

# Master's Thesis

## Application of Ensemble Prediction on a Visibility Parametrization based on Mie Scattering Theory

by Magdalena Elisabeth Haselsteiner

Freie Universität Berlin  
Fachbereich Physik

Title	Application of Ensemble Prediction on a Visibility Parametrization based on Mie Scattering Theory
Author	Magdalena Elisabeth Haselsteiner
Matriculation Number	4973395
Date of Birth	27.5.1992, Korneuburg, Austria
University/Department	Free University Berlin, Department of Physics
Course of Studies	Master studies, Physics
1 <sup>st</sup> Reviewer	Prof. Dr. Roland Netz
2 <sup>nd</sup> Reviewer	PD. Dr. Stefanie Russ
In Cooperation with	Zentralanstalt für Meteorologie und Geodynamik
Date of Submission	08.05.2018

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Background Theory</b>	<b>4</b>
2.1	Numerical Weather Prediction Models . . . . .	4
2.1.1	Primitive Equations . . . . .	4
2.1.2	Approximations . . . . .	5
2.1.3	System of Coupled Partial Differential Equations . . . . .	6
2.2	Ensemble Prediction . . . . .	7
2.2.1	The Atmosphere's Chaotic Nature . . . . .	7
2.2.2	Ensembles . . . . .	9
2.2.3	Uncertainty and Error . . . . .	11
2.2.4	Ensemble Creation by Monte-Carlo Methods . . . . .	12
2.3	Visibility . . . . .	14
2.3.1	Definitions . . . . .	14
2.3.2	Mathematical Derivation of Visibility and Visual Range . . . . .	15
2.3.3	Atmospheric Scattering . . . . .	16
<b>3</b>	<b>Model</b>	<b>19</b>
3.1	Spatial Discretization . . . . .	20
3.2	Pattern Generator . . . . .	22
3.2.1	Technical Set-Up of the Pattern Generator . . . . .	22
<b>4</b>	<b>Parametrization of Visibility</b>	<b>25</b>
<b>5</b>	<b>Methods and Set-Up</b>	<b>30</b>
5.1	Perturbation Scheme . . . . .	30
5.1.1	Set-Up of Perturbations . . . . .	32
5.2	Set-Up . . . . .	35
5.3	Evaluation . . . . .	36
5.3.1	Measurements . . . . .	37
5.3.2	Scores . . . . .	38
<b>6</b>	<b>Results</b>	<b>41</b>
<b>7</b>	<b>Conclusion and Outlook</b>	<b>53</b>
	<b>Bibliography</b>	<b>56</b>

<b>A</b>	<b>Tables</b>	<b>63</b>
A.1	Set-Up of Perturbation . . . . .	63
A.2	Variables and Abbreviations . . . . .	66
<b>B</b>	<b>Images</b>	<b>71</b>
B.1	Surface Weather Analysis . . . . .	71

# List of Figures

2.1	Deformation of Phase Space Volume . . . . .	9
2.2	Example of Ensemble Forecast . . . . .	11
2.3	Scatter Plots of Mie and Rayleigh scattering . . . . .	17
3.1	Zones in AROME's Computational Domain . . . . .	20
3.2	Comparison of Vertical Coordinate Systems . . . . .	21
3.3	Lambert Projection . . . . .	24
3.4	Example of a Small-Scale Perturbation . . . . .	24
5.1	Illustration of Perturbation Scheme . . . . .	31
5.2	Map of Observation Sites . . . . .	37
6.1	Box Plots of MSE, July 2016 . . . . .	43
6.2	Box Plots of Mean Squared Difference of Logs, July 2016 . . . . .	43
6.3	Box Plots of MSE, January 2017 . . . . .	44
6.4	Box Plots of Mean Squared Difference of Logs, January 2017 . . . . .	44
6.5	Contour Plot of Forecast Visibility . . . . .	45
6.6	Stations Marked with RPSS 27-01-2017 . . . . .	46
6.7	Satellite Image 27-01-2017, 09:00 UTC . . . . .	46
6.8	Timeline of RPSS for January 2017 . . . . .	48
6.9	Timeline of RPSS for July 2017 . . . . .	49
6.10	Stations Marked with RPSS for July 2017 . . . . .	50
6.11	Stations Marked with RPSS for January 2017 . . . . .	51
6.12	Case studies of Different Perturbation Schemes . . . . .	52
B.1	Surface Weather Analysis 06:00 UTC . . . . .	71
B.2	Surface Weather Analysis 12:00 UTC . . . . .	72

## **Abstract**

This Master's thesis aims to describe and evaluate a new parametrization for forecasting visibility within numerical weather prediction. The parametrization represents Mie scattering on atmospheric aerosols and hydrometeors and is implemented in a numerical weather prediction model, written in Fortran.

To examine the uncertainty in the parametrization and its effects, ensemble prediction is used. A new local perturbation scheme is introduced. For the verification of the visibility forecast, a logarithmic score and the Ranked Probability Skill Score are chosen to evaluate the predicted values against the observed values, which were taken at meteorological observation sites on the domain of the forecast.

The new parametrization in its current implementation outperforms the reference parametrization only in a few cases. However, we have reason to believe that further improvements the aerosol scheme will result in a significant increase of performance. The new perturbation scheme enhanced the ensemble skill remarkably in the whole verification period. The scheme can also be generalized and applied to other parametrizations or model parts of similar kind.

# Chapter 1

## Introduction

The humankind has always been exposed to nature, especially the weather. Change of climate or extreme weather events have erased whole cultures from our planet, so no wonder that since the very early days of humanity, we observed our atmosphere, tried to find patterns and predict upcoming weather events. Already in 340 BC Aristotle wrote the ‘Meteorologica’, which can be seen as the very first documented structured approach to weather forecasting. For a long time, meteorology was a purely empirical science, because there was no deeper understanding of the underlying physical processes. It was only at the beginning of the 20th century when Abbe [1] was the first one to draw the connection between large scale atmospheric processes and physics. From there on we have come a very long way to where we are today. Because weather prediction is an overlap of physics, mathematics and, more recently, computer science, meteorology has evolved hand in hand with those fields. Due to enhancements in above-mentioned disciplines, forecasts improved rapidly in the last 40 years: one day per decade. Meaning that today the accuracy of a prediction four days ahead is the same as the accuracy of a prediction three days ahead ten years ago [4]. Nevertheless researchers in universities, private enterprises, governmental and military institutions are still trying to push it further. This is because the weather affects us in so many different ways, that having skilful predictions caters to different, sometimes even opposed, needs and interests. It can be a major contributor to sustain peace and stability for humanity as a whole, especially because of its role in agriculture and food production, and it can also be a valuable tool for the prevention of economic losses due to weather events and for the management of natural disasters [69].

The basis of modern weather forecast is to mathematically describe the atmosphere as a turbulent fluid, using thermodynamics and fluid mechanics, and to solve the corresponding equations numerically [14, 3, 47].

The atmosphere is an intrinsically chaotic system and therefore it is very challenging to model and predict its behaviour [37]. It is crucial to take the full range of possible initial conditions into account by modelling different possible scenarios of one case. This requires enormous computational power and sophisticated parallelization techniques [3, 35]. The improvement of computational facilities in recent years enhanced accuracy of the models, which caused Numerical Weather Prediction

(NWP) to become the main area of research in meteorology and it is evolving rapidly.

The spatial and temporal resolution is constantly getting higher and so the NWP models need to be adapted [61, 64]. With each improvement in the resolution more processes can be calculated explicitly, instead of using inaccurate empirically gained representations, and as a result, the contribution of model error is going down [4]. Besides the efforts of adapting models to a higher resolution the second area of focus of research in NWP is ensemble forecasting: to determine and evaluate the uncertainty of the forecast. Therefore, different scenarios, of the same prediction are simulated and the output of the whole ensemble is analysed. The statistics obtained from the ensemble provide insight about what scenario can be expected with the corresponding probability [3]. One of the main contributors to model uncertainty are parametrizations [43, 35]. Parametrizations are needed, because a large number of different complex physical processes leads to a single atmospheric state [14]. Most modern NWP models use a temporal resolution of approximately 60 seconds. Due to limited computational capacities it is impossible to simulate all the underlying physical processes explicitly, for every time step for which the atmospheric state is determined. Hence, many of these simulations are replaced by mainly empirically gained representations, the parametrizations. Paired with the uncertainties of the initial state and the chaotic nature of the atmosphere, the use of statistical methods becomes inevitable [46, 7, 35].

Depending on the purpose of the forecast, different variables of NWP models are of interest: Obviously the thermodynamic state variables, but often also additional ones e.g. accumulated precipitation for estimating the risk of flooding [64], or sunshine duration to plan, when to carry out certain tasks in agriculture.

In the course of this Master's thesis, we are interested in predicting horizontal visibility, where especially for air traffic there is a demand for reliable forecasts. Visibility is a measure for the shortest horizontal distance for which an object is still visible to a human observer, considering all directions [12, 71, 49]. It is effected by a great number of non-linear processes like precipitation and radiative processes. Most parametrizations implemented in operational models are simple functions of the relative humidity, gained by fitting to measurement data, and can be evaluated analytically [20]. The coefficients of the fit functions and even the function types, can vary vastly between different parametrizations, because they cannot be directly related to any physical process. The dominating physical process that impacts visibility is Mie scattering, which is the scattering of plane electromagnetic waves on a sphere, for cases where the radius of the sphere and the wavelength of the incoming light have about the same size range [41]. For sunlight different atmospheric particles satisfy this condition [68]. Hence, the most accurate way of predicting visibility should be to resolve Mie scattering on hydrometeors and aerosols directly. But because the spatial and temporal distribution of them is difficult to predict, there is a high level of overall uncertainty when trying to predict visibility by the use of Mie theory. As mentioned, probabilistic forecasts in general and ensemble prediction in particular can help to get meaningful results, even when the associated uncertainty is high. Motivated by other attempts of probabilistic visibility forecasts [11, 52], we combined ensemble methods and the theory of atmospheric scattering to a new

scheme for probabilistic visibility prediction. This work presents and evaluates a new parametrization of visibility, based on Mie scattering theory, and how ensemble forecasting can be applied to predict visibility. It was carried out in cooperation with ‘Zentralanstalt für Meteorologie und Geodynamik’ [74], the Austrian meteorological service.

The necessary background theory to understand the basics of numerical weather prediction, the framework of this study, is provided in Chapter 2. In addition, the definition and mathematical derivation of visibility is provided and Mie theory is revised.

In Chapter 3 the model we used, is described, with special focus on the pattern generator for stochastic perturbations. The details of the new visibility parametrization and the benchmark parametrizations are explained in Chapter 4. Methods and set-up, including the evaluation strategies are the subject of Chapter 5. Then the results are presented and discussed in Chapter 6. Finally, an outlook on potential future studies and concluding remarks, about strengths and weaknesses of the visibility prediction method is given in Chapter 7.

All used variables and abbreviations are listed in Appendix A in Table A.3 and Table A.4.

# Chapter 2

## Background Theory

### 2.1 Numerical Weather Prediction Models

Numerical Weather Prediction models evolved from the approach to describe the atmosphere in terms of more general physical laws. In this section the basic assumptions and equations for NWP models are outlined according to Coiffier [14].

The core of every NWP model is based upon the ‘Primitive Equations’ which are a form of the Navier-Stokes equations describing:

- Momentum
- Mass transportation
- Energy balance

of a fluid [47].

#### 2.1.1 Primitive Equations

The following equations describe the dynamics of the atmosphere and are the core of every NWP Model.

##### Equation of Momentum

Here the acceleration of a frictionless flow of a Newtonian fluid is described. The equation is derived by taking the time derivative of the equation of momentum of a parcel of air

$$\frac{d\vec{v}}{dt} = -2\vec{\Omega} \times \vec{v} - \frac{1}{\rho} \vec{\nabla} p - \vec{\nabla} \Phi + \vec{F} , \quad (2.1)$$

where  $\vec{v}$  denotes the wind velocity,  $\vec{\Omega}$  the angular velocity of the Earth,  $\rho$  the density of air,  $p$  the pressure,  $\Phi$  the geopotential and  $\vec{F}$  dissipative forcing.

## Equation of State

The ideal gas law is used here. Equation 2.2 is the simplest equation of state of a gas and describes the constant relation of pressure  $p$ , density  $\rho$  and temperature  $T$  in an ideal gas, which is proportional to the gas constant  $R$ .

$$p = \rho RT \quad (2.2)$$

## Conservation of Energy

Equation (2.3) is derived from the First Law of Thermodynamics.

$$\frac{dT}{dt} = \frac{R}{C_p} \frac{T}{p} \frac{dp}{dt} + \frac{1}{C_p} \frac{dQ}{dt} \quad (2.3)$$

The nomenclature is the same as for Equation (2.1) and (2.2). The specific heat of air at a constant pressure  $p$  is denoted by  $C_p$  and the heat by  $Q$ .

## Continuity Equation

In general the continuity equation describes the change in time of the density of a quantity that is preserved, which in this case is mass.

$$\frac{d\rho}{dt} = -\rho \vec{\nabla} \vec{v} \quad (2.4)$$

The nomenclature remains the same as in Equation (2.1).

## Water Vapour Equation

The water vapour equation is usually not seen as part of the primitive equations, but considered in almost all atmospheric models.

$$\frac{dq_{\text{spec}}}{dt} = M \quad (2.5)$$

Here  $q_{\text{spec}}$  denotes the specific humidity and  $M$  the source or sink term for humidity.

### 2.1.2 Approximations

Depending on the model, different approximations need to be made to reduce the system of equations to a more easily solvable problem. The most common one is the so-called traditional meteorological approximation. It consists of:

- The assumption that the gravitational acceleration is constant
- The assumption that the Earth is a perfect sphere
- The replacement of the Earth's angular velocity vector  $\vec{\Omega}$  by its vertical component:  $\vec{\Omega} = \Omega \vec{e}_r \sin \phi = \frac{f}{2} \vec{e}_r$ , where  $\phi$  denotes the latitude

Setting  $\frac{f}{2}\vec{e}_r$  to  $\Omega$  is necessary, to be able to separate the wind vector into a vertical and a horizontal component.  $F$ ,  $\frac{dQ}{dt}$  and  $M$  are source and sink terms that depend on parametrizations and scale of the model. For an adiabatic, frictionless and water vapour conserving atmosphere their contributions become 0.

A very common approximation is to assume hydrostatic balance in a model. Meaning,

$$\frac{\partial p}{\partial z} = \rho g \quad (2.6)$$

needs to be satisfied. Equation (2.6) implies, that there are no vertical winds and is therefore not suitable for models that use a fine mesh size, because for many small scale processes, e.g. convection and turbulence, vertical winds are essential. For that reason, most limited area models (e.g. AROME [62]) are non-hydrostatic. However, for models with large mesh sizes, mostly global models, this simplification can help reducing the computation time and still provides meaningful results.

### 2.1.3 System of Coupled Partial Differential Equations

To get a system of coupled partial differential equations (PDEs) that can be solved numerically, we start by taking Equations (2.1), (2.2), (2.3) and (2.4) replacing the wind velocity  $\vec{v}$  by a horizontal wind vector  $\vec{V}$  and a vertical wind component  $w$  times the radial unit vector  $\vec{e}_r$  as  $\vec{v} = \vec{V} + w\vec{e}_r$ . The source and sink terms are neglected for now. Together with the approximations we introduced in Subsection 2.1.2 and by setting  $\frac{R}{C_p} = \kappa$ , what in meteorology is referred to as ‘Poisson constant’, the system of PDEs takes the form:

$$\frac{d\vec{V}}{dt} = -f\vec{e}_r \times \vec{V} - \frac{RT}{p} \vec{\nabla}_h p \quad (2.7)$$

$$\frac{dw}{dt} = -\frac{RT}{p} \frac{\partial p}{\partial z} - g \quad (2.8)$$

$$\frac{dT}{dt} = \kappa \frac{T}{p} \frac{dp}{dt} \quad (2.9)$$

$$\frac{dp}{dt} = -\frac{p}{1-\kappa} \left( \vec{\nabla}_h \vec{V} + \frac{dw}{dz} \right) \quad (2.10)$$

The sub-scripted h in the differential operators denotes only the horizontal components of the operator. These PDEs need to be solved and integrated for each time step, to get the current values of the variables  $\vec{V}, w, T$  and  $p$ . Then, the so-called ‘parametrized tendencies’ are added to the current value of the variables. The parametrized tendencies relate to source and sink terms and are the sum of the corresponding physics parametrization. Those parametrizations represent physical properties and processes that were neglected for simplification of the system of

equations, when setting up the system of PDEs and are calculated separately at each time step. By the use of this iterative method, the atmospheric system can be evolved in time and predictions for the future can be made.

The best way of solving the system of PDEs varies depending on purpose and scale of the model.

The two most common methods to numerically solve those equations in numerical weather prediction are:

- Spectral Method
- Finite Differences Method

The spectral method is only one approach of a whole set of numerical methods, which are all based on the idea of expanding the solutions in terms of base functions. The finite element and finite volume method are also of the same nature but not discussed here, because they are only of minor importance in NWP modelling [3]. In the spectral method the solutions are assumed to be global wave functions, defined on the whole domain. The main advantage is that the derivatives can be obtained analytically.

The method of finite differences, on the other hand, is a local approach. It uses the differential quotient on a discrete grid to approximate the derivative. Meaning that the numerical procedure only uses local values. Consequently, finite differences solutions lack the capability of reflecting certain global characteristics of a function accurately, but they only require local values at a time for the calculation. Thus, the finite differences method is increasingly popular, because it can make use of very efficient parallelization procedures, where the grid space is divided into sub-domains [3].

Nevertheless, the spectral method is still more common in global models, resulting from the fact that the spherical harmonics can be used as the global wave functions [3, 14]. They are the eigenfunctions of the angular dependent component of the Laplace operator and very suitable ansatz functions on a spherical domain. On other domains it is often not possible to find appropriate global wave functions and the bad approximation results in large trade-offs in numeric accuracy. These drawbacks led to the wide-spread use of the finite differences method in limited area models.

## 2.2 Ensemble Prediction

### 2.2.1 The Atmosphere's Chaotic Nature

Poincaré was the first one to describe a chaotic system, namely the solar system. When he tried to predict the planetary orbits, he started with the simplest case, a system of two large bodies, rotating each other, and discovered that even a very small third body can destabilize the system [48]. Then a long time passed, before Edward Lorenz, who is often referred to as the father of chaos theory, published his discovery, revealing why the difficulties of reliable meteorological forecasts are inherent:

Atmospheric dynamics is highly sensitive to even the smallest changes in the initial and boundary conditions. A small error can propagate and be amplified through non-linear feedback processes to a magnitude much higher than its initial value. If that is the case, completely different scenarios emerge after some time [37]. This behaviour is called chaotic and in context of atmospheric motions broadly known as ‘Butterfly Effect’. The name originates in one of Lorenz’ most famous talks, where he illustrated his complex theory with the question ‘Does the flap of a butterfly’s wings in Brazil set off a tornado in Texas?’, discussing the immense effects of extremely small changes in the initial conditions of an atmospheric state on the occurrence of extreme weather events and general predictability of future atmospheric states [38]. A more thorough mathematical definition of chaos is provided by the Lyapunov-exponents, which are briefly described in the following after Kalnay [25].

In general, all possible states of a dynamical system, that can be described by  $n$  independent variables  $x_1, \dots, x_n$  can be represented by the system’s  $n$ -dimensional phase space  $\Gamma$ . A volume in phase space is defined by a set of states  $\vec{x} = (x_1, \dots, x_n) \in \Gamma$ . To illustrate this, we use the example of the states contained in a hypersphere  $\Gamma_S$  with a set radius  $r$

$$\Gamma_S = \{\vec{x} : \|\vec{x}\| \leq r\} . \quad (2.11)$$

Then a volume  $V$  can be assigned to the set by using the indicator function  $f$  of the set  $\Gamma_S$  and the Liouville measure  $d\vec{x} = dx_1 dx_2 \dots dx_n$

$$V = \int_{\Gamma} f d\vec{x} . \quad (2.12)$$

This phase space volume  $V$  can be used to determine, whether a system can be described by Hamiltonian mechanics, or if it is necessary to move to the more complex framework of dynamical system theory, including chaos theory. The criterion for a system to be Hamiltonian, is that any initial phase space volume is conserved [28]. Let us now consider a point in phase space and apply an initial spherical infinitesimal perturbation to the point as illustrated in Figure 2.1. Then, the evolution of the sphere in time can be used to categorize the system. The deformation of the sphere is determined by its axes  $a_1, \dots, a_n$ . The axes describe the separation rates of the trajectories, that started infinitesimally close. A parameter  $\vec{\lambda} = (\lambda_1, \dots, \lambda_m)$  is introduced so that any axis becomes the initial length scaled by  $\lambda_j$  after a time  $t$

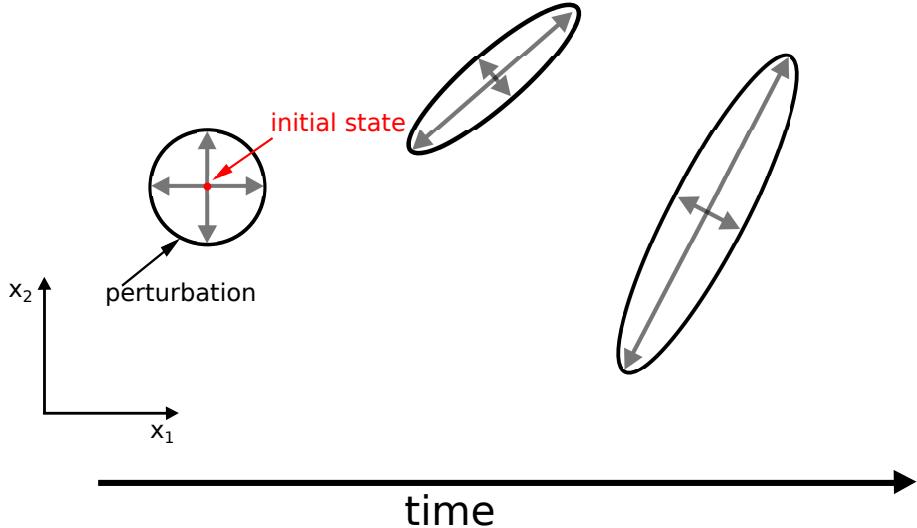
$$a_i(t) = a_i(0) e^{\lambda_j t} , \quad (2.13)$$

resulting in a deformation of the volume of the sphere

$$V(t) = V(0) e^{-\sum_{j=1}^m \lambda_j t} . \quad (2.14)$$

The coefficients  $\lambda_j$  are called Lyapunov-exponents and their indices denote their relative magnitude:

$$\forall \lambda_j : \lambda_j \geq \lambda_{j+1} . \quad (2.15)$$



**Figure 2.1:** Example of a deformation of a volume in a two-dimensional phase space: the red dot marks the deterministic initial state and the sphere the initial perturbation. The phase space volume of the initially spherical perturbation is deformed with respect to the stability properties of the dynamical system [25].

