS and AMIS contains ambulance data about dispatches/missions. They contain both unique information and information describing same things. For example, which base the helicopter are dispatched from is only included in LABAS, while the coordinates of the scene of incident only is included in AMIS. Therefore, including data from both databases are desirable, giving us as much information as possible of each dispatch/mission. LABAS is primarily used for different time stamps, but in case of missing values or obvious errors, AMIS can be used instead. This will be described in more detail later.

Since there does not exist a variable identifying missions between data in AMIS and LABAS, other information must be used in the process of merging them together. The first step is to use time of dispatch/alarm. A challenge with comparing time stamps is that they are registered in different ways. LABAS have time including minutes, while AMIS have included seconds as well, meaning we do not have identical time stamps. To handle this, we merge all dispatches/missions with a difference of (+/-) 60 seconds or less, which gives an interval of 2 minutes. This interval may contain several dispatches/mission. As a result, one dispatch/mission in LABAS may be merged together with several dispatches/missions from AMIS, where some of them will be wrongly merged together.

Next step of comparing gender and year of birth to the patient solves the problem of wrongly merged dispatches. If a dispatch is merged correctly, the gender and year of birth will match. Otherwise, at least one of gender and year of birth most likely do not match and the row is removed from the data. If there is a missing value in one of these, we just compare the other. And if both have missing values, we assume that it is merged correctly and keep the row. Additional checks is performed later to detect possible rows which is wrongly merged. Lastly, duplicates are removed, where the one with the least amount of missing values is retained.

Weather data and Interpolation

Weather measurements from the same period as the ambulance data is retrieved through the Frost Met API. Information of all weather stations is collected, both active and inactive. This includes information about the ID, altitude and longitude and latitude coordinates to each stations. We retrieve data at an hourly basis, where only active stations at the specific times is used. The available weather measurements at each station is extracted, and missing measurements registered as "NA". The inactive station are added with a "NA" value on all weather elements. All missing values are later estimated by interpolation.

Inverse distance weighing (IDW) is used to estimate the missing values. For each hour, one weather element at a time is interpolated. The stations are divided into two groups, one contain

stations with observed values, and the other contain stations with missing values. The missing values is then estimated by a weighted mean with $n = 15$ number of stations with observed values, using an inverse distance power $\beta = 2$.

When interpolating air temperature, we first recalculate air temperature to sea level. The missing values are then estimated at sea level, before we adjust them back to the correct altitude of the respective weather stations. The reason for doing this is that air temperature tends to decrease with increasing altitudes. By adjusting this before using IDW, we estimate new values to stations at the "same" altitudes. In the conversion of air temperature to sea level and back, a constant $k = 0.065$ is multiplied with the altitude and subtracted when the elevation increase, and added altitude decrease[23].

Wind direction is separated into two scalar components, u and v, in a similar way as tailwind and crosswind are calculated. They are then interpolated independently, and transformed back to wind direction again. The scalar components have continuous values.

Combine Ambulance and Weather data

When both ambulance and weather data is complete, the next step is to combine them into one data set. The ambulance data have time stamps in minutes, while the weather data is at an hourly basis. Therefore, the time stamps in the ambulance data is rounded to the closest hour to be able to combine them. The rounded time stamps is only used in the process of combining the data, and not later in the analysis.

We are interested in the weather on the route from start of the mission to arriving at the location of the patient. For each mission we know which helicopter or rapid response car is used, where we assume that the helicopter base is the starting point of all missions. We choose the three closest weather stations to the helicopter base, where a weighed mean of the weather measurements is calculated to determine the weather at the starting point. The same procedure with the three closest weather station to the patient determines the weather at end point of the mission. The mean of these is then calculated, representing the weather for the specific mission.

Calculation of new variables

From the combined data set we are able to calculate some new variable of interest. To each mission, the euclidean distance from the helicopter base to the patient is calculated using the longitude and latitude coordinates available. We assume that the helicopter will begin the mission at its own base and will fly in a relative straight line.

Different prehospital time concept such as activation time, flying/driving time and response time are also calculated. Activation time is the time from the alarm of dispatch to the mission is started. Flying/driving time is the total time from take off to landing, and response time is the total time from alarm to the ambulance reach the patient. Table 2 illustrates different prehospital time concepts. For calculation of different time concepts, data from LABAS is primarily used. In instances where the results is either unrealistic small or high, i.e. using too little time on a longer distance, or too much time on a short distance, AMIS data can be used instead.

Bearing angle, that is the initial direction of travel between two points, in our case the helicopter base and the patient, is calculated along with the mean of both wind speed and wind direction.

By using the theory in Section 3.6, wind speed and direction is decomposed to tailwind and crosswind. Tailwind have a positive value when the wind is blowing in the same direction as the travel direction, and negative if they are opposite of each other (headwind). We use absolute value for crosswind, assuming wind from both sides will have the same effect no matter the direction it comes from. The mean of the other weather elements is as well calculated using the data from start and end of the mission. The relation between air temperature and dew point temperature can be used to calculate the ceiling of the clouds, that is the altitude of the lowest cloud base. Dew point temperature should always be a lower than air temperature, meaning the ceiling can not have a negative values. If negative values for ceiling and precipitation occur, they are changed to 0.

Removing missions

For some mission, time used vs. distance of the mission can be unrealistic. This can for example be caused by wrong input in the data, or that the assumption of a starting point of the mission is the helicopter base does not hold. Therefore, the average speed for each mission is calculated, which is used when removing outliers. For helicopter mission, missions with an the average speed of less than 100 km/h and more than 400 km/h is removed. The average cruising speed for ambulance helicopter used at Rosten and Aalesund is around 250 km/h, where the helicopter at Ålesund have a little higher cruising speed [24, 25].

For the rapid response car, mission with an average speed of less than 30 km/h and more than 200km/h is removed. Missions is removed as described to avoid using wrong information when fitting models.

4.2 Building a Model

When modelling distance covered during an ambulance mission, different aspect of the mission need to be considered. Typically, a helicopter mission can be divided into three different phases:

1. Take-off, accelerating and reaching desired altitude.
2. Flying in cruising speed in direction of the patient. Cruising speed given and constant for the specific helicopter.
3. Slowing down and landing.

During each phase there exist some factors that influences the speed/time used. In our model, we focus on the effect from the weather.

A simple start to a model is looking at the relation between distance, speed and time. This gives us

$$d = v \cdot t, \tag{3}$$

where d is the distance, v is the velocity/speed and t is the time.

We want to model the distance a helicopter is able to fly given a time t . Doing a simple linear regression of Equation 3, we get

$$d = \beta_0 + \beta_1 x_t,$$

where x_t is time in hours and β_0 the intercept. The interpretation of β_1 is the distance a helicopter is able to fly during time t , which corresponds to the average speed v . From this simple model, we are able to add new variables of interest. The interpretation of each new β to a new variable (multiplied with time) is the same, i.e. how much the variable affect the average speed. A negative value decreases the speed, while a positive value increases the speed.

Wind

It is normal to consider two types of speed/velocity, *ground speed* and *air speed*. Ground speed is the speed in which the helicopter is moving compared to the ground, while air speed is the speed in which the helicopter is moving in the air. When looking at the relation between these two, wind speed, which is the speed of the air compared to the ground, plays a crucial role. Therefore, by including wind speed, we can write the relation as vector quantities by

$$v_g = v_a + w_t,$$

where v_g is the ground speed, v_a air speed and w_t wind speed (only tailwind). Further, by decomposing wind speed to tailwind and crosswind, we get

$$v_a = \sqrt{(v_g - w_t)^2 + (w_c)^2},$$

where w_t tailwind and w_c crosswind w_c . This can also be written as

$$v_g = \sqrt{v_a^2 - w_c^2} + w_t.$$

This motivates an inclusion of both tailwind and crosswind in our model.

Ceiling

Ceiling, the lowest cloud base, indicated an approximation of the highest altitude a helicopter is able to operate in. An advantage to the helicopter is that it is able to fly above the mountains, and as a result save much time compared to the rapid response car. However, if the ceiling is low, there is a reduction of this advantage.

To make an example, lets say a ceiling of 1 000 meters result in a longer distance. Then we can write the distance as

$$d = d_e + d_e \cdot c = d_e(1 + c), \begin{cases} c, & \text{if ceiling} < 1000 \\ 0, & \text{otherwise} \end{cases}$$

where d_e is the Euclidean distance and c is a constant with a unknown value which increases the distance when there is a low ceiling. In our model we look at three different categories for ceiling. Under 500 meters, between 500 meters and 1000 meters, and above 1000 meters. The affect of ceiling may vary across different periods of the year, therefore we are interested to test the affect of ceiling in different part of the year.

Precipitation

To be able to fly/drive at maximum speed, there is necessary to have a good visibility. With precipitation, the visibility is reduced, which again may result in reduced speed. Different types

of precipitation have different affect on the reduced visibility. Therefore, as with ceiling, colder periods where snow may appear, can reduce the visibility to a greater degree than periods with just rain.

Other factors

Both ceiling and precipitation can have varying affect during different phases of the year. To avoid the risk of having few observations in each month, we divide month into four categories with three month in each. One with January, February and march. The second with April, May and June. The third with July, August and September, and the last with October, November and December. The first and last category is month with colder temperatures, and the two middle ones are the warmer periods.

We look at missions from two helicopter bases with a different topographic area around the respective helicopter base. As a result, one helicopter may not be able to fly as directly as the other after take off. Therefore, adding a factor to the model of which *base* the helicopter start on, is reasonable.

Time

We use flying time in our model, which makes it possible to interpret the coefficients as the average speed of the mission. The speed is only assumed to be constant during phase 2 of a helicopter flight, while the other two phases, take off and landing, will use extra time to accelerate or brake. This will vary from mission to mission, but a guesstimate is that it takes an extra 2 - 3 minutes in total for take off and landing. The extra time used for take off and landing is assumed to be included in the intercept of the model.

The total time to reach the patient, the prehospital response time, begins when EMCC receives the emergency call. This includes time used to dispatch an unit and activation time in addition to the flying/driving time. We define the total time to reach the patient as

$$t_f = t_t - t_d - t_a,$$

where t_f is the flying time, t_t the total time from emergency call to landing, t_d time used by EMCC to dispatch and t_a the activation time.

For example, lets say the EMCC use 2 minutes from the emergency call to dispatch, and HEMS have an activation time of 6 minutes. Then the flying time to available to reach the patient within 45 minutes is $t_f = 45 - 2 - 6 = 37$ minutes. Instead of an estimation of this, we can use mean/median times from the data.

Model

To summarize, we will look at the following factors:

Value	Description
Time	Transportation time (h).
Tailwind	Wind in the direction of movement (km/h).
Crosswind	Wind from the side (km/h).
Precipitation	Amount of precipitation (mm)
Ceiling	The altitude of the lowest cloud base (meters). Categories are as following, 1 (low) : [0, 500], 2 (medium) : (500, 1000] and 0 (high) : (1000, ∞]
Location	Location of helicopter base (Rosten or Aalesund).
Season	Category where each season includes three months, 1:[Jan-Mar], ... , 4:[Oct-Dec].

How much each factor will affect the distance depends in the flying time. Longer time in the air, equals a longer period that specific factor can affect the distance. Therefore, all factors are as an interaction with time. This also means coefficients can be directly interpreted as differences in mean velocity. In the end, the following variables are included in the model:

- Time
- Time:Tailwind
- Time:Crosswind
- Time:Location
- Time:Season:Precipitation
- Time:Season:Ceiling

where ":" indicates an interaction between the different factors.

4.3 Using the model

We consider the association of different factors on the response time for helicopters. With the final model, we will visualise by drawing circles showing the area the different helicopter covers during a specific time and changes in the weather. In addition, we will use population data in base statistical units from Statistics Norway (SSB) to calculate the proportion of the population in Central Norway the two helicopters cover[16]. To decide whether a basic statistical unit is inside the circles, we check if the center of the basic statistical unit is inside or not. If it is inside, we assume that all the population in this basic statistical unit is covered.

4.4 Software

To perform the analysis, we have used R-studio version 1.4. SQL queries is used in R-studio to access ambulance data from the HEMIT database. All analyses is done on the St. Olav's University Hospital internal computer system.

R function *lm()* and the R-package *gamlss* are used when fitting the models. While R-package *ggmap* [26] is used for visualisations with maps.

5 Results

5.1 Data exploration

Table 1 presents descriptive statistics of the data used in the final models. We can observe differences both between missions with helicopter and rapid response car, and when comparing missions with a flying time with more than or less than 25 minutes and a driving time with more or less than 15 minutes. We will further explore all available data from all dispatches, and look at missions that can be removed due to containing wrong information.

