Evaluation of Cloud Representation of the high-resolution mesoscale model HARMONIE using Meteosat Second Generation observations

Master's Thesis

Koninklijk Nederlands Meteorologisch Instituut
&
Utrecht University

March - November 2016

Author:

Christiaan van Dalum

Supervisors:

Hylke de Vries (KNMI)
Wim de Rooy (KNMI)
Pier Siebesma (KNMI)
Aarnout van Delden (IMAU)

Koninklijk Nederlands Meteorologisch Instituut
Ministerie van Infrastructuur en Milieu

Universiteit Utrecht

Abstract

This study consists of two parts. In the first part we evaluate the cloud cover representation of the high-resolution mesoscale model HARMONIE using MSG (Meteosat Second Generation) observations. We assess the performance of different model versions. We also examine anomalies in time and space. Cloud cover is a difficult to model parameter and is dependent on several processes, like turbulence and convection. Verification of cloud cover in HARMONIE using satellite data has not been done systematically and is therefore needed. Although MSG cloud cover is approximately 4% higher compared with ground based observations, a significant underestimation of cloud cover in HARMONIE above sea is observed. The degree of underestimation is dependent on weather conditions. The latest version of HARMONIE correlates better with MSG than its predecessors. More advanced verification methods show a systematic underestimation of the cloud area and an increase in performance during the night. In the second part, the impact of cloud cover schemes is studied. Subgrid variability of temperature and humidity are important for cloud formation. Parametrizations are needed to link the subgrid variability of temperature and humidity to cloud cover and are called cloud schemes. Two types of cloud schemes are assessed: a relative humidity cloud scheme and a statistical cloud scheme. The relative humidity cloud scheme directly links relative humidity to cloud cover, while the statistical cloud scheme provides a probability distribution of the temperature and humidity around the mean, which can be converted to a fractional cloud cover. Each sophistication of the statistical cloud scheme contributes to better cloud cover forecasts. During the day, sophistications to the cloud scheme have a larger impact than during the night, which is due to the lack of convective and turbulent conditions during the night. Relative humidity cloud schemes perform worse than the statistical cloud schemes. The relative humidity cloud scheme does provide information about possible changes to the statistical cloud scheme.

Acknowledgements

I would like to thank my supervisors Wim de Rooy, Hylke de Vries and Pier Siebesma for their lasting support at the KNMI. I could walk into your offices whenever I had a question and could discuss the considered material in an efficient way. I want to thank you for your contribution to making this project a success. I also thank Mark Savenije in providing regridding scripts in the first few weeks of my research. Furthermore I am grateful to Jan Fokke Meirink for providing the MSG satellite data and providing adequate support. I also want to thank Bert van Ulft for running HARMONIE and helping us out when we had technical problems with the model.

Contents

1 Introduction

Clouds are a vital element of the Earth's weather and climate system. Cloud cover is not just of interest in its own, but also has a major impact on the redistribution of energy, heat and moisture. Clouds reflect the incoming solar radiation back to space. Consequently, the observed weather will be altered significantly by clouds. At night, the outgoing longwave radiation is reduced due to clouds. This impacts local weather and boundary layer formation. Clouds are also highly variable in size, location and time. The representation of clouds in numerical models proves to be one of the most challenging problems in large and mesoscale models (Haiden et al., 2015; Quaas, 2012). The resolution of these models is too coarse to distinguish all processes involved in cloud formation. Hence parametrizations are needed to represent sub grid processes.

Haiden et al. (2015) shows that ECMWF's (European Centre for Medium-Range Weather Forecasts) IFS (Integrated Forecasting System) large scale model has difficulties in producing a correct prediction beyond 24 hours for total cloud cover. The predictability of cloud cover of this model is therefore limited. For a mesocale model like HARMONIE (HIRLAM ALADIN Research on Mesoscale Operational NWP (Numerical Weather Prediction) in Euromed), the resolution is not high enough to calculate all processes involved in cloud formation. Consequently, the problems regarding cloud cover representation are comparable with large scale models and need parametrizations. These parametrizations have to be verified against observations. In this study we use satellite observations.

In the first part of this study we focus on evaluation of cloud representation of the high-resolution mesoscale numerical weather prediction model HARMONIE using Meteosat Second Generation (MSG) satellite observations. We consider different time periods: August 2006, May 2008, March 2012, October 2012 and the year 2012. HARMONIE is an operational weather model based on the ARMOME model (Seity et al., 2011) and further developed by the HIRLAM-ALADIN community (Bengtsson et al., 2017). For this study, different versions of HARMONIE will be tested for Western Europe and verified against MSG.

In the second part of this thesis we study the impact of changes made to the cloud scheme. A cloud scheme is a scheme to compute cloud cover from the turbulence and convection parametrizations, using temperature and humidity. Different methods are possible to link the temperature, humidity, turbulence and convection to cloud cover. In this part we evaluate the impact of these methods on cloud cover.

The remainder of this chapter covers important concepts of cloud formation. We also discuss the impact of clouds in the weather and climate system. Finally, we discuss a project outline including research goals.

1.1 Clouds in the weather and climate system

To be able to understand clouds in the weather and climate system, one should know the formation process. A simple cloud formation process is the adiabatic rise of warm moist air in the boundary layer. The air package has a relative low density and can only rise during unstable conditions. The air is unstable with respect to an air package, if the density of the air package is higher than the density of the environment. The package will continue to rise and cool until it is in equilibrium with its environment. This is called convection. Cold air can contain less moisture than warm air, so whenever the air package reaches the dew point temperature, the parcel is saturated and a cloud will form. The level of cloud formation is called the Lifting Condensation Level (LCL). The LCL can be interpreted as the altitude for which the air package cannot sustain its moisture, resulting in the formation of a cloud. The existence of the LCL can be explained using the Clausius-Clapeyron equation:

$$\frac{de_s}{dT} = \frac{e_s L_v}{R_v T^2}. \tag{1.1}$$

In equation 1.1, L_v and R_v are, respectively, the specific latent heat of evaporation and the gas constant for water vapour, T the temperature and e_s the partial saturation water vapour pressure. Partial pressure can be interpreted as the pressure the considered gas would have if it is the only gas present. When the partial saturation water vapour pressure is high, it means that the air can support a high concentration of water vapour. High humidity levels are reached when the partial water vapour pressure is close to the partial saturation water vapour pressure. Equation 1.1 can be rewritten to:

$$\log e_s - \log e_{s,0} = \frac{L_v}{R_v}\left(\frac{1}{T_0} - \frac{1}{T}\right), \tag{1.2}$$

with $e_{s,0}$ the partial saturation vapour pressure at $T_0 = 0°\text{C}$. Figure 1.1 shows equation 1.2, in which e_s increases when T increases. The figure shows that colder air can support less water vapour, often resulting in the saturation of air. Note in figure 1.1 that the temperature range is below freezing. This is not unusual, due to the high surface tension of water, droplets can be supercooled before freezing.

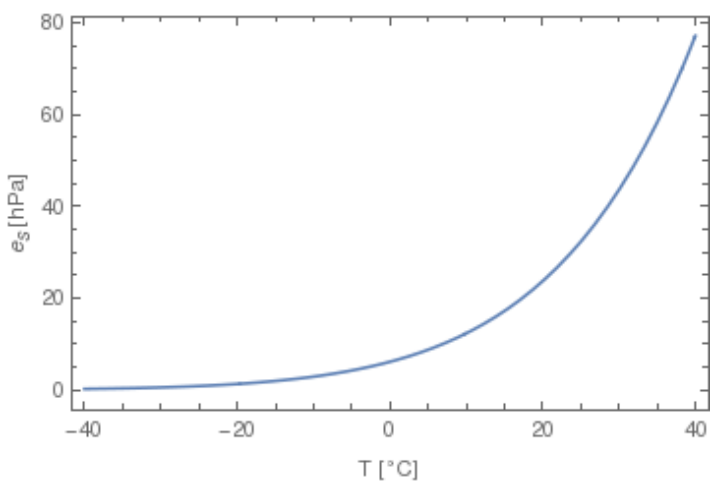


Figure 1.1: Partial saturation water vapour pressure e_s as a function of the temperature T, following equation 1.2.

Clouds do not only form due to convective lift, but also due to orographic and frontal lift. Orographic lift is the process of air lifted towards higher elevations due to rising terrain. Frontal lift occurs if a cold and a warm air mass collide, forcing the warm air mass to rise. This lift may lead to saturation

and cloud formation. The impact of turbulence on cloud formation is not negligible. Turbulence can lead to local temperature and humidity changes. This may lead to local water vapour saturation and cloud formation.

Another process involved in cloud formation is entrainment. Entrainment can be split into lateral and top entrainment. Lateral entrainment is the process of mixing surrounding air into a flow. For example, dry air can be transported into moist regions, reducing the dew point temperature. This eventually leads to changes in the convection pattern. Detrainment is the opposite effect. Detrainment represents the process of air from a flow deposited into the environment. This usually takes place at a high altitude. Entrainment and detrainment may lead to updrafts or downdrafts if the transported air is unstable. Top entrainment is the mixing of air into the boundary layer. Entrainment is difficult to model and results in an increase of uncertainty in total cloud cover. This will have its effect on the radiation budget (Knight et al., 2007).

Figure 1.2 (Wild et al., 2013) shows the global annual mean energy budget between 2001 and 2010, which is used in the IPCC report (Hartmann et al. (2013)). It shows the strong reflective and absorptive capability of clouds for short and longwave radiation. Clouds have an impact on the global annual mean energy budget between 2001 and 2010 by approximately 50 Wm^{-2}. This shows that clouds significantly impact the climate system on Earth. The impact of clouds on incoming solar radiation can be used to measure clouds. This is called the optical thickness, and is defined as the ratio of the incident and transmitted radiation.

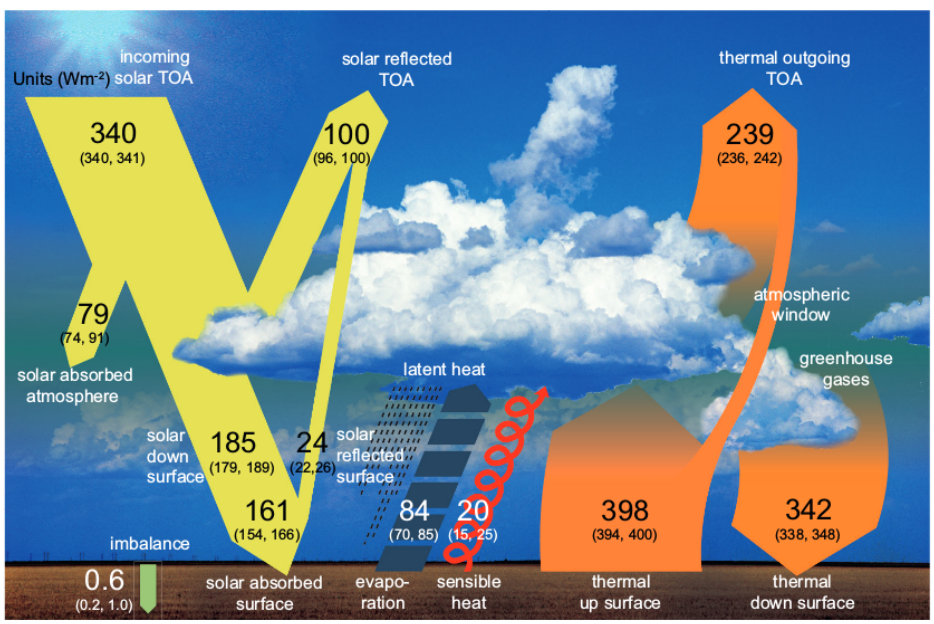


Figure 1.2: The global annual mean energy budget for the period 2001-2010 (Wm^{-2}). The broad arrows indicate the schematic flow of energy in proportion to their importance (Wild et al., 2013).

1.2 Project outline and research goals

Due to many processes involved in cloud formation, it is one of the most difficult parameters to model, and is accompagnied by large uncertainties(Stocker et al., 2013). The verification of cloud cover in numerical weather prediction models is an underdeveloped field of research and is mostly performed against ground based measurements, short time periods or small research areas, like Söhne et al. (2008), and Weniger and Friederichs (2016). An extensive research on cloud cover against satellite data on a mesoscale model for a relatively long period of time is therefore needed. The research goals of this study are:

1) Cloud cover verification of HARMONIE using MSG satellite data.
2) Evaluation and verification of different cloud scheme products.

The first part of this study we focus on the performance of different HARMONIE versions. We test the performance of HARMONIE versions against each other and study anomalies in time and space. We start this analysis with a case study. This is followed by applying some basic statistics to the data sets. We also provide a more in-depth statistical analysis. Finally we assess the model performance in time and space.

The second part consists of testing different cloud scheme products, as well as assessing the contribution of turbulence and convection. We use the statistical methods of the first part on the HARMONIE versions with a new cloud scheme to assess the performance. The goal of this part of the study is to test different parametrizations and to find a possible improvement.

This report consists of the following chapters:
Chapter 2 provides an overview of the observation method of MSG and the accompanying errors. We provide an assessment of the importance of uncertainties. Finally, we give an estimation of the error margin due to MSG to be considered when comparing with HARMONIE.
Chapter 3 contains general information about HARMONIE. We also explain the different versions of HARMONIE as well as the different modes to run the model. Section 3.1 describes the turbulence scheme as is used in HARMONIE. Section 3.2 explains different possible cloud schemes.
Chapter 4 consists of a detailed overview of the methodology. We first describe the choice of considered time periods. We also describe in detail how we merge the grids of HARMONIE and MSG together. We provide a description of the statistics that we use in this study, including basic statistics, correlation coefficients and statistical bootstrapping. We describe the Structure, Amplitude and Location (SAL) method as well.
Chapter 5 shows the results of the first part, including a SAL analysis. The year 2012 is considered in Section 5.8 and 5.9. Furthermore, differences between land and sea are investigated.
Chapter 6 covers the last part of this study, consisting of all experiments done with a new cloud scheme. This chapter also includes the investigation of the contribution of convection and turbulence on cloud cover.
In chapter 7, we wrap up the results and discuss the conclusions.