Then the condition for a Hamiltonian system of the conservation of the initial volume of the sphere in phase space can be expressed as:

$$\sum_{j=1}^m \lambda_j = 0 \quad (2.16)$$

the condition for a stable system:

$$\forall \lambda_j : \lambda_j \leq 0 \quad (2.17)$$

and the condition for a chaotic system:

$$\exists \lambda_j : \lambda_j > 0 \quad . \quad (2.18)$$

As those conditions imply, the volume contained in the sphere is conserved or can even converge to a point for a stable system. For a chaotic system at least  $\lambda_1$  has to be greater than zero, so that trajectories in phase space diverge. In the textbook by Kalnay [25] the meaning of the Lyapunov-exponents is summarized as the following: '[...]the Lyapunov-exponents describe the long-term average exponential rate of stretching or contraction in the attractor.'. The term attractor is used for the set of points that the trajectories pass multiple times.

For a stable system, reliable forecasts can be done at all times, where on the other hand, chaotic systems become unpredictable after a finite time. The Lyapunov-exponents and hence, the predictability of a system, can be calculated from theory or from experimental data [18, 66, 35].

## 2.2.2 Ensembles

Since the beginning of the 1990s ensemble prediction has become a widely-known approach to treat uncertainty in NWP and it is still considered state of the art

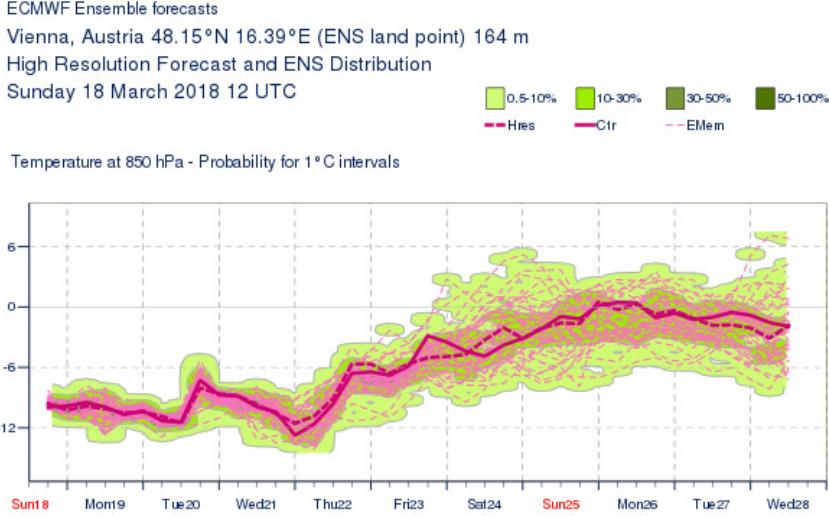
and subject of current research with increasing popularity [70, 4, 35, 55, 45]. To estimate the effect of different values for the initial and boundary condition and model parametrizations a so-called ensemble is created. Each ensemble member represents a different simulation. It is a single trajectory in the phase space of the atmospheric system. For traditional ensembles independent runs for all members are performed and the statistical properties of the whole ensemble data set are evaluated [35].

There are three main types of ensembles:

1. Multi-model ensemble: Each member uses a different NWP model [17]
2. Multi-physics ensemble: Each member uses different parametrizations or different settings of the model physics [23]
3. Traditional ensemble: Each member is obtained by perturbing one or more quantities in initial or boundary conditions, or in the model itself. If random number sampling is used, it is called Monte-Carlo forecast. [15, 70, 34]

Figure 2.2 shows an example of a traditional ensemble forecast and its components: ensemble forecasts contain several regular members and a control member, which is the deterministic scenario (the scenario considered most likely), with the same resolution as the other ensemble members. In most cases, the deterministic run is performed with a higher resolution, to get the best possible forecast. Unfortunately it would consume too much computational resources to run the whole ensemble on the highest possible resolution. Thus, a control member is computed additionally, to have a point of reference for the other ensemble members without the bias of different resolutions.

Depending on the type and source of uncertainty we want to investigate, we need to create an appropriate ensemble.



**Figure 2.2:** ECMWF ensemble: A temperature ensemble forecast of the ECMWF showing 50 regular members ('EMem'), one control member ('Ctr') and one high resolution run ('Hres'). The ensemble spread increases with time, revealing the chaotic dynamics of the atmosphere. The number of ensemble members in a certain temperature interval can be related to the probability of a certain temperature. The image was created by the ECMWF (cannot be accessed by the public).

In this study we focused on ensemble creation by stochastically perturbed parameters, a Monte-Carlo method, which is explained in greater detail in Subsection 2.2.4.

### 2.2.3 Uncertainty and Error

The sources of error can be categorized into three groups: initial conditions, model error and, for limited area models (LAM), boundary conditions. The initial conditions of the model equations have a very high level of uncertainty. Many of them are set by a best-guess gained from old model output combined with new measurements. But the number of measurement stations is usually small in comparison to the number of grid points in a model and also the measurement equipment has a limited accuracy. It is worth pointing out, that the problem is not only about the difference in number of grid points and measurement stations but also that the locations of grid points and measurement stations do not overlap.

The boundary conditions for limited area models are even more sensitive, because they are set to the output values obtained from a global model. Thus, errors can evolve and be amplified in the global model and are then passed on to the LAM, especially affecting the forecast quality on the edges of the domain.

A consequence of the limited resolution of the model is that many processes, that take place on a sub-grid scale, have to be parametrized. Since every parametrization is an approximation, all parametrized physical processes contribute to the uncertainty in a model [35]. Even if a physical process is resolved explicitly, it contributes to model error, due to the model being discrete. In nature, on the other hand, most of the concerned processes evolve continuously on the model's scales. Another source of uncertainty is that for some quantities in the model, like aerosol concentration, only climatological data is used. The reason being, that prognostic

schemes require too much computational resources and/or there is a lack of up-to-date data.

A problem specific to NWP models is the orographic mismatch. Especially in mountainous regions not all topographic structures can be resolved. This can cause a large discrepancy between reality and model.

#### 2.2.4 Ensemble Creation by Monte-Carlo Methods

In a stochastic ensemble, each member represents a different possible value of a certain quantity. How the values differ for the members is determined by a random sampling regime. Ensemble members are created by perturbing a deterministic value, that we will call the unperturbed value from now on. But it is important to acknowledge that the unperturbed value shall never be seen as the ‘real’ value of a variable, but the most likely value. A value being categorized as most likely can be based on climatological data, knowledge of the underlying physics or current measurement data. Because of limited accuracy of measurement equipment and the spatial resolution, it is inherently impossible to determine a ‘real’ value of a variable in an atmospheric system. Only a certain range for the values can be determined. We distinguish the different perturbation schemes by the perturbed quantity:

1. Perturbing the initial value of a variable
2. Perturbing the tendency of a variable (SPPT)
3. Perturbing constant parameters in a parametrization (SPP)

The SPPT and the SPP perturbation schemes are used to take model error into account. Perturbation of the initial values, on the other hand, incorporates the errors contributed by the set-up of initial and/or boundary conditions.

The different schemes are briefly outlined in the following.

##### Perturbation of initial conditions

The simple random perturbation of initial conditions is the most basic perturbation scheme. Let  $e_j$  be the time integration of the model equations (‘prognostic variable’)

$$\frac{\partial e_j}{\partial t} = A(e_j, t) + P(e_j, c_0, \dots, c_N) , \quad (2.19)$$

where  $A$  are the non-parametrized and  $P$  are the parametrized contributions, which depend on  $e_j$ , the time  $t$  and  $N$  constant parameters  $c_0, \dots, c_N$ .

Now only the initial conditions are perturbed

$$e_j(t=0) \equiv e_{j,0} + \delta e_j(t=0) , \quad (2.20)$$

where  $e_{j,0}$  is the deterministic value at  $t = 0$  and  $\delta e_j$  is the perturbation. From investigation of Equation (2.19) and (2.20) it is obvious that the difference in the evolution of different ensemble members has its origin solely in the perturbation of initial conditions. The simplest way to set-up the perturbation is to only perform random sampling. But today more sophisticated techniques, like the use of Lyapunov-exponents, are more common than a plain Monte-Carlo approach [44, 7].

## Stochastically Perturbed Parametrized Tendencies (SPPT)

In the model equations the tendencies of all model variables are calculated at each time step and used for the integration. The tendency of a variable is defined as its change in one time step. It is the sum of the change of the variable calculated from the Primitive Equations  $A(e_j, t)$  and the change contributed by parametrized processes  $P(e_j, c_0, \dots, c_N)$ . Thus,  $P(e_j, c_0, \dots, c_N)$  is called parametrized tendency hereafter. We start as before, but now the parametrized tendency  $P$  is perturbed and denoted by  $P'$

$$\frac{\partial e_j}{\partial t} = A(e_j, t) + P'(e_j, c_0, \dots, c_N) . \quad (2.21)$$

For each grid point the perturbed parametrized tendency is defined as

$$P' \equiv r_j(\lambda_{\text{lon}}, \phi, t)P , \quad (2.22)$$

where  $r_j$  is a random number. It is taken from a spatially and temporally correlated pattern consisting of randomly sampled numbers, chosen from a suitable distribution [7]. Regarding Equation (2.21), it is noteworthy that only errors contributed by parametrizations are considered.

## Stochastically Perturbed Parameters (SPP)

We start as before, but now the constant parameters  $c_0, \dots, c_N$  in the parametrized tendency  $P$  are perturbed. Therefore the resulting parametrized tendencies are denoted by  $P''$

$$\frac{\partial e_j}{\partial t} = A(e_j, t) + P''(e_j, c_0r_0, \dots, c_Nr_N) . \quad (2.23)$$

For each grid point the perturbed parametrization is defined as

$$P'' \equiv P''(e_j, c_0r_0(\lambda_{\text{lon}}, \phi, t), \dots, c_Nr_N(\lambda_{\text{lon}}, \phi, t)) , \quad (2.24)$$

where again,  $r_j$  is a random number, taken from a spatially and temporally correlated pattern consisting of randomly sampled numbers. In this regime the uncertainties can be directly related to single physical parameters [43].

The choice of scheme depends on what we want to examine and how the sources of error should be weighted.

For example, if we want to investigate the effect of the variation of initial conditions due to limited accuracy in measurements and hence, in data assimilation, then the initial value of the variable needs to be perturbed. To examine the effects of uncertainty due to model error, introduced by spatial and time discretization, the tendencies of the variables should be perturbed as well. There are many parametrized processes contributing to the tendencies, so perturbing the tendencies can be an effective way to increase model spread based on acknowledging the effect of uncertainties in different parametrizations.

If we obtained a parametrization e.g. by fitting data, then, to analyse the impact of the uncertainty of the coefficients, the constant parameters can be perturbed.

To decide which variables, parameters or tendencies should be perturbed, one has to identify those, which have the highest level of uncertainty and which are most influential to the model output. Moreover, one must consider what is computationally affordable and stable.

## 2.3 Visibility

Since this study focuses on forecasting visibility, the most important definitions, mathematical derivations and physical processes related to it, are presented in this Section.

### 2.3.1 Definitions

The World Meteorological Organization (WMO) sets all standards and definitions for meteorological measurements, forecasts and procedures, including visibility observation. It provides three important definitions in the context of visibility [71]:

- The **meteorological optical range** is the length of path in the atmosphere required to reduce the luminous flux in a collimated beam from an incandescent lamp, at a colour temperature of 2700 K, to 5% of its original value, the luminous flux being evaluated by means of the photometric luminosity function of the International Commission on Illumination (CIE).
- **Visibility, meteorological visibility (by day) and meteorological visibility at night** are defined as the greatest distance at which a black object of suitable dimensions (located on the ground) can be seen and recognized when observed against the horizon sky during daylight or could be seen and recognized during the night if the general illumination were raised to the normal daylight level.
- **Visual range (meteorological)**: Distance at which the contrast of a given object with respect to its background is just equal to the contrast threshold of an observer.

Regarding their definitions, the meteorological visibility and the visual range are closely related. The meteorological visibility is mostly relevant for human observers, denoting always the lowest value considering all horizontal directions. Although the WMO sets the meteorological optical range as the measurement standard, the meteorological visibility is the recommended approximation for human observers, since it can be understood intuitively. Because most visibility measurements are still performed by human observers, meteorological visibility is the measure for visibility in praxis.

Koschmieder was the first one to derive a mathematical expression for visibility, using the visual range [29].

Also in most models the visual range is used, because it provides a more thorough

definition than the meteorological visibility and is therefore easier to be implemented, but still relates to the observed visibility. The mathematical derivation is briefly outlined in the following section according to Price [49], who refers to the works of Koschmieder [29].

### 2.3.2 Mathematical Derivation of Visibility and Visual Range

Let the contrast  $C$  be defined as

$$C = \frac{B_b - B_a}{B_b} , \quad (2.25)$$

where  $B_b$  denotes the brightness of the background and  $B_a$  the brightness of the object, and let the change of the brightness on an infinitesimal line segment along a horizontal line  $x$  be

$$\frac{dB}{dx} = -\beta_{\text{tot}} B + B_s , \quad (2.26)$$

where  $\beta_{\text{tot}}$  denotes the total atmospheric extinction coefficient and  $B_s$  denotes the light that is being scattered in the line. Equation (2.26) can also be applied to  $B_b$ . Since the brightness of the background is independent of the distance, the following condition is satisfied:

$$\frac{dB_b}{dx} = 0 . \quad (2.27)$$

Then, the change of contrast along an infinitesimal line segment can be written as

$$\frac{dC}{dx} = -\frac{1}{B_b} \frac{B_a}{dx} . \quad (2.28)$$

Evaluating Equation (2.26) for the brightness of the background we gain the expression

$$\beta_{\text{tot}} B_b = B_s . \quad (2.29)$$

When plugging (2.29) into Equation (2.26) and evaluating Equation (2.28) we get

$$\frac{dC}{dx} = -\beta_{\text{tot}} C . \quad (2.30)$$

As Equation (2.30) implies, we can use an exponential ansatz and get the necessary boundary conditions by investigation of the case of a perfectly black object in front of a white background. In the considered scenario, the contrast of the object at the point of the observer has to be 100%, because all light is absorbed. Also, the contrast should go to 0 for the object being infinitely far away. Hence, the following equations need to be satisfied:

$$C(x = 0) = 1 \quad (2.31)$$

$$\lim_{x \rightarrow \infty} C(x) = 0 \quad (2.32)$$

With these boundary conditions the coefficients of the contrast are derived as an exponential function of the distance between the observer and the object

$$C(x) = \exp(-\beta_{\text{tot}} x) . \quad (2.33)$$

From this we can derive the maximum distance for which the contrast is still above the contrast threshold value. This distance is called the ‘visual range’. Since the atmospheric extinction can vary for different directions, the visual range can vary too. The minimum of the visual range, considering all directions, is approximately the visibility  $vis$ . For reasons of computational simplicity, we only use the atmospheric extinction of the point where the visibility is forecast. Hence, the visual range and the visibility are equivalent in this study and defined by

$$vis = -\frac{\ln(\varepsilon)}{\beta_{tot}} . \quad (2.34)$$

The contrast threshold  $\varepsilon$  should be chosen under the consideration of the purpose and usage of the visibility data. Most authors suggest to set the contrast threshold to 2% (e.g. Ballard, Wright and Golding [2]) while other consider 5% to be more accurate (e.g. Claxton [13]). The differing values are due to the fact that the contrast is chosen with respect to what our eyes are gauged to. But as we experience in our everyday life, this can vary for different people, shapes and colours [29]. As this implies, the constant setting of  $\varepsilon$  is an approximation for the sake of simplicity.

### 2.3.3 Atmospheric Scattering

Mie and Rayleigh scattering are special cases of scattering of electromagnetic waves on a homogeneous sphere [24]. Both scattering regimes together are sufficient to describe the most important scattering processes in the atmosphere. Mie scattering, for example, is responsible for visibility reduction due to fog and for clouds being perceived as white. To observe the effects of Rayleigh scattering, one must simply look at the blue colour of the clear sky [68]. Generally, if we aim to describe a radiative process with one of the two theories, we need to verify first, if it is suitable for the process’ size range. To be able to do so, a size parameter  $\alpha$  is introduced. It depends on the radius of the sphere  $r$  and the wavelength of the incoming light  $\lambda$  and is defined as

$$\alpha = \frac{2\pi r}{\lambda} . \quad (2.35)$$

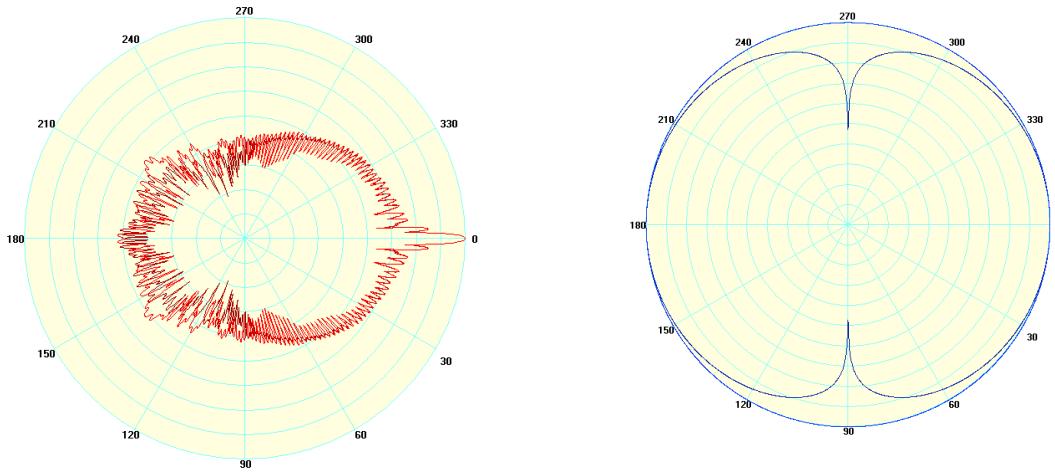
For Rayleigh scattering the condition

$$\alpha \ll 1 \quad (2.36)$$

must hold. In other words, the radius of the object must be much smaller than the wavelength of the incoming radiation. In the Earth’s atmosphere Rayleigh scattering is mostly contributed by scattering on air molecules and much less variable in time and space than the contribution by Mie scattering.

Another difference between the two regimes is that Rayleigh scattering strongly favours shorter wavelengths and therefore in the visible spectrum the colour blue. Mie scattering on the other hand, has a very weak dependence on the wavelength [50]. As a result we see most Mie-scattered sunlight as white, because it contains all wavelengths of the sun’s spectrum. The Mie scattering regime can be applied for all cases where

$$0.1 < \alpha < 50 \quad (2.37)$$



(a) Scattering of visible red light ( $\lambda=650\text{nm}$ ) on a water drop.

Used parameters: particle radius:  $1.0\text{e-}05\text{m}$ ; refractive index:  $1.33257 + \text{i}1.67\text{e-}08$ ; scale: logarithmic, value at outer circle:  $2.18\text{e+}07$ , value at inner circle:  $2.18\text{e-}02$ .

(b) Scattering of visible blue light ( $\lambda=400\text{nm}$ ) on an oxygen molecule.

Used parameters: particle radius:  $1.52\text{e-}10\text{m}$ ; refractive index:  $1.000152$ ; scale: logarithmic, value at outer circle:  $2.5\text{e-}24$ , value at inner circle:  $2.18\text{e-}34$ .

**Figure 2.3:** Scatter plot of Mie and Rayleigh scattering: Intensity versus scattering angle.  
Created by the use of *MiePlot* (*A computer program for scattering of light from a sphere using Mie theory and the Debye series*)

holds.

For visible light, which has a wavelength from approximately 390nm to 700nm, the condition is satisfied for scattering on vapour, hydrometeors and aerosols [68, 9]. Due to the dominating physical process being Mie scattering [49], there is a strong dependency of the total atmospheric extinction coefficient on the specific humidity, precipitation, cloud water in liquid and ice form and aerosols. The relations can be derived by characterizing the hydrometeors and aerosols in the air as a suspension of perfect spheres with different refractive indices and radii in a gas [32]. Figure 2.3a illustrates an example of atmospheric scattering: visible red light on a water drop. The plot shows a peak for forward scattering, typical for the angular intensity distribution of Mie scattering.

The scattering characteristics of a material are described by their extinction, scattering and absorption coefficients. In the course of predicting visibility, we are only interested in the extinction coefficient to be able to quantify the total reduction the light's intensity [22]. The extinction coefficient  $Q_{ext}$  is the dimensionless equivalent of the extinction cross section  $C_{ext}$  and defined as

$$Q_{ext} = \frac{C_{ext}}{G} \quad , \quad (2.38)$$

where  $G$  is the geometric factor that is set to  $\pi r^2$  for spherical particles, because of their circular cross section.

The extinction cross section is a measure for the total difference in radiant flux through a defined surface  $S$ . The two compared cases are, when a particle is present

at a location and when it is not. The extinction cross section is defined as

$$C_{ext} = \frac{1}{E^i} \int_{4\pi R^2} (E^i - E) dS , \quad (2.39)$$

where  $E^i$  denotes the electric field vector when there is no particle at the location, and  $E$  when a particle is present. The boundary surface  $S$  is set to the surface of a sphere because we are interested in the case of scattering on spherically approximated particles. For aerosol scattering the electric field vectors can be obtained by the use of Mie scattering theory. The theory was first developed by Gustav Mie [41] and is derived by applying boundary and continuity conditions for light approaching a sphere to Maxwell's equations.