	Helicopter(flying time)		Rapid response car(driving time)	
	<25 min (N=3286)	>25 min (N=1924)	<15 min (N=1120)	>15 min (N=252)
Helicopter base				
Rosten	2082 (63 %)	1392 (72 %)	856 (76 %)	198 (79 %)
Aalesund	1204 (37 %)	532 (28 %)	264 (24 %)	54 (21 %)
Activation time (min)				
Mean (SD)	7.5 (± 4.1)	9.2 (± 5.9)	2.9 (± 1.8)	5.0 (± 6.2)
Flying/driving time (min)				
Mean (SD)	17 (± 4.7)	33 (± 7.6)	8.4 (± 2.7)	22 (± 9.4)
Distance (km)				
Mean (SD)	51 (± 21)	99 (± 28)	7.6 (± 4.0)	26 (± 18)
Tailwind (km/h)				
Mean (SD)	0.26 (± 8.9)	-0.29 (± 9.2)	1.3 (± 8.8)	-1.5 (± 9.8)
Crosswind (km/h)				
Mean (SD)	6.5 (± 5.2)	7.2 (± 5.3)	6.1 (± 5.8)	6.7 (± 5.5)
Ceiling (meters)				
< 500	281 (9 %)	155 (8 %)	146 (13 %)	58 (23 %)
500-1000	617 (19 %)	381 (20 %)	263 (23 %)	60 (24 %)
> 1000	2388 (73 %)	1388 (72 %)	711 (63 %)	134 (53 %)
Precipitation (mm)				
Mean (SD)	0.090 (± 0.32)	0.090 (± 0.28)	0.14 (± 0.49)	0.20 (± 0.42)

Table 1: Table with descriptive statistics of missions included in the final model. Mission are differentiated by helicopter of rapid response car, and again by flying/driving time of 25/15 minutes respectively.

Missions

We look at all dispatches for primary missions available from 2014-2021. This includes both completed missions, aborted missions, and rejected missions. Missions performed by the crew

at Rosten and Ålesund are included, where either the helicopters or the rapid response cars are used. By looking at Table 4, we can see a little variation in the number of mission each year, where 2017-2019 is the period with the highest number of dispatches/missions. As a trend for most of the years, Rosten have around twice as many dispatches as Ålesund and the helicopter is used between three to four times as much as the rapid response car.

	2014	2015	2016	2017	2018	2019	2020	2021	Sum
Rosten	1 122	1 193	1 257	1 321	1 230	1 023	976	1 043	9 165
Ålesund	400	458	504	562	621	567	611	562	4 285
Total	1 522	1 651	1 761	1 883	1 851	1 851	1 587	1 605	13 450
Helicopter	1 159	1 217	1 352	1 446	1 437	1 266	1 256	1 262	10 395
Rapid response car	363	434	409	437	414	324	331	343	3 055
Total	1 522	1 651	1 761	1 883	1 851	1 851	1 587	1 605	13 450

Table 2: Number of dispatches/mission each year grouped by base or helicopter/rapid response car.

The result of each dispatch is presented in Table 3. The majority of the dispatches is either completed with out any deviations, interrupted before they reached the patient or rejected. The most common reasons for being interrupted or rejected are no longer any need (2 338), bad weather (930) and concurrency conflicts (738).

Status	Sum	%
Completed	8 227	61.16
Rejected	2 610	19.41
Interrupted	2 117	15.74
Delayed	370	2.75
Other	106	0.79
Change of destination	20	0.15
Total	13 450	100

Table 3: Summary Statistics of the dispatch/mission status for all missions.

The relation between time and distance in completed missions for helicopter and rapid response car are shown in Figure 4a and Figure 5a. The red line represent a simple linear model with time and distance. We can see that the majority of missions seems to follow a linear trend, but with some outliers. The relation distance/time describes the average speed on the mission. Some of these are "unrealistic", either to slow or to fast. As described in Section 4.1, missions with a too low/high average speed is removed to avoid using mission that may include wrong information, such as missions that do not satisfies the assumption of having it's own base as the starting point of the mission. Looking at Figure 4b and Figure 5b, where some mission are removed as described in Section 4.1, we can see a more visible linear trend. However, as time increase, we observe an increasing variance as well.

From now, only the remaining missions in Figure 4b and Figure 5b are included in the analysis.

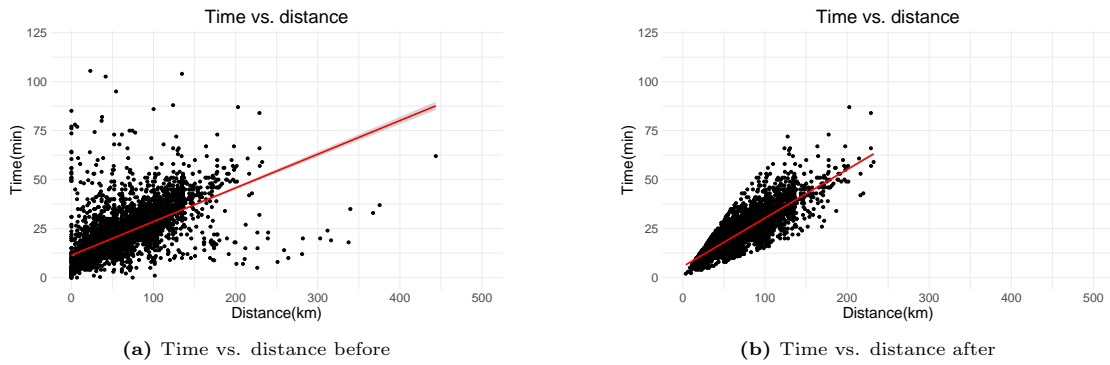


Figure 4: Time vs distance for helicopter missions before and after some missions where removed.

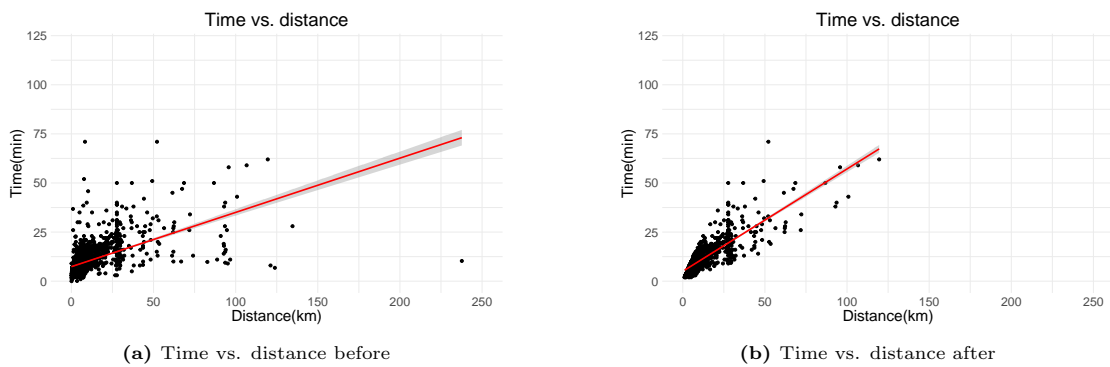


Figure 5: Time vs distance for rapid response car missions before and after some missions where removed.

A comparison between response time vs distance between the helicopters and rapid response car can be seen in Figure 6. A simple linear model between response time and distance is used to construct the lines. Response time includes the activation time and flying/driving time from helicopter base to patient. Over shorter distances, the rapid response car is faster due to shorter activation time. However, by looking at the slope we see that response time for rapid response car increases faster than for the helicopter. Panel b) have zoomed in around the area of intersection between the two lines. We observe that the helicopter have regained the lead to the rapid response care after ≈ 16 minutes / ≈ 15 kilometers. The results includes all remaining missions, and we have not separated the mission with different conditions. Therefore, the time needed for the helicopter to regain the lead of rapid response car may change in different conditions.

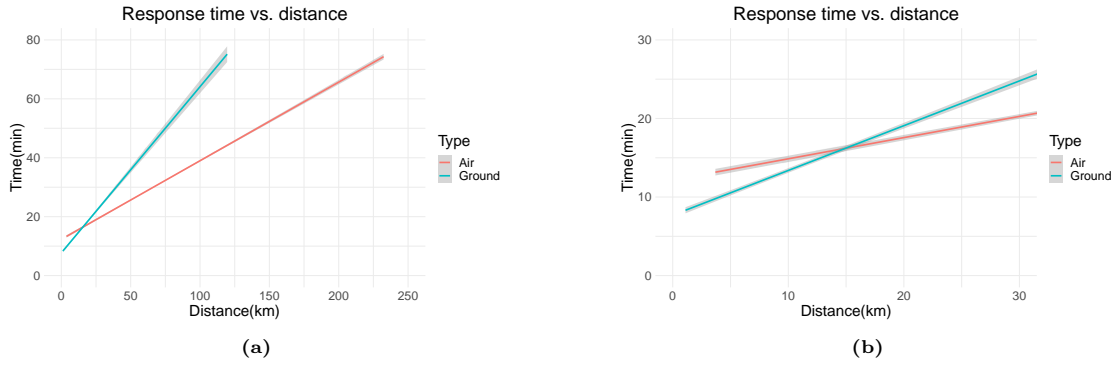


Figure 6: Comparison of response time vs. distance for helicopter and rapid response car. Response time is time from alarm until arriving at the patient. Figure b) is a zoomed in version of Figure a).

Table 4 have the number of completed missions after removal of the missions with an unrealistic time vs. distance. From a total of 8 227 completed primary mission, we now have 6 604 left. The relation between the number of missions at the different helicopter bases and type of mission looks to be the same. Rosten have around double the amount as Ålesund, and the helicopter is used around three times as much as the rapid response car.

	2014	2015	2016	2017	2018	2019	2020	2021	Sum
Rosten	586	585	602	628	565	515	505	558	4 544
Ålesund	216	211	223	288	322	248	302	250	2 060
Total	802	796	825	916	887	763	807	808	6 604
Helicopter	619	602	638	722	702	625	648	654	5 210
Rapid response car	183	194	187	194	185	138	159	154	1 394
Total	802	796	825	916	887	763	807	808	6 604

Table 4: Number of completed mission each year grouped by base or helicopter/car.

Type	Base	Activation time			Flying time			Response time			Distance		
		25%	50%	75%	25%	50%	75%	25%	50%	75%	25%	50%	75%
Air	Rosten	5	8	10	16	22	27	24	30	39	45.2	70.3	95.2
Air	Aalesund	4	7	10	14	19	27	20	27	36	30.2	50.0	75.7
Ground	Rosten	2	3	4	6	8	12	9	11	15	3.7	7.1	10.0
Ground	Aalesund	2	3	4	6	9	13	9	12	16	3.6	6.8	10.2

Table 5: Summary Statistics for times and distance for completed missions with the following quantiles; 25%, 50% (median) and 75%. Air: helicopter, Ground: Rapid Response car. Activation time, flying time and response time are in minutes, and distance are in kilometers.

Summary statistics regarding times and distances for completed missions are presented in Table 5. There is a clear difference in activation time used depending on type of mission, which is natural since helicopter missions require more preparation than missions with rapid response

Weather element	Min	25%	Median	75%	Max
Air temperature (°C)	-19.3	2.8	7.9	13.3	30.6
Precipitation amounts (mm)	0.0	0.0	0.0	0.1	8.5
Tailwind (km/h)	-45.9	-5.4	0.3	5.7	45.1
Crosswind (km/h)	0.0	2.7	5.4	9.2	43.9
Ceiling (m)	0	880	1 660	2 780	9 860

Table 6: Summary Statistics of weather elements for the remaining completed primary missions.

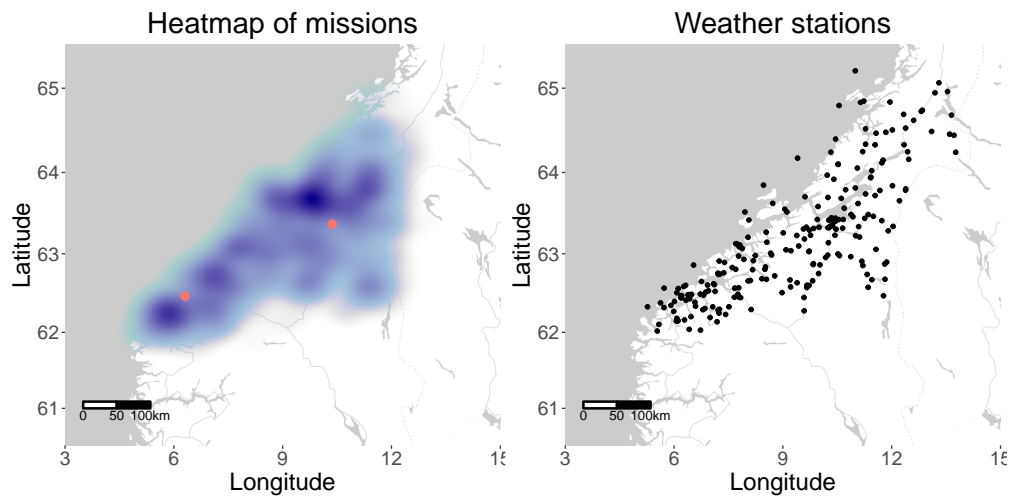
car. Comparing the different bases, there seems to be similar for both helicopter and rapid response car.