2 Observations

2.1 Introduction

In the past, the cloud cover representation of HARMONIE has been verified against ground measurements. To further assess the performance of cloud cover forecast in HARMONIE, an extensive analysis using satellite data is needed. In this study, the model output is verified with total cloud cover satellite data generated by the Meteosat Second Generation (MSG) satellites (Schmetz et al., 2002). The quality of the satellite data is dependent on the errors in the measurements. It is important to know the uncertainty of the observational data sets in order to determine if the results of HARMONIE deviate significantly. In this chapter we provide further information about the MSG satellite and present a detailed analysis about uncertainties in the data. From this uncertainty analysis follows an estimation of the uncertainty in the observational data sets that we use in this thesis.

2.2 MSG Cloud Products

The MSG project consists of satellites in geostationary orbit and is produced by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT). The satellites have been specifically designed to deliver data needed for numerical weather predictions as well as for climate research (Legendre et al., 2010; Schmetz et al., 2002). The Spinning Enhanced Visible and InfraRed Imager (SEVIRI) has been mounted on the satellites and measures passively multiple variables including cloud cover in twelve spectral bands, ranging from the visible (0.6 μm) to the infrared spectrum (13.4 μm) (Schulz et al., 2008). During the night, only infrared is used to compute cloud cover. Hence SEVIRI is expected to perform better during the day, when all spectral bands can be used. SEVIRI delivers a full disk image of Europe every 15 minutes with a resolution of 4 by 7 km. This is sent to the Satellite Application Facilities (SAF) and processed at the European Space Operations Centre in Darmstadt, Germany. Apart from SEVIRI, the satellites also carry Geostationary Earth Radiation Budget (GERB) instrument on board. This measures the short and longwave radiation coming from the Earth. SEVIRI's cloud cover data is compared to ground measurements and NASA's spectrometer The Moderate Resolution Imaging Spectroradiometer (MODIS). Differences can be quite large and up to 10% (Finkensieper et al., 2016).

MODIS is an imaging instrument mounted on the sun synchronous Terra and Aqua satellites. Terra passes the equator north to south at 10:30 local time and Aqua passes the equator south to north at 13:30 local time. Time taken to fully scan the surface of the Earth is up to 2 days, in 36 spectral bands. Like SEVIRI, MODIS performs measurements passively. MODIS proves to be stable, has a relatively high resolution of 1 by 1 degree and have been measuring for the complete MSG time period Platnick et al. (2003). For the verification of MSG cloud products, we use the MODIS satellite.

2.3 Observational uncertainties

The measurement angle of SEVIRI is important for the quality of the data. The measurement angle increases with latitude, because the satellites are in geostationary orbit. The bias in figure 2.1 reveals an underestimation of SEVIRI for total cloud cover when compared with ground based measurements for viewing angles $< 40°$ and an overestimation for viewing angles $> 40°$. The statistics in regions with a low viewing angle are not mature due to the lack of ground observations. The vertical dashed lines indicate the considered latitude range of HARMONIE. In this domain, the bias ranges from 0% to +6%. An overestimation of the total cloud cover of SEVIRI is therefore likely.

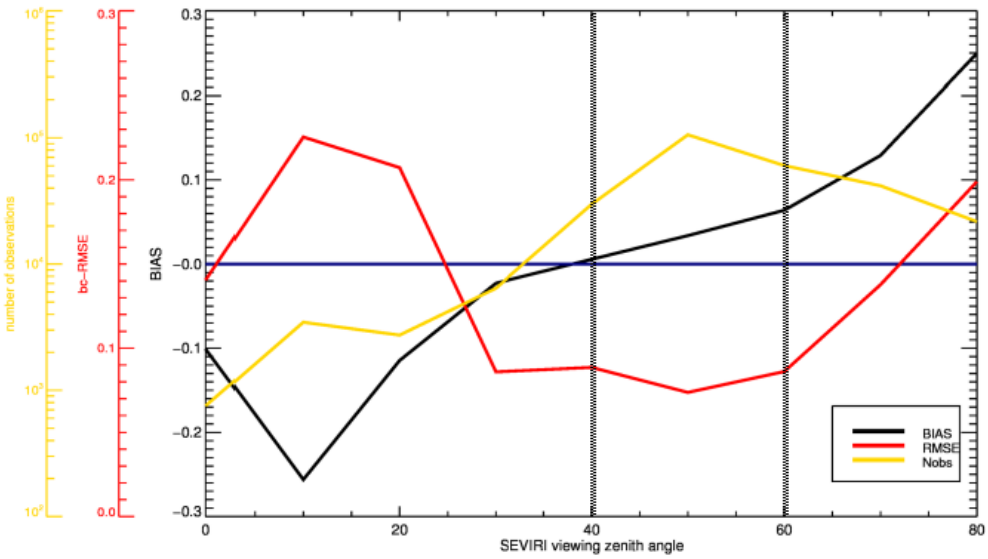


Figure 2.1: Ground observations, bias between SEVIRI and ground observations and root mean square error against the SEVIRI zenith angle (Finkensieper et al., 2016). The vertical lines indicate the considered domain in HARMONIE.

Figure 2.2 presents ground based measurements against SEVIRI cloud cover data (Finkensieper et al., 2016) in a scatter plot and a frequency distribution. It shows the monthly mean cloud cover data between 2004 and 2015. Figure 2.2 shows a clear overestimation of the total cloud cover by SEVIRI during cloudy conditions (cloud fraction of 0.5 - 0.85) when compared with ground observations, with a mean bias of 0.0369. The overestimation of clouds is reduced when overcast conditions are dominant, because ground observations and satellite data provide almost the same results. The overestimation of clouds is also reduced during dominantly clear sky conditions, but with some small clouds present. These small clouds are overestimated, but the impact of these errors are small due to the general lack of clouds. As a consequence, cloud cover conditions in between the two extremes provide the largest overestimation (to be more precise, the largest overestimation corresponds with a cloud fraction of 0.5 - 0.85).

With larger viewing angles, the location error of clouds will increase due to the parallax effect. Figure 2.3 shows the concept. When the satellite measures under an angle, it localizes the cloud as if it is on a different location. It is translated with respect to the actual location of the cloud, which is directly linked to the viewing angle. No correction is applied for this effect in SEVIRI.

SEVIRI struggles with the correct detection of opaque clouds. Opaque clouds are interpreted as overcast and can lead to a strong overestimation by SEVIRI. This overestimation is mostly due to high measurement angles for Western Europe. It results in an overestimation of the optical path

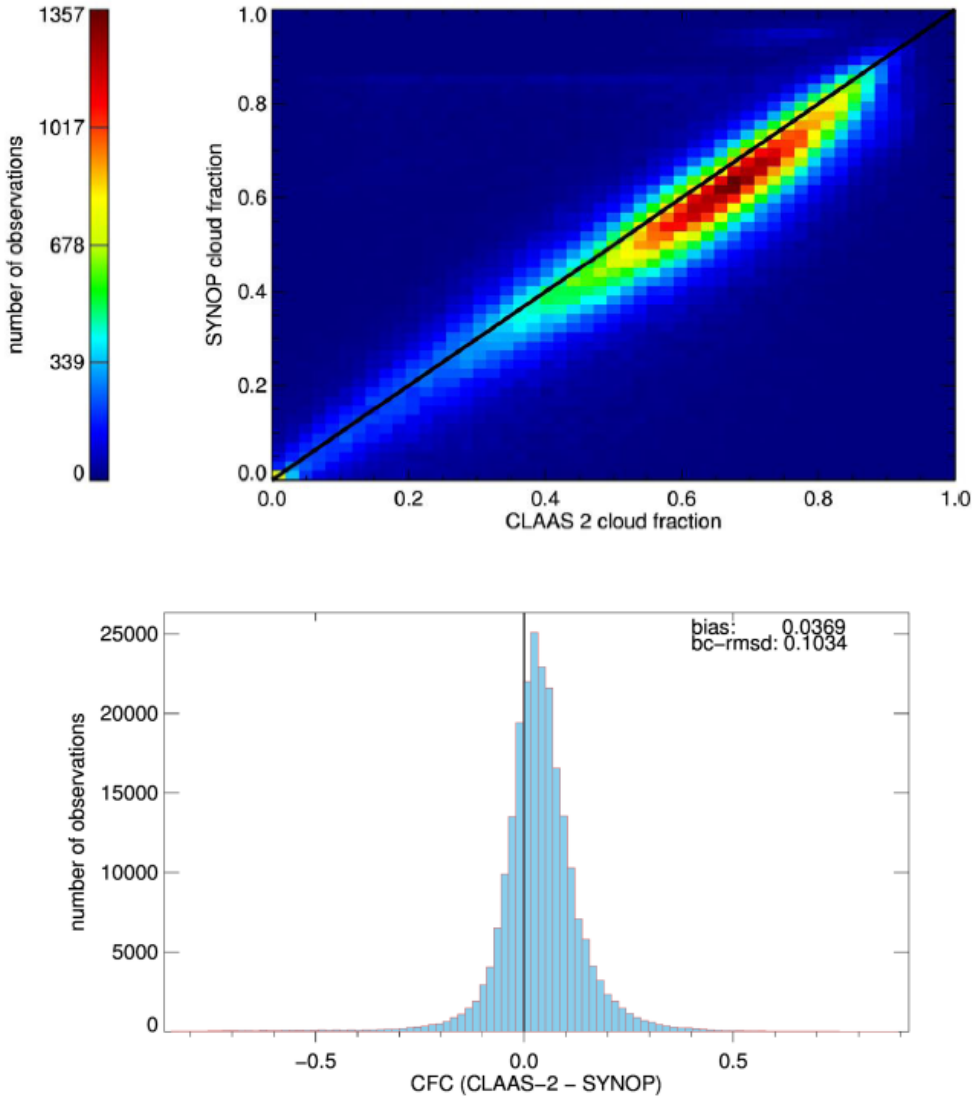


Figure 2.2: Ground observations and SEVIRI cloud cover data in a scatter plot and a frequency distribution (Finkensieper et al., 2016).

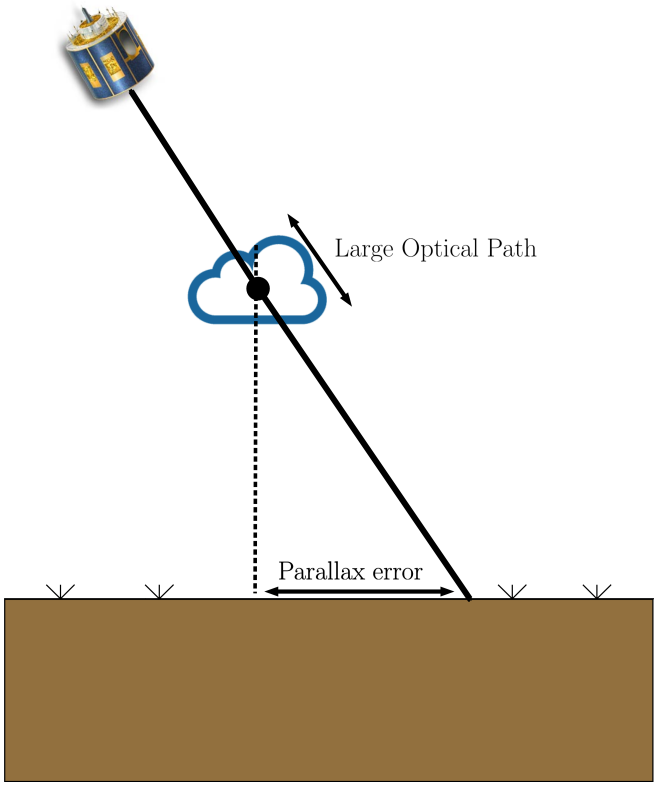


Figure 2.3: Schematic illustration of the parallax error when measuring the location of a cloud and overestimation of the optical path. An increasing measurement angle of the satellite with respect to the cloud will cause a larger error.

and therefore cloud thickness. The upper part of figure 2.3 schematically shows this effect. The quality of cloud detection is not only dependent on the measurement angle, but also on the surface albedo. Whenever the surface albedo is high, clouds are more difficult to detect. The satellite then has difficulties to distinguish clouds and surface, due to the high albedo of both. In Western Europe, this is only a small effect. However, it may become important during extensive snow cover periods.

Figure 2.4 shows the process responsible for the overestimation of the area of clouds during cloudy, but not overcast conditions. Due to a non zero viewing angle, the estimated cloud area is larger than the real area of the cloud. The structure is therefore likely to be overestimated. The impact of the error in the estimated cloud area is dependent on the size of a cloud. The impact is higher for small clouds than for large clouds. For large clouds, the cloud area is extensive and the overestimation is relatively small. During overcast conditions, the error in the area is therefore reduced. For small clouds the overestimated area is relatively large compared with the real area.

The detection of clouds by SEVIRI is based on the brightness temperature. When a threshold is exceeded, a cloud is detected (Derrien et al., 2012). The threshold depends on the brightness temperature, albedo, time of day, land or sea and the viewing angle. The brightness temperature is defined as the temperature the object would have if it is in equilibrium with its surrounding and is a black body radiator with the same intensity as is observed. However, different near-infrared bands and thresholds are used above the ocean to detect clouds. This may lead to differences in quality of observations above land and sea. The differences of detection between land and sea can be quite large (Derrien et al., 2012). Clouds are generally easier to detect above water than land due to the large albedo difference between the surface and the cloud. A detection difference between land and water is ex-

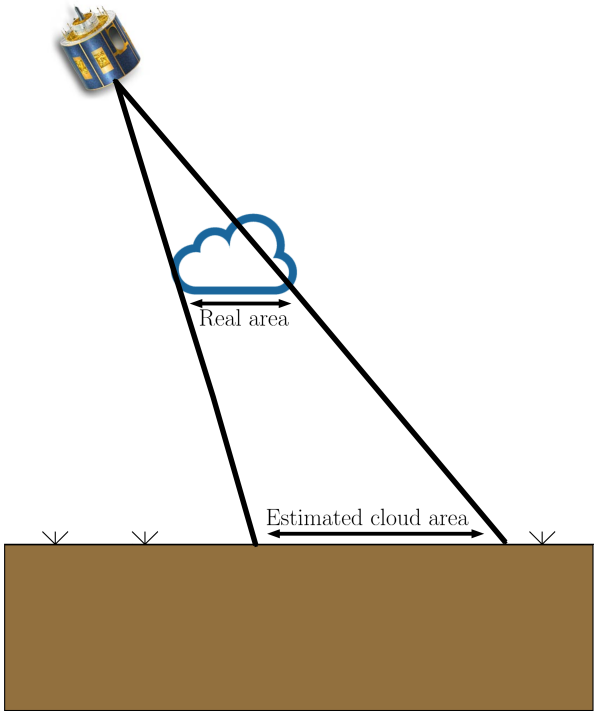


Figure 2.4: Schematic illustration of the structure overestimation of clouds estimated by SEVIRI. The degree of overestimation is linked to the viewing angle.

pected. The separation between snow and ice and water clouds is done by looking at low reflectance in the 1.6 μm and 3.9 μm spectrum. This restriction may be insufficient to distinguish snow and clouds, especially during night (Derrien et al., 2012). The determination of opaque clouds also follows a brightness temperature threshold. This threshold is only applied to cloud contaminated regions.