The resulting wave functions are in form of power series composed of Bessel functions and Legendre polynomials [26]. When inserting these in Equation (2.39) we derive the following mathematical expression for the extinction cross section:

$$C_{ext} = -\frac{2\pi}{k_0^2} \Re \left( \sum_{n=1}^{\infty} (2n+1)(a_n^s + b_n^s) \right) , \quad (2.40)$$

where  $k_0$  is the wave number in vacuum and  $a_n^s$  and  $b_n^s$  are the Mie coefficients. Those are derived as part of the solution when solving the scalar wave equation. The Mie coefficients are composed of Riccati-Bessel functions and include a dependency on the refractive index. The entire detailed derivation is presented in several textbooks e.g. [73], [24]. Regarding Equation (2.40), we see that only an approximation of  $C_{ext}$  can be calculated numerically and the accuracy depends mainly on the number of terms computed. Many different algorithms for the computations were developed and Wriedt [72] provides a good overview with special focus on the application on aerosols.

As mentioned, the second most relevant scattering process in the atmosphere is Rayleigh scattering. Regarding the criterion presented in Equation (2.36), it is obvious that Rayleigh scattering takes place, if the size of the particle is much smaller than the incoming wavelength. It is a result of the polarizability of the gas molecules: the incoming electromagnetic wave excites the molecules, so that each molecule acts as Hertzian dipole. Figure 2.3b illustrates an example of visible blue light being scattered on an oxygen molecule and shows the characteristic angular intensity distribution of a Hertzian dipole. It radiates with the same wavelength as the incoming light. This is also true for Mie scattering. Such processes, without a shift between the wavelength of the scattered light and the incident light, are called elastic, because the change in energy of the photons is negligibly small.

# Chapter 3

## Model

The model we used for this study, is called ‘Application of Research to Operations at Mesoscale’ (AROME) and was developed by the ALADIN consortium [40]. At the ‘Zentralanstalt für Meteorologie und Geophysik’ (ZAMG), the model has been operational since 2014, which means that it is not only used for research, but also for forecasts on a daily basis. If not mentioned otherwise, all settings remained the same as for the operational configuration of the model at ZAMG.

As the name implies, AROME is a limited area model, written in Fortran. The version operationally used at ZAMG (C40T1) has a horizontal resolution of 2.5km on a 432-by-600 grid and 90 vertical layers. The highest layer is at a pressure level of 10hPa, which is the equivalent of approximately 25km. One time step is set to 60 seconds and the geographical model domain of central Europe is shown in several graphics in the following (e.g. Figure 5.2 , 3.4 ). It is a convection permitting and non-hydrostatic model [62]. Convection permitting means that ascending and descending convective currents are resolved, opposed to the strategy of simply parametrizing their effects, commonly done in global models with a lower resolution.

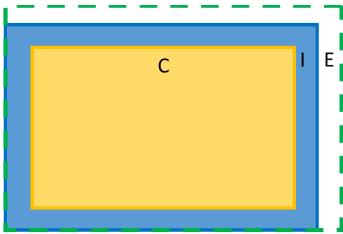
Since AROME is a LAM, it has to be nested in a model with a larger domain, to set its boundary values. At ZAMG the used model is the global European ‘Integrated Forecast System’ (IFS), version C41R2, developed by the European Centre for Medium-Range Weather Forecasts (ECMWF) [19]. The IFS model is operationally run with a 51 member ensemble at the ECMWF. ZAMG uses a subset of 16 members and the control member of the ECMWF-ensemble.

AROME was forked from the original ALADIN (Aire Limitée Adaptation Dynamique Développement International) model, which was developed by Météo-France. Most parts of the code are contained in the common code package IAAA (IFS, ALADIN, AROME, ARPEGE) where AROME is a canonical configuration of the package [62] and uses the so-called ‘MESO-NH’ physics package [31].

For its dynamical core AROME uses a semi-implicit semi-Lagrangian scheme [62, 54]. The Lagrangian and the Eulerian schemes are different methods for treating flow fields in continuum mechanics. A Lagrangian scheme describes a system by the movements of the parcels of the fluid. The counter-model of that, the Eulerian scheme, uses the change rates at specific locations to capture the dynamics of the field. In a semi-Lagrangian scheme an Eulerian grid is used and at each time step the trajectories that brought the parcels of the fluid to their current locations are cal-

culated [16]. The main reason why semi-Lagrangian schemes very common in NWP models is, because when combined with semi-implicit integration, they remain stable even for long time steps [67].

### 3.1 Spatial Discretization



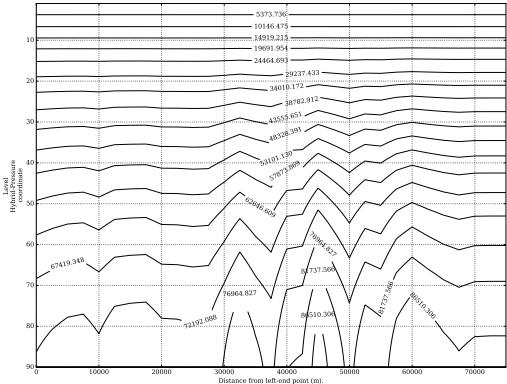
**Figure 3.1:** Different Zones in AROME model domain: C - Central Zone, I - Intermediate Zone, E - Extension Zone [62]

The AROME Model mainly uses spectral representations of the fields, with the exception of the grid point representation of humidity. Each time, fields of different types need to interact, a Fast Fourier Transformation (FFT) is performed.

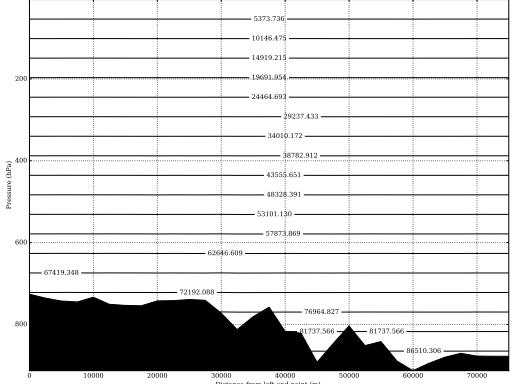
The spectral horizontal discretization is a bi-Fourier spectral representation with elliptical truncation on a conical Lambert-projection, as illustrated in Figure 3.3. The vertical discretization uses finite differences and finite elements [62, 5].

Due to the bi-Fourier spectral representation, different zones (as shown in Figure 3.1) had to be introduced in the model. The outermost zone E is called extension zone and is the actual computational domain of the spectral method. Because of the periodic nature of the Fourier Series the computational domain describes a torus. Consequently, it can never be congruent with the geographical LAM domain. To resolve this issue, the corresponding geographical domain is a cut-out of the extension zone to avoid physically nonsensical periodic values at the edges of the domain. In the Intermediate Zone I the boundary conditions from the global model get sourced in. Only in the Central Zone C can the physics and dynamics evolve completely freely. This zone can be seen as the equivalent of the geographical domain [62].

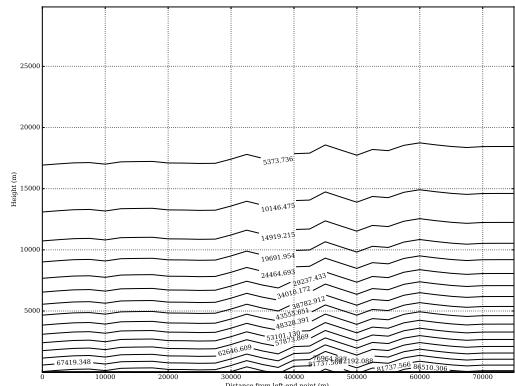
The vertical coordinate is a hybrid pressure terrain-following coordinate. It is preferred over a simple pressure coordinate, because as illustrated in Figure 3.2, geographical barriers like mountains are considered.



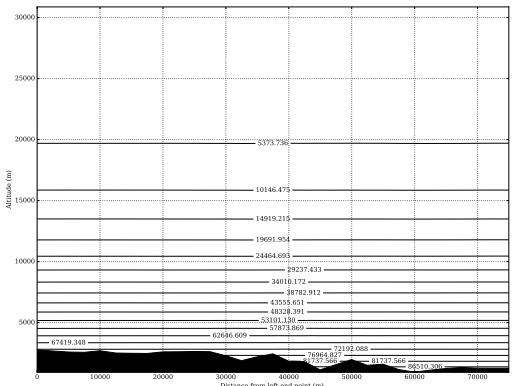
(a) Hybrid Pressure-Terrain-following coordinates: layer number on Y-axis



(b) Pressure coordinates: pressure [hPa] on Y-axis



(c) Terrain-following coordinates: height above ground [m] on Y-axis



(d) Traditional Height Coordinates: height above sea-level [m] on Y-axis

**Figure 3.2:** Isobars in different vertical coordinate systems: 18 isobars (in [Pa]) of a cross section are shown in four different coordinate systems. The Y-axis is the vertical coordinate, different in each plot, and the X-axis is the horizontal distance from the starting point of the cross section. The cross sections starts at [lon:10.13, lat:46.85] (left end) and ends at [lon:10.88, lat:47.29], showing a part of the Alps in western Austria to illustrate how topography is represented in different coordinate systems. Topological structures that cannot be assigned a value assigned to are indicated as black surface.

## 3.2 Pattern Generator

To perform Monte-Carlo forecasts by creating a stochastic physics ensemble a random number generator is needed. In most cases spatially and/or temporal correlations are necessary to represent realistic deviations from the deterministic scenario. To do so, a pattern generator that provides suitable randomly sampled spatial and temporal patterns, which can be applied to the relevant fields in the model, is the tool of choice. In the AROME model the stochastic pattern generator generates a two-dimensional pattern, where the points are sampled from a normal distribution. Since the ranges of the quantities we want to perturb are restricted, the ranges of the perturbations need to be restricted as well. Therefore a so-called clip ratio  $x_{\text{clip}}$  is defined in the namelist. Then, as described in the following by Equation (3.1), the distribution is cut off at the value of  $x_{\text{clip}}$  times the standard deviation  $\sigma_{2D}$ . In this case ‘cut off’ means, that if we define  $a = x_{\text{clip}}\sigma_{2D}$ , all sampled values outside the interval  $[-a, a]$  are set to  $-a$  or respectively  $a$ . This leads to the undesired result that the probability density function  $f(p)$  does not have its minima, as desired at the edge of the mathematical domain. In fact, if the value for  $\sigma_{2D}$  is too small, it can happen that the  $f(p)$  has its maxima at the edges of the interval.

Another problem is that the pattern generator can only produce normal distributions. If the variable is suspected to be distributed differently, the only way to construct other distributions of randomly sampled numbers, is by use of the normally distributed pattern (e.g. taking the exponential of the numbers for lognormal distribution). Thus, one is severely limited in the selection of distribution.

### 3.2.1 Technical Set-Up of the Pattern Generator

The pattern is constructed in spectral space and the corresponding coefficients  $\hat{a}_{n,m}$ , presented in Equation (3.1), are the coefficients of a double Fourier series. The pattern generator was originally implemented and described by Palmer et al. [46] for the global ECMWF IFS model, where  $\hat{a}_{n,m}$  relates to the spherical harmonics. The different coefficients are a consequence of the different representation of spectral fields in the ECMWF IFS and the AROME model. When the pattern generator is used on a LAM model the statistical properties of the fields are conserved, but the LAM fields are not congruent with the LAM domain of the global fields. The spectral coefficients are obtained by random sampling and evolved by an auto-regressive process as described in Equation (3.1).

$$\hat{a}_{n,m}(t + \Delta t) = \Phi_t(t)\hat{a}_{n,m} + \sigma_{2D}(n)\epsilon_{m,n}\Theta(x_{\text{clip}} - \epsilon_{m,n})\Theta(\epsilon_{m,n} - x_{\text{clip}}) \quad (3.1)$$

$$\epsilon_{m,n} \in \mathbb{C} : \Re(\epsilon_{m,n}) \sim N_1(\mu = 0, \sigma = 1); \Im(\epsilon_{m,n}) \sim N_2(\mu = 0, \sigma = 1)$$

Here  $\Theta$  denotes the Heaviside step function which is defined as:

$$\begin{aligned} \Theta : \mathbb{R} &\rightarrow \{0, 1\} \\ x < 0 : \Theta(x) &= 0 \\ x \geq 0 : \Theta(x) &= 1 \end{aligned}$$

The random number generator samples the real and the imaginary part of  $\epsilon_{m,n}$  independently from a Gaussian distribution with variance  $\sigma = 1$  and mean  $\mu = 0$ .  $\epsilon_{m,n}$  is sampled independently for each  $m, n$  and  $t$ , where  $n$  is the total wave number and  $m$  is the zonal wave number, as defined in Equation (3.2). The zonal wave number is the wave number for periodic boundary conditions at a given latitude  $\phi$

$$m\lambda = 2\pi R_E \cos(\phi) , \quad (3.2)$$

where  $R_E$  is the Earth's radius,  $\phi$  the latitude and  $\lambda$  the wavelength. Then the sample gets cut off below and above the clip ratio  $x_{\text{clip}}$  and scaled by  $\sigma_{2D}(n)$ .

Each pattern  $\Psi$  is characterized by its spatial correlation  $L_\Psi$  and its temporal correlation  $\tau_\Psi$ , because those two parameters determine  $\Phi_t(t)$  and  $\sigma_{2D}(n)$  which define the spectral coefficients of the pattern, as demonstrated by Equation (3.1). The dependencies of  $\Phi_t(t)$  and  $\sigma_{2D}(n)$  on the correlation scales are non-linear and defined as the following:

$$\Phi_t = \exp\left(\frac{-\Delta t}{\tau_\Psi}\right) \quad (3.3)$$

$$\sigma_{2D}(n) = F_0 \chi(n) \quad (3.4)$$

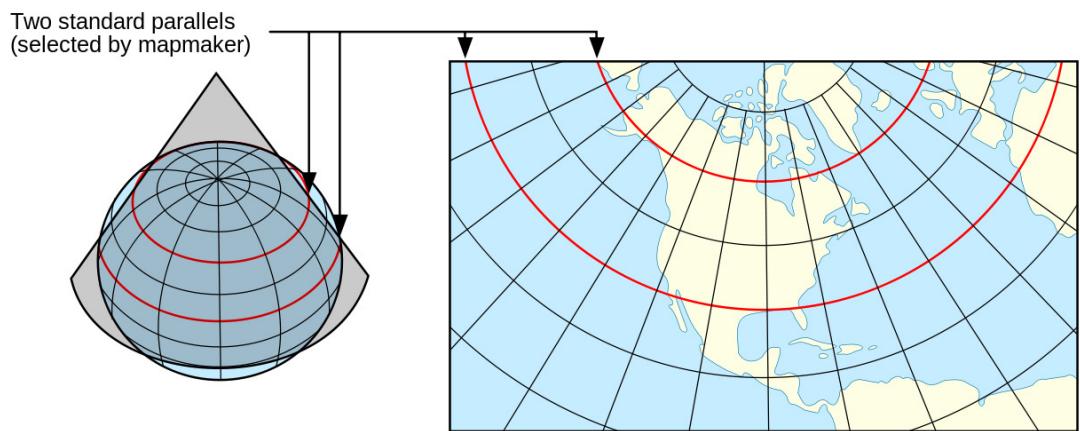
$$F_0 = \frac{\sqrt{1 - \Phi_t^2} \sigma_{\text{nl}}}{\gamma} \quad (3.5)$$

$$\chi = \exp(-0.5 \zeta n (n + 1)) \quad (3.6)$$

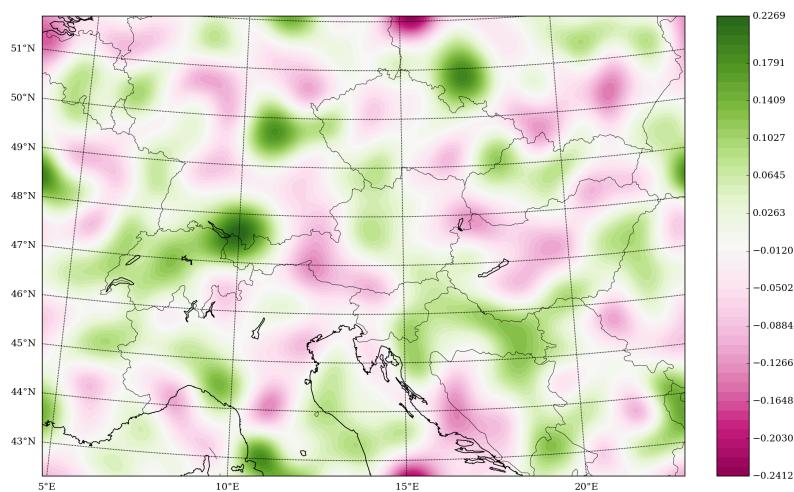
$$\zeta = 0.5 \left(\frac{L_\Psi}{R_E}\right)^2 \quad (3.7)$$

$$\gamma = \sum_n \sqrt{2n + 1} \chi(n) \quad (3.8)$$

The scaling parameters  $\Phi_t(t)$ ,  $L_\Psi$ ,  $x_{\text{clip}}$  and  $\sigma_{\text{nl}}$  (relates to the variance of the distribution) are all set in the model configuration, called namelist hereafter.



**Figure 3.3:** Illustration of a conic conformal Lambert projection. Image by U.S. Geological Survey [65]



**Figure 3.4:** Example of a normally distributed small-scale perturbation generated by the pattern generator

# Chapter 4

## Parametrization of Visibility

For the newly introduced parametrization, two different existing approaches are merged: One for Mie scattering by hydrated aerosols introduced by Clark et al. [12] and the other one by Stoelinga and Warner [58], where the parametrization of the extinction coefficient is directly dependent on the mass concentration of snow, rain, cloud water and cloud ice.

This ‘merged’ parametrization is called ‘Clark+’ hereafter and is defined as

$$vis = -\ln(\varepsilon) \left( \beta_R + \sum_{i=1}^{N_h} \beta_{h,i} + \sum_{j=1}^{N_a} \beta_{a,j} \right)^{-1}, \quad (4.1)$$

where  $N_h$  denotes the number of different species of hydrometeors. Usually hydrometeor means all forms of precipitation, but in this case we will not only call snow and rain hydrometeors, but liquid cloud water and cloud ice too, to be consistent with nomenclature in the models. The number of different aerosol species is denoted by  $N_a$ . In general, an aerosol is defined as air with suspended particles, but in meteorological terminology the term aerosols only refers to the particles in the suspension, excluding hydrometeors [50]. The extinction coefficients of the different hydrometeors are taken from the parametrization by Stoelinga and Warner [58], who refer to Rutledge and Hobbs [53] and Kunkel [30], who define them as

$$\beta_{h,i} = a_{0,i} (c_i \rho \times 10^3)^{a_{1,i}} , \quad (4.2)$$

where  $c_i$  stands for the mass concentration of the hydrometeors and  $\rho$  for the density of air. In general the coefficients  $a_{0,i}$  and  $a_{1,i}$ , shown in Table 4.1, are related to the surface-volume-ratio of a hydrometeor. That is because the larger the surface-to-volume ratio of a hydrometeor is, the more likely is the light to be scattered for the same mass concentration ( e.g. a snowflake has a higher extinction efficiency than a water drop).

The extinction coefficients of the different aerosols  $\beta_{a,j}$  depend mainly on the mass attenuation coefficients  $\alpha_j$  of the aerosols:

$$\beta_{a,j} = \alpha_j(q_{\text{rel}}(q_{\text{spec}}, T, p)) c_j \rho_{\text{dry}} . \quad (4.3)$$

As  $\alpha_j(q_{\text{rel}}(q_{\text{spec}}, T, p))$  implies,  $\beta_{a,j}$  depends also on the humidity. The reason being, that the particles’ sizes and therefore the scatter properties of aerosols, are affected

Surface-Volume-Coefficients		
Type of Hydrometeor	$a_{0,i}$	$a_{1,i}$
rain	1.1	0.75
snow	10.4	0.78
liquid cloud water	144.7	0.88
cloud ice	1.1	1

**Table 4.1:** Surface-Volume-Coefficients

by hygroscopic growth [32], which is considered in the parametrization.

In the parametrization five different aerosol species are considered: sea salt, desert dust, black carbon, organic matter and sulphates. From these five aerosols only sea salt, sulphates and organic matter undergo hygroscopic growth and it can be neglected for desert dust and black carbon, due to their predominantly hydrophobic chemical composition. The mass attenuation coefficients, presented in Table 4.3, are gained from a Mie scattering algorithm implemented and described by Reddy et al. [51] and take the humidity into account by adapting the particle radius and the mass attenuation. All values were chosen for the wavelength of 555nm for the incoming light, because the human eye has a peak in sensitivity at this wavelength [22, 71].

The contrast threshold  $\varepsilon$  determines when an object is considered to be visible. As discussed in Section 2.3 different values have been used in the past, but in agreement with Clark et al. [12] it was set to 2% for this study.

The layers which are used for calculating the average visibility can be set, depending on the purpose of the forecast. The default values, also used for this study, are set to average over the lowest 10 layers. This corresponds to a height from 0 to 240–300m above the ground. The range of the upper boundary is due to the horizontal variation of the maximum height within one layer caused by the usage of hybrid-pressure-terrain-following coordinates (see Figure 3.2).

The data of the aerosol concentration and its spatial distribution are taken from the NASA database [[https://gacp.giss.nasa.gov/data\\_sets/transport/](https://gacp.giss.nasa.gov/data_sets/transport/)]. This aerosol data set was generated by simulating sinks and sources of different aerosols according to the works of Tegen et al. [60], Chin et al. [10], Lioussse et al. [36] and Tegen and Fung [59], and is a climatological data set. Thus, high accuracy and currentness of the data is not given. This increases the uncertainty in the visibility parametrization. The aerosols data is provided in the unit of optical depth. To calculate the mass concentration the scattering cross section, a reference height and the density of air is needed. The scattering cross sections are also included in the NASA data set [42] and reference height in the corresponding works [60, 10, 36, 59]. The extinction coefficient of the Rayleigh contribution represents the scattering by atmospheric gases. As suggested by Clark et al. [12] it was set to  $\beta_R = 0.0391$ , the equivalent of 100km visibility.