Naturally, the helicopter covers a larger distance than the rapid response car on their respective missions. In addition, there is a large difference when looking at distances covered by the helicopter at Rosten and Ålesund. Median distance for Rosten is 20 km longer than Ålesund. Although there is a large difference in distance, there is only a three minute difference in flying time.

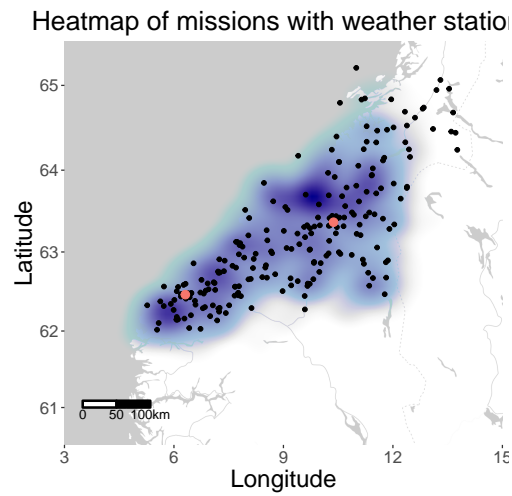
Weather

In Figure 7 we have three plots visualizing heat maps of missions and location of weather stations. Naturally, we can see from Figure 7a) that mission are in the area around each base. Darker blue indicates a higher density of missions, which appear relative close to each of the helicopter bases. A location of the bases close to larger cities in Central-Norway may be a natural reason since areas with a higher population will result in more incidents. Further, the two helicopters operate as expected in areas around themselves and between each other. From Figure 7c), we see that the weather stations cover the missions well. Therefore, all mission should have weather stations nearby.

Summary statistics of the weather for all completed primary mission used when fitting the models are shown in Table 6. The majority of mission have none or very little precipitation, light tailwind and crosswind, and a high ceiling.



(a) Heat map of primary missions without any deviations. Red circles indicates the helicopter bases, Ålesund (left) and Rosten (right). (b) Map with the location of weather stations.



(c) Heat map of primary mission without any deviations, along with the helicopter bases (red) and weather stations (black).

Figure 7: Figure including 3 plots showing heat maps of missions and location of weather stations used to retrieve weather data.

5.2 Models

Linear model: Ordinary least square

As a baseline model, we will start with a linear model using ordinary least squares (OLS) with distance as response and variables as described in Section 4.2. This model will be referred to as "OLS" model. Results from the "OLS" model can be seen in Table 7. Looking at the F-statistic and corresponding p-value, we can conclude with a significant regression. A R^2 value of 0.752 is high, indicating that the regression model explained a large share of the variance in the observed data.

The estimated coefficients tells the impact from each variable on the distance, and can be interpreted as a change in average speed when dividing on time. Time as a variable alone obviously increases the distance as time increase, but will not increase the average speed. Therefore, $\beta_{time} = 184.1$ tells the constant average speed without any impact from other variables.

Tailwind is an other significant variable, which as well have a positive impact on the response. For an increase by one, the average speed increases by 0.76 km/h, which is almost a one to one (1:1) ratio with tailwind and average speed. This makes sense as a positive tailwind pushes the helicopter forward, while a negative tailwind is the same as headwind and will slow down the helicopter. Crosswind does not have a substantial impact in this model.

Location of the helicopter base, and then the starting point for the mission, makes up a big difference. When comparing Ålesund with Rosten, we observe that Ålesund reaches ≈ 14 kilometers less during one hour in Euclidean distance from the base.

The two other variables, precipitation and ceiling have a negative impact on the distance/average speed. Regarding precipitation, the negative impact is larger during season 1 and season 4, which is the month with colder temperature. On average, the respective seasons have a decreases of average speed by 13.5 km/h (95% confidence interval: [-25.7 , -1.3]) and 12.9 km/h (95% confidence interval: [-20.3 , -5.6]) respectively for each millimeter with precipitation.

A ceiling of above 1 000 meters, category 0 is used as reference group for the interaction term including ceiling. We observe that both categories of ceiling, under 500m (low) and between 500 m and 1 000 m (medium) have a negative impact compared to ceiling above 1 000 m. A low ceiling gives on average a reduction of around 15 km/h in average speed for all seasons except for the period April to June. A medium ceiling have the same effect, but around half the reduction. This indicated that a lower ceiling, gives a shorter (Euclidean) distance, which again gives a lower average speed compared to a high ceiling.

Residuals to the "OLS" model is shown in Figure 8. The mean is around zero, and looks to be constant. However, the variance does not fluctuate around the mean, but increases as the fitted values/distance increases. This indicates heteroskedasticity in our model, with further evidence when looking at the Breuch-Pagan test in Table 8.

Variable	Estimate	Std. Error	t-value	Pr(> t)
(Intercept)	1.6484	0.5977	2.7579	0.0058
Time	184.0892	1.7767	103.6105	0.0000
Time:Tailwind	0.7572	0.0632	11.9880	0.0000
Time:Crosswind	0.1144	0.1122	1.0199	0.3078
Time:LocationAalesund	-14.2277	1.2830	-11.0892	0.0000
Time:Precipitation:Factor(season)1	-13.4811	6.2305	-2.1637	0.0305
Time:Precipitation:Factor(season)2	-3.8501	5.6228	-0.6847	0.4935
Time:Precipitation:Factor(season)3	-0.9329	2.7933	-0.3340	0.7384
Time:Precipitation:Factor(season)4	-12.9149	3.7539	-3.4404	0.0006
Time:Factor(season)1:Factor(ceiling)1	-14.6398	5.9667	-2.4536	0.0142
Time:Factor(season)2:Factor(ceiling)1	-0.6006	6.0182	-0.0998	0.9205
Time:Factor(season)3:Factor(ceiling)1	-13.0478	3.6146	-3.6097	0.0003
Time:Factor(season)4:Factor(ceiling)1	-15.7908	3.1336	-5.0392	0.0000
Time:Factor(season)1:Factor(ceiling)2	-6.9318	3.1256	-2.2177	0.0266
Time:Factor(season)2:Factor(ceiling)2	-8.7713	2.9966	-2.9271	0.0034
Time:Factor(season)3:Factor(ceiling)2	-1.2985	2.4758	-0.5245	0.6000
Time:Factor(season)4:Factor(ceiling)2	-4.9431	2.5570	-1.9332	0.0533
$R^2 = 0.7515$		$R^2_{adj} = 0.7507$		
F-statistic: 981.5 (df:16 and 5193)		p-value: < $2 \cdot 10^{-16}$		

Table 7: Summary of the linear model with OLS predicting distance(km). x:y represents an interaction between x and y. df:degree of freedom.

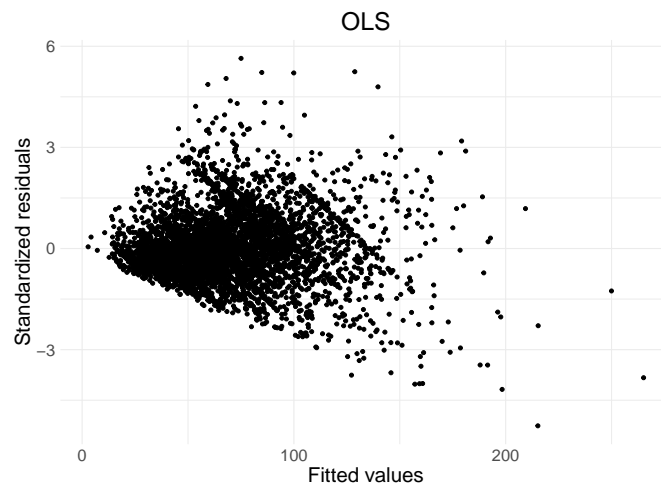


Figure 8: Standardized residuals against fitted values from a classical linear model with OLS.

Model	Breuch-Pagan	df	p-value
OLS	495.89	16	$< 2 \cdot 10^{-16}$
Log response	30.86	16	0.014
WLS	14.37	16	0.571
WLS 2	8097.5	16	$< 2 \cdot 10^{-16}$

Table 8: Results from Breuch-Pagan test on different models to test heteroscedasity. OLS is a linear model with Ordinary least square. Log model is a model with $\log(\text{distance})$ as response. WLS linear model with weights from the two-stage method, and WLS2 uses $1/\text{time}$ as weights.

Alternative models

Due to evidence of heteroskedasticity in the model, we try to fit some alternative models treating heteroskedasticity as described in Section 3.3. The first alternative is a "log" model with a log transformation of the response. Next is the "WLS" model using the two-stage method to estimate weights and use weighted least squares. "WLS2" is a similar model, but in this we assume the weights are known and do not need to be estimated. Instead, we use $1/\text{time}$ as weights. Lastly, we have a "GAMLSS" model where we use simultaneous estimation of β , α and σ^2 .

The first alternative model is a "log" model. This model is harder to interpret by just looking at the coefficients, and as well difficult to compare since the values are much lower. We can observe a lower R^2 value, indicating that the model is not as good as the previous model. However, by looking at the estimated coefficients, they have the same positive or negative value as the "OLS" model.

The last three models, "WLS", "WLS2" and "GAMLSS" takes a non constant variance into consideration, but have the same interpretation as the "OLS" model. All three models produces fairly similar results as each other and compared to the "OLS" model. Time and tailwind have positive values, while the rest have negative values with some exceptions. Time, tailwind, location and ceiling under 500 in season 1 and 4 are statistical significant in all three models, with some others significant in one or two of them. "WLS2" have the largest R^2_{adj} value, but this is as well similar with "WLS" and "OLS" model.

The summary of the "GAMLSS" model do not produce the R^2 value, but it have estimates for σ . When estimating σ we have not used all variables in the model, only time and interaction between time and location. We observe that the intercept is ≈ 2 , which gives the value of σ when the other are zero. A unit increase for time, increases the variance by ≈ 1.6 and ≈ 0.4 in addition if the helicopter is located in Ålesund. Hence, as time increases, the variance in the model increases.

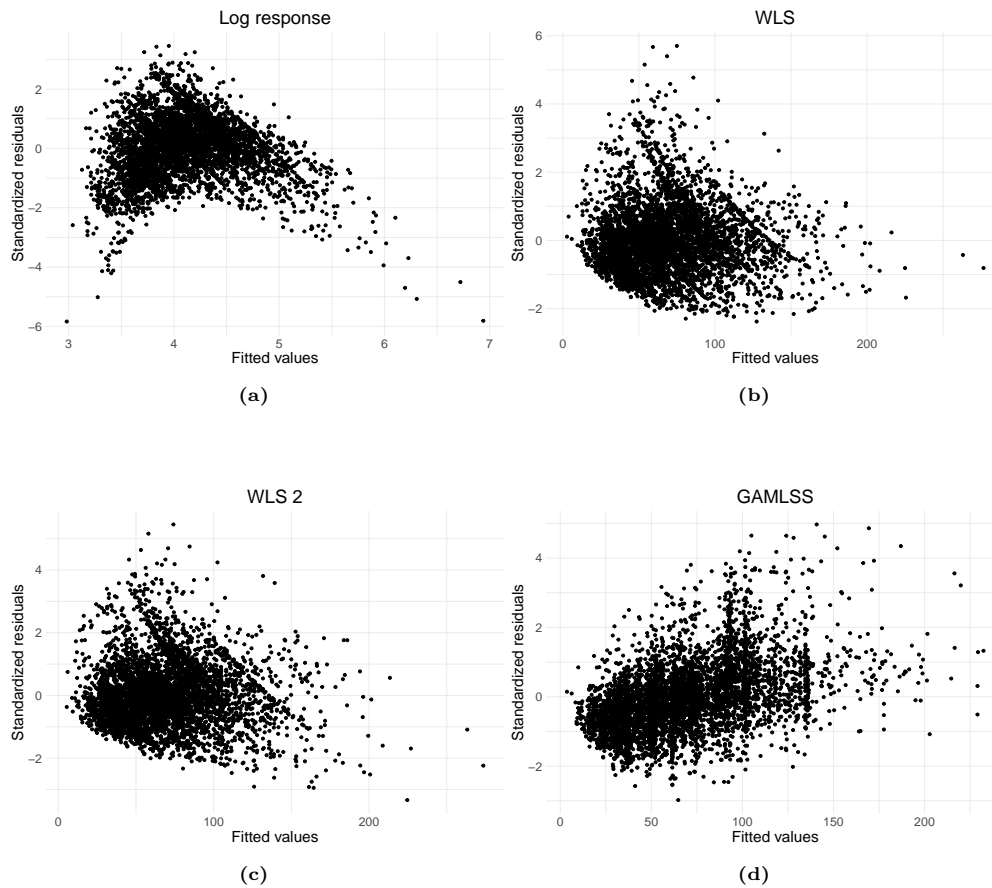


Figure 9: Standardized residuals plotted against fitted values for the alternative models.