To summarize, for the considered domain in HARMONIE, cloud cover is overestimated during cloudy conditions. The uncertainties are dependent on the viewing angle. After the analysis of this chapter, we estimate the uncertainty of MSG cloud cover to be approximately +4% against ground observations. In this study we use this 4% uncertainty range. This uncertainty range is mostly caused by the area error, and in a lesser extent the location and amplitude. We have also seen that SEVIRI has difficulties with estimating cloud cover over high albedo surfaces. A difference in quality between day and night is also observed. Cloud detection above land and sea can be significant, especially for high, optically thin clouds.

One should also note that the data provided by SEVIRI is a binary cloud mask, resulting in a 1.0 if a cloud is present and 0.0 if none is present. For Western Europe, SEVIRI provides data on a 4 by 7 km grid. Although SEVIRI performs the relevant cloud measurements, in this study we call the satellite observations simply MSG.

3 Model

In this study, we use different versions of HARMONIE (HIRLAM ALADIN Research on Mesoscale Operational NWP in Euromed). HARMONIE is a high resolution mesoscale model, which we run on the domain of Western Europe. It is a non-hydrostatic model with a resolution of 2.5 by 2.5 km. HARMONIE consists of 65 vertical levels. The thickness of each layer is variable, with relatively thin layers at the surface. In the model, some processes can be calculated directly on the grid. These are called resolved processes. The theoretical relations between quantities can be implemented. One of the resolved processes in the model is large scale convection. Other processes, like turbulence, are unresolved. Unresolved processes take place on a subgrid scale, meaning that they cannot be calculated directly using the equations of motion with the resolution of the model. Parametrizations are needed to incorporate these subgrid scale processes. Parametrizations are simplified representations of the considered process. For parametrizations, resolved quantities are used to approximate unresolved quantities. Subgrid scale processes thus can be estimated using parametrizations. Turbulence is important for cloud cover, but is an unresolved process. Parametrizations are therefore needed to represent turbulence. In the following section we discuss the turbulence parametrization.

3.1 Turbulence scheme

To understand the turbulence scheme implemented in HARMONIE, we start with the potential temperature and moisture equation for a turbulent flow (Stull, 1988). For the full derivation of these type of equations, see appendix A.

$$\frac{\partial \bar{\theta}}{\partial t} + \bar{u}\frac{\partial \bar{\theta}}{\partial x} + \bar{v}\frac{\partial \bar{\theta}}{\partial y} + \bar{w}\frac{\partial \bar{\theta}}{\partial z} = -\frac{\partial \overline{\theta' w'}}{\partial z} + S_\theta, \tag{3.1}$$

$$\frac{\partial \bar{q}}{\partial t} + \bar{u}\frac{\partial \bar{q}}{\partial x} + \bar{v}\frac{\partial \bar{q}}{\partial y} + \bar{w}\frac{\partial \bar{q}}{\partial z} = -\frac{\partial \overline{q' w'}}{\partial z} + S_q. \tag{3.2}$$

The grid boxed averaged potential temperature and moisture are respectively $\bar{\theta}$ and $\bar{q}$ and represent the resolved quantities. The perturbation potential temperature and moisture are θ' and q' and represent the unresolved quantities. Other source and sink terms are represented in S_q and S_θ. The grid boxed averaged horizontal velocity are $\bar{u}$ and $\bar{v}$ and in the vertical $\bar{w}$. The perturbation vertical velocity in the z direction is w'. The advection terms (element 2 - 4 on the left hand side (l.h.s.) in equations 3.1 and 3.2) are resolved by the model. The first term on the right hand side (r.h.s.) of equation 3.1 is the vertical turbulent potential temperature flux and the first term on the r.h.s. of equation 3.2 is the vertical turbulent moisture flux. Both terms are unresolved by the model. For these equations, Reynolds decomposition has been used. For example, θ has been decomposed to:

$$\theta = \bar{\theta} + \theta', \tag{3.3}$$

with $\bar{\theta'} = 0$. Whenever turbulent fluctuations are present, the first term on the r.h.s. in equations 3.1 and 3.2 is non zero. Stable conditions prevail when $\frac{\partial \overline{\theta' w'}}{\partial z} > 0$, what results in cooling. We do

not consider the horizontal convergence of turbulence and heat, because it is often several orders of magnitude smaller than the vertical component. Equations 3.1 and 3.2 have more variables than equations and thus a closure problem occurs. We discuss different closure schemes in the following sections.

3.1.1 First order closure scheme

The first order closure is often used to close relations of the form of equations 3.1 and 3.2 (Perov and Gollvik, 1996). The first order closure scheme directly links turbulent fluxes to the mean gradient of the considered quantity. Vertical turbulent fluxes are then described by:

$$\overline{\phi' w'} = -K \frac{\partial \bar{\phi}}{\partial z}. \tag{3.4}$$

The considered quantity is ϕ and K is the eddy viscosity coefficient with units m^2/s. The term on the r.h.s. of equation 3.4 represents the resolved part and is linked to the unresolved part on the l.h.s.. The vertical turbulent flux is now directly dependent on a resolved quantity. The parameter K can be considered as a length scale (L) times a vertical velocity scale (W):

$$K = W \cdot L \tag{3.5}$$

The value of K depends on the considered variable and atmospheric conditions. If $K > 0$ and $\frac{\partial \bar{\phi}}{\partial z} < 0$ it results in $\bar{\phi' w'} > 0$, what means that the turbulent flux is flowing down-gradient. In the atmospheric boundary layer (the layer for which the atmosphere experiences friction from the bottom), the first order K-theory leads to the often observed Ekman spiral (Ekman, 1905).

The quality of the first order K-theory is undermined by its basic assumptions made (Perov and Gollvik, 1996). First of all, turbulence is assumed to be in equilibrium. Hence a time memory of the turbulent structure is not included. The turbulence is also generated locally and cannot be transported to other regions with advection. A higher order closure scheme is able to solve some of the problems occurring in the first order closure scheme. In the next section we therefore focus on the one-and-a-half order closure scheme.

3.1.2 One-and-a-half order closure scheme

The difference between the one-and-a-half order closure scheme and the first order scheme, is that it not only retains the zero-order variables, but also the equations for the variance of the most relevant quantities. It is not a second order closure scheme, because we do not calculate the variance of all variables. More processes are incorporated when using higher order closure schemes, but then new unknowns occur. For the one-and-a-half order closure scheme, we have to introduce a new variable, which is calculated directly. This variable is the turbulent kinetic energy (TKE, or e), defined as $\text{TKE} = e = \frac{1}{2}\left(\overline{u'^2} + \overline{v'^2} + \overline{w'^2}\right)$. The perturbation velocity in the x, y and z direction are respectively u', v' and w'. The TKE represents kinetic energy per mass of eddies and provides information about turbulence. The TKE equation can be used to find the vertical velocity W of equation 3.4 to calculate K. The TKE equation for the one-and-a-half order closure scheme is given by (Stull, 1988):

$$\frac{\partial \bar{e}}{\partial t} = -\overline{u'w'}\frac{\partial \bar{u}}{\partial z} - \overline{v'w'}\frac{\partial \bar{v}}{\partial z} + \overline{w'\theta'}\frac{g}{\bar{\theta}} - \frac{\partial \left[\overline{w'(p'/\rho + e)}\right]}{\partial z} - \epsilon. \tag{3.6}$$

$$\tag{3.7}$$

The density of air is represented by ρ. The production occurs in the first two terms on the r.h.s.. The third term on the r.h.s. represents the buoyancy flux, the fourth term the pressure and TKE diffusion and the last term dissipation. The TKE provides information about the turbulence and can therefore be used for the parametrization. Using one of the possible parametrizations, the equations then become, splitting K into a heat (K_H) and momentum (K_m) part:

$$\overline{u'w'} = -K_m\left(\bar{e}, \overline{\theta'^2}\right)\frac{\partial \bar{u}}{\partial z} \tag{3.8}$$

$$\overline{v'w'} = -K_m\left(\bar{e}, \overline{\theta'^2}\right)\frac{\partial \bar{v}}{\partial z} \tag{3.9}$$

$$\overline{\theta'w'} = -K_H\left(\bar{e}, \overline{\theta'^2}\right)\frac{\partial \bar{\theta}}{\partial z} - \gamma\left(\bar{e}, \overline{\theta'^2}\right) \tag{3.10}$$

$$\overline{w'\left[p'/\bar{\rho} + e\right]} = \frac{5}{3}\Lambda_3 e^{-1/2}\frac{\partial \bar{e}}{\partial z} \tag{3.11}$$

$$\epsilon_R = 0,\ \epsilon = \frac{\bar{e}^{3/2}}{\Lambda_1},\ \epsilon_\theta = \frac{\bar{e}^{1/2}\overline{\theta'^2}}{\Lambda_2}. \tag{3.12}$$

The pressure is p, u and v are the horizontal wind velocity, γ is the mixing length, which can be considered as the ability of turbulence to cause mixing and ϵ, ϵ_R and ϵ_θ are dissipation terms. The terms K_m, K_H and γ are dependent on $\bar{e}$ and $\overline{\theta'^2}$. The Λ factors are length-scale parameters. The TKE provides an estimation of the vertical velocity W of equation 3.4. The Λ term provides the length scale L. The K coefficient can therefore be approximated with:

$$K = \Lambda \bar{e}^{1/2}. \tag{3.13}$$

The one-and-a-half order closure is similar to the first order closure, but it is now dependent on TKE. Equations 3.8 - 3.12 are complex relations for which no analytical solutions are available. These equations can nonetheless be solved numerically, as is done in HARMONIE (Cuxart et al., 2000). Cuxart et al. (2000) can be consulted for the numerical implementation of the one-and-a-half order closure scheme. Using one-and-a-half order closure has some important advantages (Stull, 1988). This scheme creates well mixed layers that increase in depth during the day, as is expected due to increased turbulence during daytime. Secondly, a nocturnal stable boundary layer forms. Parametrizations of condensation and cloud formation can now be computed. Entrainment is also possible to model with this scheme. This turbulence scheme is implemented in the version of HARMONIE called HARATU (HArmonie with RAcmo TUrbulence). This turbulence scheme leads to enhanced turbulent mixing compared with the older versions of HARMONIE with a first order closure scheme.

3.2 Cloud Schemes

To connect the turbulence calculated from the turbulence scheme to cloud cover, a cloud scheme is necessary. The cloud scheme consists of parametrizations necessary to determine the cloud cover within a grid box. Turbulence can cause local saturation of water vapour within a grid box. This leads to clouds smaller than the resolution of the model. Cloud schemes are therefore necessary to incorporate subgrid cloud cover and leads to a fractional cloud cover. Figure 3.1 shows a schematic illustration of subgrid variability that can lead to cloud cover. The figure shows the total water specific humidity within a grid box. The total water specific humidity is defined as:

$$q_t = q_v + q_l + q_i. \tag{3.14}$$

The liquid water specific humidity is represented with q_l and q_i is the ice water specific humidity. The saturation specific humidity q_s is defined as the maximum humidity the air can sustain. When $q_t > q_s$, a cloud will form. The dashed area of figure 3.1 represents the region within a grid box where

saturation occurs and a cloud will form. These regions within the grid may have clouds, while the non-dashed areas do not. The fraction of the saturated and unsaturated parts of the grid box then provides the fractional cloud cover for the grid box. The cloud scheme is necessary to incorporate these subgrid variabilities. In this study, we provide a relative humidity cloud scheme and a statistical cloud scheme. In the following section we discuss the relative humidity scheme.

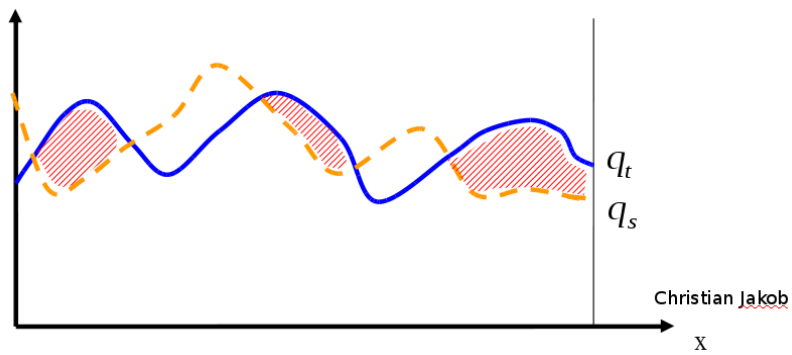


Figure 3.1: Schematic illustration of the subgrid variability of cloud cover. The total specific humidity and saturation specific humidity are drawn within a gridbox. A cloud will form when $q_t > q_s$ and is indicated by the dashed area.

3.2.1 Relative Humidity Cloud Scheme

A simple cloud scheme, which has been used in the past and is still used in various models (Quaas, 2012), is based on a simple relation between relative humidity and cloud cover. Relative humidity is defined as:

$$RH \equiv \frac{q_v}{q_s}. \tag{3.15}$$

The relative humidity is directly connected to cloud cover. When the relative humidity is 1.0, saturation occurs and a cloud will form. With a relative humidity scheme, we assume that relative humidity is affected by subgrid processes. This results in possible cloud formation even if the relative humidity is below 1.0. The subgrid processes are incorporated in the scheme with a probability density function (PDF) for the relative humidity. The relative humidity can then be converted to a fractional cloud cover, starting cloud formation from a critical relative humidity (RH_{crit}), until a maximum relative humidity (RH_{max}). In this study, we use a relative humidity scheme with a rectangular PDF, as is shown in figure 3.2a.