Since so many physical processes that contribute to visibility cannot be resolved

explicitly, there are many different approaches of parametrizing visibility [13, 12, 56, 58, 20, 21, 11]. The simplest and computationally least expensive way to forecast visibility is to find an analytical function of some model variables by fitting to a set of measured data.

For benchmarking, seven other parametrizations of that kind were implemented, which were gathered from Gultepe and Milbrandt [20] and Gultepe, Müller and Boybeyi [21] and are presented in Table 4.2.

Abbreviation	Reference	Parametrization
Gul	Gultepe, Müller and Boybeyi(2006) [21], post-processing	$\min [0.0219 * \text{LWC}^{-0.9603} \times 10^{-3}, 67700 * (1 - q_{\text{rel}})^{0.67} \times 10^{-3}]$
Clark+	Clark et al. (2008)[12], Stoelinga and Warner(1999)[58]	$-ln(\epsilon) \left( \beta_R + \sum_i^{N_{\text{hydro}}} \beta_{\text{hydro},i} + \sum_j^{N_{\text{aero}}} \beta_{\text{aero},j} \right)^{-1}$
RUC	Smirnova, Benjamin and Brown(2000)[56]	$60 \exp \left( \frac{-2.5(q_{\text{rel}} \times 10^2 - 15)}{80} \right)$
FRAMC	Gultepe, Müller and Boybeyi(2006)[21]	$-41.5 \ln (q_{\text{rel}} \times 10^2) + 192.3$
AIRS	Gultepe, Müller and Boybeyi(2006)[21]	$-0.0177 (q_{\text{rel}} \times 10^2)^2 + 1.46q_{\text{rel}} + 30.80$
FRAML5	Gultepe and Milbrandt(2010)[20]	$-0.000114 (q_{\text{rel}} \times 10^2)^{2.7} + 27.45$
FRAML50	Gultepe and Milbrandt(2010)[20]	$-5.19 \times 10^{-10} (q_{\text{rel}} \times 10^2)^{5.44} + 40.10$
FRAML95	Gultepe and Milbrandt(2010)[20]	$-9.68 \times 10^{-14} (q_{\text{rel}} \times 10^2)^{7.19} + 52.20$

**Table 4.2:** Overview of all implemented parametrizations, including references and used abbreviations.

Mass Attenuation Coefficients of Different Aerosols $\alpha_j[m^2/g]$												
Relative Humidity	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.85	0.9	0.95
Sulphates	4.94247	4.94247	4.94247	4.94247	6.85	7.74471	8.92815	10.6356	13.5782	16.2277	21.0072	35.4261
Organic Matter	3.07043	3.07043	3.07043	3.07043	4.27185	4.84567	5.60681	6.70989	8.6229	10.3565	13.5060	23.1454
Sea Salt	0.03889	0.03889	0.03889	0.03889	0.07921	0.09226	0.10549	0.122930	0.149650	0.17160	0.21005	0.31026
Desert Dust	0.42744											
Black Carbon	9.412290											

**Table 4.3:** List of mass attenuation coefficients of different aerosols. Sulphates, organic matter and sea salt undergo hygroscopic growth and resulting in a different mass attenuation for different values of relative humidity. The relative humidity is categorized in different bins. The values shown in this table, starting from the left to the right, denoting the edges of the bins. The upper edge of the last bin is missing, obviously being 1.

# Chapter 5

## Methods and Set-Up

### 5.1 Perturbation Scheme

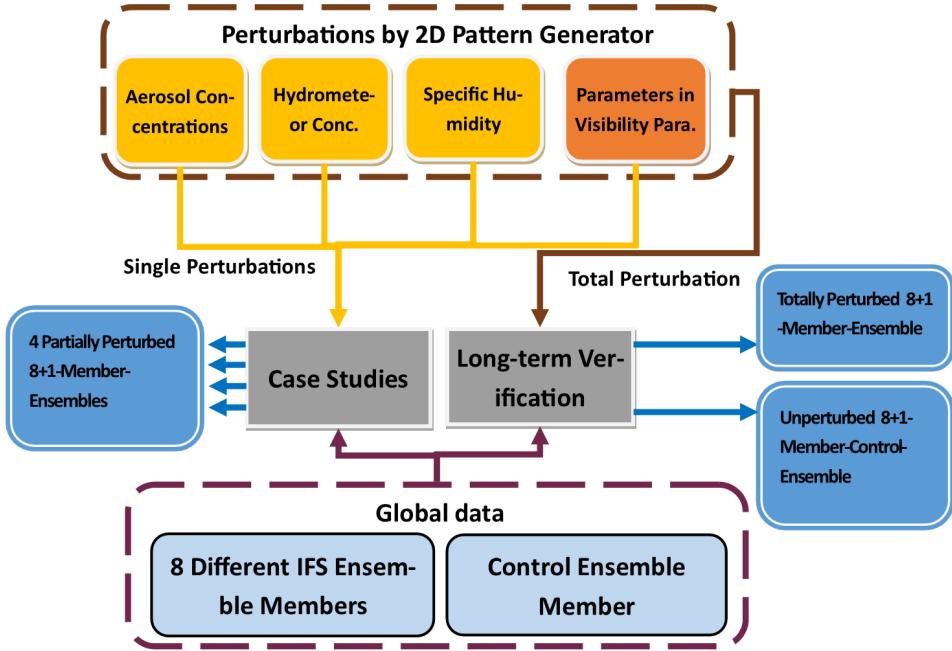
For the evaluation of the parametrization, several forecasts with different perturbation schemes, also called NWP experiments, were performed. To investigate the influence of different variables on visibility, the following versions of perturbations, presented in Figure 5.1, were tested:

1. Control ensemble
2. Perturbation of aerosol concentrations
3. Perturbation of hydrometeors
4. Perturbation of the parameters in the visibility parametrization
5. Perturbation of specific humidity
6. Total perturbation (aerosols, hydrometeors, specific humidity, parameters)

Figure 5.1 illustrates the listed perturbed variables and which combinations were run as case studies and for the long-term verification. No additional perturbation was applied to the control ensemble. It constitutes of an AROME ensemble, which was created by using data of eight perturbed members and one control member of the IFS ensemble to set different boundary and initial conditions for each member. The measured visibility lies often outside the ensemble spread of the control ensemble, mainly, because this ensemble does not take model error into account. Especially for visibility, many of the key parameters are hard to predict reliably and often not varied in the standard perturbation schemes [11]. To tackle this underdispersion of the control ensemble, additional local perturbations were introduced.

In this case, local means, the perturbations were chosen and set-up to be applied only in the visibility module. The other parts of the model remained untouched.

For the perturbed ensemble, the current values are perturbed with the corresponding spatial-temporal pattern at each time step and the perturbed quantities are passed on to the visibility parametrization and thus, do not affect any future states of the model. This method forbids the perturbations to be amplified in time, and therefore the default amplitudes must be higher in comparison to a perturbation scheme that



**Figure 5.1:** Illustration of the perturbation scheme: Showing the different perturbed quantities at the top and the used ensemble members from the IFS model on bottom.

is applied to the overall model and can evolve. The reason why a local perturbation scheme is used, is that no global perturbation scheme for AROME is available, where all the relevant parameters and variables can be perturbed. Perturbing all of them globally, would lead to a change in a number of non-linear feedback processes. The model could not handle those perturbations, because the effects of different perturbations can have conflicting outcomes for different variables. Since we are only interested in investigating the uncertainty of the visibility forecast, a local perturbation is sufficient.

Although we can estimate the ranges for all parameters quite well, we have very little a priori knowledge about the compatibility of the values of different variables after a certain time.

For a locally applied perturbation, the distance to the original trajectory in phase space can be restricted and chosen small enough to have tolerable errors. If we would allow them to propagate, we are unable to predict how the perturbed trajectory will evolve. There is no guarantee that it will lead to a realistic outcome at a later time, because the tolerable error can be amplified to a completely nonsensical scenario. There are several other attempts of probabilistic forecasting of visibility [11, 52], but we believe that the advantage of our method is its comprehensible nature. It is not purely statistical but relates to single physical quantities.

The details of the perturbation scheme are outlined in the following subsections.

### 5.1.1 Set-Up of Perturbations

All perturbations were generated by use of the pattern generator described in Section 3.2. The spatial correlation length was only roughly estimated in terms of large-scale and meso-scale processes, because it, is only a namelist setting and has a complex relation to the actual spatial scale. It cannot be directly related to a length scale, because it is scaled with respect to the Earth's radius, since it was originally set-up for a global model. All fields, with exception of the aerosol concentrations, are assumed to be meso-scale perturbations. Temporal correlations were all set to 72 hours, because the time scales have a negligible impact, since the local perturbations cannot evolve in time.

A weakness of the implementation of the pattern generator is that when multiple fields are perturbed that perturbations can vary in spatial and temporal correlation length, but all of them must have the same clip ratio  $x_{\text{clip}}$ , which was set to 1. Same holds for  $\sigma_{\text{nl}}$ , which was set to 0.05.

To compensate for this, we introduced scaling factors that are applied to the perturbations in the appropriate way. They are different for each quantity and explained later in this section, when the details of the perturbations for each quantity are provided.

Depending on the perturbed quantity and the perturbation scheme, we needed to define a suitable range for each quantity, implemented by means of the scaling factor. For some quantities, like the aerosol concentrations, global data and its extrema can be used (see Table 5.1) to determine the range. In other cases there are no global extrema to be considered, but a physical constraint, like a quantity being greater than zero (e.g. the surface-volume-coefficients). If the related phenomena are strongly local, like precipitation, local climatological extrema are appropriate gauge values for the perturbation ranges. The different constituents of the perturbation scheme are presented in Table 5.1 with the corresponding restriction of the perturbation range. Table 5.1 shows all considered uncertainties of the scheme and the type of the errors, where the uncertainties originate from. It provides an overview of the different restrictions of the quantities, which limit the ranges and are discussed in the following.

Contributor	Restrictions	Type
IFS Global Model	8 Members	b.c.s/i.c.s *
Specific Humidity	10% of $q_{\text{spec}}$	Model Error
Hydrometeors	Local Climatological Extrema	Model Error
Aerosols	Global Climatological Extrema	Model Error
$c_i$ in the <i>vis</i> parametrization	Physical Considerations	Model Error

**Table 5.1:** Constituents of the Perturbation Scheme: Different contributors differ in the source of error and the restrictions for the ranges.

\*b.c.s=boundary conditions; i.c.s=initial conditions

The namelist settings for the pattern generator for each field or variable can be found in Table A.1 in Appendix A.

## Perturbation of IFS data

The control ensemble is created in AROME by using eight different members and the control member of the global IFS model. The IFS-ensemble provides different initial and boundary conditions, that were gained by the application of the ECMWF perturbation scheme documented in [39] and [19]. In all other perturbation regimes the perturbation is applied in addition to the varying initial and boundary conditions taken from the IFS.

## Perturbation of Hydrometeors

When analysing the uncertainty of hydrometeor concentration, we find two contributors: the location of the hydrometeors and the actual concentrations.

The discrepancy of model and reality is considered worse, when on certain locations no occurrence of hydrometeors is forecast, but in reality some clouds, or precipitation are reported, than in comparison of wrong estimation of precipitation just in terms of heaviness. Hence, the perturbation scheme has to include variance in location in a way, that also on locations where the unperturbed fields of the concentrations are zero, they can be non-zero on the perturbed fields. The tendencies of the hydrometer concentrations were used to create such a perturbation scheme that also relates to the current weather situation, by using the accumulated tendencies from previous time steps.

To understand the perturbation regime, one has to understand the calculation of the unperturbed hydrometeor concentrations

$$c_i(t_n) = c_i(t_{n-1}) + T_{c_i}(t_n) \quad (5.1)$$

first. The concentration  $c_i$  at a time  $t_n$  is the sum of the concentration at the previous time step of the hydrometeor  $c_i(t_{n-1})$  and the tendency of the concentration of the hydrometeor  $T_{c_i}$  at  $t_n$ . The tendency  $T_{c_i}$  is the sum of the output of different physical parametrizations. These parametrizations take other model variables (at the current time step) as input, e.g. humidity and temperature, so that hydrometeor concentration and distribution are related to the updated overall atmospheric state.

The perturbations  $P_{c_i}$  are set up as the past accumulated tendencies multiplied by the exponential of a random number  $r_j$  summed up over a time span  $t_{\text{acc}}$ . For this study the time span is set to two hours.

$$P_{c_i}(t_N) = \sum_{t_n=t_N-t_{\text{acc}}}^{t_N} T_{c_i}(t_n) \exp(r_j) \quad (5.2)$$

As described by Equation (5.3), the perturbed concentration  $\tilde{c}_i(t_n)$  is set by adding the perturbation to the unperturbed quantity  $c_i(t_n)$ .

$$\tilde{c}_i(t_n) = c_i(t_n) + s P_{c_i}(t_n) \quad (5.3)$$

The perturbation is scaled by the factor  $s$ . The scaling factor  $s$  was gained by testing different values under the criterion, that the perturbed concentrations cannot exceed

the doubled monthly extrema on the domain  $L$ . This restrain was chosen, because the climatological data is a good reference in terms of what can be considered as in the norm. Global extrema are irrelevant, since precipitation depends strongly on the climate zone. The advantage of using the tendencies of a time span, close to the current time, is that the change in location of the hydrometeors is related to local weather conditions that promote the presence of hydrometeors.

## Perturbation of Aerosols

Aerosol distributions are usually evolving on a global scale [60, 10, 36, 59]. Hence, it is reasonable to use global extrema to set the ranges for the perturbed aerosol concentrations  $\tilde{c}_j$  on the LAM domain. To do so, some general considerations about local and global extrema in NWP and stochastic physics ensembles are made:

When the unperturbed quantity  $\xi$ , in a LAM has a local maximum  $\max_{\text{local}}$  and local minimum  $\min_{\text{local}}$  on a local domain  $L$ , the perturbation range must be chosen in a way, that the perturbed quantity  $\tilde{\xi}$  cannot exceed any known global extrema  $\max_{\text{global}}$  or  $\min_{\text{global}}$ . To make sure that the perturbed quantity  $\tilde{\xi}$  is in a physically reasonable range on  $L$  for any possible case, the following condition must hold at all times:

$$\min_{\text{global}} < \tilde{\xi} < \max_{\text{global}}. \quad (5.4)$$

This leads to the criteria for the range  $a$  for log-normally distributed perturbations:

$$a_{\log}(\max) = \ln \left( \frac{\max_{\text{global}}}{\max_{\text{local}}} \right) \quad (5.5)$$

$$a_{\log}(\min) = \ln \left( \frac{\min_{\text{local}}}{\min_{\text{global}}} \right) \quad (5.6)$$

Since in most cases the range criteria vary, depending on choosing a restriction with respect to minimum or maximum, one has to decide, if one extremum can be used and the other one be exceeded or not even reached.

We decided to set the range with respect to the maximum. Thus, the global minimum is not reached, because the amplitude of the perturbation is too small. This is justified by the greater importance of the maximum for a realistic estimate of the worst case scenario of the lowest possible visibility.

The aerosols concentrations  $c_j$  are perturbed by multiplying the concentration by an exponential function:

$$\tilde{c}_j = c_j \exp(sr_j) . \quad (5.7)$$

The scaling factor  $s$  was calculated for each aerosol as described by Equation (5.5) and all values are presented in Table A.2 in Appendix A.

Because the climatological aerosol data varies monthly [42], the scaling factor was calculated for each month. Because the dynamics of aerosols happens on larger scales than most other atmospheric transport processes and by investigation of the global aerosol data set [42], the spatial correlation length for the pattern generator was estimated to be 200000km. It is important to keep in mind that this correlation length does not correspond directly to the scale of the process but in a complex non-linear manner as described in Section 3.2.

## Perturbation of Physical Parameters

The surface-volume-coefficients of hydrometeors, scattering cross sections and mass attenuation coefficients of aerosols were perturbed. The motivation is that these parameters were set to one ‘best-guess’ value, but in reality they depend on the radii of the particles, which can vary significantly. The extinction caused by Rayleigh scattering is also perturbed because the changes in distributions of the composition of the gaseous components of air are neglected in the model. All these physical parameters need to be greater than zero at all times. Additionally, the physical assumptions for different relevant processes in the related works of Stoelinga and Warner [58] and Tegen et al. [60] were used as well as the set-up of the ECMWF stochastic physics scheme described by Ollinaho et al. [43]. Then test cases were run to study the impact of the perturbations and empirically adapt the perturbations to the values presented in Table A.1 in Appendix A.

## Perturbation of Specific Humidity

The works of Stensrud, Bao and Warner [57] was used as a point of reference to set the range for specific humidity. Stensrud, Bao and Warner provide statistics about the RMSE values, which let us estimate an initial perturbation of 10%. Then, again, test cases were run to study the impact of the perturbations and empirically adapt the perturbations to their final values presented in Table A.1 in Appendix A.

## 5.2 Set-Up

Model Setting	Value	Effect
Lag Time	36h	Total Spread
Coupling Frequency	3h	Dependence on Global Model
Forecast Duration	18h	Available Outputs, Total Spread
Base Time	00:00	Available Outputs, Total Spread
Long Term Verification	July 2016 January 2017	Evaluation
Case Studies	10.07.2016 14.07.2016 08.01.2017 21.01.2017 29.01.2017	Impacts of Single Quantities

**Table 5.2:** Overview of the model’s settings, the chosen verification periods and case studies.

The model output was created from the totally perturbed ensemble, which incorporates all single perturbations, and the control ensemble for two whole months to obtain a large enough data set for the verification process. The months January

2017 and July 2016 were chosen, to be able to analyse data of summer and winter conditions. In addition, all the different versions of single perturbations were run for the following selected dates: **10.07.2016, 14.07.2016, 08.01.2017, 21.01.2017, 29.01.2017**.

The days for the case studies were selected with having in mind, the want of investigating a broad variation of weather conditions. The case studies are of great importance to gain a deeper insight into the impact of each parameter and variable.

Since the overall ensemble spread is known to be too low, the model settings presented in Table 5.2 are chosen under the aspect of amplifying it.

The **base time**, which denotes the starting time of a forecast, was set to **00:00 UTC** and the **forecast duration to 18 hours**. UTC stands for Universal Time, Coordinated and means that only one global time zone is used. The setting was chosen, because visibility measurements are taken from 06:00 to 18:00 UTC. A base time of 00:00 UTC assures that the ensemble can evolve before the first value for verification is taken. For every run eight randomly selected ensemble members from the global IFS ensemble were taken, to investigate the effect of varying initial conditions. This number was chosen based upon what is computationally affordable and still gives a reasonable spread [8]. For each member a different perturbation pattern was applied, by passing on a different seed to the random number generator. In addition every parameter has a different pattern. The purpose of this is to model the realistic scenario that different components in a system can either amplify or damp one and another.

The coupling of the global and the limited area model is something that has great influence on the model spread. The lag time is the parameter that defines the difference of the base time of the run of the global model and the base time of the limited area model. It is known that for short-range and medium-range forecasts on average, the longer a forecast period lasts, the higher the spread is [6, 4]. Hence, the lag time has an impact, because if we take different ensemble members after a long forecast period the initial conditions have a much higher variation, than if they were taken at an earlier stage. We set the **lag time to 36 hours**. The coupling frequency is the model parameter that determines how often the boundary conditions should be updated with respect to the global model. The coupling is especially important, to avoid that the limited area model can evolve over longer periods of time in a way that is in total disagreement with the global model, and hence nonsensical. So the best is to couple as often as possible, with regards to availability of global data and computational affordability. For this study the **coupling frequency is set to three hours**.

### 5.3 Evaluation

For the verification of the forecast, the model output data was evaluated against visibility measurements on the forecast domain.

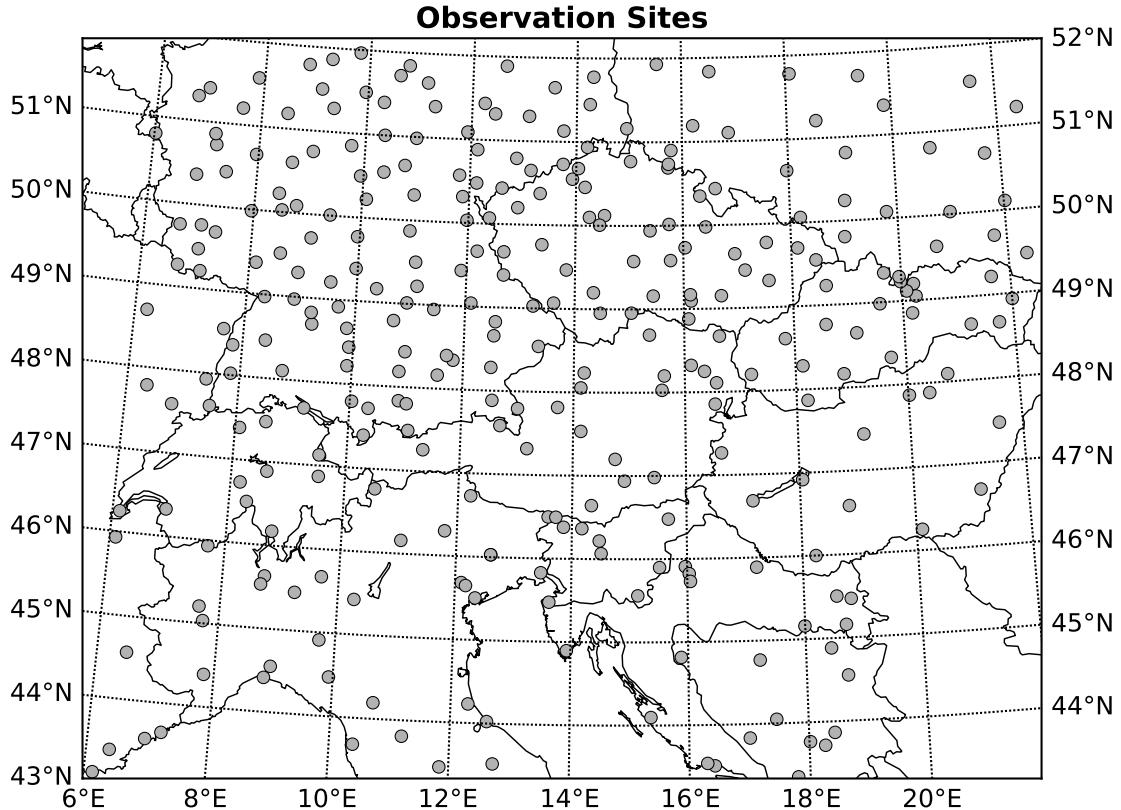
In this section the measurement procedure for visibility is explained and the different scores used for the evaluation of the parametrizations are presented.