The standardized residuals plotted against the fitted values for the alternative models can be seen in Figure 9. We can observe that the "log" model do not have a constant mean, and have a polynomial form of degree two. "WLS" and "WLS 2" have a similar form, almost identical. They have a constant mean around zero, with a sort of constant variance except some high values when the fitted values are less than 100. The "GAMLSS" model have a constant variance, but a mean which increases. Compared to the "OLS" model, the standardized residuals for all of "WLS", "WLS 2", and "GAMLSS" seems to be more homoscedastic.

Comparison of models

k-fold Cross validation is used when comparing "OLS", "WLS", "WLS2" and "GAMLSS" models with $k = 10$. Root mean square error (RMSE) and mean absolute error (MAE) is calculated for each fold, and the mean values are presented in Table 13. We observe that "OLS" have the lowest RMSE followed by "GAMLSS", but there is not a large difference between the models. "GAMLSS" and "WLS" have the lowest MAE score, but here to there is a little difference between the models. In addition, AIC and BIC values from the cross validation is shown in Table 14, where we can see that "GAMLSS" gives the best results.

Feature	Estimate	Std. Error	t-value	Pr(> z)
(Intercept)	3.0853	0.0104	296.3194	0.0000
Time	2.8105	0.0310	90.8073	0.0000
Time:Tailwind	0.0104	0.0011	9.4615	0.0000
Time:Crosswind	-0.0022	0.0020	-1.1177	0.2638
Time:LocationAalesund	-0.1992	0.0223	-8.9138	0.0000
Time:Precipitation:Factor(season)1	-0.1703	0.1085	-1.5690	0.1167
Time:Precipitation:Factor(season)2	-0.1519	0.0979	-1.5513	0.1209
Time:Precipitation:Factor(season)3	0.0267	0.0487	0.5494	0.5828
Time:Precipitation:Factor(season)4	-0.1436	0.0654	-2.1964	0.0281
Time:Factor(season)1:Factor(ceiling)1	-0.1637	0.1039	-1.5750	0.1153
Time:Factor(season)2:Factor(ceiling)1	0.0606	0.1048	0.5784	0.5630
Time:Factor(season)3:Factor(ceiling)1	-0.2033	0.0630	-3.2281	0.0013
Time:Factor(season)4:Factor(ceiling)1	-0.2430	0.0546	-4.4522	0.0000
Time:Factor(season)1:Factor(ceiling)2	-0.1303	0.0544	-2.3934	0.0167
Time:Factor(season)2:Factor(ceiling)2	-0.1306	0.0522	-2.5023	0.0124
Time:Factor(season)3:Factor(ceiling)2	-0.0307	0.0431	-0.7126	0.4761
Time:Factor(season)4:Factor(ceiling)2	-0.0615	0.0445	-1.3810	0.1673
$R^2 = 0.6946$	$R^2_{adj} = 0.6937$			
F-statistic: 738.2.7(df:16 and 5193)	p-value: < $2 \cdot 10^{-16}$			

Table 9: Summary of the linear model with OLS and logarithm of distance as response.

Feature	Estimate	Std. Error	t-value	Pr(> z)
(Intercept)	-1.6789	0.5352	-3.1367	0.0017
Time	195.2636	1.9215	101.6191	0.0000
Time:Tailwind	0.6682	0.0675	9.8942	0.0000
Time:Crosswind	-0.0536	0.1208	-0.4434	0.6575
Time:LocationAalesund	-15.5264	1.3228	-11.7377	0.0000
Time:Precipitation:Factor(season)1	-9.9527	6.1399	-1.6210	0.1051
Time:Precipitation:Factor(season)2	-8.4852	5.9811	-1.4187	0.1561
Time:Precipitation:Factor(season)3	1.0136	2.8162	0.3599	0.7189
Time:Precipitation:Factor(season)4	-8.9350	3.6403	-2.4545	0.0141
Time:Factor(season)1:Factor(ceiling)1	-17.5399	5.6518	-3.1034	0.0019
Time:Factor(season)2:Factor(ceiling)1	-3.2178	5.9134	-0.5442	0.5864
Time:Factor(season)3:Factor(ceiling)1	-15.1031	3.9039	-3.8688	0.0001
Time:Factor(season)4:Factor(ceiling)1	-14.3111	3.3146	-4.3176	0.0000
Time:Factor(season)1:Factor(ceiling)2	-9.8320	3.3449	-2.9394	0.0033
Time:Factor(season)2:Factor(ceiling)2	-5.0009	3.2157	-1.5552	0.1200
Time:Factor(season)3:Factor(ceiling)2	-3.6965	2.6935	-1.3723	0.1700
Time:Factor(season)4:Factor(ceiling)2	-4.4399	2.7380	-1.6216	0.1050
$R^2 = 0.7352$	$R^2_{adj} = 0.7344$			
F-statistic:901(df:16 and 5193)	p-value: < $2 \cdot 10^{-16}$			

Table 10: Summary of the linear model with WLS where the weights is estimated by the two-stage method.

Feature	Estimate	Std. Error	t-value	Pr(> z)
(Intercept)	-0.8482	0.4922	-1.7231	0.0849
Time	191.8136	1.7811	107.6927	0.0000
Time:Tailwind	0.6960	0.0669	10.4072	0.0000
Time:Crosswind	-0.0029	0.1186	-0.0246	0.9804
Time:LocationAalesund	-15.2429	1.3306	-11.4555	0.0000
Time:Precipitation:Factor(season)1	-11.3008	6.4744	-1.7455	0.0810
Time:Precipitation:Factor(season)2	-7.6137	5.9542	-1.2787	0.2011
Time:Precipitation:Factor(season)3	0.5208	2.8457	0.1830	0.8548
Time:Precipitation:Factor(season)4	-9.5509	3.9060	-2.4452	0.0145
Time:Factor(season)1:Factor(ceiling)1	-16.7388	5.8957	-2.8392	0.0045
Time:Factor(season)2:Factor(ceiling)1	-2.1472	5.8981	-0.3640	0.7158
Time:Factor(season)3:Factor(ceiling)1	-14.9208	3.9157	-3.8105	0.0001
Time:Factor(season)4:Factor(ceiling)1	-15.0043	3.3720	-4.4497	0.0000
Time:Factor(season)1:Factor(ceiling)2	-8.7211	3.3229	-2.6245	0.0087
Time:Factor(season)2:Factor(ceiling)2	-5.7786	3.2209	-1.7941	0.0729
Time:Factor(season)3:Factor(ceiling)2	-2.8791	2.6397	-1.0907	0.2755
Time:Factor(season)4:Factor(ceiling)2	-4.7719	2.7152	-1.7575	0.0789
$R^2 = 0.7688$	$R_{adj}^2 = 0.7681$			
F-statistic:1079(df:16 and 5193)	p-value:< $2 \cdot 10^{-16}$			

Table 11: Summary of the linear model with WLS where 1/time is used as weights.

Feature	Estimate	Std. Error	t-value	Pr(> z)
Mu estimates				
(Intercept)	-1.6247	0.5338	-3.0438	0.0023
Time	195.1561	1.9132	102.0066	0.0000
Time:Tailwind	0.7003	0.0692	10.1205	0.0000
Time:Crosswind	-0.0387	0.1210	-0.3201	0.7489
Time:LocationAalesund	-16.0623	1.4232	-11.2862	0.0000
Time:Precipitation:Factor(season)1	-10.2978	6.6151	-1.5567	0.1196
Time:Precipitation:Factor(season)2	-7.5177	6.0662	-1.2393	0.2153
Time:Precipitation:Factor(season)3	0.4202	2.7792	0.1512	0.8798
Time:Precipitation:Factor(season)4	-7.8112	4.0043	-1.9507	0.0511
Time:Factor(season)1:Factor(ceiling)1	-17.8920	5.6716	-3.1547	0.0016
Time:Factor(season)2:Factor(ceiling)1	-2.6508	5.6002	-0.4733	0.6360
Time:Factor(season)3:Factor(ceiling)1	-15.4445	3.8666	-3.9943	0.0001
Time:Factor(season)4:Factor(ceiling)1	-14.5434	3.3287	-4.3691	0.0000
Time:Factor(season)1:Factor(ceiling)2	-8.7994	3.3013	-2.6655	0.0077
Time:Factor(season)2:Factor(ceiling)2	-5.6151	3.2326	-1.7370	0.0824
Time:Factor(season)3:Factor(ceiling)2	-3.7224	2.6574	-1.4008	0.1613
Time:Factor(season)4:Factor(ceiling)2	-4.7358	2.7103	-1.7473	0.0806
Sigma estimates				
(Intercept)	2.0852	0.0273	76.4423	0.0000
Time	1.6138	0.0681	23.7034	0.0000
Time:LocationAalesund	0.3941	0.0538	7.3217	0.0000

Table 12: Summary of GAMLSS regression model, where both μ and σ are estimated.

Model	RMSE	MAE
OLS	16.8084	12.3882
WLS	16.9100	12.3191
WLS2	17.1737	12.3623
GAMLSS	16.8896	12.3191

Table 13: RMSE and MAE values for the different models after 10-fold cross validation.

Model	AIC	BIC
OLS	44187.80	44305.85
WLS	43477.17	43595.22
WLS2	43410.50	43528.55
GAMLSS	39104.37	39233.43

Table 14: AIC and BIC values for the different models after 10-fold cross validation.

However, the biggest difference between the models is how they handle the variance. Figure 10 have four panels, one with the true observations, and the last three predictions on constructed data with a 95% prediction interval using the "OLS", "WLS" and "GAMLSS" models. The data used for prediction is constructed such that location, season and ceiling is constant for the observations used, while tailwind, crosswind and precipitation have a little variation similar to the real data.

The black line is the mean predictions, and the purple lines are the prediction interval. We observe that the mean predictions are similar for all models, while there is a bigger difference in the predictions interval. "OLS" have a constant mean, which looks to be to large for small values of time. "WLS" have very small prediction interval in the beginning, almost equal to the mean predictions, and then start to increase after 30 min. "GAMLSS", as well, has a relative small prediction interval in the beginning, but increases gradually as time increase. Compared to the true observed values, the prediction interval to "GAMLSS" looks to be the closest.

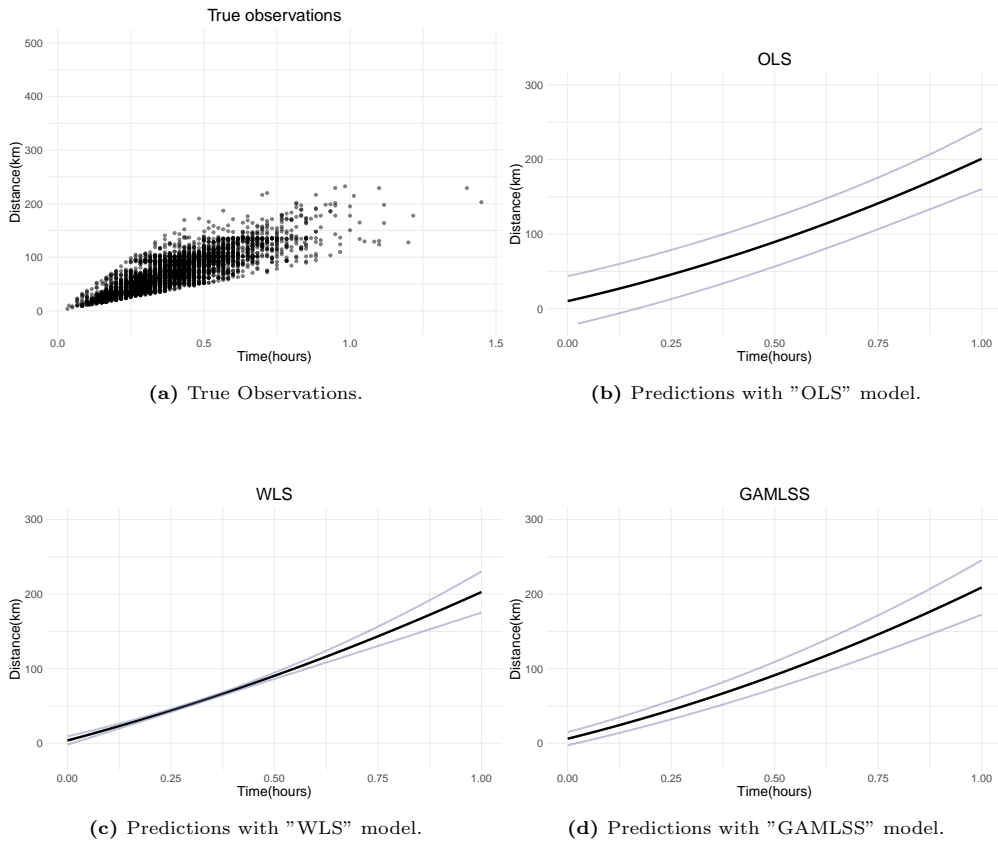


Figure 10: Predictions of distance plotted against time using different models including all variables/features. In plot b), c) and d), the black points are predictions (mean) and the purple points are predictions with added variance. Plot a) is the true observations, while b), c) and d) have been predicted on simulated data.