In this cloud scheme, ice and liquid water are neglected, q_t then reduces to q_v. Clouds begin to form from a certain critical relative humidity RH_{crit}, which is lower than 1.0. The width of the PDF determines RH_{crit} and corresponds to the maximum possible subgrid variability. The width is represented with Δq. The width can be calculated using figure 3.2b. This figure shows the PDF when clouds start to form for the first time within the grid. The q_s line then matches with the $q_v + \Delta q$ line, so

$$q_s = q_t + \Delta q \tag{3.16}$$

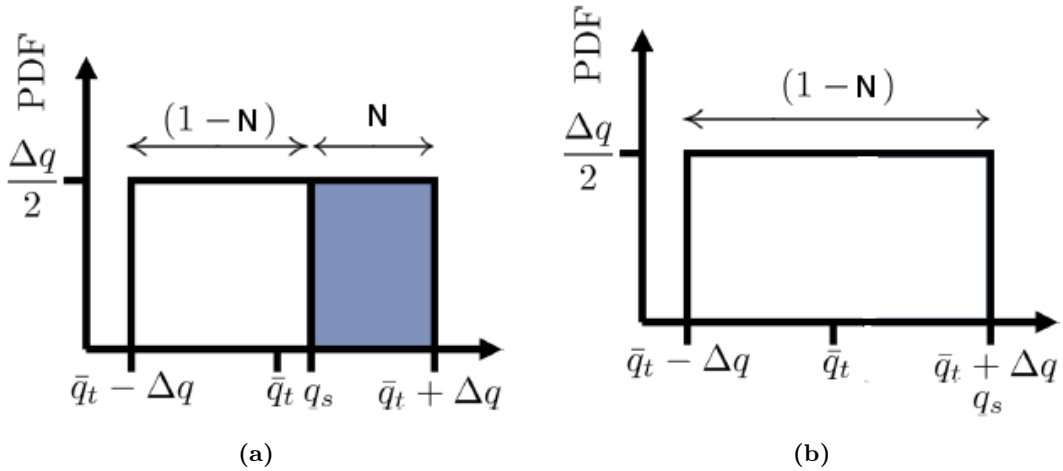


Figure 3.2: (**a**) The uniform probability density function of total water specific humidity as is used for the relative humidity cloud scheme. N is the cloud cover (Quaas, 2012). (**b**) The probability density function of total water specific humidity when clouds start to form. Then $RH = RH_{crit}$ and $q_t + \Delta q = q_s$.

$$\Delta q = q_s - q_v \tag{3.17}$$

$$= \left(1 - \frac{q_v}{q_s}\right) q_s \tag{3.18}$$

For this case, we defined the relative humidity to be RH_{crit}. Equation 3.18 then becomes:

$$\Delta q = (1 - RH_{crit})\, q_s. \tag{3.19}$$

This equation shows that the choice of RH_{crit} impacts the width of the PDF and therefore the fractional cloud cover. A large RH_{crit} corresponds with large subgrid variability. From this PDF, we calculate the cloud cover. The cloud cover in a rectangular PDF is represented with the blue area of figure 3.2a. This area equals:

$$N = \frac{q_v + \Delta q - q_s}{2\Delta q} \tag{3.20}$$

$$= \frac{RH + (1 - RH_{crit}) - 1}{2(1 - RH_{crit})} \tag{3.21}$$

The grid box mean relative humidity ($\overline{RH}$) can be split into a clear sky part and a cloudy sky part. We assume the relative humidity to be 1.0 in the cloudy part. The grid box mean relative humidity is then:

$$\overline{RH} = f \cdot 1 + (1 - f) RH_{clear}. \tag{3.22}$$

The clear sky relative humidity RH_{clear} can be rewritten, using the PDF of 3.2a:

$$
\begin{aligned}
RH_{clear} &= \frac{q_s + q_v - \Delta q}{2q_s} && (3.23)\\
&= \frac{1 + q_v/q_s - (1 - RH_{crit})}{2} && (3.24)\\
&= \frac{RH + (1 - RH_{crit}) - 1 - 2(1 - RHcrit) + 2}{2} && (3.25)\\
&= N(1 - RH_{crit}) + 1 - (1 - RH_{crit}). && (3.26)
\end{aligned}
$$

This can be used in equation 3.22:

$$
\begin{aligned}
\overline{RH} &= N + (1 - N)\, RH_{clear} && (3.27)\\
&= N + N\,(1 - RH_{crit}) + 1 - 1 + RH_{crit} - N^2\,(1 - RH_{crit}) - N + N\,(1 - RH_{crit}) && (3.28)\\
&= N + (1 - N)\,((1 - RH_{crit})\, N + 1 - (1 - RH_{crit})) && (3.29)\\
&= N + (1 - N)\,((1 - RH_{crit})\,(N - 1) + 1) && (3.30)\\
&= 1 - (1 - RH_{crit})\,(1 - N)^2\,. && (3.31)
\end{aligned}
$$

The cloud fraction then becomes:

$$
N = 1 - \sqrt{\frac{1 - \overline{RH}}{1 - RH_{crit}}}. \qquad (3.32)
$$

From now on, we omit the bar above RH for readability. Figure 3.3 shows the cloud cover profile as a function of the relative humidity. The dashed vertical lines represent from left to right RH_{crit} and RH_{max}. We have chosen $RH_{crit} = 0.8$ as an example.

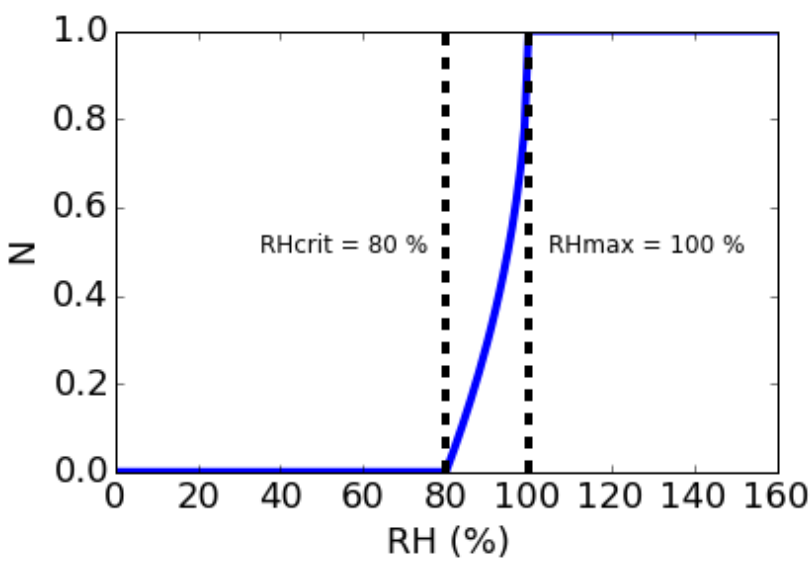
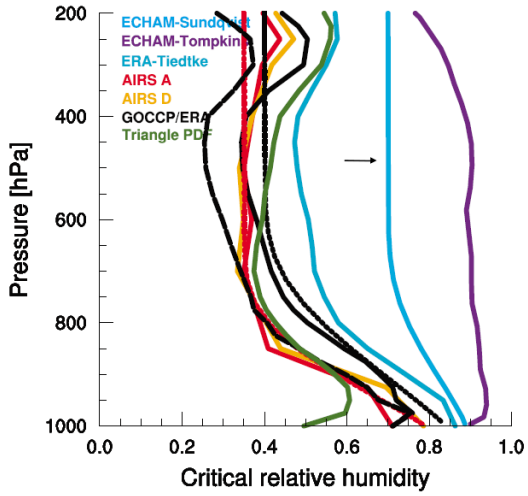


Figure 3.3: Cloud cover as a function of RH, following equation 3.32. The dashed lines correspond with RH_{crit} and RH_{max}.



(a)

Figure 3.4: (**a**) Global mean pressure profiles of RH_{crit} as measured (red, orange and black) and modelled. The arrow shows the profile of (Sundqvist et al., 1989), which we use in this study. The dashed lines show fits. (Quaas, 2012).

Variance of the subgrid-scale total water is not a constant, but height dependent. Note that RH_{crit} is directly related to the subgrid variability and thus the variance. Variance of temperature and humidity is produced by turbulence and convection. Turbulence causes for example mixing, what increases the variance. Clouds in higher regions of the atmosphere have a longer lifetime and are transported into the grid box, increasing the variance relative to q_s, therefore reducing the critical relative humidity for clouds. Figure 3.4a shows RH_{crit} measurements, using ERA interim (Dee et al., 2011) as well as Atmospheric Infra-Red Sounder (AIRS) mounted on the Aqua satellite (Susskind et al., 2003) to compute a vertical profile of RH_{crit} (Quaas, 2012). The figure shows the decrease of RH_{crit} with height. Our relative humidity scheme will follow ECHAM-Sundqvist (Sundqvist et al., 1989) from figure 3.4a, which is highlighted with the arrow. It consists of a gradual decrease of RH_{crit} in the boundary layer and above, followed by the stabilization of RH_{crit} from the middle troposphere and up. The measurement sets from AIRS and ERA-interim (red, orange, light blue and black) show a lower critical relative humidity profile compared to the profiles parametrizations of the models ECHAM-Sundqvist and ECHAM-Tompkins (blue and purple). A different choice of PDF also impacts the critical relative humidity profile, as is shown with the triangle PDF. In this study, we experiment with the profile of Sundqvist et al. (1989):

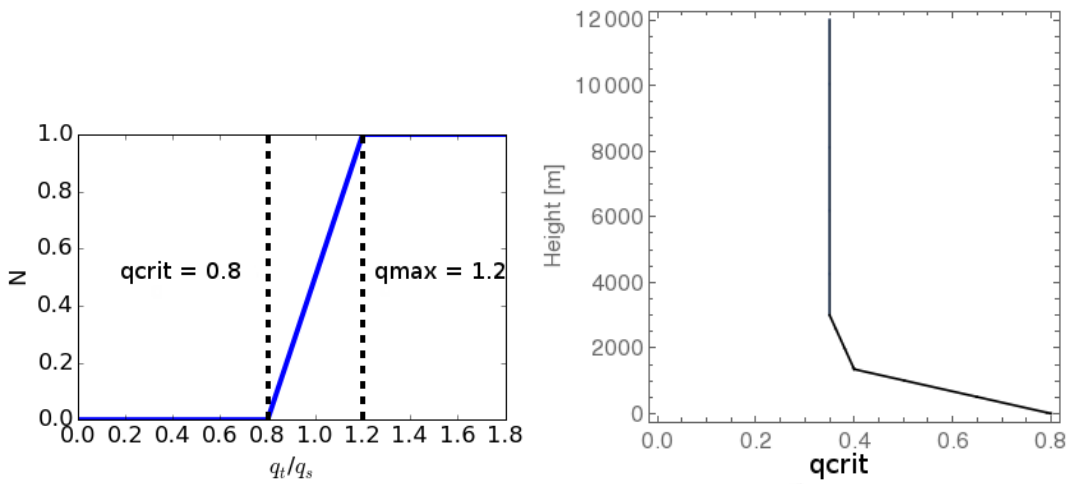
$$RH_{crit} = RH_{0,top} + (RH_{0,surf} - RH_{0,top}) \exp\left(1 - \left(\frac{p_s}{p}\right)^n\right). \tag{3.33}$$

The $RH_{0,top} = 0.7$, $RH_{0,surf} = 0.9$, $p_s = 1000$ hPa and $n = 4$ are respectively the top and surface relative humidity, surface pressure and fitting parameter. Note that this empirical estimate of the vertical profile is based on a global mean profile, which may vary in time, space and resolution (higher resolution decreases subgrid variability and hence increases RH_{crit}) and may in general be a too simple representation of the atmosphere. Also note that the considered models, ECHAM-Sundqvist and ECHAM-Tompkins, do have higher RH_{crit} values than the observations.

Another relative humidity scheme that we consider includes q_i and q_l of equation 3.14 in the cloud scheme. If a rectangular PDF is considered, then due to ice and liquid water, q_t/q_s can be larger than 1.0. Cloud cover will start to grow when $q_t/q_s = q_{crit}$. The cloud cover then increases from

q_{crit} to q_{max}, as is shown in figure 3.5a. This figure shows the cloud fraction N as a function of q_t/q_s with this cloud scheme. Clouds begin to form when $q_t/q_s = q_{crit}$, and overcast conditions occur when $q_t/q_s = q_{max}$. When a rectangular PDF is considered, then $q_t/q_s = 1.0$ is in the middle of the linear relation, which corresponds in this scheme with a cloud cover of $N = 0.5$. When $q_t/q_s = 1.0$, overcast conditions are present for 50% of the grid box and clear sky conditions for the other 50%, resulting in $N = 0.5$. Due to the linear relation between q_{crit} and q_{max}, and that q_t/q_s always corresponds with $N = 0.5$, q_{max} is linked to q_{crit}. If q_{crit} is reduced, then is q_{max} is increased similarly in order to keep $N = 0.5$ correspond with $q_t/q_s = 1.0$. Beyond q_{max}, the variance is not high enough to cause q_t/q_s to drop below 1.0 anywhere, resulting in complete overcast.

For this relative humidity scheme, we use an approximated q_{crit} profile following the critical relative humidity measurements of AIRS A and AIRS B (Quaas, 2012) from figure 3.4a. Figure 3.5b shows the implemented q_t profile profile as a function of height. Keep in mind that the chosen q_{crit} profile is linked to critical relative humidity, which does not include q_l and q_i in the cloud scheme. The critical relative humidity profile does provide a first guess about the choice of q_{crit}. This relative humidity scheme provides information about the impact of q_i and q_l in the cloud scheme.



(a) **(b)**

Figure 3.5: (**a**) Cloud cover N as a function of q_t/q_s at ground level, when also considering ice an liquid. The vertical lines correspond with RH_{crit} and RH_{max}. (**b**) Estimated dependence of RH_{crit} on height, using the measurements of AIRS A and AIRS D (Quaas, 2012) of figure 3.4a.