### 5.3.1 Measurements

Visibility measurements are taken regularly at 6:00, 9:00, 12:00, 15:00 and 18:00 UTC, at each of the observation sites marked in Figure 5.2. The measurement is performed by a human observer, who identifies the most distant mark, which still can be seen [71]. All horizontal directions need to be considered and the lowest value determines the visibility. The number-coded system for the categorization is presented in Table A.5 in Appendix A and decreases in accuracy with increasing visibility.

The observational data is continuously stored in a joint database, where all national meteorological offices worldwide push their data.

Observations encoded by a different number-coded system, as found in Bosnia-Herzegovina and Serbia, were discarded. The data was further pre-selected by only using stations for the monthly average, which stably provided at least three measurements a day during the verification period. Additionally, it was tested if stations placed in a mountainous terrain showed a worse skill, because they were suspected of not being representative due to orographic mismatch, orographic shadowing or strong variations within a grid box. This turned out to have a negligible effect and as a result topological height was discarded as a selection criterion.



**Figure 5.2:** Model domain and all observation sites where visibility measurements are taken

### 5.3.2 Scores

For the computations of the scores only certain values on the model domain are used: The model output data for the forecast visibility is a two-dimensional field. For every station where visibility is measured, the data at the point of the coordinates of the measurement station was extracted from the two-dimensional field and evaluated. To obtain values for all coordinates of the stations, a nearest neighbour interpolation was performed. This was done for all points in time, when measurements are taken. Thus, the data can be indexed by time up to the number of all measurements within the verification period at one location  $N_{\text{Mes}}$  and by location up to the number of considered observations sites  $N_S$ .

The different scores, we used for the evaluation of the goodness of the visibility prediction, are described in the following.

#### Root Mean Squared Difference of the Logarithmic values (RMSLog)

The average of the root-mean-square of the difference of the logarithm of the forecast and observed value is mathematically expressed as the following:

$$\text{RMSLog}_n(\lambda_{\text{lon}}, \phi) = \frac{1}{N_{\text{Mes}}} \sum_{m=1}^{N_{\text{Mes}}} \frac{\sum_{i=1}^{N_{M_{\text{tot}}}} (\log(vi_{s_{o,n,m}}) - \log(vi_{s_{f,i,n,m}}))^2}{N_{M_{\text{tot}}}} \quad (5.8)$$

$$\text{RMSLog}_{\text{total}} = \sqrt{\frac{\sum_{n=1}^{N_S} \text{RMSLog}_n}{N_S}}. \quad (5.9)$$

To obtain a local score for each station  $\text{RMSLog}_n$ , first, the average squared differences of the logarithmic values of the forecast visibility  $vi_{s_{f,i,n,m}}$  and the observed visibility  $vi_{s_{o,n,m}}$  are calculated. Here  $i$  stands for the number of the member,  $m$  indexes the time of the  $m$ th measurement and  $n$  the location of the  $n$ th observation site. First, the scores are averaged over the whole ensemble, where  $N_{M_{\text{tot}}}$  stands for the total number of ensemble members. Second, the average over all measurements in time is taken, where  $N_{\text{Mes}}$  stands for the number of measurements in the time span of interest.

To compute the total score  $\text{RMSLog}_{\text{total}}$ , as described by Equation (5.9), the root-mean-square of the mean of the local score  $\text{RMSLog}_n$  of the total number of stations  $N_S$  is taken.

The score, as described by Equation (5.9) and (5.8), is computed from the logarithmic values with respect to the magnitude of the observed values: A forecast value that is off by 20km should be more weighted, when the observed value is 2km than, when it is 70km. Another reason for using a logarithmic score is that the observations are categorized with increasingly large intervals, so this way the different bin sizes of the measurement data (see Table A.5) can be acknowledged.

## Ranked Probability Score (RPS)

The ranked probability score evaluates the probabilities gained from ensemble forecasting instead of only a mean value [8]. Thus, it is seen as a more accurate measure to test, if the perturbation scheme is well chosen. It is a commonly used score and described in the following after the textbook by Wilks [70].

To calculate the RPS, first, the forecast values need to categorized into  $K$  different bins. The number of ensemble members  $N_{\text{bin}_j, M}$ , whose predicted values are categorized in one bin, needs to be counted for a number of specified points in space and time for all bins. The points are chosen with respect to the available measurements, usually overlapping with their location and timing. As described by Equation (5.10), the condition for a member being included in the count, the forecast values  $f$  (in our case  $\text{vis}_{f_i, n, m}$ )  $\text{bin}_{j, \min} < f < \text{bin}_{j, \max}$  needs to be satisfied.

$$N_{\text{bin}_j, M}(\lambda_{\text{lon}}, \phi, t) = |\{\text{mem} : f(\lambda_{\text{lon}}, \phi, t) \in [\text{bin}_{j, \min}, \text{bin}_{j, \max}]\}| \quad (5.10)$$

From that, the forecast probability of an observation that lies within the interval of the  $j$ th bin is calculated as

$$p_{\text{bin}_j}(\lambda_{\text{lon}}, \phi, t) = \frac{N_{\text{bin}_j, M}(\lambda_{\text{lon}}, \phi, t)}{N_{M_{\text{tot}}}} . \quad (5.11)$$

For the RPS, the accumulated probabilities

$$F_J = \sum_{j=1}^J p_{\text{bin}_j} \quad (5.12)$$

and accumulated observations

$$O_J = \sum_{j=1}^J o_{\text{bin}_j} \quad (5.13)$$

are used, where  $o_j$  stands for the occurrence of an event. So if the observed value lies in the interval of the  $j$ th bin it equals 1, otherwise 0. A consequence of calculating a score from the accumulated probabilities is, that the higher the bin index  $j$  is, the less is the accuracy of the bin weighted in the score. Like for the RMSLog, this is a desired property, because the accuracy of the visibility forecasts is more important for low values. Finally, we can define the RPS as

$$\text{RPS} = \sum_{J=1}^K (F_J - O_J)^2 . \quad (5.14)$$

The conditions  $\sum_{j=1}^K p_{\text{bin}_j} = 1$  and  $\sum_{j=1}^K o_{\text{bin}_j} = 1$  must hold and consequently, the last term of Equation (5.14) always vanishes [70].

The bin edges for the RPS were defined as: [0.0, 0.5, 1.0, 5.0, 10.0, 15.0, 20.0, 30.0, 50.0, 80.0]. The increasing bin size relates to the decrease in accuracy in the number-coded system for the measurement data (Table A.5 in Appendix A). RMSLog and RPS show a high sensitivity to distance.

A major concern when evaluating new parametrizations is, to not assess the performance of the whole model, but only of the parametrization.

This problem arises, because for certain days and weathers situations, the forecasts as a whole are bad, and since the visibility forecast uses other model variables as input, the total scores in visibility reflect also the overall model performance. The unbiased evaluation of parametrizations can never be achieved to perfection, but a thorough preselection of data can help. For this, the scores can be computed using only data of dates, where the overall model skill was high, or the correlations of the skill of different model variables and the forecast variable we want to assess can be analysed.

Another way to tackle this problem is, to use a relative score. Meaning, if a comparable parametrization already exists, the quality of the forecasts is only quantified in terms of relative improvement. The Gul-parametrization was so far the only tool for visibility prediction and already implemented in a post-processing package for AROME at ZAMG. Therefore, we decided to evaluate the other parametrizations against the Gul-parametrization (see Table 4.2).

In order to do so, a relative version of the RMSLog and the RPS are introduced. For the RMSLog and the RMSE, the relative score is expressed in percentage of the total score of the Gul-parametrization. For the RPS, the relative version is given by the Ranked Probability Skill Score (RPSS). The RPSS is a traditional skill score. A skill score SS is generally defined as

$$SS = \frac{A - A_{\text{ref}}}{A_{\text{perf}} - A_{\text{ref}}} \quad (5.15)$$

(e.g. Wilks [70] ), where  $A$  stands for the accuracy.  $A_{\text{ref}}$  is the accuracy of the reference case and  $A_{\text{perf}}$  the accuracy of a perfect forecast. That means that the goodness of the parametrization is rated with respect to a reference parametrization that has a ranked probability score  $RPS_{\text{ref}}$ . The RPS of the other parametrization is evaluated against it, as described by

$$RPSS = 1 - \frac{RPS}{RPS_{\text{ref}}} \quad , \quad (5.16)$$

since for the RPS the perfect accuracy  $A_{\text{perf}}$  equals 0. Regarding Equation (5.16) it is evident that the RPSS has to lie within the interval  $[-\infty, 1]$  and all values greater than zero indicate a better performance, than the one of the reference.

# Chapter 6

## Results

In this chapter the findings of this thesis are presented. An inter-comparison of the performance of all implemented parametrizations is shown. Then, the different skills for different scores of the Clark+-parametrization are discussed with respect to the Gul-parametrization, which was used as a reference. In addition, seasonal variations and the influence of the perturbation scheme and the different weather situations on the forecast performance are illustrated and analysed.

### Overall Performance and Inter-comparison of all Parametrizations

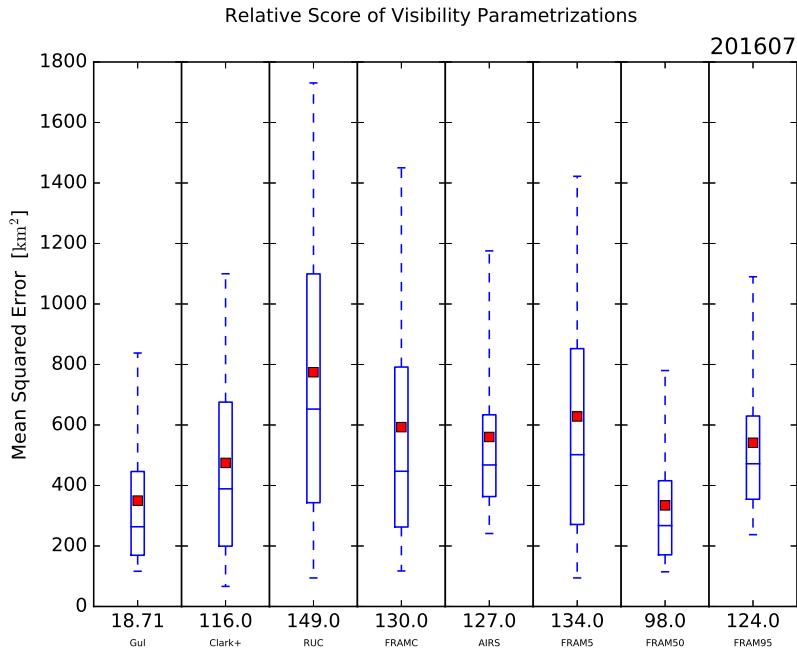
An overview of the relative average RMSE and RMSLog is provided in Table 6.1. The different scores vary significantly depending on the parametrization and season. Concerning the RMSE for January 2017, presented in Figure 6.3, four out of seven parametrizations, including the Clark+-parametrization, show a better performance than the reference parametrization (Gul). When comparing this to the RMSLog, presented in Figure 6.4, the number decreases to three out of seven, because the Clark+-parametrizations fails by 1% to outmatch the reference (Table 6.1). These three parametrizations have a worse skill, when evaluated for RMSLog than for RMSE, indicating a decrease in performance for low visibility cases. For the period of July 2016 the overall scores, shown in Figures 6.2, 6.1 and 6.9 are significantly worse. When analysing the relative RMSLog score in Figure 6.2, we cannot find a single parametrization outperforming the Gul-parametrization. Similar can be said for the RMSE (Figure 6.1). Only the FRAM50-parametrization appears to perform well at first glance, but its poor skill for low visibilities is exhibited by its RMSlog value (Figure 6.2) and its RPSS values (Figure 6.9). This seasonal variation is discussed in more detail later in this section.

The scores in Table 6.1, show that on average none of the other parametrizations could outperform the Gul-parametrization. Of all implemented parametrizations the Clark+ and FRAMC show the lowest error by 3.75% and 1.75%. Although FRAMC has a lower average, we still consider Clark+ as the best alternative to Gul, because its scores show lower overall spread, implying a more stable performance. When comparing the RPSS (Figures 6.9 and 6.8) and the RMSLog (Figures 6.2 and 6.4) a strong correlation can be found, since both are weighted scores, as explained in Section 5.3. However, the skill of Clark+ is significantly better, when assessed by the relative RMSLog than by the RPSS. This is probably a result of the different

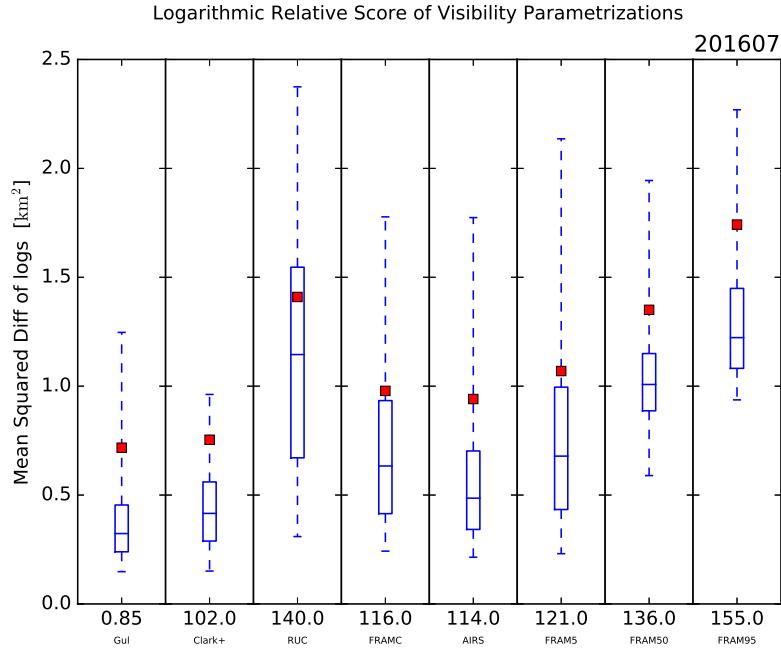
<b>Relative Score</b>	<b>Clark+</b>	<b>RUC</b>	<b>FRAMC</b>	<b>AIRS</b>	<b>FRAM5</b>	<b>FRAM50</b>	<b>FRAM95</b>
<b>RMSE July 2016</b>	116	149	130	127	134	98	124
<b>RMSLog July 2016</b>	102	140	116	114	121	136	155
<b>RMSE January 2017</b>	96	76	79	152	77	108	181
<b>RMSLog January 2017</b>	101	77	82	117	98	103	133
<b>Average</b>	103.75	110.5	101.75	127.5	107.5	111.25	148.25

**Table 6.1:** Relative Score Overview: the monthly averaged relative RMSE and RMSLog values for all different parametrizations is presented. The Gul-parametrization was used as reference and all values denote the score in percentage with respect to it.

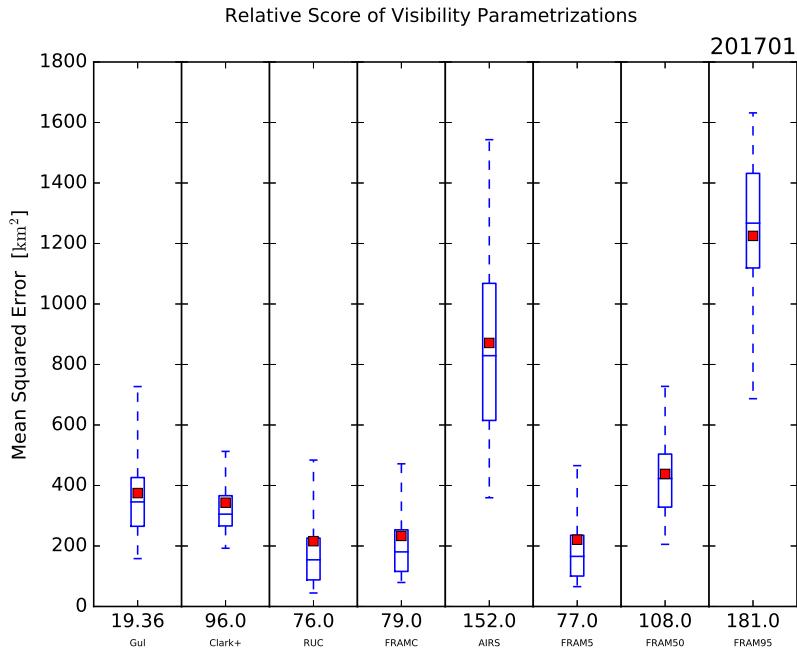
weighting. The RPSS weighting is based on a set of bins, customized to the study of visibility, and is thus, believed to be the more accurate score. Nevertheless, our results show that the verification profits from the usage of different scores. Because although the RPSS is a more sophisticated score, and therefore in the following mainly used for the analysis of the results, error estimation in percentage is easier to understand intuitively.



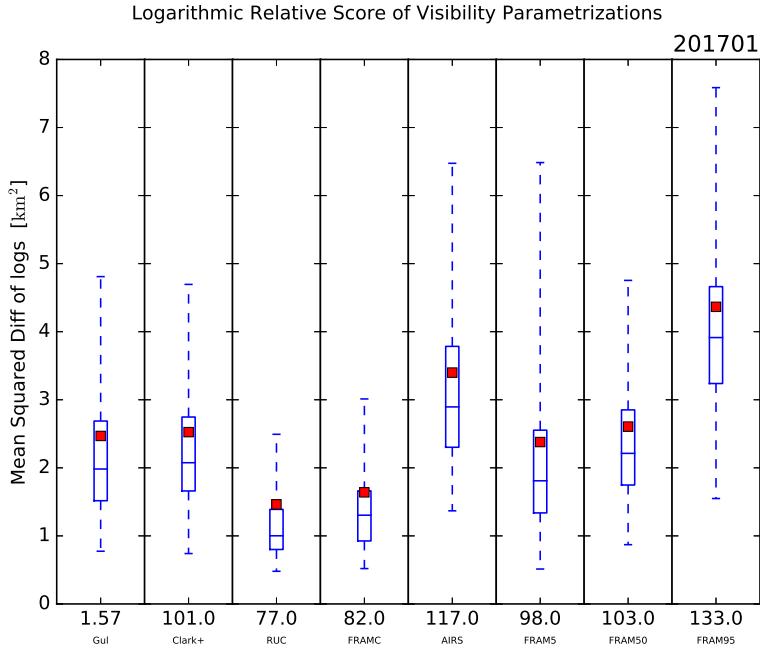
**Figure 6.1:** Box plots of the ensemble average Mean Squared Error, average for July 2016. Whiskers denoting the 5th- and the 95th-percentiles, the red squares the mean values. The first value on the X-axis is the average RMSE value of the reference parametrization ‘Gul’ and the other values denote the Root Mean Squared Error relative to the ‘Gul’ parametrization in percentage.



**Figure 6.2:** Box plots of the ensemble average mean squared difference of the logarithmic observed and forecast values, average for July 2016. Whiskers denoting the 5th- and the 95th-percentiles, the red squares the mean values. The first value on the X-axis is the average RMSLog value of the reference parametrization ‘Gul’ and the other values denote the RMSLog relative to the ‘Gul’ parametrization in percentage.



**Figure 6.3:** Box plots of the ensemble average Mean Squared Error, average for January 2017. Whiskers denoting the 5th- and the 95th-percentiles, the red squares the mean values. The first value on the X-axis is the average RMSE value of the reference parametrization ‘Gul’ and the other values denote the Root Mean Squared Error relative to the ‘Gul’ parametrization in percentage.



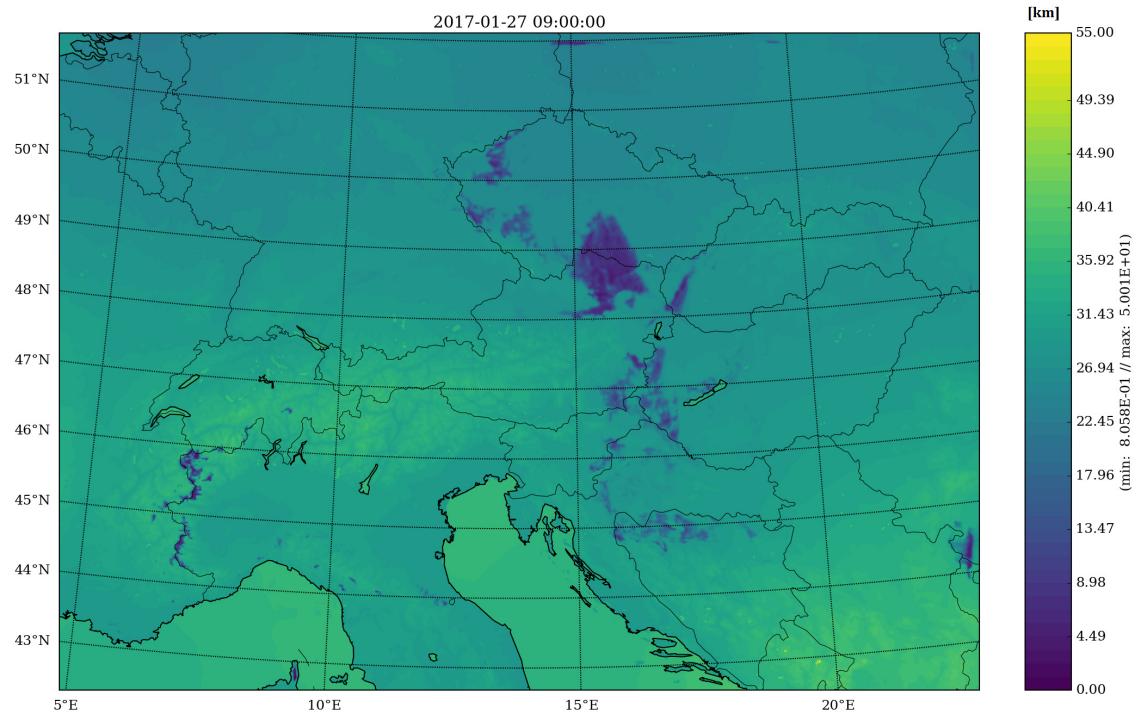
**Figure 6.4:** Box plots of the ensemble average mean squared difference of the logarithmic observed and forecast values, average for January 2017. Whiskers denoting the 5th- and the 95th-percentiles, the red squares the mean values. The first value on the X-axis is the average RMSLog value of the reference parametrization ‘Gul’ and the other values denote the RMSLog relative to the ‘Gul’ parametrization in percentage.