5.3 Visualizations in different weather conditions

Using "GAMLSS" as our final model, we can visualize the impact different conditions have on the radius from helicopter base the helicopter can reach with in time t .

By using approximately 10 minutes before take off, the helicopters have a flying time of around 35 minutes to reach the patient inside the threshold of 45 minutes. Figure 11a) shows the area which the different helicopters covers during 35 minutes with normal conditions, i.e. no wind, no precipitation and a high ceiling. To plot the circles we have used a high ceiling with no precipitation and no wind during the period January to March. We can observe that the two helicopters cover the majority of Central Norway with a flying time of 35 minutes.



(a) The red points are helicopter bases at Rosten and Alesund.

(b)

Figure 11: Map with circles showing the area the two different helicopters cover during a flying time of 35 minutes, along with the center of all basic statistical units in Central Norway (black dots).

In Figure 11b) we have the same circles, but in addition included the center of all basic statistical units in Central Norway. By calculating the sum of inhabitants in each basic statistical unit inside the circles, the two helicopters cover 91.5% of the inhabitants with a flying time of 35 minutes.

Further, by changing the variables in the model, we can investigate how the circles change in different conditions. We will look the helicopter base Rosten, and vary tailwind, ceiling and precipitation.

Tailwind

Figure 12 shows the influence of different strengths of tailwind for different flying times. The arrow show direction of the wind. We observe that the circle moves in the direction of the wind as expected, and more for longer flying times. The difference in distance between the circles are the same. For 15 minutes, the maximum distance is 3.5 km. 5.8 km for 25 minutes and 8.2 km for 35 minutes. That means for 35 minutes, the difference in distance for no wind and 40 km/h tailwind is 16.2 km.

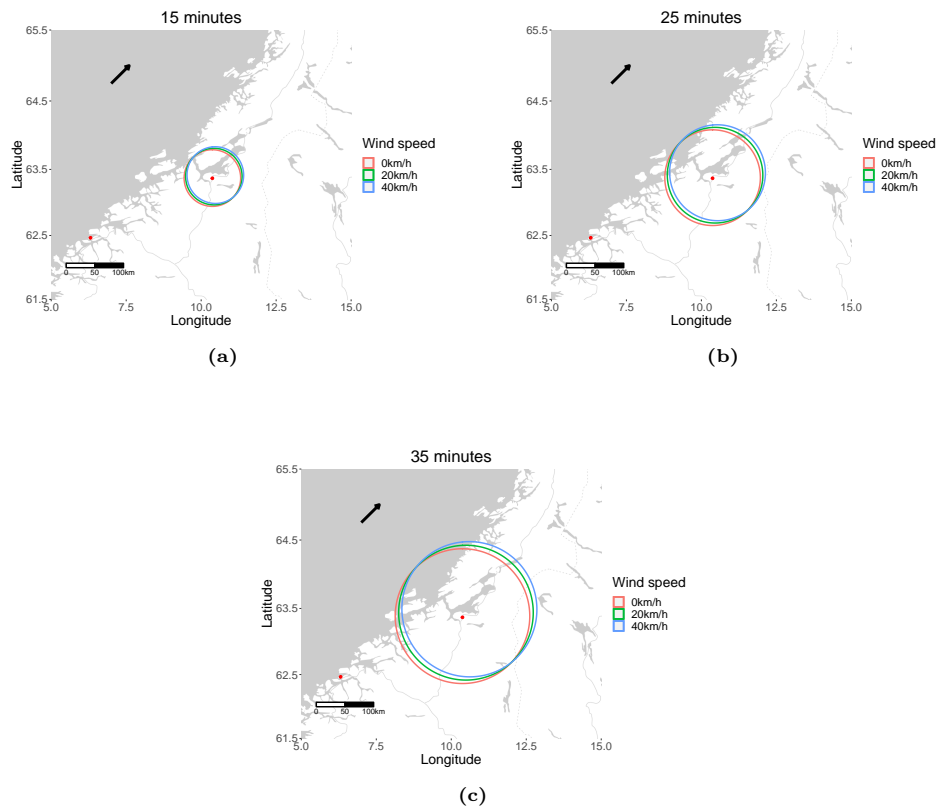


Figure 12: Circles showing the areas that the helicopter at Rosten can cover during 15min, 25min and 35 min with different tailwind. The arrow indicates wind direction.

Ceiling

Differences in ceiling are shown in Figure 13. We can see that difference is not that big, especially for 15 minutes, but increases for longer flying times. The largest difference between the respective panels are 2.3 km, 3.8 km and 5.3 km, where the difference between the circles in the same panel are the same. A lower ceiling, gives a shorter distance given the estimated coefficients.

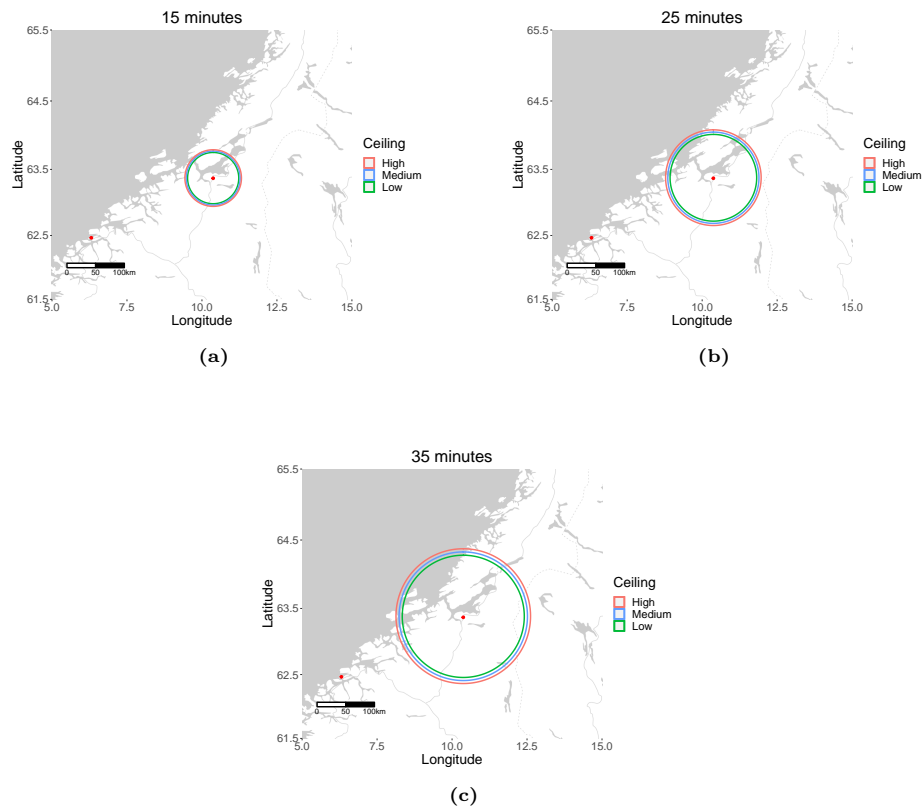


Figure 13: Circles showing the areas that the helicopter at Rosten can cover during 15 min, 25 min and 35 min with low, medium or high ceiling (in season 1).

Precipitation

Next we look at precipitation, where we have used 0 mm, 1 mm and 3 mm with precipitation in season 1 (Jan.-Mar.). We observe that an increase with two units instead of one makes the difference twice as big as increasing with one unit. The difference between the two outer circles in the respective panels are 2.6 km, 4.3 km and 6.0 km, with a difference twice as big between the two inner circles.

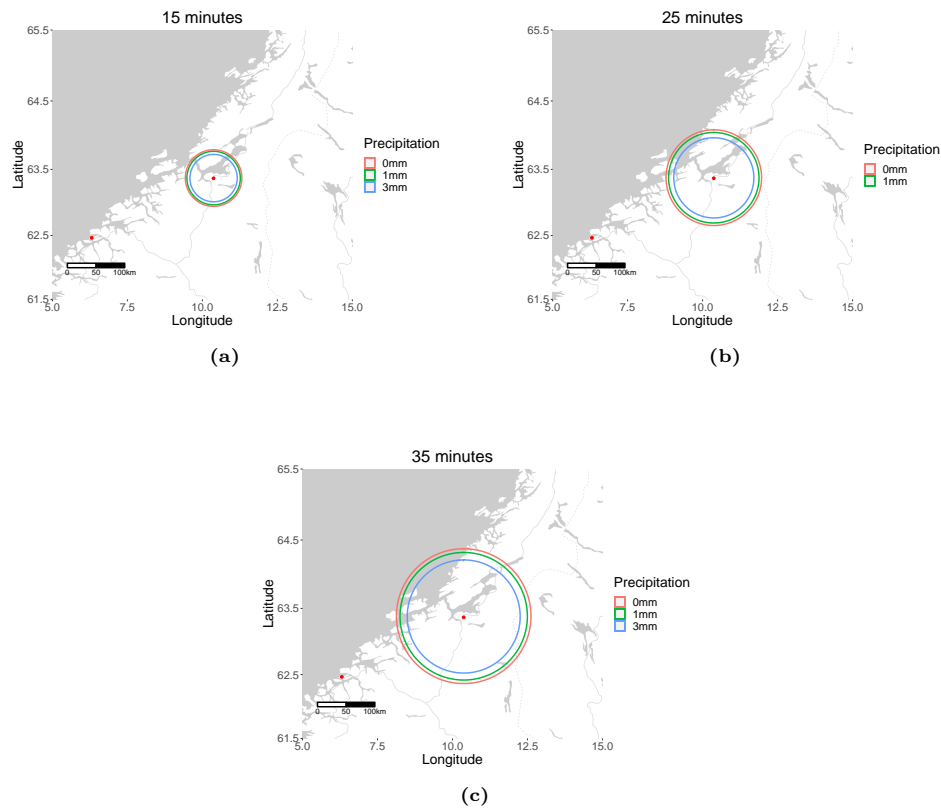


Figure 14: Circles showing the areas that the helicopter at Rosten can cover during 15 min, 25 min and 35 min with a precipitation amount of 0 mm, 1 mm and 3 mm (in season 1).

Combination

As seen from the estimated coefficients in the "GAMLSS" model will tailwind increase the distance in the direction of the wind, while a low ceiling and precipitation decreases the distance. However, a change in several factors at the same time might give a better picture of how the weather affect the helicopter than just changing one at a time.

The first combination (red circle) have wind speed of 20 km/h in the direction of the red arrow, no precipitation and a high ceiling. The second combination (blue circle) have the same wind speed, but in the other direction, a low ceiling and 1 mm precipitation. Both have a flying time of 25 minutes, and in the period January to March.

We observe that the blue circle covers a smaller area than the red, which is caused by a low ceiling and precipitation. The wind moves the circles in each direction, which makes them reach approximate the same distance in one direction, but a large difference in the other. This show the difference between having positive tailwind or negative tailwind (headwind).

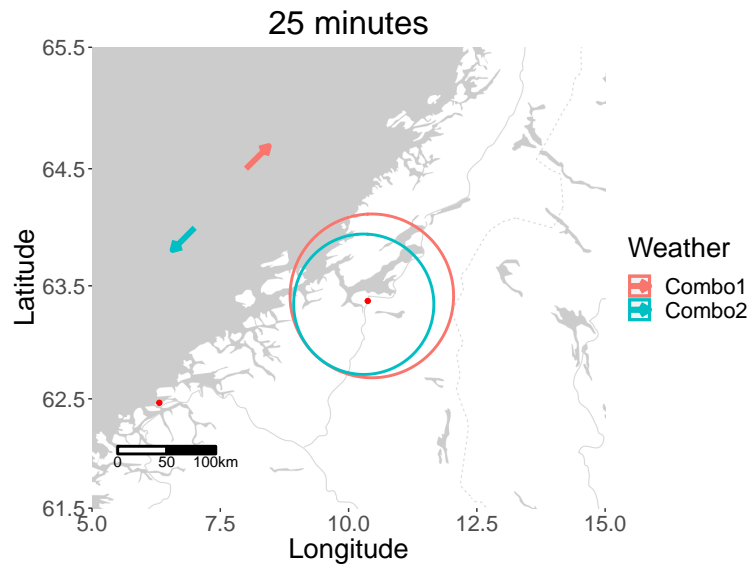


Figure 15: Circles showing area covered during a flying time of 25 minutes in two different sets of weather during season 1. Red circle is wind speed of 20 km/h (red arrow shows direction), 0 mm precipitation and a high ceiling. Blue circle is wind speed of 20 km/h (blue arrow shows direction), 1 mm precipitation and a low ceiling.