3.2.2 Statistical Cloud Scheme

The statistical cloud scheme is the default cloud scheme in HARMONIE. It is a more complex cloud scheme than the relative humidity scheme. The statistical cloud scheme has a strong theoretical basis. The scheme assumes a probability distribution around the grid box mean temperature and moisture, which represents the subgrid variability. This PDF determines the fractional cloud cover. The most important difference between the different schemes is the way of obtaining the variances of q_t and θ_l. The variable θ_l is given by:

$$\theta_l = \theta - \frac{L_v}{c_p \Pi} q_l, \tag{3.34}$$

with c_p the specific heat at constant pressure for moist air, L_v the latent heat and Π the Exner function. The Exner function is defined as:

$$\Pi = \left(\frac{p}{p_{ref}}\right)^{\frac{R}{c_p}} \tag{3.35}$$

$$= \frac{T}{\theta}, \tag{3.36}$$

with R the gas constant for dry air and p_{ref} a reference pressure. One can view the cloud parametrization problem as a probability density problem. Suppose variability in θ_l and q_t are described by the PDF in a grid volume. Then the cloud cover N and the liquid water specific humidity q_t can be determined using integration from minus to plus infinity for θ_l and q_t over the PDF, using the Heaviside function:

$$N = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H(q_t - q_s)P(\theta_l, q_t)\mathrm{d}q_t\mathrm{d}\theta_l \tag{3.37}$$

$$q_l = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (q_t - q_s)\, H(q_t - q_s)P(\theta_l, q_t)\mathrm{d}q_t\mathrm{d}\theta_l. \tag{3.38}$$

The probability density function corresponds with P and $H(x) = 0$ if $x < 0$ and $H(x) = 1$ if $x > 0$. Let us call $s = q_t - q_s(p, T)$. This quantity can be interpreted as how far the total specific humidity is from the saturation curve. Figure 3.6 shows the saturation curve as a function of q_t and T. From a certain mean temperature and total specific humidity $(\bar{T}, \bar{q}_t)$, the shortest distance to q_s is $\bar{s}$.

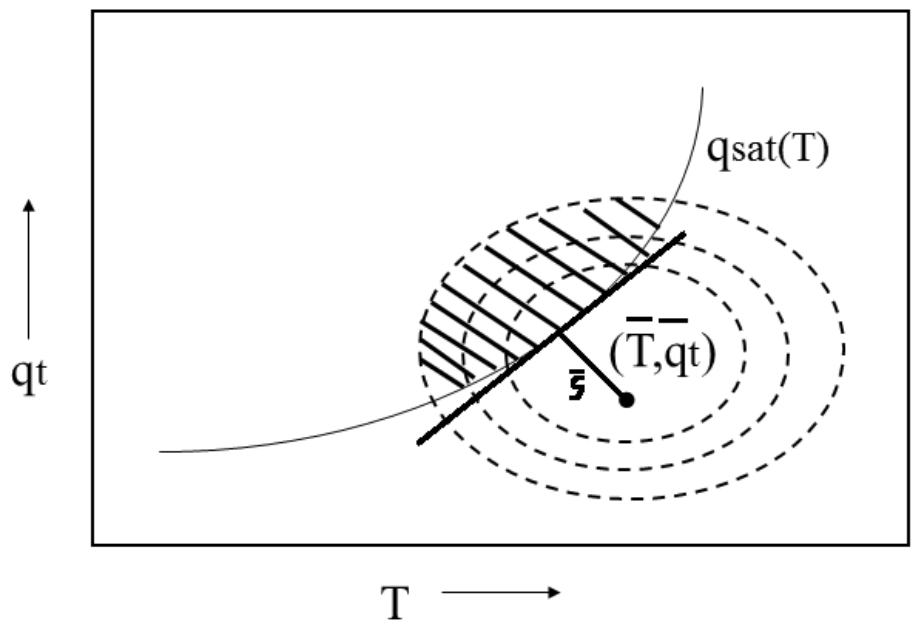


Figure 3.6: The saturation curve as a function of q_t and T (De Rooy et al., 2014). The variation around the grid box mean $(\bar{T}, \bar{q}_t)$ are the dashed PDF isolines. The shaded area is the part of the PDF that contributes to the fractional cloud cover. The shortest distance to the saturation curve from $(\bar{T}, \bar{q}_t)$ is given by $\bar{s}$.

Whenever a cloud is present, the saturation mixing ratio q_s can be estimated using a first order Taylor expansion around $\overline{T_l}$, which is the mean liquid water temperature. The saturation specific humidity q_s then becomes:

$$q_s(T) \simeq q_s(\overline{T_l}) + \frac{\partial q_s(\overline{T_l})}{\partial T}\left(T - \overline{T_l}\right) + \mathcal{O}\left(\left(T - \overline{T_l}\right)^2\right). \tag{3.39}$$

Equation 3.34 can be rewritten in terms of the temperature and using equation 3.36.

$$T = T_l + \frac{L}{c_p} q_l \tag{3.40}$$

Equation 3.39 then becomes, using $T_l = \overline{T_l} + T_l'$, and using equation 3.38:

$$q_s(T) \simeq q_s(\overline{T_l}) + \frac{\partial q_s(\overline{T_l})}{\partial T}\left(T_l + \frac{L}{c_p} q_l - \overline{T_l}\right) \tag{3.41}$$

$$q_s(T) \simeq q_s(\overline{T_l}) + \frac{\partial q_s(\overline{T_l})}{\partial T}\left(T_l' + \frac{L}{c_p} q_l\right) \tag{3.42}$$

$$q_s(T) \simeq q_s(\overline{T_l}) + \frac{\partial q_s(\overline{T_l})}{\partial T}\left(\theta_l'\Pi + \frac{L}{c_p} q_l\right) \tag{3.43}$$

$$q_s(T) \simeq q_s(\overline{T_l}) + \frac{\partial q_s(\overline{T_l})}{\partial T}\left(\theta_l'\Pi + \frac{L}{c_p} H(s) s\right). \tag{3.44}$$

$$\tag{3.45}$$

Now s can be approximated as follows:

$$s = q_t - q_s(\overline{T_l}) + \frac{\partial q_s(\overline{T_l})}{\partial T}\theta_l'\Pi - \frac{\partial q_s(\overline{T_l})}{\partial T} s H(s) \frac{L}{c_p}, \tag{3.46}$$

which can be split up, depending on the sign

$$s = \begin{cases} \left(1 + \left(\frac{L}{c_p}\right)\frac{\partial q_s(\overline{T_l})}{\partial T}\right)^{-1}\left[q_t - q_s(\overline{T_l}) - \frac{\partial q_s(\overline{T_l})}{\partial T}\Pi\theta_l'\right] & \text{if } s > 0; \\ q_t - q_s(\overline{T_l}) - \frac{\partial q_s(\overline{T_l})}{\partial T}\Pi\theta_l' & \text{if } s < 0. \end{cases}$$

The distance to the saturation curve s can be interpreted as a parameter that controls saturation within the grid, because saturation is possible whenever $q_l = s \geq 0$. The parameter s and q can be split into a grid box mean and a perturbation part: $s = \overline{s} + s'$ and $q_t = \overline{q_t} + q_t'$. Saturation takes place whenever $s' \geq -\overline{s}$. s' is given by:

$$s' = \frac{q_t' - \frac{\partial q_s(\overline{T_l})}{\partial T}\Pi\theta_l'}{1 + \left(\frac{L}{c_p}\right)\frac{\partial q_s(\overline{T_l})}{\partial T}} \tag{3.47}$$

The standard deviation of s is the root of the mean of the square of s':

$$\sigma_s = \sqrt{\overline{s'^2}} = \left(1 + \left(\frac{L}{c_p}\right)\frac{\partial q_s(\overline{T_l})}{\partial T}\right)^{-1}\sqrt{\overline{q_t'^2} - 2\frac{\partial q_s(\overline{T_l})}{\partial T}\overline{q_t'\theta_l'} + \left(\frac{\partial q_s(\overline{T_l})}{\partial T}\right)^2 \overline{\theta_l'^2}}. \tag{3.48}$$

To get a dimensionless quantity that has a one-to-one correspondence with cloud fraction for Gaussian PDF's, one needs to normalize s', which will be called $t \equiv s'/\sigma_s$. Saturation is present when $t \geq -\frac{\overline{s}}{\sigma_s}$. The normalized averaged saturation deficit Q_1 (Sommeria and Deardorff, 1977) is then:

$$Q_1 = \frac{\overline{s}}{\sigma_s}. \tag{3.49}$$

The PDF of t is called $G(t)$. Clouds form whenever saturation is reached. Thus the total cloud cover of the grid cell is the integral from the lowest value possible for clouds (when $t = 0$) to infinity:

$$N = \int_0^{+\infty} G(t)\mathrm{d}t. \tag{3.50}$$

When assumed Gaussian, $G(t)\mathrm{d}t = \frac{1}{\sqrt{2\pi}} \exp\left(-t^2/2\right) \mathrm{d}t$. Equation 3.50 then has a standard solution in the form of:

$$N = 1/2 \left(1 + \mathrm{erf}(Q_1/\sqrt{2})\right), \tag{3.51}$$

in which erf is the error function:

$$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} \mathrm{d}t. \tag{3.52}$$

The relation of equation 3.51 has been tested against measurements and large eddy simulation models (LES) (Cuijpers and Bechtold, 1995), as is shown in figure 3.7. The distribution of N follows approximately equation 3.51, but does not fit exactly. A best fit is performed by Cuijpers and Bechtold (1995) and is empirical in nature. The cloud cover in HARMONIE is calculated using the estimation done by Cuijpers and Bechtold (1995):

$$N = 0.5 + 0.36 \arctan(1.55 Q_1). \tag{3.53}$$

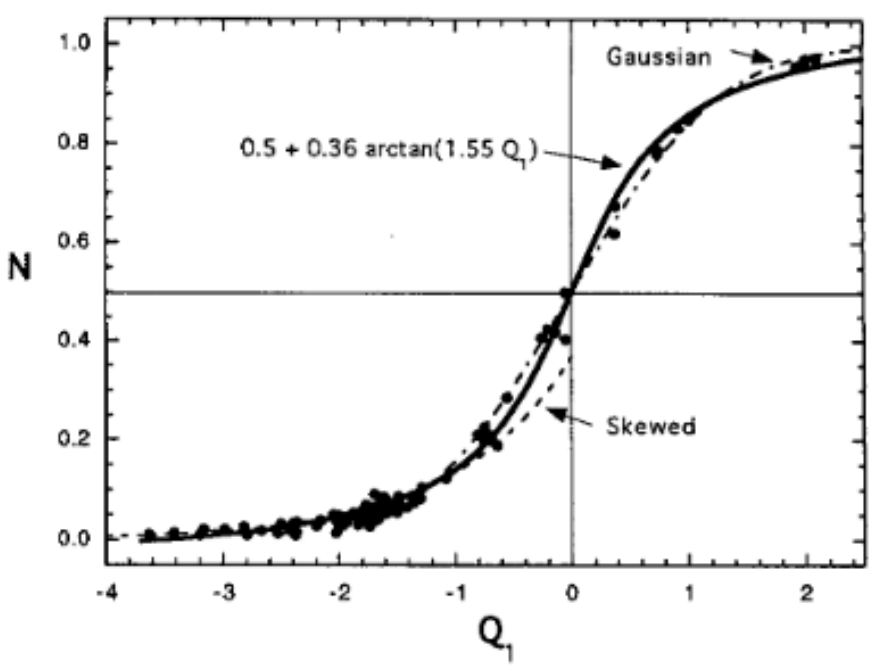


Figure 3.7: Cloud cover N as a function of Q_1, as performed by Cuijpers and Bechtold (1995). The full line is the parametrized cloud fraction. The dots are measurements.

Figure 3.7 shows the cloud cover N as a function of Q_1 (Cuijpers and Bechtold, 1995), with the parametrization of equation 3.53, which is the best fit following observations. With $Q_1 < -2$, almost

no cloud cover is present, following a sharp increase to an almost complete cloud cover field with $Q_1 > 2$. Skewness of the profile is limited, thus the deviation of the empirically determined curve varies only slightly with the Gaussian cloud cover curve. In HARMONIE, N is only dependent on Q_1 and is therefore determined whenever Q_1 is known.

Figure 3.6 also shows how the statistical cloud scheme can be interpreted. The dashed circles around the point $(\bar{T}, \bar{q}_t)$ show the variation and are isolines of a joint PDF. The area of the circle beyond the saturation curve contribute to the fractional cloud cover. The fraction of the shaded area against the total area of the circles can therefore be interpreted as the cloud fraction.

The standard deviation σ_s of equation 3.48 consists of three parts in HARMONIE: the convection part (σ_{con}), turbulence part (σ_{turb}) and an extra term (σ_{extra}). The turbulence part is produced by the turbulence scheme, as is described in section 3.1. The convection term is produced in the convection scheme. The convection term includes entrainment and detrainment (De Rooy and Pier Siebesma, 2010). The extra variance part is included to incorporate any relevant processes not included in the cloud scheme, like transport of clouds at high altitudes. The extra variance term is dependent on q_s. This results in a direct dependence of Q_1 with RH in this term. The variable q_s is temperature and pressure dependent, thus it varies with height. Due to the profile of q_s of figure 3.6, the extra variance term will usually decrease with height. The standard deviation σ_s is then calculated by:

$$\sigma_s = \sqrt{a\left(\sigma_{con}^2 + \sigma_{turb}^2\right) + b\sigma_{extra}^2}. \tag{3.54}$$

The terms a and b are tuning parameters. The tuning parameters can be used to assess the impact of the different terms of equation 6.1.

3.3 Versions of HARMONIE

In this study, we run HARMONIE in three different modes: hindcast (HC), climate (CL) and data assimilation (DA). In the HC mode, the model is initialized from ERA-Interim data (Dee et al., 2011) at the boundaries and in the atmosphere every 24 hours, followed by a forecast of 36 hours. However, the conditions of the surface are initialized by the previous 24 hour forecast. For the first forecast, ERA-interim data is used for the surface. The first 12 hours are disregarded due to spin-up effects. In the CL mode, the model is initialized only once after which it runs freely without any reinitializations. In both models the lateral boundary conditions are provided by ERA-Interim. Due to the long forecasting time, the climate run will have difficulties in producing the correct individual weather events. A statistical analysis of the weather is needed to interpret the CL mode (it is therefore called the climate mode). The operational version of HARMONIE uses the DA mode. The DA mode makes a forecast every 3 hours, using the previous forecast combined with observations (it is therefore called the data assimilation mode). The boundaries are provided by the ECMWF. The DA mode is computationally intensive. We focus on the computationally less intensive, but otherwise similar HC mode.