## Correlation of Weather Conditions and Visibility Forecast

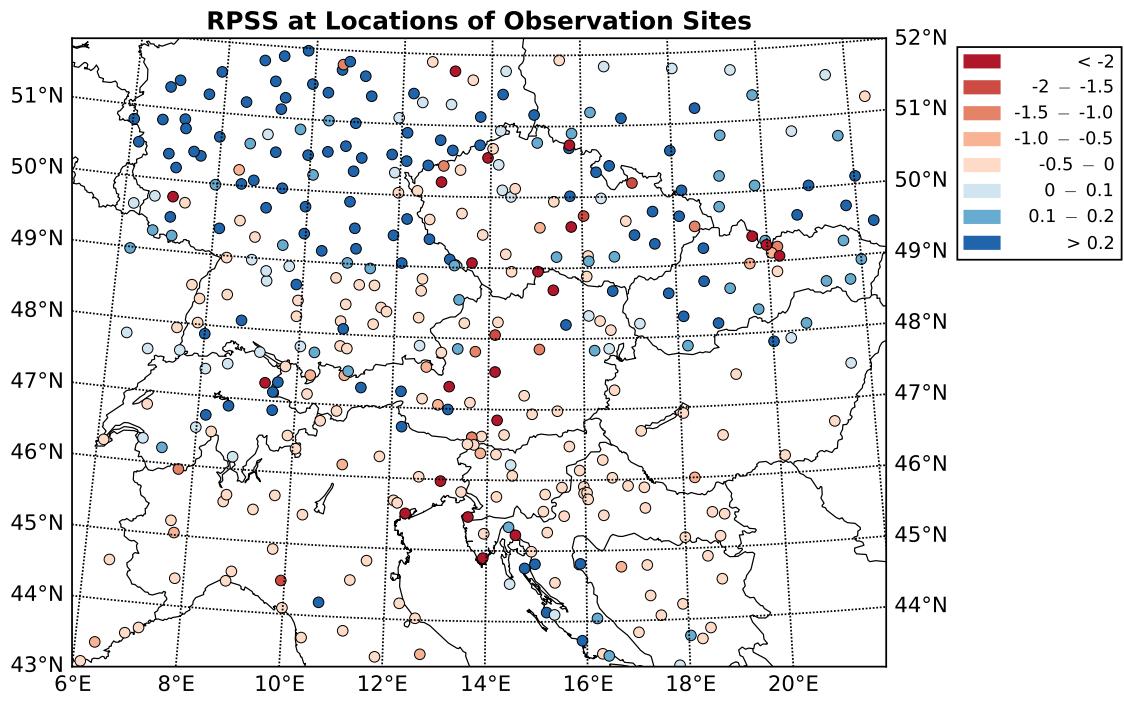
To understand the correlation of present weather conditions and the forecast visibility, the visibility forecast of 27<sup>th</sup> January 2017, 09:00 UTC, presented in Figure 6.5, is discussed.

As shown in Figure 6.8, it is one of the days with a comparably high RPSS in the verification period. The satellite image in Figure 6.7 is used to analyse the weather conditions. In addition to it, the surface weather analysis in Figure B.1 and B.2 in Appendix B provide a qualitative overview for an enhanced interpretation of the correlation of weather conditions and forecast visibility.

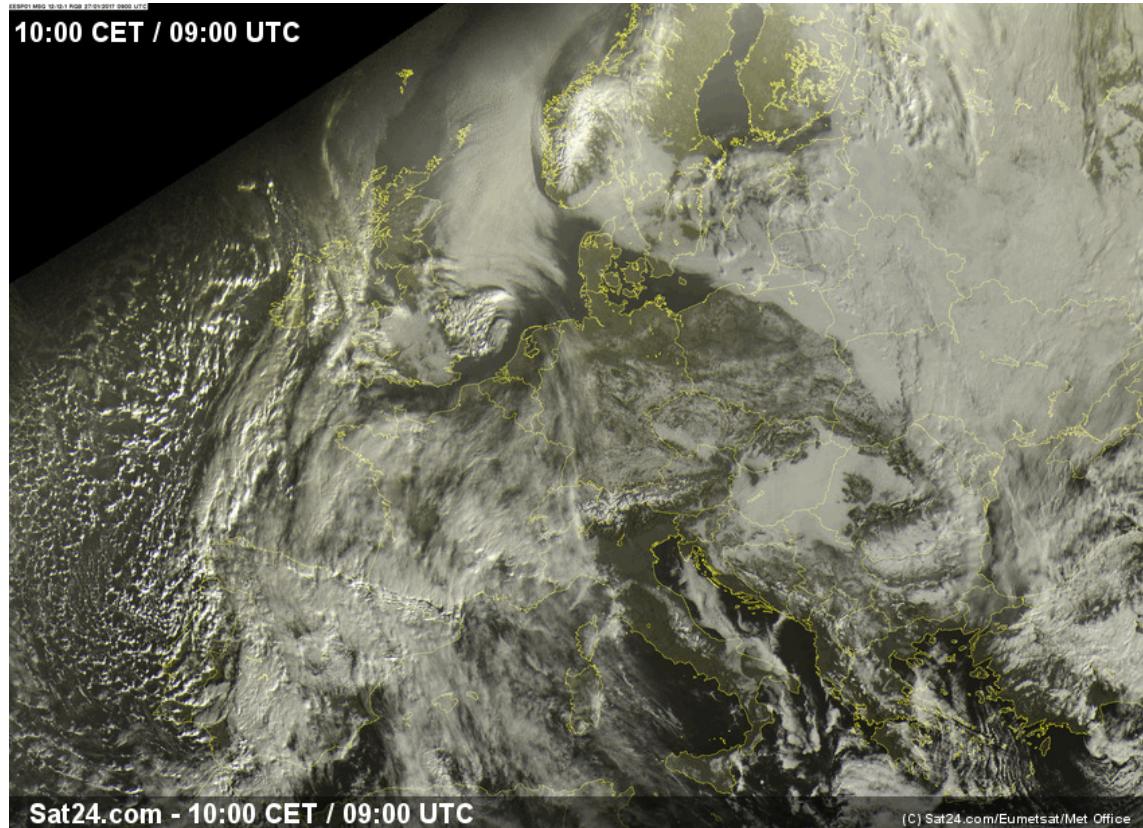
The local RPSS values of the Clark+-parametrization for 27<sup>th</sup> January 2017 are shown in Figure 6.6. The good scores in the northern part of the domain are probably the result of the stable and sunny weather conditions. The low visibility in some areas like the Austrian-Czech-border is due to fog. In the satellite image (Figure 6.7) fog can be identified as the very homogeneous grey areas. Fog is forecast as liquid cloud water in the model, generally hard to predict and extremely sensitive to initial conditions [75]. The analysis of the results of days with a particularly good or poor performance revealed a strong correlation between the skill of the Clark+-visibility-forecast and the skill of the forecast of liquid cloud water close to the surface. Since in this case the fog forecast had a high accuracy, the visibility forecast had a high accuracy as well. Looking at Figure 6.7, B.1 and B.2, we find a presence of fog in Hungary, where no reduced visibility is indicated in Figure 6.5. The successful prediction for the northeast of Austria and the poor forecast for Hungary are reflected in the local skills illustrated in Figure 6.6.



**Figure 6.5:** Contour plot of forecast visibility by the Clark+-parametrization on 27-01-2017, 09:00 UTC. The forecast is taken from the control run from the control ensemble.



**Figure 6.6:** RPSS of the Clark+-parametrization marked and colour-coded on the considered observation sites on the Austrian AROME domain on 27-01-2017

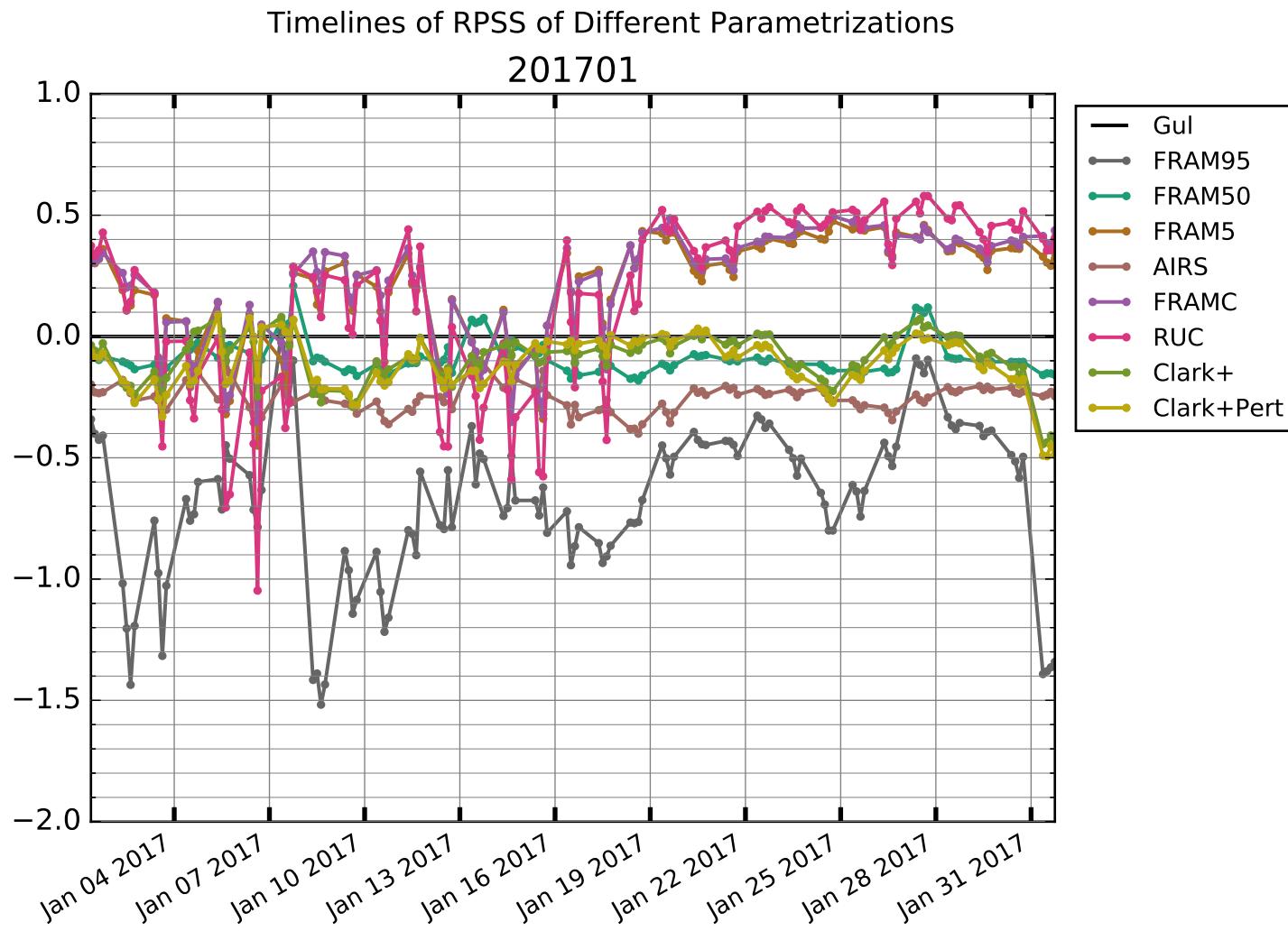


**Figure 6.7:** Satellite Image 27-01-2017, 09:00 UTC 27-01-2017, 09:00 UTC of Central Europe. Image from the online archive of ‘The Meteo Company B.V.’ [63]

## Seasonal Variations in Skill

When comparing Figure 6.8 and Figure 6.9 we see a significant difference in skill: The overall RPSS of the Clark+-parametrization is much better in January, whereas the Gul-parametrization seems to be unimpaired by seasonal changes (Figure 6.1, 6.2, 6.3, 6.4). This can be explained by the greater impact of the hydrometeors rain, snow, liquid cloud water and cloud ice in the Clark+-parametrization. In winter, the formation of fog is more likely due to the low saturation humidity. Moreover, the forecast of precipitation, another parameter that impairs visibility, is more skilful in winter [27]. This is suspected to be also the cause for the different effects of the perturbations observed in the case studies (Figure 6.12). Furthermore, the case studies give reason to believe that if the perturbation scheme is adapted, an increase of the RPSS of about 0.1 in January can be expected. Regarding Figure 6.8, it can be seen that this would result in the Clark+-parametrization outperforming the Gul-parametrization for most days in winter.

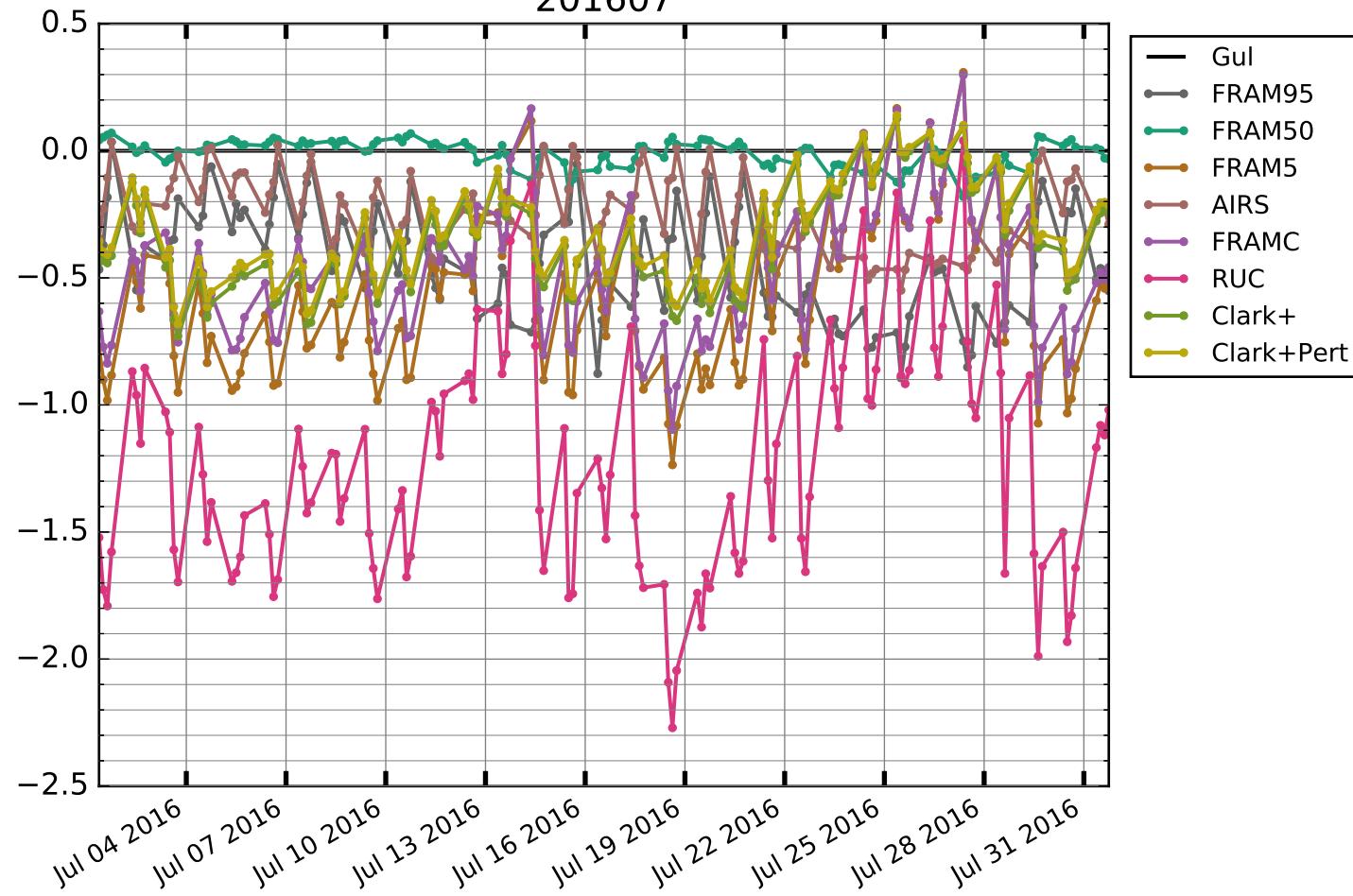
In summer, on the other hand, the clouds form higher in the atmosphere and the concentration of liquid cloud water is decreased correspondingly. In the lower levels, from which we take the input variables for the visibility module, almost no hydrometeors are present, except for rain in some cases. Although there is a dependency on the relative humidity in the Clark+-parametrization, humidity impacts the visibility only by hygroscopic growth of aerosols. But since climatological aerosol data is used, large discrepancies between the aerosol distributions in the model and the currently present aerosol distributions in the atmosphere can arise. Especially near the coast, the concentration of sea salt can vary strongly in time. Organic matter and sulphates are also effected by hygroscopic growth, but their concentrations are more stable in space and time.



**Figure 6.8:** RPSS versus time for January 2017. The RPSS of all parametrizations is shown: the unperturbed ‘Clark+’, the totally perturbed ‘Clark+Pert’ and the seven benchmark parametrizations. The black line, always 0, stands for the ‘Gul’-parametrization, which is the corresponding reference of the RPSS.

### Timelines of RPSS of Different Parametrizations

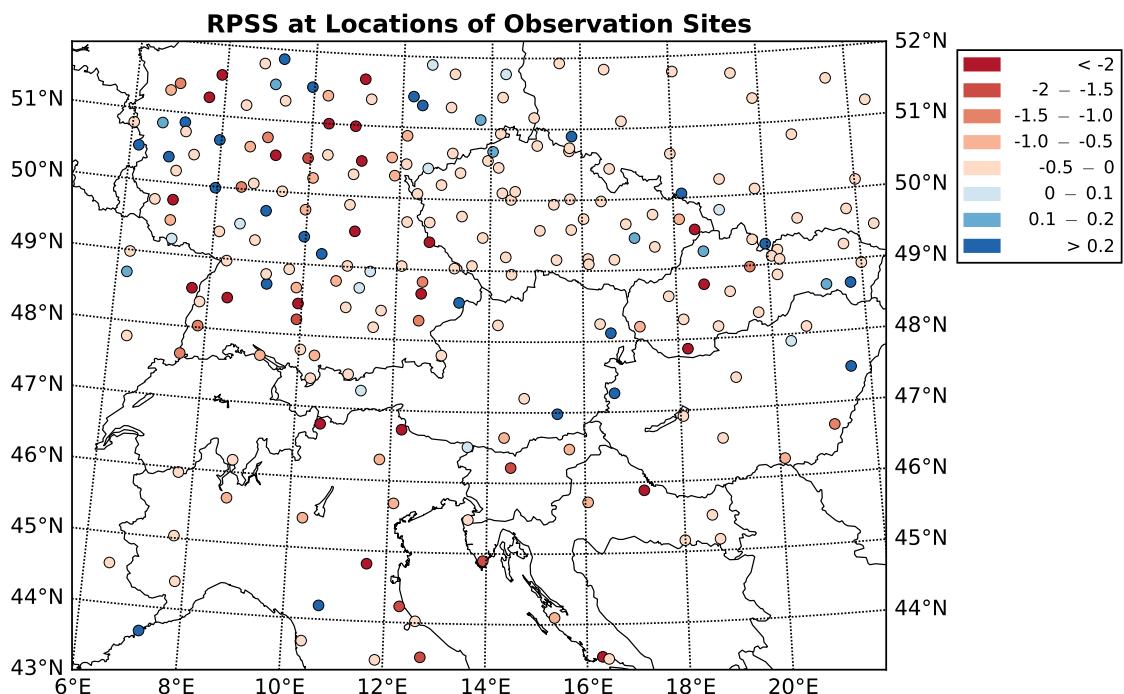
201607



**Figure 6.9:** RPSS versus time for July 2016. The RPSS of all parametrizations is shown: the unperturbed ‘Clark+’, the totally perturbed ‘Clark+Pert’ and the seven benchmark parametrizations. The black line, always 0, stands for the ‘Gul’-parametrization, which is the corresponding reference of the RPSS.

## Geographical Bias on the Forecast Performance

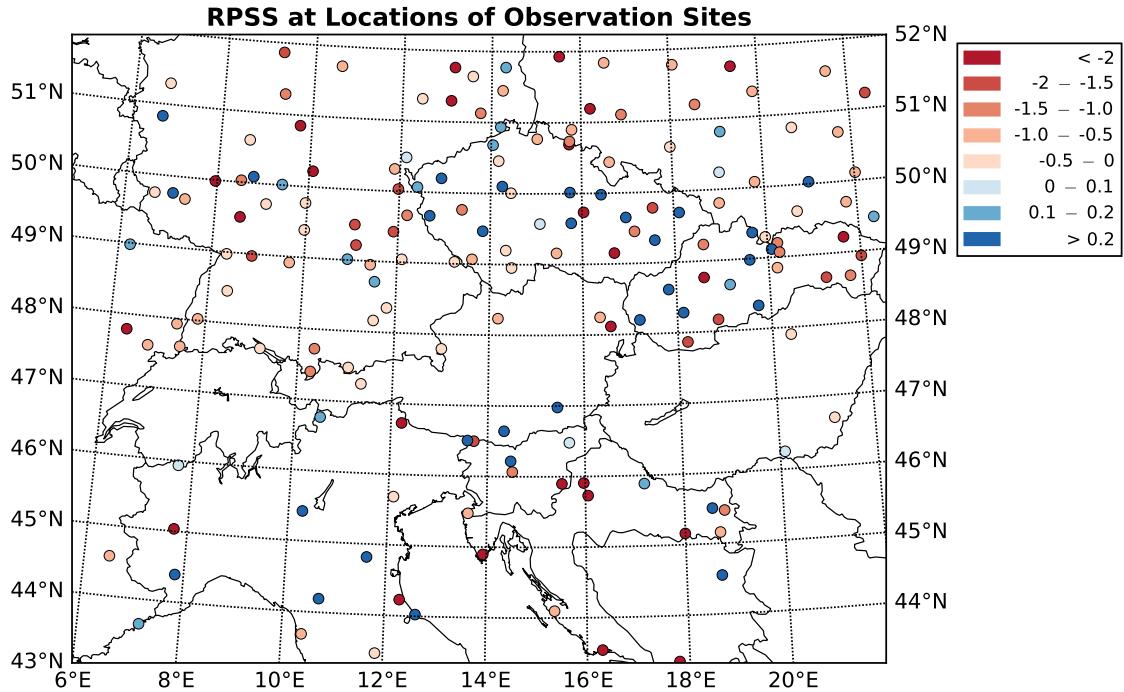
When looking at Figure 6.8 and 6.9, it is evident that for some locations the chances of a skilful forecast are higher than for others. Especially along the Adriatic coast, the Clark+-parametrization performs poorly regarding to its RPSS. This is probably, because the Gul-parametrization has a direct dependence on relative humidity, which is a key parameter close to the sea. As mentioned in the discussion of the seasonal bias, sea salt concentration strongly varies, what mainly affects coastal areas. While this is the case along the coast it is noteworthy that locations in land-locked countries like in Austria or the Czech republic profit of the more complex Clark+-parametrization.