Coverage

We have previous seen that under normal conditions, the helicopters covers 91.5% of the population in Central Norway with a flying time of 35 minutes. However, due to more difficult conditions, this can change. With a medium or low ceiling, the helicopters reaches respectively 89.4% and 87.8%. With 1 mm precipitation, the helicopter covers 89.1%, and with the combination of low ceiling and 1 mm precipitation they cover 86.4% of the population in Central Norway. These numbers are calculated during season 1, including January, February and March, and is the period where the weather have the largest effect on the helicopter performance.

6 Discussion and Conclusion

In this thesis we have looked at how the performance of helicopters can differ in different conditions, with an aim of providing decision support to EMS dispatchers and to better plan organisation of services in the future. The performance is studied based on the time from take off to arrival at the location of an incident. Data from primary missions performed by HEMS in Central Norway, combined with weather data are used in the analysis. We wanted to identify factors which affected the distance a helicopter could reach during a specified time limit. What type of influence each factor has is generally well-known, but we wanted to estimate how big the influence is, and how it can be quantified with available weather data. Differences were then visualised with circles on a map. In addition, we wanted to look at the distance in which the helicopter becomes a faster type of EMS compared to the rapid response car.

Several different models have been tested, where it seems like the "GAMLSS" model has the best performance. The different models we have tested have a similar prediction, but handle heteroscedastic errors differently. The "OLS" model seems to work well as well, which is a simpler model.

By a simple linear model with time and distance, the helicopter was faster than the rapid response car after 15 kilometers. Further, our final model identified, as expected, an increase in distance as time increased and with a positive tailwind. For precipitation and a lower ceiling, the distance decreased, with a larger reduction during the winter. Under normal conditions, our final model predicts a coverage of 91.5% of the population in Central Norway for both helicopters with a flying time of 35 minutes (response time of 45 minutes). However, with a low ceiling and 1 mm precipitation, the coverage was reduced to 86.4%.

In the comparison between the helicopter and the rapid response car, the helicopter was faster after 15 kilometers. This is similar to the findings of Chen et al.(2018)[3]. By using less time to react, the rapid response car is faster at shorter distances, but the higher speed gives the helicopter an advantage on longer distances. Our result is based on a mean of all completed primary missions, but in practice, the distance before the helicopter regains the lead of the rapid response car will differ in different conditions. The difference can be caused by weather, which we later introduced in our models, but as well other factors such as traffic for the rapid response car.

By interpreting the estimates of the "GAMLSS" model, we could see a positive effect from tailwind, flying in the same direction as the wind. On the other hand, crosswind does not have a substantial impact in this model whilst flying with constant speed. However, crosswind may have a bigger influence during take off and landing. We assumed that the extra time used during take off and landing was covered by the intercept in the model, which again may have reduced the estimated coefficients of crosswind.

Further, we have observed that precipitation and lower ceiling give and reduced distance and average speed. By categorising precipitation and low ceiling as bad weather and difficult flying conditions, our models have the same conclusion as Chen et al.[3] that bad weather reduces the performance of the helicopter and gives a longer response time. In addition to their conclusion, we have as well seen an difference in influence of bad weather during the winter. Mainly, the influence seems to be greater in colder periods where January to March are the most vulnerable months. Estimates for precipitation, low ceiling and medium ceiling give the most reduction of distance/average speed in this period.

Further, we identified a difference between the helicopters based in Ålesund and Rosten. A mission starting in Ålesund covered less distance over the same time, although the helicopter in Ålesund has a higher cruising speed. This indicates a difference in the accessibility of HEMS between geographical regions, which was one of the conclusions in Zakariassen et al.[4].

In practise, we have to look at factors as a combination of all, not just one factor at a time. If there is precipitation, there is likely to be low or medium ceiling as well. By having low ceiling and 1 mm precipitation in January with a flying time of 20 minutes (1/3 hours), the distance is reduced by

$$\frac{1}{3} \cdot 1 \cdot (-10.3) + \frac{1}{3} \cdot I(\text{ceiling1} = \text{True}) \cdot (-17.9) = -9.4 \text{ kilometers}$$

in our model. A reduction of almost 10 kilometers can make a big difference, which again shows the reduction in coverage and the vulnerability for the population in the border areas.

The coverage of 91.5% of the population with a flying time of 35 minutes (response time 45 minutes) we got, is a little less than the 96.9% coverage from Røislien et al. (2018) [7]. These percentages are however not directly comparable. We have only used two helicopters in our calculations, and looked at Central Norway. If other nearby helicopters (Brønnøysund, Dombås and Ørlandet) were included, we might have gotten a higher coverage. However, we could estimate how weather affected the helicopter's performance, a reduction of 5%-points with low ceiling and 1 mm precipitation. In practice, if the weather is bad enough, the helicopter is not able to fly at all and then covers 0% of the population. This is a scenario where the rapid response car, along with the GEMS, have to cover as much as possible.

A central limitation of our models, is that only completed missions are included. Consequently, results can only be interpreted as relating to weather that is typically encountered by the helicopter.

The predictions from the models may be imprecise, which may be a result of assumptions about the data and the kind of data available. We assumed constant weather along the route of the mission, and that the helicopter travelled in a straight line. In reality, the weather can change quickly and differ in areas close to each other, and routes adapt to weather conditions and topography. In addition, some of the weather measurements are interpolated. Weather measurements were only available at ground level where wind may be less strong than at flight level. These assumptions must be taken into account when interpreting the coefficients in the model. Although we assumed constant weather, we have used several weather stations when calculating the weather for each mission. This gives a picture of how the weather is in the area, which seems to work well given the estimates from our model.

There is also some uncertainty in the data from the HEMIT data warehouse. Both when combining data from AMIS and LABAS, and in some extent the different time stamps. Some time stamps are manually registered, which in some cases might be wrong due to human error. Also, we have no information on where the helicopter was at the start of a mission. By removing missions which seem obviously wrong we solve most of this problem. However, the majority of missions are registered correctly. By just removing the outliers, we avoid removing data which would result in removing variation of interest.

Another limitation in the study is assuming an identical effects in every direction. In some parts of Central Norway there are a good amount of mountains, which would have an impact on the flight time. If topography was included, weather elements may have an other effect as well. For

example, a lower ceiling will primarily have a negative influence around mountains since it forces the pilot to fly a longer path. With no mountains, there may not be a difference between a low or high ceiling. In our model, we have partly solved this problem by including starting point in our model. By doing this, we were able to differentiate the topographical differences around Ålesund and Rosten to some extent.

When analysing coverage, we only included inhabitants in Central Norway, although the population is served by helicopters in neighboring parts of the country. If, however, weather affects all resources similarly, the size of the reduction in coverage with bad weather may be indicative of the countrywide effect. Also, by not including missions rejected because of severe weather, we may underestimate the impact of weather on coverage.

Despite some limitations, our model appears to produce estimates that may be of use when dispatching EMS or planning services. We believe that a model that handles the limitations with more complex solutions may provide more accurate estimates.

6.1 Impact and further work

The work done in this thesis can be expanded in several ways. An obvious next step would be to include more detailed information about the weather and topography from helicopter base to scene of incident. With our model, we are able analyse the effect of weather and other factors during a helicopter mission. By improving with more detailed information and data, these predictions can be less affected by uncertainty.

Currently, the air ambulance crew can access real-time pictures of the weather from different points between base and patient. This is used to determine if it is possible to fly or not, which already is a tool that saves time. Not just in time to reach the patient, but as well regarding the duty time. By using real-time pictures to decline dispatches, they avoid using unnecessary working time which could make the helicopter with crew unavailable due to the rest time restrictions described in Section 1.2.

Combining real-time weather data with a model predicting distance/time may save additional time. For example, if there has been an incident and dispatching the air ambulance is considered: From the helicopter base to the patient, there may exist different alternative routes,

- A) Shortest distance, over mountains and vulnerable for difficult conditions.
- B) Longer distance, around mountains and less vulnerable for difficult conditions.

In different types of conditions, both of these routes can be the optimal choice. Therefore, a tool that can predict time used for both routes, taking weather and topographic data into consideration would be useful. This could provide great help in the decision of dispatching the best unit, and as well choosing the optimal route from helicopter base to patient with a reduced risk of having to abort the mission.

Future development to better address the need for information when dispatching EMS could also address conditions not suitable for flying, which our model does not address. This can include identifying challenging weather conditions as well as conditions for landing.

A potential for such models is that changing weather conditions due to climate change can be taken into account when planning the future organisation of services. Another direction for

future work may be to predict changes in emergency preparedness over time.

6.2 Conclusion

There is a big potential in using weather data when looking at performance of HEMS. We have seen that the helicopter performs different in bad conditions, resulting in a shorter distance reached during the same time compared to flying in better conditions. Using such information can be a big advantage, and save time if used properly and with some improvements on the method, data and model. Time is one of several factors involved when dispatching, therefore, a tool like this should not be used by itself. Instead, it could be used as one of several tools, where it can give an indication of the best route and unit to dispatch based on time used.

Bibliography

- [1] *Concept variable: Basic statistical unit*. <https://www.ssb.no/a/metadata/conceptvariable/vardok/135/en>. Accessed: 2022-06-22.
- [2] *Værkameraene viser været langs ruten*. <https://norskluftambulanse.no/innovasjon/vaerkameraene-viser-vaeret-langs-flyruten/>. Accessed: 2022-06-16.
- [3] Xilin Chen et al. ‘Logistics of air medical transport: When and where does helicopter transport reduce prehospital time for trauma?’ In: *Journal of Trauma and Acute Care Surgery* 85.1 (2018), pp. 174–181.
- [4] Erik Zakariassen, Oddvar Uleberg and Jo Røislien. ‘Helicopter emergency medical services response times in Norway: do they matter?’ In: *Air Medical Journal* 34.2 (2015), pp. 98–103.
- [5] *Nøkkeltall om ambulansetjenesten*. <https://helse-midt.no/om-oss/nokkeltall/nokkeltall-om-ambulansetjenesten>. Accessed: 2022-04-27.
- [6] *About the National Air Ambulance Services of Norway*. <http://www.luftambulanse.no/about-national-air-ambulance-services-norway>. Accessed: 2021-11-10.
- [7] Jo Røislien et al. ‘Comparing population and incident data for optimal air ambulance base locations in Norway’. In: *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine* 26.1 (2018), pp. 1–11.
- [8] *Gjennomgang av basestruktur for ambulanshelikopter, 2020*. <https://helse-vest.no/seksjon/styresaker/Documents/2020/28.09.2020/Sak%2010820%20Vedl.%201%20-%20Rapport%20gjennomgang%20basestruktur.pdf>. Accessed: 2021-12-01.
- [9] *Prosedyre for varsling og kommunikasjon mellom AMK og LA-tjenesten*. <https://www.helsedirektoratet.no/veiledere/nodnett-i-helsetjenesten/vedlegg/prosedyre-for-varsling-og-kommunikasjon-mellom-amk-og-la-tjenesten>. Accessed: 2022-05-02.
- [10] *Rekvirering, varsling, koordinering og flight following av luftambulanshelikopter*. <https://kokom.no/wp-content/uploads/2019/09/20190607-Modul-6-Versjon-2-med-f%C3%B8lgebrev.pdf>. Accessed: 2022-05-02.
- [11] *Definisjonskatalog for den akuttmedisinske kjede*. http://luftambulanse.no/system/files/internett-vedlegg/Defkat-akuttmed_kjede-v1.1-31072012.pdf. Accessed: 2022-06-08.
- [12] *Organisering av luftambulansetjenesten*. <https://www.regjeringen.no/contentassets/f01597ca73e648f3aab3a512f8cno/pdfs/organisering-av-luftambulansetjenesten.pdf>. Accessed: 2022-06-22.
- [13] *Forskrift om luftfartsoperasjoner*. <https://lovdata.no/dokument/SF/forskrift/2013-08-07-956>. Accessed: 2022-06-22.
- [14] *What is Frost?* <https://frost.met.no/index.html>. Accessed: 2021-10-01.
- [15] *Weather and Climate Elements(Frost API)*. <https://frost.met.no/elementtable>. Accessed: 2021-10-01.
- [16] *Befolkning: Grunnkretsenes befolkning*. <https://www.ssb.no/statbank/table/04317/>. Accessed: 2022-06-22.
- [17] Ludwig Fahrmeir et al. ‘Regression’. In: *Models, Methods and Applications*. Springer, 2013.
- [18] James Gareth et al. *An introduction to statistical learning: with applications in R*. Spinger, 2013.