From time to time, HARMONIE will be updated following recent studies. Different operational versions of HARMONIE are indicated by cycles. The current operational version of HARMONIE at the KNMI is cycle 36. For this study, we consider different versions of cycle 38 up to the most recent version cycle 40. More verification is needed for a smooth transition to a new cycle. As a reference, HARMONIE cycle 38 with inhomogeneity factor set to 1 is used and is called HARMONIE cyc38. The cloud inhomogeneity parameter χ is defined as the ratio of the exponential of the logarithmic average of the cloud optical thickness to the linear average of the cloud optical thickness Gleeson et al. (2015):

$$\chi = \frac{e^{\overline{\log \tau}}}{\bar{\tau}}. \tag{3.55}$$

Table 3.1: Table of the considered model versions with its description.

Model version	Model description
HARMONIE cyc38	Reference HARMONIE cycle 38 version
HARMONIE SM	HARMONIE cycle 38 with simple micro physics
HARATU	HARMONIE with RACMO turbulence scheme
HARATU update	HARATU with updated turbulence and new radiation scheme
HARATU update DA	As HARATU update, but in data assimilation mode

For which τ is the optical thickness, defined as:

$$\tau = \log\left(\frac{\Phi_i}{\Phi_t}\right). \tag{3.56}$$

The optical thickness is therefore the logarithm of the fraction of the received and transmitted radial flux (respectively Φ_i and Φ_t). The choice of the inhomogeneity parameter has its impact on the radiation budget of the model. For all considered HARMONIE versions, the inhomogeneity factor is set to 1.0, which is a reasonable value for a 2.5 by 2.5 km resolution.

Another model version assumes only simple micro physics, which we call HARMONIE SM. Micro physics of clouds describe the properties and structures of clouds on the microscopic scale. This means that cloud microphysics include processes happening on the droplet scale. It includes the growth and precipitation of droplets. Cloud microphysics happens on a subgrid scale, so parametrizations are needed. A microphysics scheme is incorporated in HARMONIE, but now we turn off some recent updates made to this scheme. Some of the turned off changes are: tuning of the conversion from ice to water vapour and vice versa, and a separate ice cloud fraction, contributing to the total cloud cover.

HARATU (HArmonie with RAcmo TUrbulence) is an important parametrization to test. The turbulence scheme HARATU is different than is used in HARMONIE cycle 38. This turbulence scheme has been developed for the regional climate model RACMO Neggers et al. (2004), but has also been implemented in HARMONIE (De Rooy et al., 2014). RACMO includes changes in boundary layer turbulence. It includes enhanced mixing of turbulence, resulting in reduced cloud cover if compared to older versions of HARMONIE. The new turbulence scheme calculates the TKE and turbulent mixing length scale L of equation 3.5 at half vertical levels (De Rooy et al., 2014). Prognostic variables at full levels are determined by the vertical divergence of the turbulent fluxes at half levels and is now calculated by HARATU. Previously, TKE and L were calculated at full levels and thus interpolation was necessary.

Another version we use in this study, is an updated version of HARATU and a new radiation scheme (here called HARATU update). Cloud liquid optical properties have been simplified by Nielsen et al. (2014), especially in the Mie scattering range (when the size of aerosols is comparable with the wavelength of the incoming light). The update of the turbulence scheme also provides a better estimation of the zonal 10 meter wind speed under stormy conditions. As a result of the update in the turbulence scheme, slightly more mixing takes place near the surface and considerably more mixing at intermediate heights. We also run HARATU update in data assimilation mode, which we call HARATU update DA. Table 3.1 shows an overview of the different HARMONIE versions.

4 Methodology

4.1 Coarse graining

Before MSG and HARMONIE data can be compared, it has to be converted to the same grid. For Western Europe, MSG observes clouds with a resolution of 4 by 7 km, while HARMONIE provides data on a 2.5 by 2.5 km grid. MSG and HARMONIE will be interpolated to the same regular latitude-longitude grid. The new resolution is then 0.05 by 0.05 degrees. For Western Europe, this corresponds with a grid size of approximately 3.5 by 5.5 km. The interpolation may lead to a fractional cloud cover of MSG. If a data set misses data for a certain location or time, it will be omitted from all other sets. For some calculations it is necessary to coarse grain even further. The impact of the binary cloud cover data provided by MSG is reduced on a coarser grid and fractional cloud cover can occur. Errors are also expected to average out with a coarser grid. The results can be easily compared with lower resolution models. Figure 4.1 shows how the initial grid, called 1 x 1, is converted to a coarser grid. An average from neighbouring grid points is taken, which becomes a new grid point. For example, for 3 x 3, the initial grid point and its eight surrounding points are taken to compute a mean. A grid of 5 x 5 or more can be calculated using the same method. Any further calculations are done on the newly calculated grid.

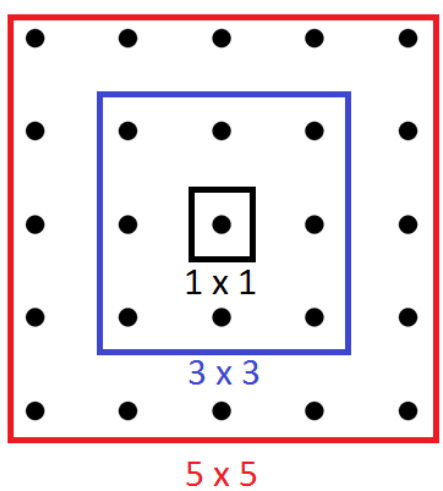


Figure 4.1: Illustration of coarse graining. Every dot represents a grid point, with the initial grid corresponding to 1 x 1 (black box). To create a coarser grid, averages will be taken from neighbouring grid points. For example, for 3 x 3 (blue box), the initial grid point and its eight surrounding points are taken to compute a mean. The mean is then taken as a new grid point. Any further calculations are done on the newly calculated grid. The red box shows the points that are used for calculating a new grid point using 5 x 5.

4.2 Statistics

To be able to quantify the observed differences between MSG and HARMONIE, a statistical analysis is needed. We start with the bias of the data set.

Let us consider two hypothetical data sets, model output x and observational set y. Both data sets range from 1 to n. We define the bias as the mean difference between the model output and the observational data:

$$\text{BIAS} = \frac{\sum_{i=1}^{n} (x_i - y_i)}{n}. \tag{4.1}$$

The bias can be used to locate regions of over- or underestimation of the considered variable by the model. The bias does not provide information about the cause of the model and observational differences. Overestimation may also cancel underestimations. To prevent this, it is convenient to calculate the absolute bias. The absolute bias shows how strongly the model and observations deviate. A small absolute bias shows that the observation and data set are similar. The absolute bias can be calculated for different model versions. The difference between these biases then provide a measure of the performance of the different model versions against MSG. A large absolute bias difference shows that one version performs better when compared with MSG than the other.

The root-mean-square error (RMSE) is another method for measuring the correctness of a model for a particular variable. The observations (x) and model outcome (y) are compared in the following way:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - y_i)^2}{n}}. \tag{4.2}$$

The RMSE has the same units as the considered variable and is therefore easy to relate to. It can also be interpreted as the mean distance between a model point and the corresponding measurement point. The RMSE is expected to be large when comparing a data set that provides fractions with a binary data set. In this study, HARMONIE provides fractional cloud cover while MSG provides mostly binary data. On each grid point and on every time step, model and observational data are compared. On average, the difference between model and observations is large due to comparison of fractional data set with this binary data set. As a result, the RMSE is expected to be large.

To quantify the dispersion of a data set, the standard deviation can be used. The standard deviation is independent of other data sets and represents the spread of the variable in the considered data set. Data with a small standard deviation is close to the expected value and shows the spread around the mean. The standard deviation is defined as:

$$\sigma_X = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^{n} (x_i - X)^2}. \tag{4.3}$$

The mean of data set x is represented with X. The standard deviation can be enhanced to assess the mutual change of two data sets. This results in the covariance. The covariance shows how two variables variate together. It is defined as:

$$\sigma_{X,Y} = \frac{1}{n(n-1)} \sum_{i=1}^{n} (x_i - X)(y_i - Y). \tag{4.4}$$

The covariance can be used, along with the standard deviation, to quantify the linear dependence of variable x and y. The correlation coefficient is defined as the ratio between the covariance and the standard deviations in x and y. The correlation coefficient r equals:

$$r = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y} \tag{4.5}$$

$$r = \frac{\sum_{i=1}^{n} (x_i - X)(y_i - Y)}{\sqrt{\sum_{i=1}^{n} (x_i - X)^2 \sum_{i=1}^{n} (y_i - Y)^2}}. \tag{4.6}$$

The correlation coefficient $r \in$ [-1,1], in which $r = 1$ corresponds with a perfect positive correlation and $r = -1$ with a perfect negative correlation. The correlation coefficient can be used to show the degree of correlation between two data sets. It is an important method to test the performance of the used models. To compare the calculated correlation coefficients of two different model versions, the theory of Fisher (1921) and Steiger (1980) about statistical significance between dependent and independent correlation coefficients is used (see appendix B.1 and B.2).

The standard deviation, RMSE and correlation coefficient can be combined into a single diagram, which is called the Taylor diagram (Taylor, 2001). The Taylor diagram is a diagram used for comparing the same variable for different models. The Taylor diagram consist of the standard deviation on the x and y axis. The observational standard deviation of a certain variable is given as a star on the x axis, and follows the quarter circle toward the y axis. The standard deviation is therefore the radial distance to the origin. If the spread in the model is estimated correctly, it should be comparable with the observations and should be on the dashed line. Points closer (further) to (from) the centre underestimate (overestimate) the spread. Circles from the observational point on the x axis are drawn, representing the RMSE. The correlation coefficient is related to the angle with the x axis, with highest correlation coefficients closest to the x axis.

Figure 4.2 shows an example Taylor diagram, with three hypothetical models. The performance of the different models can be easily compared. Model 1 has a relatively high RMSE and low correlation coefficient. The standard deviation is underestimated, meaning that the spread of model 1 is too low. Model 3 severely overestimates the uncertainty, resulting in a high RMSE. Model 2 performs best, with a standard deviation comparable with the observational data set, low RMSE and high correlation coefficient. The Taylor diagram is thus an easy method to compare multiple statistical methods.

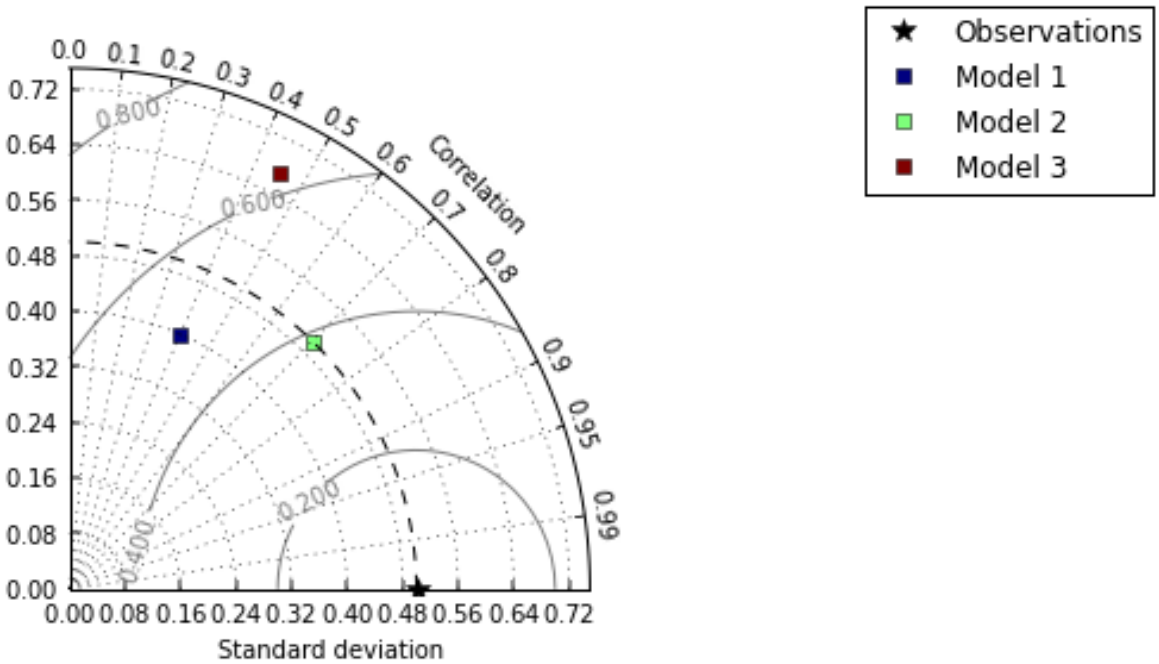


Figure 4.2: Example Taylor diagram (Taylor, 2001) for hypothetical models and an observational data set. The observations are denoted with a star. The standard deviation is on the x and y axis, and follows the quarter circle. The correlation coefficient is dependent on the angle with the x axis. The RMSE corresponds with the circles around the observations.

In this study, we use the statistical method of bootstrapping to determine significance. Bootstrapping consists of generating a randomly sampled list of data to approximate the statistics of the considered quantity. Most importantly, the bootstrapped dataset can provide the standard deviation. The significance of the observed value can be determined when comparing with the bootstrapped error margin. For this study, the 2σ, or 5% boundary, is used as a decent margin for significance. Any value above the bootstrapped uncertainty range of 2σ is considered as a significant difference. Statistical bootstrapping can be used for different calculations, including bias calculations.

4.3 Structure, Amplitude and Location (SAL) Analysis

Quality analyses are done with simple to use and quickly to implement methods like the RMSE and comparison of correlation coefficients. However, spatial and structural information of cloud cover is not measured by these quantities. The following example is useful to illustrate how grid point based error measures, such as the RMSE, can state that a model calculates cloud fraction very poorly, while it is only slightly off. Let us consider a cloud, which is correctly calculated in the model in terms of amplitude and structure, but not in location. The grid point based error calculations will rate the model outcome as very poor. Under such conditions, the RMSE is comparable to a completely misplaced event in the model, while in fact it is only slightly off compared to the observations. The RMSE does not identify the nature of the error. A statistical verification method which considers the amplitude, structure and location of cloud cover is therefore imperative.
The World Meteorological Organization (WMO) recommends (Mittermaier et al., 2012) spatial verification methods for verifying cloud-related forecasts. In this study we use the Structure, Amplitude and Location (SAL) method (Wernli et al., 2008). This method is developed for precipitation, but has also been used for cloud cover verification (Crocker and Mittermaier, 2013; Weniger and Friederichs,

2016). The SAL method considers the structure (S), amplitude (A) and location (L) of the cloud fraction in the given domain. The SAL method can also be used to rank the performance of models compared to the observations.