**Figure 6.10:** RPSS of the Clark+-parametrization marked and colour-coded by value on the considered observation sites on the Austrian AROME domain for July 2016. Only observations sites for which three flawless measurements per day were available were taken into account, and are therefore shown on this plot.

## Evaluation of the Perturbation Scheme

The impact of the different variables can be investigated by analysis of the graphs in Figure 6.12. It shows the different RPSS results of the Clark+-parametrization, for the case studies, where different variables and parameters were perturbed, and also the control ensemble, where no additional perturbation was applied. Especially for the test cases in January, significant improvements can be achieved by using a suitable perturbation. But when analysing the data, we find that in some cases the added perturbation can result in a worse skill, than the one of the control ensemble. The exclusive perturbation of the aerosol concentrations ‘AerPert’ is clearly superior to all others on the two days in winter. This was probably induced by very stable



**Figure 6.11:** RPSS of the Clark+-parametrization marked and colour-coded by value on the considered observation sites on the Austrian AROME domain for January 2017. Only observations sites for which three flawless measurements per day were available were taken into account, and are therefore marked.

weather conditions that led to a better than average predictability of the model. Then, the uncertainties of forecast specific humidity and hydrometeor concentrations go down, but for climatological aerosol data and most constant parameters they remain the same. Figure 6.12 also reveals that in winter the perturbation of the hydrometeors seems to have a bad effect on the ensemble skill, whereas in summer, it causes the ensemble skill to improve. This is due to the fact that the uncertainty of the forecast concentration of hydrometeors undergoes seasonal variations, because of the different scale of precipitation and cloud formation in winter and summer. In summer small-scale convective processes are much more influential to hydrometeor distributions than in winter [27].

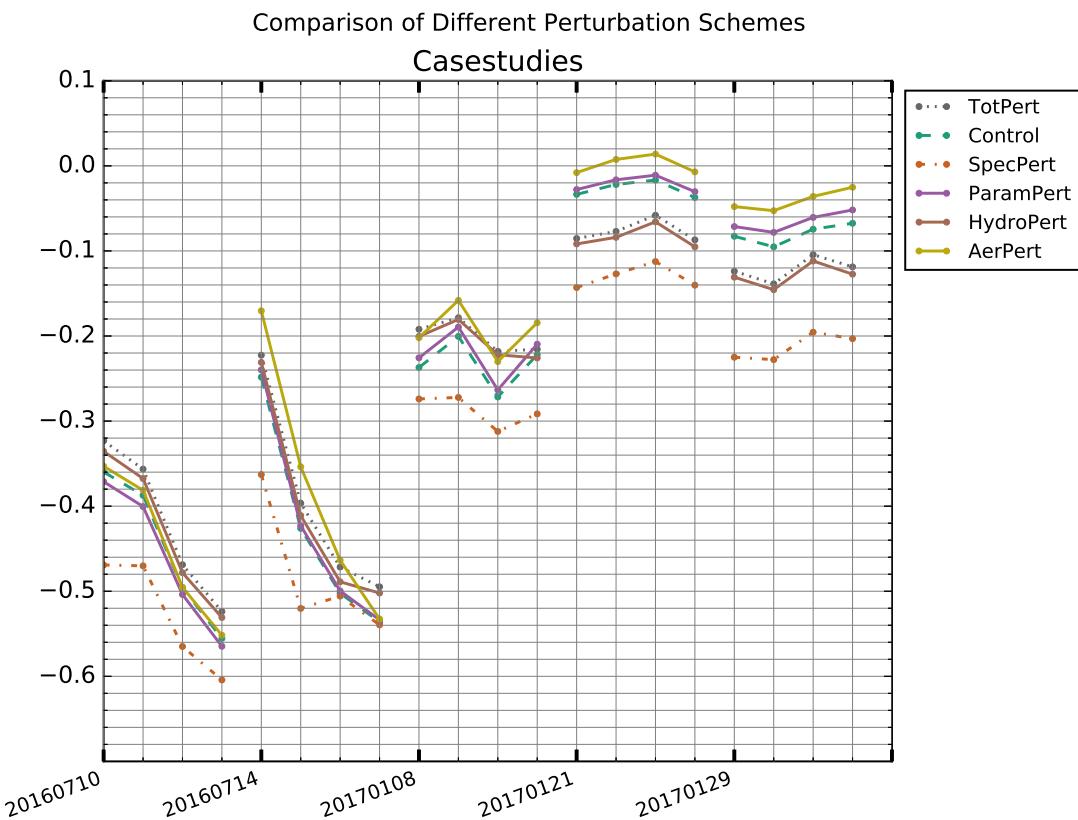
Future studies will likely benefit from further development of the local perturbation scheme by adjusting the amplitude of the perturbation to the present situation differently. For this study monthly climatological extrema were used to set the perturbation amplitude for each month, but the seasonally differing dynamics of precipitation, like its scale and velocities, were neglected. The major difficulty for incorporating them lies within the determination of the adjustment factor, which would require further investigation to be set properly.

Regarding the test cases in July, on the other hand, the totally perturbed ensemble, ‘TotPert’, outperforms all others on average. We suspect that this behaviour is due to a greater contribution of small-scale processes to the overall atmospheric state in summer than in winter, so the increased uncertainties let the model profit from a perturbation [27]. When comparing the different single perturbations with the total

perturbation, the total perturbation seems well balanced, although the perturbation of hydrometeors has clearly the strongest effect, since we can observe a significant correlation between the two ensembles.

The perturbation of the specific humidity always decreased the skill, except for two measurements in July, where it had no measurable effect. This is thought to be the result of an overestimation of uncertainty for the specific humidity. The specific humidity seems to be already forecast quite well by the model, which also leads to the good skill of the Gul-parametrization.

Furthermore, the results presented in Figure 6.12 show, that even a locally applied perturbation scheme can lead to more skilful ensemble forecasts and can be an alternative in cases, where a global perturbation scheme cannot be implemented. Although this was only tested for the visibility parametrization, there is no obvious reason why this should not be generalized for different parametrizations and model components.



**Figure 6.12:** RPSS of the Clark+-parametrization for five dates in July 2016 and January 2017: the label positions of the on the x-axes indicates the skill score corresponding to the 09:00 UTC measurements and the other three following up to 12:00, 15:00 and 18:00 UTC. The x-axes is linear for each day, but the time-frame where no measurements are taken is cut out.

# Chapter 7

## Conclusion and Outlook

Although the overall performance of the established Gul-parametrization is in most cases better than of the newly introduced one, we have reason to believe, that with the general advancement of the model, the explicit Clark+ visibility forecast will be able to outmatch it in the future. On the long run the overall improvement of AROME and in particular prognostic aerosols and advances in data assimilation should provide a decrease of the uncertainties and approximations, which are mainly responsible for the rather low scores of the Clark+-parametrization.

Unfortunately, including data assimilation, would have exceeded the scope of this Master's thesis and was therefore not done. Data assimilation mainly improves forecasts within the first few hours. Since the outputs used in the evaluation were taken the earliest after nine hours of forecast duration, we suspect that data assimilation would not have had any effects. However, for short-range forecasts data assimilation reduces uncertainties and thus, can lead to different results for the skill of the visibility parametrization, because of the increased accuracy of relative humidity and hydrometeor concentrations. It seems likely, that data assimilation would have a major influence on how well a certain perturbation scheme performs too. Especially within the first two hours, the uncertainties of some quantities decrease, whereas of others, e.g. aerosol distribution, they remain the same. As a consequence the ranges of the perturbation patterns will need to be adapted, if data assimilation is used. Testing the effect of data assimilation on the performance in short-range forecasts should be the subject of future studies, to investigate, if the new visibility parametrization can profit from it and might be used in nowcasting systems (NWP within six hours from base time).

The results confirm that the severest weakness of our set-up is that the used aerosol data is climatological. Thus, more exceptional events, e.g. sudden increase of Saharan dust over Europe and the like, are not treated in the model. Also changes in the distribution of sea salt, which heavily contributes to the total aerosol concentration in coastal areas and varies strongly, is neglected. Since the new parametrization relies heavily on aerosols, it is thought to provide significant improvement in performance if a prognostic aerosols scheme would be used. This is further confirmed by the large increase in ensemble skill, when aerosols are included in the perturbation

scheme, as shown in the case studies. Tests with aerosol data from a high resolution aerosol transport model were run and also showed promising results. These tests were carried out by a different implementation of the parametrization in a python program that used AROME output and the mentioned aerosol data as input. The aerosol transport model is based on the weather research and forecasting (WRF) model and has not been coupled with AROME so far. Coupling them could be a way to incorporate high quality aerosol forecasts in the model, but is challenging for the operational mode due to the associated computational costs. For the operational mode the MESO-NH package, which is partially used in AROME and open-source, already includes a simple prognostic aerosol scheme that could be harmonized and calibrated for use in the AROME model in future studies.

A change that is expected to improve the parametrization itself is, to not only consider the forecast visual range at the current location, but also at the adjacent grid points. This would lower the discrepancy of forecast and observation, because the human observers also analyse all directions, when taking measurements. But because of how the model domain is split up for parallelization purposes, we were not able to implement this in the current version. It might be worth considering implementing such a method as part of the post-processing of the output data. Similar techniques are already applied to probabilistic precipitation forecasts [64]

Concerning the perturbation scheme, we found that a local perturbation can be a good strategy to introduce a more realistic spread for model parts or parametrizations, when global perturbations cannot be applied to the needed extent. This is not limited to the forecast of visibility, but can be applied to other forecast variables as well. Especially when the model parts of interest have a large number of different complex dependencies, it is in many cases not feasible to implement a global perturbation of the corresponding quantities. Then local perturbation can help to still be able to obtain an estimate of reliability and the probability distribution of the forecast variable.

Another advantage of a local perturbation scheme is that for diagnostic quantities like visibility it can also be used in post-processing, if the prediction is not contained in the model. The perturbation scheme could also be further improved by fine tuning the set-up and a better pattern generator.

As illustrated and discussed in Chapter 6, the perturbation of specific humidity resulted in a decrease of overall skill. Hence, decreasing the perturbation range drastically or simply excluding specific humidity from the perturbation scheme, is expected to increase the skill of the totally perturbed ensemble. It should also lead to a stronger correlation of the totally perturbed ensemble and the perturbed aerosol and the perturbed parameter ensemble, which seem the most promising indicated by their high overall performance, especially in winter.

The severe shortcomings of the pattern generator, discussed in Section 3.2, are briefly revised: Due to the fact that the namelist settings cannot be directly related to spatial scales, scaling the perturbations is a process of trial and error. Additionally, the cut-off of sampled values distorts the probability density functions.

This could be resolved by re-sampling values that lie outside the range. Solving these issues is challenging, but worth it, since any future experiments that include perturbations would profit from it.

# Bibliography

- [1] C Abbe. ‘The physical basis of long-range weather forecasts’. In: *Monthly Weather Review* 29.12 (1901), pp. 551–561.
- [2] SP Ballard, BJ Wright and BW Golding. ‘Diagnosis of visibility in the UK Met Office mesoscale model and the use of a visibility analysis to constrain initial conditions’. In: *Forecasting Research Scientific Paper* 4 (1992).
- [3] A Bátkai, P Csomós, I Faragó, A Horányi and G Szépszó. *Mathematical Problems in Meteorological Modelling*. Springer, 2016.
- [4] P Bauer, A Thorpe and G Brunet. ‘The quiet revolution of numerical weather prediction’. In: *Nature* 525.7567 (2015), pp. 47–55.
- [5] F Bouttier. *The Météo-France NWP system*. Apr. 2010. URL: <http://www.umr-cnrm.fr/gmap/nwp/nwpreport.html>.
- [6] R Buizza. ‘Potential Forecast Skill of Ensemble Prediction and Spread and Skill Distributions of the ECMWF Ensemble Prediction System’. In: *Monthly Weather Review* 125.1 (1997), pp. 99–119.
- [7] R Buizza, M Milleer and TN Palmer. ‘Stochastic representation of model uncertainties in the ECMWF ensemble prediction system’. In: *Quarterly Journal of the Royal Meteorological Society* 125.560 (1999), pp. 2887–2908.
- [8] R Buizza and T N Palmer. ‘Impact of ensemble size on ensemble prediction’. In: *Monthly Weather Review* 126.9 (1998), pp. 2503–2518.
- [9] A Chandrasekar et al. *Basics of atmospheric science*. PHI Learning Pvt. Ltd., 2010.
- [10] M Chin et al. ‘A global three-dimensional model of tropospheric sulfate’. In: *Journal of Geophysical Research: Atmospheres* 101.D13 (1996), pp. 18667–18690.
- [11] RM Chmielecki and AE Raftery. ‘Probabilistic visibility forecasting using Bayesian model averaging’. In: *Monthly Weather Review* 139.5 (2011), pp. 1626–1636.
- [12] PA Clark et al. ‘Prediction of visibility and aerosol within the operational Met Office Unified Model. I: Model formulation and variational assimilation’. In: *Quarterly Journal of the Royal Meteorological Society* 134.636 (2008), pp. 1801–1816.
- [13] BM Claxton. ‘Using a neural network to benchmark a diagnostic parametrization: the Met Office’s visibility scheme’. In: *Quarterly Journal of the Royal Meteorological Society* 134.635 (2008), pp. 1527–1537.

- [14] J Coiffier. *Fundamentals of numerical weather prediction*. Cambridge University Press, 2011.
- [15] J Du. ‘Uncertainty and ensemble forecast’. In: *Science and Technology Infusion Lecture Series* (2007).
- [16] DR Durran. *Numerical methods for fluid dynamics: With applications to geophysics*. Vol. 32. Springer Science & Business Media, 2010.
- [17] EE Ebert. ‘Ability of a Poor Man’s Ensemble to Predict the Probability and Distribution of Precipitation’. In: *Monthly Weather Review* 129.10 (2001), pp. 2461–2480.
- [18] J-P Eckmann, S Oliffson Kamphorst, D Ruelle and S Ciliberto. ‘Liapunov exponents from time series’. In: *Physical Review A* 34.6 (1986), p. 4971.
- [19] European Centre for Medium-Range Weather Forecasts. *Official IFS Homepage*. 2017. URL: <https://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model/ifs-documentation> (visited on 20/03/2018).
- [20] I Gultepe and JA Milbrandt. ‘Probabilistic parameterizations of visibility using observations of rain precipitation rate, relative humidity, and visibility’. In: *Journal of Applied Meteorology and Climatology* 49.1 (2010), pp. 36–46.
- [21] I Gultepe, MD Müller and Z Boybeyi. ‘A New Visibility Parameterization for Warm-Fog Applications in Numerical Weather Prediction Models’. In: *Journal of Applied Meteorology and Climatology* 45.11 (2006), pp. 1469–1480.
- [22] H Horvath. ‘Atmospheric visibility’. In: *Atmospheric Environment (1967)* 15.10-11 (1981), pp. 1785–1796.
- [23] PL Houtekamer, L Lefavre, J Derome, H Ritchie and HL Mitchell. ‘A System Simulation Approach to Ensemble Prediction’. In: *Monthly Weather Review* 124.6 (1996), pp. 1225–1242.
- [24] HC van de Hulst. *Light scattering by small particles*. Courier Corporation, 1957.
- [25] E Kalnay. *Atmospheric modeling, data assimilation and predictability*. Cambridge university press, 2003.
- [26] M Kerker. *The scattering of light and other electromagnetic radiation*. Elsevier, 2016.
- [27] C Kidd et al. ‘Intercomparison of High-Resolution Precipitation Products over Northwest Europe’. In: *Journal of Hydrometeorology* 13.1 (2012), pp. 67–83.
- [28] A Knauf. *Mathematical Physics: Classical Mechanics*. Springer, 2018.
- [29] H Koschmieder. ‘Theorie der horizontalen Sichtweite’. In: *Beitrage zur Physik der freien Atmosphare* (1924), pp. 33–53.
- [30] BA Kunkel. ‘Parameterization of droplet terminal velocity and extinction coefficient in fog models’. In: *Journal of Climate and applied meteorology* 23.1 (1984), pp. 34–41.

- [31] JP Lafore et al. ‘The Meso-NH atmospheric simulation system. Part I: Adiabatic formulation and control simulations’. In: *Annales Geophysicae*. Vol. 16. 1998, pp. 90–109.
- [32] N Lang-Yona et al. ‘Interaction of internally mixed aerosols with light’. In: *Physical Chemistry Chemical Physics* 12.1 (2010), pp. 21–31.
- [33] P Laven. *MiePlot (A computer program for scattering of light from a sphere using Mie theory and the Debye series)*. 2011. URL: <http://www.philiplaven.com/mieplot.htm>.
- [34] CE Leith. ‘Theoretical skill of Monte Carlo forecasts’. In: *Monthly Weather Review* 102.6 (1974), pp. 409–418.
- [35] M Leutbecher and TN Palmer. ‘Ensemble forecasting’. In: *Journal of Computational Physics* 227.7 (2008), pp. 3515–3539.
- [36] C Lioussse et al. ‘A global three-dimensional model study of carbonaceous aerosols’. In: *Journal of Geophysical Research: Atmospheres* 101.D14 (1996), pp. 19411–19432.
- [37] E Lorenz. ‘Deterministic nonperiodic flow’. In: *Journal of the atmospheric sciences* 20.2 (1963), pp. 130–141.
- [38] E Lorenz. ‘The butterfly effect’. In: *WORLD SCIENTIFIC SERIES ON NON-LINEAR SCIENCE SERIES A* 39 (2000), pp. 91–94.
- [39] European Centre for Medium-Range Weather Forecasts. ‘PART V: ENSEMBLE PREDICTION SYSTEM’. In: *DOCUMENTATION-Cy41r2, IFS* (2016).
- [40] Météo-France. *Official Homepage of the Aladin Consortium*. 2017. URL: <http://www.umr-cnrm.fr/aladin/> (visited on 31/01/2018).
- [41] G Mie. ‘Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen’. In: *Annalen der Physik* 330.3 (1908), pp. 377–445.
- [42] NASA Goddard Institute of Space Studies. *NASA aerosol transport modelling*. 2013. URL: [https://gacp.giss.nasa.gov/data\\_sets/transport/](https://gacp.giss.nasa.gov/data_sets/transport/) (visited on 20/01/2018).
- [43] P Ollinaho et al. ‘Towards process-level representation of model uncertainties: stochastically perturbed parametrizations in the ECMWF ensemble’. In: *Quarterly Journal of the Royal Meteorological Society* 143.702 (2017), pp. 408–422.
- [44] TN Palmer. ‘Predicting uncertainty in forecasts of weather and climate’. In: *Reports on Progress in Physics* 63.2 (2000), p. 71.
- [45] TN Palmer et al. ‘Ensemble prediction’. In: *Proceedings of the ECMWF Seminar on Validation of models over Europe*. Vol. 1. 1993.
- [46] TN Palmer et al. ‘Stochastic parametrization and model uncertainty’. In: *ECMWF Tech. Memo* 598 (2009), pp. 1–42.
- [47] NA Phillips. ‘On the problem of initial data for the primitive equations’. In: *Tellus* 12.2 (1960), pp. 121–126.

- [48] H Poincaré. ‘Sur une forme nouvelle des équations du problème des trois corps’. In: *Acta Mathematica* 21.1 (1897), p. 83.
- [49] JD Price. ‘An advanced model for predicting clear-air visibility’. In: *Meteorological Applications* 14.4 (2007), pp. 435–439.
- [50] W Raith. *Erde und Planeten, Band 7 der Reihe Bergmann-Schaefer: Lehrbuch der Experimentalphysik*. Walther de Gruyter, Berlin-New York, 2001.
- [51] M Shekar Reddy et al. ‘Estimates of global multicomponent aerosol optical depth and direct radiative perturbation in the Laboratoire de Meteorologie Dynamique general circulation model’. In: *Journal of Geophysical Research: Atmospheres* 110.D10 (2005).
- [52] S Roquelaure, R Tardif, S Remy and T Bergot. ‘Skill of a Ceiling and Visibility Local Ensemble Prediction System (LEPS) according to Fog-Type Prediction at Paris-Charles de Gaulle Airport’. In: *Weather and Forecasting* 24.6 (2009), pp. 1511–1523.
- [53] SA Rutledge and P Hobbs. ‘The mesoscale and microscale structure and organization of clouds and precipitation in midlatitude cyclones. VIII: A model for the “seeder-feeder” process in warm-frontal rainbands’. In: *Journal of the Atmospheric Sciences* 40.5 (1983), pp. 1185–1206.
- [54] Y Seity et al. ‘The AROME-France convective-scale operational model’. In: *Monthly Weather Review* 139.3 (2011), pp. 976–991.
- [55] V Shemyakin and H Haario. ‘Online identification of large-scale chaotic system’. In: *Nonlinear Dynamics* (2018).
- [56] TG Smirnova, SG Benjamin and JM Brown. ‘Case study verification of RUCS/MAPS fog and visibility forecasts’. In: *Conference on Aviation, Range, and Aerospace Meteorology, 9 th, Orlando, FL*. 2000, pp. 31–36.
- [57] DJ Stensrud, J-W Bao and TT Warner. ‘Using Initial Condition and Model Physics Perturbations in Short-Range Ensemble Simulations of Mesoscale Convective Systems’. In: *Monthly Weather Review* 128.7 (2000), pp. 2077–2107.
- [58] MT Stoelinga and TT Warner. ‘Nonhydrostatic, mesobeta-scale model simulations of cloud ceiling and visibility for an East Coast winter precipitation event’. In: *Journal of Applied Meteorology* 38.4 (1999), pp. 385–404.
- [59] I Tegen and I Fung. ‘Contribution to the atmospheric mineral aerosol load from land surface modification’. In: *Journal of Geophysical Research: Atmospheres* 100.D9 (1995), pp. 18707–18726.
- [60] I Tegen et al. ‘Contribution of different aerosol species to the global aerosol extinction optical thickness: Estimates from model results’. In: *Journal of Geophysical Research: Atmospheres* 102.D20 (1997), pp. 23895–23915.
- [61] P Termonia et al. ‘The ALADIN System and its Canonical Model Configurations AROME and ALARO’. In: *Annales Geophysicae*. Vol. 16. 1. Springer. 1997, pp. 90–109.