-
- [19] Feng-Wen Chen and Chen-Wuing Liu. ‘Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan’. In: *Paddy and Water Environment* 10.3 (2012), pp. 209–222.
- [20] K Hartmann, J Krois and B Waske. ‘E-learning project SOGA: Statistics and geospatial data analysis’. In: *Department of Earth Sciences, Freie Universitaet Berlin* 33 (2018).
- [21] *Crosswind and headwind calculation*. https://mediawiki.ivao.aero/index.php?title=Crosswind_and_Headwind_calculation. Accessed: 2022-04-06.
- [22] *Annex 2 to the Convention on International Civil Aviation: Rules of the air*. https://www.icao.int/Meetings/anconf12/Document%20Archive/an02_cons%5B1%5D.pdf. Accessed: 2022-06-17.
- [23] *Temperatur og fuktighet*. <https://www.uib.no/gfi/57482/temperatur-og-fuktighet>. Accessed: 2022-06-17.
- [24] *Luftambulansebasen i Trondheim*. <http://www.luftambulanse.no/baser/trondheim>. Accessed: 2022-02-16.
- [25] *Luftambulansebasen i Ålesund*. <http://www.luftambulanse.no/baser/alesund>. Accessed: 2022-02-16.
- [26] David Kahle and Hadley Wickham. ‘ggmap: Spatial Visualization with ggplot2’. In: *The R Journal* 5.1 (2013), pp. 144–161. URL: <https://journal.r-project.org/archive/2013-1/kahle-wickham.pdf>.
- [27] *R example script*. https://frost.met.no/r_example.html. Accessed: 2021-11-10.

Appendix

A Confidence interval

A.1 OLS model

Table with a 95% confidence interval of the estimated values for the OLS model.

Features	2.5 %	97.5 %
(Intercept)	0.4766	2.8202
Time	180.6060	187.5723
Time:Tailwind	0.6333	0.8810
Time:Crosswind	-0.1055	0.3344
Time:LocationAalesund	-16.7429	-11.7124
Time:Precipitation:Factor(season)1	-25.6955	-1.2667
Time:Precipitation:Factor(season)2	-14.8730	7.1729
Time:Precipitation:Factor(season)3	-6.4090	4.5432
Time:Precipitation:Factor(season)4	-20.2742	-5.5556
Time:Factor(season)1:Factor(ceiling)1	-26.3371	-2.9425
Time:Factor(season)2:Factor(ceiling)1	-12.3987	11.1976
Time:Factor(season)3:Factor(ceiling)1	-20.1340	-5.9617
Time:Factor(season)4:Factor(ceiling)1	-21.9339	-9.6476
Time:Factor(season)1:Factor(ceiling)2	-13.0593	-0.8043
Time:Factor(season)2:Factor(ceiling)2	-14.6459	-2.8966
Time:Factor(season)3:Factor(ceiling)2	-6.1520	3.5551
Time:Factor(season)4:Factor(ceiling)2	-9.9558	0.0697

Table 15: Confidence interval to estimated coefficients for the OLS model.

A.2 GAMLSS model

Table with a 95% confidence interval of the estimated values for the GAMLSS model.

Features	2.5 %	97.5 %
Mu estimates		
(Intercept)	-2.6708157	-0.5785040
Time	191.4063287	198.9058238
Time:Tailwind	0.5646955	0.8359491
Time:Crosswind	-0.2758971	0.1984276
Time:LocationAalesund	-18.8517083	-13.2729518
Time:Precipitation:Factor(season)1	-23.2631955	2.6675110
Time:Precipitation:Factor(season)2	-19.4072079	4.3718175
Time:Precipitation:Factor(season)3	-5.0270249	5.8673414
Time:Precipitation:Factor(season)4	-15.6594104	0.0370925
Time:Factor(season)1:Factor(Ceiling)1	-29.0080983	-6.7759170
Time:Factor(season)2:Factor(Ceiling)1	-13.6269653	8.3253272
Time:Factor(season)3:Factor(Ceiling)1	-23.0228542	-7.8660490
Time:Factor(season)4:Factor(Ceiling)1	-21.0675027	-8.0193220
Time:Factor(season)1:Factor(Ceiling)2	-15.2697709	-2.3290224
Time:Factor(season)2:Factor(Ceiling)2	-11.9509665	0.7206780
Time:Factor(season)3:Factor(Ceiling)2	-8.9309167	1.4860548
Time:Factor(season)4:Factor(Ceiling)2	-10.0477929	0.5762778
Sigma estimates		
(Intercept)	2.0317255	2.1386530
Time	1.4803520	1.7472315
Time:LocationAalesund	0.2886367	0.4996553

Table 16: Confidence interval to estimated coefficients for the GAMLSS model.

B R code

This appendix have some code from R-studio showing some steps from retrieving data, fitting the models and visualisation of the "GAMLSS" model.

B.1 Extract weather data

Example of code used for extracting weather data from Frost API. Uses data of all stations in "Trøndelag" and "Møre & Romsdalen". Some part of the code are from an example tutorial of how to use Frost Met[27].

```
# Libraries needed
library(jsonlite)
library(tidyr)
library(dplyr)
library(lubridate)

#weather stations in central norway
stations_trd = read.csv("FILE_NAME")
stations_mr = read.csv("FILE_NAME")

get_weather_data = function(sources, referenceTime, timeres, weather_elements){
  #Function that retrieve data of weather elemtns during a chosen time
  → periode
  # Insert your own client ID here
  client_id = 'ac97e322-7f9c-4bf7-b143-417ff524d3e6'
  #Collaps weather elements into one string
  elements = paste(as.character(weather_elements), collapse=", ")
  sources = paste(as.character(sources), collapse=", ")
  endpoint <- paste0("https://", client_id,
                    "@frost.met.no/observations/v0.jsonld")
  # Build the URL to Frost
  url <- paste0(
    endpoint, "?",
    "sources=", sources,
    "&referencetime=", referenceTime,
    "&elements=", elements,
    "&timeresolutions=", timeres
  )
  # Issue an HTTP GET request and extract JSON data
  xs <- try(fromJSON(URLEncode(url),flatten=T))

  # Check if the request worked, print out any errors
  if (class(xs) != 'try-error') {
    df <- unnest(xs$data, cols = c("sourceId","referenceTime","observations"))
    print("Data retrieved from frost.met.no!")
  }
}
```

```

} else {
  print("Error: the data retrieval was not successful!")
}
# These additional columns will be kept
columns=
  ↪ c("sourceId","referenceTime","elementId","value","unit","timeOffset")
df2 = df[columns]
# Convert the time value to something R understands
df2$referenceTime <- as.character(strptime(df2$referenceTime,
                                         format = "%Y-%m-%dT%H:%M:%S", tz =
                                         ↪ ""))
df2$sourceId <- gsub(':', '', df2$sourceId)
return(df2)
}

sort_weather_data = function(data, weather_elements){
  #'Function that sorts the data in better structure. One row for all weather
  #'elements to one statin for one hour
  count = 0
  for(i in weather_elements){
    if(count == 0){
      test_df = data[data$elementId == i,]
      if(nrow(test_df) != 0){
        test_df = select(test_df, -c("elementId", "unit", "timeOffset"))
        names(test_df)[names(test_df) == 'value'] <- i
        count += 1
      }
    } else {
      test2_df = data[data$elementId == i,]
      if(nrow(test2_df) != 0){
        test2_df = select(test2_df, -c("elementId", "unit", "timeOffset"))
        names(test2_df)[names(test2_df) == 'value'] <- i

        test_df = merge(test_df, test2_df, by =
          ↪ c("sourceId","referenceTime"),
          all = TRUE)
      }
    }
  }
  names(test_df)[names(test_df) ==
    "sum(precipitation_amount PT1H)"] = "prec_amount"
  return(test_df)
}

get_wearer_data_in_periode = function(sources,startDate,endDate,
                                     timeres, weather_elements){
  #'Function that retrieves data from a chosen time periode
  df = data.frame(NA)

```

```

#Takes 20 days at a time
daysPlus = days(20)
while(startDate < endDate){
  #Make referenceTime to "get_weather_data" function
  if(startDate + daysPlus > endDate){
    time = paste(startDate, endDate, sep="/")
  }else{
    time = paste(startDate, startDate + daysPlus, sep="/")
  }
  #increase startDate to next iteration
  startDate = startDate + daysPlus
  #Get data in this time period
  wthr_df = get_weather_data(sources = sources,
                             weather_elements = weather_elements,
                             timeres = timeres, referenceTime = time )
  new_df = sort_weather_data(data = wthr_df,
                             weather_elements = weather_elements)
  #Rounding of to closest hour if some time stamps have minutes
  new_df$referenceTime = format(round(as.POSIXlt(new_df$referenceTime),
                                           units="hours"), format="%Y-%m-%d %H:%M:%S")

  #New data frame for first iteration,
  #add the other iterations to this data fram
  if(is.na(df[1,1])){
    df = new_df
  }else{
    df = rbind(df, new_df)
  }
  print(time)
}
return(df)
}

#Set parameters for functions
startDate = ymd("2020-01-01")
endDate = ymd("2020-01-31")
weather_elements = c("sum(precipitation_amount PT1H)", "air_temperature",
                    "wind_speed", "wind_from_direction",
                    "dew_point_temperature")
timeres <- 'PT1H'

#Retrive data from weather stations in Møre og Romsdalen
sources_mr = paste(as.character(stations_mr$id), collapse=", ")
weather_mr = get_weater_data_in_periode(sources = sources_mr,
                                       startDate = startDate, endDate =
                                       ↪ endDate,
                                       timeres = timeres,
                                       weather_elements = weather_elements)
weather_mr = merge(x = weather_mr, y = stations_mr, by.x = "sourceId",

```

```

        by.y = "id", all.x = TRUE)

#Retrieve data from weather stations in Trøndelag
sources_trd = paste(as.character(stations_trd$id), collapse=", ")
weather_trd = get_wewater_data_in_periode(sources = sources_trd,
                                           startDate = startDate, endDate=
                                           ↪ endDate
                                           , timeres = timeres,
                                           weather_elements = weather_elements)
weather_trd = merge(x = weather_trd, y = stations_trd,
                    by.x = "sourceId", by.y = "id", all.x = TRUE)

#Combine the retrieved data
weather_data = rbind(weather_trd, weather_mr)

```

B.2 Models

Some of the code used when fitting different models. Data used is the merged data of air ambulance missions and the weather. Some variables have been used to calculate new variables.

```

library(caret)
library(gamlss)

#Load data
df = read.csv("FILE_NAME")
#Add time as hours
df$tid_transport_til_pas_h = df$tid_transport_til_pas/60

#Set negative rain and ceiling to "0"
df[df$mean_ceiling < 0,]$mean_ceiling = 0
df[df$mean_prec_amount < 0,]$mean_prec_amount = 0

#Set month into periods
df$season = 0
df[!is.na(df$month) & df$month < 4,]$season = 1
df[!is.na(df$month) & df$month > 3 & df$month < 7,]$season = 2
df[!is.na(df$month) & df$month > 6 & df$month < 10,]$season = 3
df[!is.na(df$month) & df$month > 9,]$season = 4

#Set ceiling categories
df$ceiling = 0
df[!is.na(df$mean_ceiling) & df$mean_ceiling <= 500,]$ceiling = 1
df[!is.na(df$mean_ceiling) & df$mean_ceiling > 500 &
   df$mean_ceiling <=1000,]$ceiling = 2

#Split in air and ground ambulance

```

```

df_luft= df[(df$avvik == 1) & (df$type == "Air") &
            !(is.na(df$tid_transport_til_pas)),]
df_bil = df[(df$avvik == 1) & (df$type == "Ground") &
            !(is.na(df$tid_transport_til_pas)),]

#Features of interest
features = c("dist_til_pas", "act_tid", "tid_transport_til_pas",
            ↪ "tid_transport_til_pas_h", "resp_tid", "mean_prec_amount",
              "prec_type", "crosswind", "crosswind_kmh", "tailwind",
              "tailwind_kmh", "mean_air_temp", "mean_dew_point_temp",
              "mean_ceiling", "lokasjon", "year", "month", "week", "day",
              "season", "ceiling")
#Dataframe With features of interest
heli_df = df_luft[,features]
bil_df = df_bil[,features]
bil_df = bil_df[bil_df$tid_transport_til_pas < 1000,]

#Remove observations where the helicopter have an average speed of <100km/t
↪ and >400km/t
heli_df2 = heli_df
heli_df2$speed = heli_df2$dist_til_pas/(heli_df2$tid_transport_til_pas/60)
#Remove rows with "wrong" speed
new_heli_df = heli_df2[heli_df2$speed > 100 & heli_df2$speed < 400,]

#Formula for regression model
formula = dist_til_pas ~ tid_transport_til_pas_h +
          tid_transport_til_pas_h:tailwind_kmh +
          tid_transport_til_pas_h:crosswind_kmh +
          tid_transport_til_pas_h:mean_prec_amount:factor(season)+
          lokasjon:tid_transport_til_pas_h+
          factor(ceiling):tid_transport_til_pas_h:factor(season)

#### Baseline model #####
base_mod = lm(formula, data = new_heli_df)
sum_lm_mod = summary(base_mod)
sum_lm_mod

#Confidenc interval to OLS model
ols_confint = as.data.frame(confint(base_mod))

##### Log of the reponse #####
log_heli_mod = lm(log(dist_til_pas) ~ tid_transport_til_pas_h +
                  tid_transport_til_pas_h:tailwind_kmh +
                  tid_transport_til_pas_h:crosswind_kmh +
                  tid_transport_til_pas_h:mean_prec_amount:factor(season)+
                  lokasjon:tid_transport_til_pas_h+
                  factor(ceiling):tid_transport_til_pas_h:factor(season),