4.3.1 Objects

For the computation of the location and structure components, individual cloud fields have to be identified and are called objects. These fields are separated between each other and are essential for the calculation of S and L. A threshold is used to identify objects. Cloud cover is only considered if the cloud fraction is larger than the threshold. The choice of this threshold f to get the best results is difficult, and dependent on time and model version. Two different fractions are tested, one for the MSG data and one for the models. The threshold for the MSG data is expected to be high due to the tendency of the satellite to overestimate the cloud cover relative to ground measurements (see chapter 2). It is important to choose f large enough to filter out the overestimation of high cirrus clouds. The choice of f may have large consequences on the SAL calculations (Wernli et al., 2008). Figures 4.3 and 4.4 show an example of the objects in the lower figures when $f = 0.5$ and $f = 0.25$ respectively during 25 March 2012, 13 UTC. The figures show the cloud cover distribution of this moment according to MSG and HARATU. Every cloud object receives a number, ranging from 1 to the total amount of objects on this moment. Figures 4.3 and 4.4 show the importance of the fraction choice. Especially small cloud fields and a possible interconnection between cloud fields are dependent on the threshold choice. The definition of the components of SAL is given in the following sections.

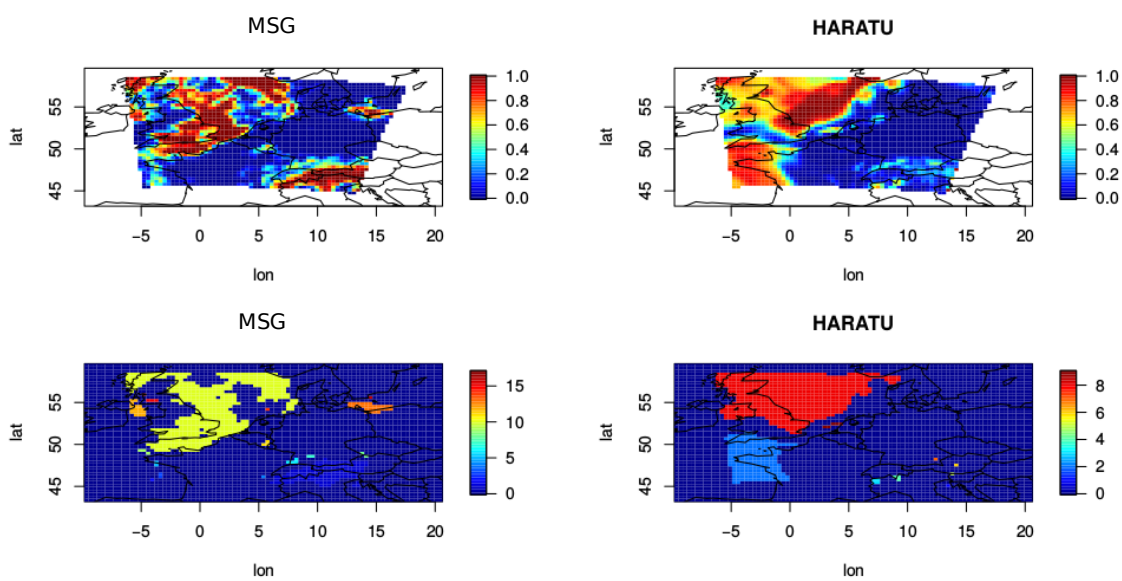


Figure 4.3: Cloud cover distribution (upper figures) and identification of objects (lower figures) for MSG and HARATU during 25-03-2012 13 UTC, with $f = 0.5$.

4.3.2 Amplitude A

The amplitude component of SAL is the normalized difference of the domain averaged cloud cover, between the model and the observations. This is given by the following equation:

$$A = \frac{D(R_{mod}) - D(R_{obs})}{0.5[D(R_{mod} + D(R_{obs})]}, \tag{4.7}$$

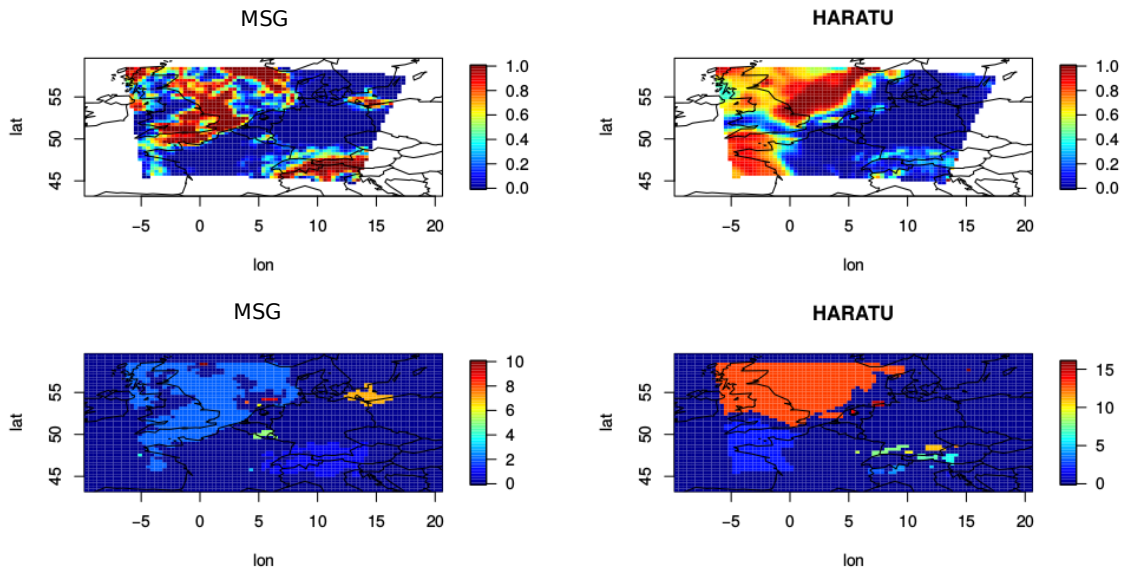


Figure 4.4: Cloud cover distribution (upper figures) and identification of objects (lower figures) for MSG and HARATU during 25-03-2012 13 UTC, with $f = 0.25$.

with:

$$D(R) = \frac{1}{N} \sum_{(i,j)\in d} R_{ij}. \tag{4.8}$$

The cloud cover field is given by R and $D(R)$ refers to the domain average of R in the domain d. R_{ij} are cloud cover grid point values. The amplitude component of SAL provides only information on the total amount of cloud cover in d, in which $A \in$ [-2,2]. An over- (under)estimation of the domain averaged cloud cover by a factor three corresponds with $A = 1$ ($A = -1$) (Wernli et al., 2008). If $A = 0$, the calculated amplitude of the model is the same as the observations. Note in equation 4.7 that A is independent of objects.

4.3.3 Structure S

The structure of a cloud field is calculated using the volume of the objects. If the volume of the objects of MSG and the models differ, then $S \neq 0$. No different vertical layers are used, hence the volume is an area. The weighted mean of the volume of all objects (V) is:

$$V(R) = \frac{\sum_{n=1}^{M} R_n^2}{\sum_{n=1}^{M} R_n}. \tag{4.9}$$

The sum of the cloud cover of all grid points within an object n is defined as R_n. For the calculation of V, R_n is squared and summed from the first object, to the last object called M, divided by the sum of R_n over all objects. S is defined as the normalized difference of the volume V, and is calculated as described by Wernli et al. (2008):

$$S = \frac{V(R_{mod}) - V(R_{obs})}{0.5[V(R_{mod}) + V(R_{obs})]}. \tag{4.10}$$

If S is large, then cloud cover is widespread in the domain. In HARMONIE, parts of the cloud field often have a cloud cover of 1.0. Therefore, one would expect S and A to correlate with each other. This means that an increase of S is expected to correspond with an increase of A. Like A, $S \in$ [-2,2]. When $S = 0$, it means that the model and the observations produce exactly the same structure.

4.3.4 Location L

The location component L consists of two equally important parts; L_1 and L_2 ($L_1, L_2 \in$ [0,1]). L_1 is defined as the normalized location difference of the centre of mass of the cloud field and is independent of objects. The first location component L_1 is given by:

$$L_1 = \frac{|\mathbf{x}(R_{mod}) - \mathbf{x}(R_{obs})|}{\delta}. \tag{4.11}$$

In this equation, δ is the largest distance between two possible boundary points of the domain d, and $\mathbf{x}$ is the location of the centre of mass of the cloud cover field R. If $L_1 = 0$, the centre of mass is predicted correctly. It does not necessarily mean the location of all clouds are predicted correctly. One large cloud in the middle of the domain may have the same centre of mass as two clouds on opposing sides. The other part of L is chosen to distinguish these cases. It is not always possible to compare corresponding objects in the observations and the model. The location of objects will therefore be calculated with respect to the centre of mass of the total cloud cover. The normalized average distance of the centre of mass of the different objects to the centre of mass of the total cloud field is assessed in L_2. The averaged distance r is calculated as follows:

$$r = \frac{\sum_{n=1}^{M} R_n \, |X - X_n|}{\sum_{n=1}^{M} R_n}. \tag{4.12}$$

The location of the centre of mass of the whole cloud cover field is X, and X_n the location of the centre of mass of cloud object n. The absolute difference between the modelled and observed r is then used to calculate L_2:

$$L_2 = \frac{2\,|r(R_{mod}) - r(R_{obs})|}{\delta}. \tag{4.13}$$

The factor 2 in equation 4.13 is chosen such that $L_1 \in [0, 1]$ and $L = L_1 + L_2 \in$ [0,2], in which 0 means a perfect prediction. The calculation of each component is relative to observations, in this study relative to MSG satellite data.

4.3.5 Idealized situations

Idealized cases have been sketched for SAL interpretation in figure 4.5 (Wernli et al., 2008). The shown situations are useful to pinpoint the strength and weakness of the SAL analysis. Figure 4.5 shows from B to I: B is a cone shaped object; C is a larger cone shaped object; D is the same object as B, but displaced; E is a smaller cone, with lower amplitude; F is two connected cones; G is two cone shaped object side to side; H is a high amplitude object, and I is a peaked cone. The threshold f of object identification have been taken to equal 0.15. Table 4.6 shows the SAL values for the different scenarios when compared to each other, with rows corresponding with the forecast, columns with the observations. Do keep in mind that a positive (negative) S or A component corresponds with a over-(under)estimation of the total cloud cover by the model with respect to the observations. Let us first consider BC. One would expect a negative structure and amplitude, due to the underestimation of B compared to C. However, the location of the object is correct, thus L is around zero. According to table 2, $S = -0.88$, $A = -0.88$ and $L = 0$, which is expected. D is translated with respect to B,

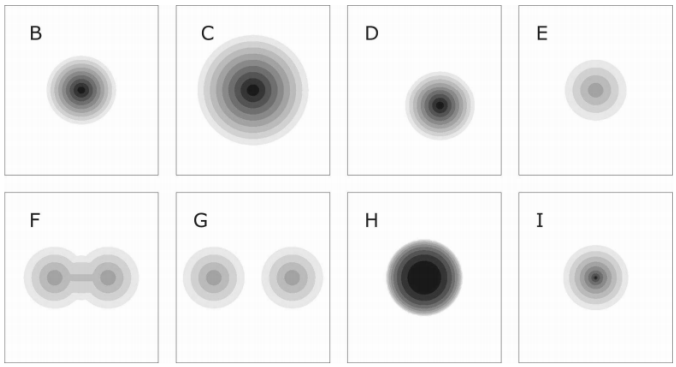


Figure 4.5: Contour plots of idealized objects for evaluation of SAL. The sub figures range from B to I (not A due to possible confusion with the amplitude component). The darker the color, the higher the amplitude (Wernli et al., 2008).

	B	C	D	E	F	G	H	I
B	0/0/0	−0.88/−0.88/0	0/0/0.14	0/0.67/0	−0.67/0/0	0/0/0.18	−0.68/−0.67/0	0.40/0.38/0
C		0/0/0	0.88/0.88/0.14	0.88/1.35/0	0.24/0.88/0	0.88/0.88/0.18	0.23/0.25/0	1.17/1.16/0
D			0/0/0	0/0.67/0.14	−0.67/0/0.14	0/0/0.32	−0.68/−0.67/0.14	0.40/0.38/0.14
E				0/0/0	−0.67/−0.67/0	0/−0.67/0.18	−0.68/−1.20/0	0.40/−0.31/0
F					0/0/0	0.67/0/0.18	−0.01/−0.67/0	1.00/0.38/0
G						0/0/0	−0.68/−0.67/0.18	0.40/0.38/0.18
H							0/0/0	1.01/0.98/0
I								0/0/0

Figure 4.6: SAL values for all possible combinations of figure 4.5. The used format is $S/A/L$. The lower half of the table is not shown for readability. Rows correspond with forecasts, columns with observations (Wernli et al., 2008). For example, BC has a negative S and A. C is considered the observational set and B the model. B underestimates the structure and amplitude with respect to C, hence S and A are negative.

and therefore only the L component is non zero. BE only has an amplitude difference and BF only a structure difference (due to underestimation of the object). H shows an overestimation and I an underestimation of S and A, which is caused by the amplitude and volume difference. However, the difference between F and G when compared to B shows the importance of the choice of the threshold f. The structure of G compared to B is correct, while the location is not. If f is taken lower, G could become similar to F, due to the identification of only a single object. This results in a correct position L, but an incorrect structure S when compared to B. Thus the different L components of F and G, is completely caused by L_2 of equation 4.13.

4.3.6 Object identification algorithms

The stability of the threshold f is linked to uncertainties in the data (see chapter 2). Observational uncertainties should not be omitted in spatial verification methods, and are especially important in the sensitivity of object identification. Errors due to spatial interpolation of model and satellite outcome will also contribute to the uncertainty.