- [62] P Termonia et al. ‘The ALADIN System and its Canonical Model Configurations AROME CY41T1 and ALARO CY40T1’. In: *Geoscientific Model Development Discussions* 2017 (2017), pp. 1–45.
- [63] The Meteo Company B.V. *online archive of satellite images*. 2018. URL: [www.sat24.com](http://www.sat24.com) (visited on 06/04/2018).
- [64] SE Theis, A Hense and U Damrath. ‘Probabilistic precipitation forecasts from a deterministic model: a pragmatic approach’. In: *Meteorological Applications* 12.3 (2005), pp. 257–268.
- [65] U.S. Geological Survey. *Illustration of a conic conformal Lambert projection*. 2008. URL: <http://erg.usgs.gov/isb/pubs/MapProjections/projections.html> (visited on 06/05/2008).
- [66] S Vannitsem. ‘Predictability of large-scale atmospheric motions: Lyapunov exponents and error dynamics’. In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 27.3 (2017), p. 032101.
- [67] A Vincent. *CRM Proceedings and Lecture Notes: Numerical Methods in Fluid Mechanics*. Vol. 16. Centre de Recherche Mathématique, Université de Montréal, 1998.
- [68] JM Wallace and PV Hobbs. *Atmospheric science: an introductory survey*. Vol. 92. Academic press, 2006.
- [69] PJ Webster. ‘Meteorology: Improve weather forecasts for the developing world’. In: *Nature* 493.7430 (2013), p. 17.
- [70] DS Wilks. *Statistical methods in the atmospheric sciences*. Vol. 100. Academic press, 2011.
- [71] World Meteorological Organization. ‘Chapter 9 - Measurement of visibility’. In: *Guide to Meteorological Instruments and Methods of Observation* 8 (2014), pp. 291–309.
- [72] T Wriedt. ‘Light scattering theories and computer codes’. In: *Journal of Quantitative Spectroscopy and Radiative Transfer* 110.11 (2009), pp. 833–843.
- [73] W Zdunkowski, T Trautmann and A Bott. *Radiation in the atmosphere: a course in theoretical meteorology*. Cambridge University Press, 2007.
- [74] Zentralanstalt für Meteorologie und Geodynamik. *Official homepage and online archive of surface weather analysis*. 2018. URL: [www.zamg.ac.at](http://www.zamg.ac.at) (visited on 06/04/2018).
- [75] B Zhou, J Du, I Gultepe and G Dimego. ‘Forecast of low visibility and fog from NCEP: Current status and efforts’. In: *Pure and applied geophysics* 169.5–6 (2012), pp. 895–909.

## **Statement of Originality**

I, Magdalena Elisabeth Haselsteiner, born on 27.05.1992 in Korneuburg, hereby declare in front of Freie Universität Berlin that I have written the present Master's thesis independently without the use of sources and aids other than those specified. The present work has not been submitted to any other university in this or similar form and has not been published yet.

Magdalena Elisabeth Haselsteiner

## Acknowledgements

Firstly, I want to express my sincere gratitude to all members of the ‘model development’-group at ‘Zentralanstalt für Meteorologie und Geodynamik’ and in particular to my scientific supervisor Dr. Clemens Wastl. Everyone in the department was incredibly patient, especially with my initial lack of knowledge in meteorology, offering insightful explanations and great support.

Florian Weidle and Florian Meier helped me a lot to get on with the peculiarities of using the high performance computer at the ECMWF and Josef Kemetmüller was never tired of demystifying the underlying mathematics of the model for me. Moreover, I am grateful to have had the opportunity to profit from Jasmin Vural’s experience and her advice on scientific writing. Thanks to them, but also to Christoph Wittmann and Clemens Wastl, I felt welcome at the department from day one. Dr. Wastl’s door was always open and he helped me to stay on track during my research phase, without interfering too much, letting me create a work based on my own ideas.

I would also like to acknowledge Prof. Roland Netz at Free University of Berlin and Mag. Christoph Wittmann and Dr. Yong Wang at ZAMG for making this cooperation possible and letting me carry out this thesis in the form as it is.

Finally, I want to thank my family for their support throughout my studies, especially my brother Andreas Haselsteiner. He was not only always willing to discuss the topics of my research, but also to help in times of struggle of more personal nature.

Special thanks goes to all my dear friends, in particular to Clemens Heistracher and Christa Lichtenberger, who have always been a source of encouragement, even when being on the other side of the globe. Without you, my studies, including working on this thesis, wouldn’t have been half as much fun and as successful, thanks to our shared studying sessions and the sometimes badly needed distractions.

# Appendix A

## Tables

### A.1 Set-Up of Perturbation

Perturbation settings			
Variable	spatial correlation [km]	distribution	scaling factor
$c_{ice}$	60000	lognormal	0.25
$c_{snow}$	60000	lognormal	0.25
$c_{rain}$	60000	lognormal	0.25
$c_{lcw}$	60000	lognormal	0.25
$c_{om}$	200000	lognormal	Table A.2
$c_{bc}$	200000	lognormal	Table A.2
$c_{ss}$	200000	lognormal	Table A.2
$c_{sul}$	200000	lognormal	Table A.2
$c_{dd}$	200000	lognormal	Table A.2
$q_{spec}$	60000	normal	2.00
$\alpha_{om}$	60000	normal	1.00
$\alpha_{bc}$	60000	normal	1.00
$\alpha_{ss}$	60000	normal	1.00
$\alpha_{sul}$	60000	normal	1.00
$\alpha_{dd}$	60000	normal	1.00
$\sigma_{om}$	60000	normal	0.1
$\sigma_{bc}$	60000	normal	0.1
$\sigma_{ss}$	60000	normal	0.1

$\sigma_{\text{sul}}$	60000	normal	0.1
$\sigma_{\text{dd}}$	60000	normal	0.1
$c_{0,\text{rain}}$	60000	normal	0.20
$c_{0,\text{lcw}}$	60000	normal	0.20
$c_{0,\text{ice}}$	60000	normal	0.20
$c_{0,\text{snow}}$	60000	normal	0.20
$c_{1,\text{rain}}$	60000	normal	0.20
$c_{1,\text{lcw}}$	60000	normal	0.20
$c_{1,\text{ice}}$	60000	normal	0.20
$c_{1,\text{snow}}$	60000	normal	0.20

**Table A.1:** Namelist settings defining the patterns, constructed by the pattern generator, of the perturbed quantities. The spatial correlation provided, is only a namelist setting and has a complex relation to the actual spatial scale as described in Section 3.2. It cannot be directly related with a length scale, because it is scaled with respect to the Earth’s radius, since it was originally set-up for a global model.

Month	Scaling Factors of Aerosol Perturbation											
	1	2	3	4	5	6	7	8	9	10	11	12
Sulphate	1.3357	1.5475	2.1818	2.8114	2.7717	3.1702	3.6221	3.4091	3.6193	3.0257	3.9347	4.4281
Organic Matter	4.3013	4.2449	4.0527	3.6794	4.3204	4.4870	3.7650	3.3974	3.0685	3.3869	3.5936	4.0124
Sea Salt	3.6922	3.3340	2.2834	2.0473	1.5699	1.4266	1.8392	1.7394	2.3499	2.7272	3.5206	3.1504
Desert Dust	3.4771	2.2858	2.5117	3.4750	4.1395	4.0536	3.5781	3.6066	3.9279	3.1142	2.4776	2.4406
Black Carbon	5.2838	5.3458	5.0817	4.7262	5.4137	5.6549	4.7267	4.2958	3.9782	4.2889	4.5131	4.9752

**Table A.2:** Scaling factors of the different aerosols. Set with respect to their global maxima

## A.2 Variables and Abbreviations

Used Variables		
Variable	Name	Unit
$A$	non-parametrized tendency	varies
$a_{0,i}, a_{1,i}$	surface volume coefficients	[m <sup>2</sup> kg <sup>-1</sup> ], [none]
$a_n^s, b_n^s$	Mie coefficients	[none]
$B_a$	brightness of the object	[cd m <sup>-2</sup> ]
$B_b$	brightness of the background	[cd m <sup>-2</sup> ]
$B_s$	brightness of scattered light	[cd m <sup>-2</sup> ]
$c_{bc}$	mass concentration of black carbon	[μg kg <sup>-1</sup> ]
$C_{ext}$	extinction cross section	[m <sup>2</sup> ]
$c_j$	constant parameters	varies
$c_{ice}$	mass concentration of ice	[kg kg <sup>-1</sup> ]
$c_{lcw}$	mass concentration of liquid cloud water	[kg kg <sup>-1</sup> ]
$C$	Contrast	[none]
$C_p$	specific heat of air for constant pressure	[kg m <sup>2</sup> K <sup>-1</sup> s <sup>-2</sup> ]
$c_{snow}$	mass concentration of snow	[kg kg <sup>-1</sup> ]
$c_{rain}$	mass concentration of rain	[kg kg <sup>-1</sup> ]
$c_{om}$	mass concentration of organic matter	[μg kg <sup>-1</sup> ]
$c_{ss}$	mass concentration of sea salt	[μg kg <sup>-1</sup> ]
$c_{sul}$	mass concentration of sulphates	[μg kg <sup>-1</sup> ]
$c_{dd}$	mass concentration of desert dust	[μg kg <sup>-1</sup> ]
$E$	electric field vector	[V m <sup>-1</sup> ]
$\vec{e}_r$	radial unit vector	[none]
$e_j$	prognostic model variabels	varies
$\vec{F}$	dissipative forcing	[kg m <sup>-1</sup> s <sup>-2</sup> ]
$g$	gravitational constant	[kg m <sup>-1</sup> s <sup>-2</sup> ]
$G$	geometric factor	[m <sup>2</sup> ]
$L_\Psi$	spatial correlation length	[km]

$m$	zonal wave number	[none]
$M$	source/sink term for humidity	[kg kg <sup>-1</sup> s <sup>-1</sup> ]
$n$	total wave number	[none]
$T$	temperature	[° K]
$T_x$	tendency of x	varies
$p$	pressure	[Pa]
$P$	parametrized tendency	varies
$P_x$	perturbation of x	varies
$Q$	heat	[J]
$Q_{ext}$	extinction coefficient	[none]
$q_{rel}$	relative humidity	[none]
$q_{spec}$	specific humidity	[kg kg <sup>-1</sup> ]
$r$	particle radius	[m]
$\vec{r}$	radial vector	[m]
$r_j$	random number	[none]
$R$	gas constant	[kg m <sup>2</sup> s <sup>-2</sup> K <sup>-1</sup> mol <sup>-1</sup> ]
$R_E$	average radius of the Earth	[km]
RPS	Ranked Probability Score	[none]
RPSS	Ranked Probability Skill Score	[none]
$s$	scaling factor	[none]
SS	Skill Score	[none]
$t$	time	[sec]
$\vec{v}$	wind velocity	[m s <sup>-1</sup> ]
$\vec{V}$	horizontal wind vector	[m s <sup>-1</sup> ]
$vv_{score}$	model score	[km]
$vis_{o_n}$	observed visibility value	[km]
$vis_{f_n}$	forecast visibility value	[km]
$z$	height	[m]
$w$	vertical wind component	[m s <sup>-1</sup> ]

$\alpha$	scattering size parameter	[none]
$\alpha_j$	attenuation coefficient of aerosols	[m <sup>2</sup> g <sup>-1</sup> ]
$\beta_{a,j}$	atmospheric extinction coefficient of different aerosols	[km <sup>-1</sup> ]
$\beta_{h,i}$	atmospheric extinction coefficient of different hydrometeors	[km <sup>-1</sup> ]
$\beta_R$	atmospheric extinction coefficient by Rayleigh scattering	[km <sup>-1</sup> ]
$\beta_{\text{tot}}$	atmospheric extinction coefficient	[m <sup>-1</sup> ]
$\varepsilon$	liminal contrast	[none]
$\kappa$	Poisson constant	[mol <sup>-1</sup> ]
$\rho$	density of air	[kg m <sup>-3</sup> ]
$\rho_{dry}$	density of dry air	[kg m <sup>-3</sup> ]
$\sigma_j$	scattering cross section of aerosols	[m <sup>2</sup> ]
$\lambda$	wavelength	[m]
$\vec{\lambda}$	Lyapunov vector	[s <sup>-1</sup> ]
$\lambda_j$	Lyapunov exponent	[s <sup>-1</sup> ]
$\lambda_{\text{lon}}$	longitude	[°]
$\phi$	latitude	[°]
$\Phi$	geopotential	[m <sup>2</sup> s <sup>-2</sup> ]
$\Phi_t$	time dependent phase factor	[none]
$\vec{\Omega}$	angular velocity	[none]
$\vec{\omega}$	vorticity	[s <sup>-1</sup> ]

**Table A.3:** List of used variables

Abbreviation and Symbols	Meaning
AROME	Application of Research to Operations on Mesoscale
ALADIN	Aire Limitée Adaptation Dynamique Développement International
ECMWF	European Centre for Medium-Range Weather Forecasts
EDA	Ensemble Data Assimilation
FFT	Fast Fourier Transformation
IFS	Integrated Forecast System
LAM	Limited Area Model
NWP	Numerical Weather Prediction
UTC	Universal Time, Coordinated
RMSE	Root Mean Squared Error
RMSLog	Root Mean Squared Difference of the Logarithmic values
RPS	Ranked Probability Score
RPSS	Ranked Probability Skill Score
WMO	World Meteorological Organization
WRF	Weather Research and Forecasting
$\Im$	Imaginary part of a complex number
$\Re$	Real part of a complex number

**Table A.4:** List of abbreviations and symbols in this thesis

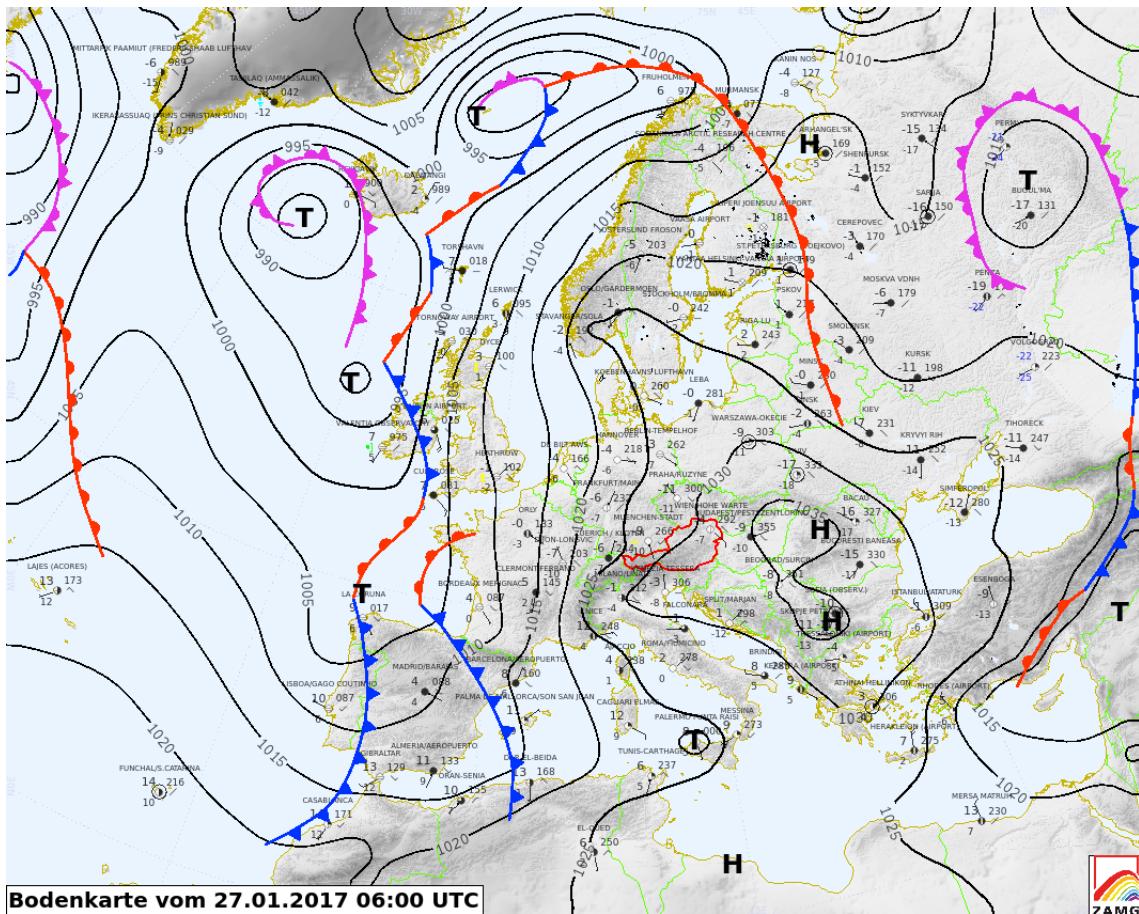
<b>SYNOP key</b>	<b>visibility [km]</b>
0	0.00
1	0.10
2	0.20
...	..
49	4.90
50	5.00
51	none
...	..
55	none
56	6.00
57	7.00
...	..
79	29.00
80	30.00
81	35.00
...	..
88	70.00
89	more than 70

**Table A.5:** SYNOP keys: Numeric system used by the human observers to encode visibility

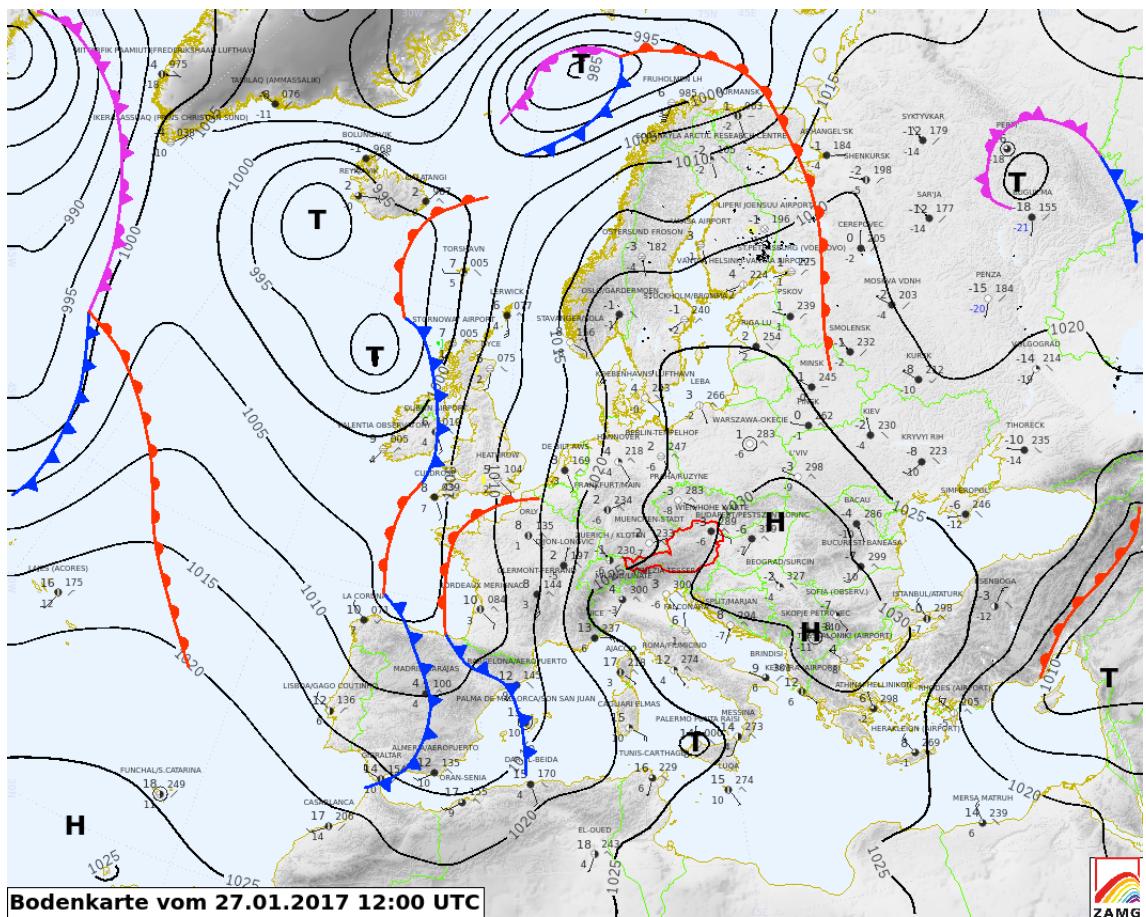
# Appendix B

## Images

### B.1 Surface Weather Analysis



**Figure B.1:** Surface weather analysis 27-01-2017, 06:00 UTC of Europe. Image from the online archive of ‘Zentralanstalt für Meteorologie und Geodynamik’ [74].



**Figure B.2:** Surface weather analysis 27-01-2017, 12:00 UTC of Europe. Image from the online archive of ‘Zentralanstalt für Meteorologie und Geodynamik’ [74].