```

```

        data = new_heli_df)
sum_log_mod = summary(log_heli_mod)

#####Weighted linear regression (lin.reg with the new weights) #####
#calculate weights: two stage method
cal.weights <- 1 / exp(lm(log(base_mod$residuals^2) ~
                           base_mod$fitted.values)$fitted.values)
#fit model
wls_mod = lm(formula,
              data = new_heli_df, weights = cal.weights)

sum_wls_mod = summary(wls_mod)

##### weights: 1/time #####
wls_mod2 = lm(formula,
              data = new_heli_df, weights = 1/(tid_transport_til_pas_h))
sum_wls_mod2 = summary(wls_mod2)

##### GAMLSS model#####
#formula for sigma
formula_sigma = ~tid_transport_til_pas_h +
  tid_transport_til_pas_h:factor(lokasjon)
#fit model
gamlss_mod <- gamlss(formula = formula,
                    sigma.formula = formula_sigma,
                    family = NO(), data = new_heli_df)
sum_gamlss_mod = as.data.frame(summary(gamlss_mod))

##### Cross validation #####
#Set seed
set.seed(123)
train_control<- trainControl(method = "cv",
                             number = 10)

#Cross-validation
cv_ols = train(formula,
               data = new_heli_df, trControl=train_control, method = "lm")
cv_wls<- train(formula,
               data = new_heli_df, trControl=train_control, method = "lm",
               weights = cal.weights)
cv_wls2<- train(formula,
                data = new_heli_df, trControl=train_control, method = "lm",
                weights = 1/dist_til_pas)

print(cv_ols$results)
print(cv_wls$results)
print(cv_wls2$results)

```



```

##Cross validation for GAMLSS model
#set.seed(123)
cv_data = new_heli_df
a = createFolds(cv_data$dist_til_pas, k = 10, list = T, returnTrain = FALSE)

cv_data$predsgamlss = 1000000
gamlss_rmse = rep(1000, 10)
gamlss_mae = rep(1000, 10)
gamlss_r2 = rep(1000, 10)
for(i in 1:length(a)){
  train_set = new_heli_df[-a[[i]],]
  test_set = new_heli_df[a[[i]],]
  gamlss_mod <- gamlss(formula = formula,
                      sigma.formula = formula_sigma,
                      family = NO(), data = train_set)
  gamlsspred = predict(gamlss_mod, newdata = test_set, what="mu",
                      type="response")

  cv_data[a[[i]],]$predsgamlss = gamlsspred
  rmse_i = RMSE(gamlsspred, new_heli_df[a[[i]],]$dist_til_pas)
  gamlss_rmse[i] = rmse_i
  mae_i = MAE(gamlsspred, new_heli_df[a[[i]],]$dist_til_pas)
  gamlss_mae[i] = mae_i
  r2_i = R2(gamlsspred, new_heli_df[a[[i]],]$dist_til_pas)
  gamlss_r2[i] = r2_i
}

cv_gamlss = data.frame("RMSE" = mean(gamlss_rmse),
                      "Rsquared" = mean(gamlss_r2),
                      "MAE" = mean(gamlss_mae))
#Combine results from all 4 model in one data frame
cv_results = rbind(cv_ols$results[2:4], cv_wls$results[2:4],
                  cv_wls2$results[2:4], cv_gamlss)
rownames(cv_results) = c("OLS", "WLS", "WLS2", "GAMLSS" )
cv_results

```

B.3 Visualizations

Example of code used when visualizing the figures in Sections 5.3. We have used the "GAMLSS" model when predicting distances used in the figures. We have included functions made, and one example plot. Map plotted with R-package *ggmap* [26].

```

library(ggplot2)
library(ggmap)
library(ggsn) #devtools::install_github('oswaldosantos/ggsn')

```

```

library(gamlss)
library(geosphere)

#library(grid)
#library(gridExtra)

#Load data
heli_df = read.csv("FILE_NAME")

#Make the model
#Formula for regression model
formula = dist_til_pas ~ tid_transport_til_pas_h +
  tid_transport_til_pas_h:tailwind_km_h +
  tid_transport_til_pas_h:crosswind_km_h +
  tid_transport_til_pas_h:mean_prec_amount:factor(season)+
  lokasjon:tid_transport_til_pas_h+
  factor(ceiling):tid_transport_til_pas_h:factor(season)

formula_sigma = ~tid_transport_til_pas_h +
  tid_transport_til_pas_h:factor(lokasjon)

## GAMLSS mod
gamlss_mod <- gamlss(formula,
                     sigma.formula = formula_sigma,
                     family = NO(), data = heli_df)
summary(gamlss_mod)

gamlss_mod_car <- gamlss(formula,
                         sigma.formula = formula_sigma,
                         family = NO(), data = car_df)
summary(gamlss_mod_car)

#MAPS
base_map = function(min_lon, max_lon, min_lat, max_lat){
  mapImage <- get_stamenmap(bbox = c(min_lon,min_lat,max_lon,max_lat),
                           source = "stamen",
                           maptype = "toner-background",
                           zoom = 7,
                           color = "bw")

  # change opacity of basemap
  mapatt <- attributes(mapImage)
  map_transparent <- matrix(adjustcolor(mapImage, alpha.f = 0.2),
                            nrow = nrow(mapImage))
  attributes(map_transparent) <- mapatt
#Make plot
plot = ggmap(map_transparent) +
  theme(panel.grid = element_blank(),
        panel.background = element_rect(fill = "white")) +

```

```

    xlab("Longitude") + ylab("Latitude") + scalebar(x.min = min_lon+0.5,
                                                    x.max = max_lon,
                                                    y.min = min_lat+0.5,
                                                    y.max = max_lat,
                                                    location = "bottomleft",
                                                    transform =T,
                                                    dist = 50,dist_unit =
                                                    ↪ "km",
                                                    model = "WGS84",

                                                    ↪ box.fill=c("black","white"),
                                                    st.size = 4 )

  return(plot)
}
#Make plots
map = base_map(3,15,60,65)
map_trd = base_map(5,15,61.5,65.5)
map_ale = base_map(3,12,60.5,64.5)
#Points at the location to the bases
bases = geom_point(aes(x = longitude, y = latitude),data = data,
                   color = "#ff0000")

predict_distance_gamlss = function(model, wind_speed, wind_dir, time,
                                   lokasjon, ceiling, prec_amount, season){
  #'Predict distance the helicopter can reach given some
  #'parameters with GAMLSS model
  #'Wind_speed: positive value of the speed of wind
  #'wind_dir: The dirction which the wind is COMMING from, i.e, 180 comes
  #'from south and blows north
  #'time: minutes
  #'Lokasjon: Rosten og Aalesund
  #'Ceiling: positive value(checks if under 1000 or not)
  #'prec_amount: precipitation in millimeters
  #'season: 1(jan-mar), ...,4(oct-dec)

  angles = seq(0,2*pi,length=200) *180/pi

  alpha = angles - wind_dir - 180
  twind = wind_speed*cos((alpha) *(pi/180))
  cwind = abs(wind_speed*sin((alpha) *(pi/180)))

  dd = data.frame("tid_transport_til_pas_h"=rep(time,200),
                  "tailwind_km_h"= twind, "crosswind_km_h"=cwind,
                  "ceiling"= rep(ceiling,200), lokasjon = rep(lokasjon, 200),
                  "mean_prec_amount" = rep(prec_amount,200),
                  "season" = rep(season,200))
  pred_mu_gamlss <- predict(gamlss_mod, what = "mu", newdata = dd,
                           type = "response")
  pred_mu_sigma = predict(gamlss_mod, what = "sigma", newdata = dd,

```

```

                                type = "response", level = 0.95)
dd$pred= pred_mu_gamlss
dd$lci = pred_mu_gamlss - pred_mu_sigma
dd$uci = pred_mu_gamlss + pred_mu_sigma
dd$angles = angles
return(dd)
}

make_circles = function(pred,lokasjon,mean = F){
  #'Function that construct circles where the helicopter can reach
  #'given some predictions
  #'pred: predicted distances
  #'lokasjon: which helicopter base ("Rosten", "Aalesund")
  #'mean: FALSE if predicted distances is used. TRUE if the mean of
  #'predictions is used, i.e. wind_speed = 0km/t
  #'Only use rows to chosen helicopter
  pred = pred[pred$lokasjon == lokasjon,]
  nPoints = length(pred$pred)
  if(mean == T){
    print("Mean")
    pred$pred = mean(pred$pred)
  }else{
    pred$pred = (pred$pred)
  }
  #DF with coordinates to helicopter base
  data = data.frame(
    ID = c("Rosten", "Aalesund"),
    latitude = as.numeric(c(63.368, 62.464226)),
    longitude = as.numeric(c(10.374, 6.312053))
  )
  data = data[data$ID == lokasjon,]
  Circle <- data.frame(ID = rep(data$ID, nPoints))

  #Empry df to be filled with new lon lat coordinats
  lonlat = data.frame(lon = vector("numeric", nPoints),
                      lat = vector("numeric", nPoints))

  for(i in 1:nPoints){
    lonlat[i,] = destPoint(cbind(data$longitude[1], data$latitude[1]),
                          pred$angles[i], (pred$pred[i])*1000,
                          a=6378137, f=1/298.257223563)
  }

  Circle$lon = lonlat[,1]
  Circle$lat = lonlat[,2]
  return(Circle)
}

wind_arrow = function(wind_speed, wind_dir, x_start = 5, y_start = 65){

```

```

#'Function that makes an arrow showing direction and "speed" of the wind
#'wind_speed: positive value >= 0
#'wind_dir: direction the wind is coming from on the interval [0,360]
#'x_start, y_start: lon,lat coordinates where the arrow begins
#end_x = (wind_speed/50) * cos((wind_dir - 180) *(pi/180))
#end_y = (wind_speed/50) * cos((wind_dir - 180) *(pi/180))
end = destPoint(cbind(x_start, y_start), wind_dir - 180,
                (wind_speed)*1000, a=6378137, f=1/298.257223563)
x_end = end[1]
y_end = end[2]

coords = data.frame(x_start =x_start, y_start = y_start,
                    x_end = x_end,y_end = y_end )
return(coords)
}

#DF with coordinates to helicopter base
data = data.frame(
  ID = c("Rosten", "Aalesund"),
  latitude = as.numeric(c(63.368, 62.464226)),
  longitude = as.numeric(c(10.374, 6.312053))
)

pred_x_min_gamlss = function(xmin, gamlssmod, wind_speed, wind_dir,ceiling,
                             prec_amount, season){
  #'Predict distance with a spevified model
  #'xmin : Number of minutes that the helicopter is in the air
  trd_xmin = predict_distance_gamlss(gamlssmod, wind_speed, wind_dir,
                                     time = xmin/60, "Rosten", ceiling,
                                     prec_amount, season)
  ale_xmin = predict_distance_gamlss(gamlssmod, wind_speed, wind_dir,
                                     time = xmin/60, "Aalesund", ceiling,
                                     prec_amount, season)
  pred_xmin = rbind(trd_xmin,ale_xmin)
  return(pred_xmin)
}

pred_to_circle = function(preds, mean = F){
  #'Use prediction(preds) to make lon/lat coordinates
  trd_circle = make_circles(preds, lokasjon = "Rosten", mean)
  ale_circle = make_circles(preds, lokasjon = "Aalesund",mean)
  circles = rbind(trd_circle, ale_circle)
  return(circles)
}

#Parameters
wind_speed = 40 #km/t
wind_dir = 45 + 180 #Direction which the wind is COMMING from
time = (25/60) #min/60 to get hours

```

```

ceiling = 0
prec_amount = 0
season = 1
#model som brukes til predictions
gamlssmod = gamlss_mod

#Arrow for wind speed/direction
arrow = wind_arrow(wind_speed, wind_dir,x_start = 5, y_start = 64)
arrow_trd = wind_arrow(wind_speed, wind_dir, x_start = 7, y_start = 64.75)
arrow_ale = wind_arrow(wind_speed, wind_dir, x_start = 10, y_start = 62.5)

#Predict distance from base in all directions
pred35_gamlss = pred_x_min_gamlss(35,gamlssmod, wind_speed, wind_dir, ceiling,
                                prec_amount, season)

#Make predictions to circles around base
circles35 = pred_to_circle(pred35_gamlss, mean=F)
#Plot circle with map and base
map + bases +
  geom_polygon(data = circles35, aes(lon, lat, group = ID, color = ID),
              size = 1, alpha = 0) +
  theme(plot.title = element_text(hjust = 0.5),text = element_text(size=20)) +
  labs(color="Helikopter")

```