Different algorithms may be used for identifying objects. Wernli et al. (2008) only used a single threshold, derived by a rough estimate of the measurement errors. This approach is rather simplistic, but has a clear physical interpretation. No filtering of small objects have been used, what leads to substantial differences between the SAL analysis and other verification methods in conditions with a low

cloud cover. The location and structure component may be way off, while model predictions of cloud cover are almost correct (Morales and Calvo, 2013). This error arises especially with small objects. Under more cloudy conditions, the objects are much larger and expected to be much closer to the observations (Weniger and Friederichs, 2016). Another algorithm is smoothing the data over a window (for example 3 x 3, Weniger and Friederichs (2016)), using a smoothing radius. This method depends on the choice of the radius. The third algorithm considers only objects of a minimum size. If only considering the median of the SAL distribution, the threshold and smoothing algorithms are sufficient, although the smoothing algorithm is often more stable than the threshold algorithm. Figure 4.7 shows the correlation coefficient for changing parameter values for the different algorithms (threshold, smoothing parameter and minimum size (Weniger and Friederichs, 2016)). The figure displays the weakness of the threshold algorithm, which shows correlation coefficients between 0.4 and 0.8. In figure 4.7, the smoothing parameter method (convthresh) has the highest correlation coefficient compared to the other methods.

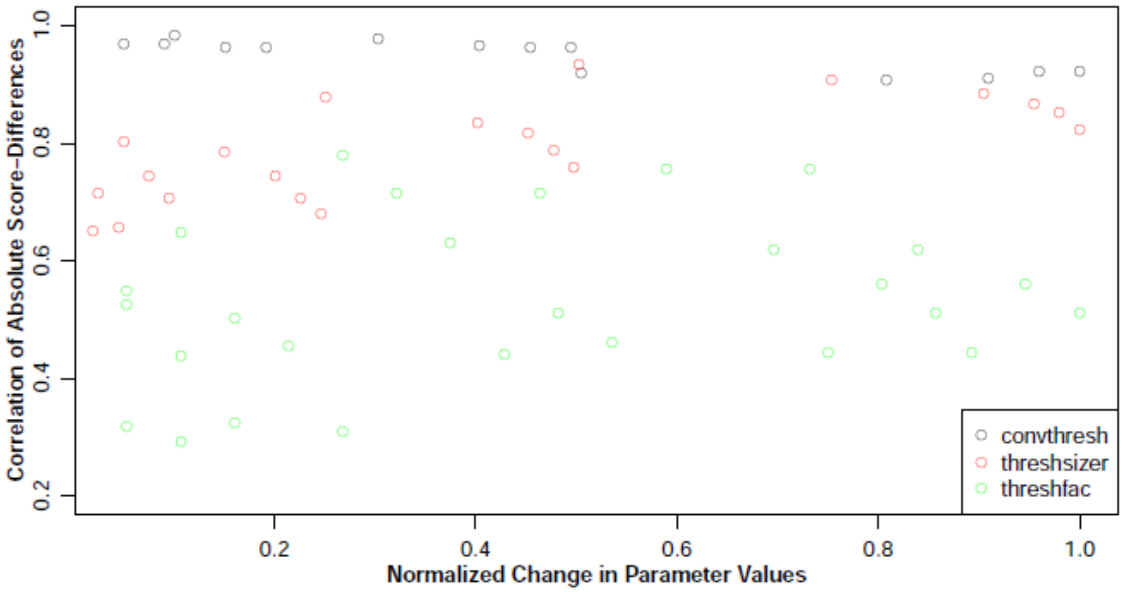


Figure 4.7: Illustration of the correlation coefficient against changing parameter values between S and L_2 for a test case, as performed by Weniger and Friederichs (2016). Thresfac is the threshold method, convthresh the smoothing parameter method and the threshsizer the minimum size method.

Weniger and Friederichs (2016) tests the SAL analysis with hypothetical data sets. These data sets are converted sets of their original data. Weniger and Friederichs (2016) concludes that the ability of SAL to distinguish between data sets depends on the data. Thus it cannot be stated a-priori if SAL is able to calculate S, A and L correctly for a data set.

4.4 Experiments

In this study we use the statistical methods, as described in this chapter, on MSG and the HARMONIE versions of table 3.1. This study covers a case study, the mean cloud cover and an application of all named statistics of this section, including the SAL analysis. Differences between land and sea, day and night and seasonal variations are studied as well. We also compare model output in the climate mode with MSG. The second part consists of a study of different cloud schemes, following section 3.2. This part includes an analysis of the different components contributing to σ_s of equation 3.48 as a function of height.

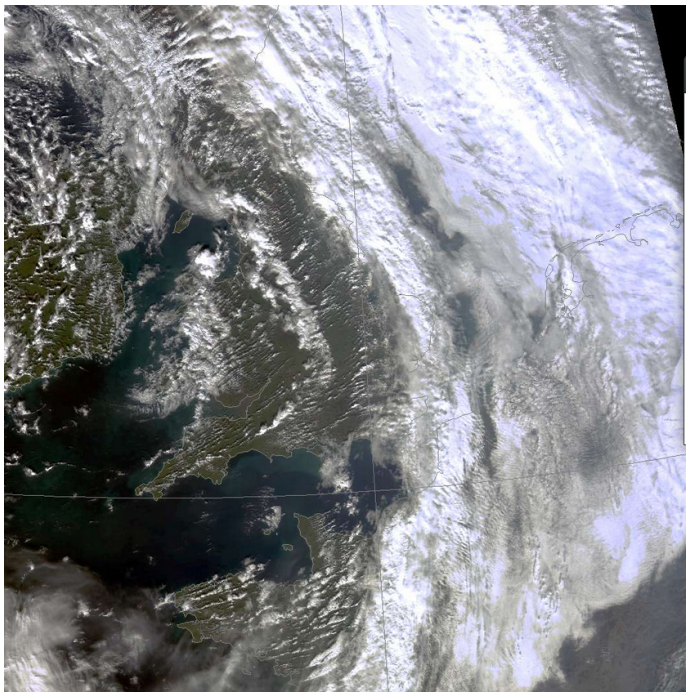
For this study, we consider the following time periods: August 2006; May 2008; March 2012; October 2012 and the year 2012. August 2006 is characterized by record high precipitation for the Netherlands, May 2008 by unusual sunny and warm weather, and March 2012 by slippery road conditions. October 2012 is considered to evaluate any seasonal signals when compared to March 2012. The year 2012 is used to investigate the performance of the versions of HARMONIE on a longer time scale.

5 Results part I: Verification of cloud cover with MSG

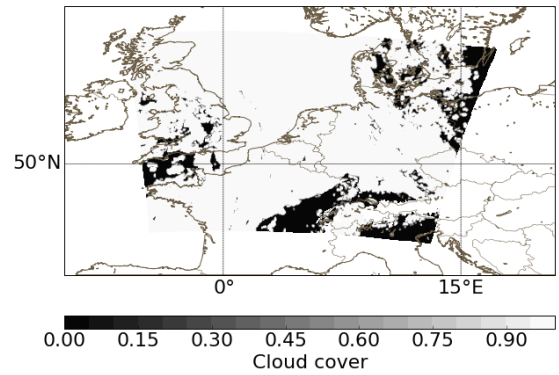
In the following chapter, all results are in HC mode, unless stated otherwise. The results are split in two parts. Part I consists of the evaluation of HARMONIE using MSG. Part II consists of a study regarding cloud schemes.

5.1 Case study

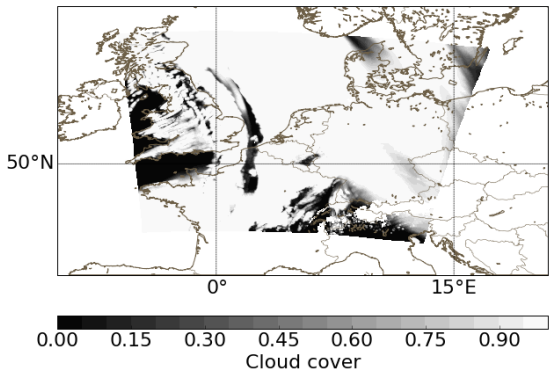
We start the result chapter with two cases. Cases are important to illustrate problems that would be hard to spot with a long term statistical analysis. A snapshot may also depict MSG cloud cover retrieval errors, as is discussed in chapter 2. A case study can be used to locate differences between MSG observations and other measurement sets. Figure 5.1a shows a true color satellite image taken by MODIS on 3 March 2012, 13 UTC (Dundee, 2016). MODIS measures cloud cover and its optical properties, radiation budget and processes in the lower atmosphere (Platnick et al., 2003). This moment is characterized by a mostly closed cloud field in the east and a broken cloud field in the west, alongside some clear sky areas. This image is compared with MSG and HARMONIE data in figure 5.1. Figure 5.1b shows the cloud cover according to MSG for this moment. Figure 5.1c, 5.1d ,5.1e shows respectively HARATU update, HARATU update DA and HARMONIE cyc38. Do note that the domain and the projection of figure 5.1a is different than figures 5.1b - 5.1e. This day is a common day in March. MSG shows in the western part of the domain some overestimations. Both HARATU update and HARATU update DA perform well with respect to MODIS, especially the clear sky band off the east coast of England. This band is missed by HARMONIE cyc38 and is almost absent for MSG. Let us consider another moment that shows what could go wrong with MSG and HARMONIE.



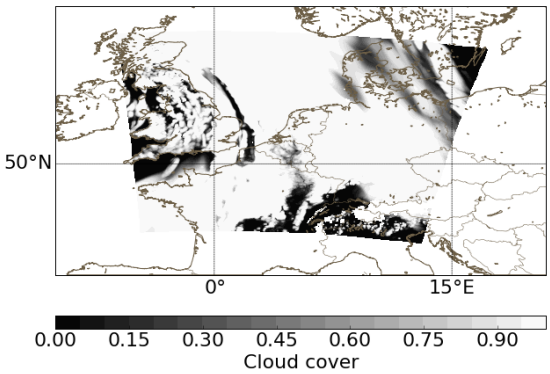
(a)



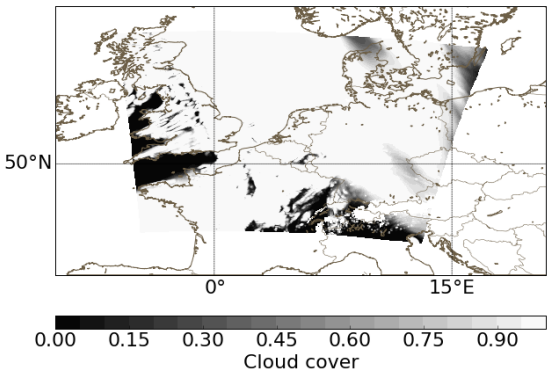
(b)



(c)



(d)



(e)

Figure 5.1: True color satellite image taken by MODIS on 3 March 2012, 13 utc (**a**). Cloud cover during the same moment according to MSG (**b**), HARATU update (**c**), HARATU update DA (**d**), HARMONIE cyc38 (**e**).

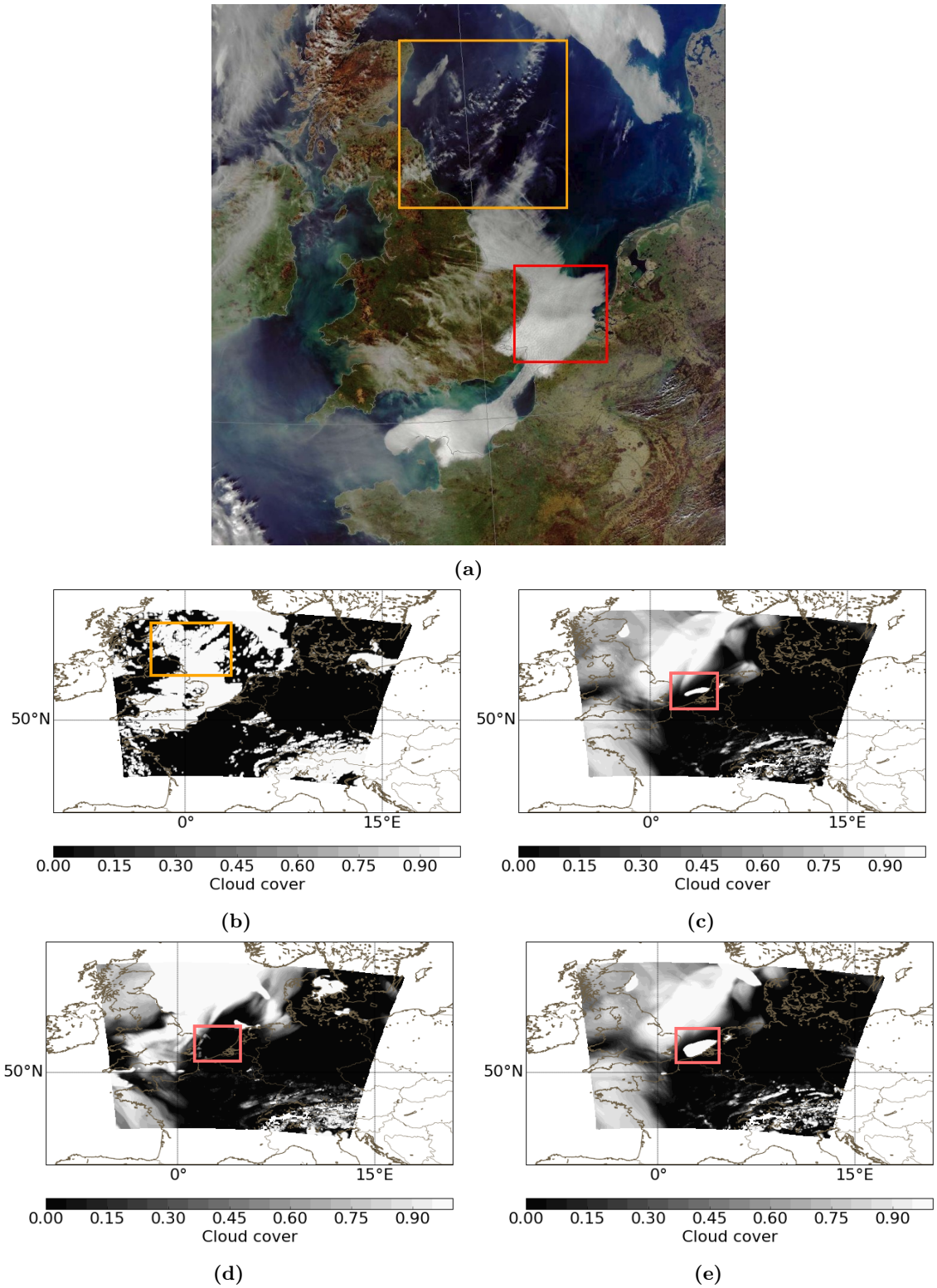
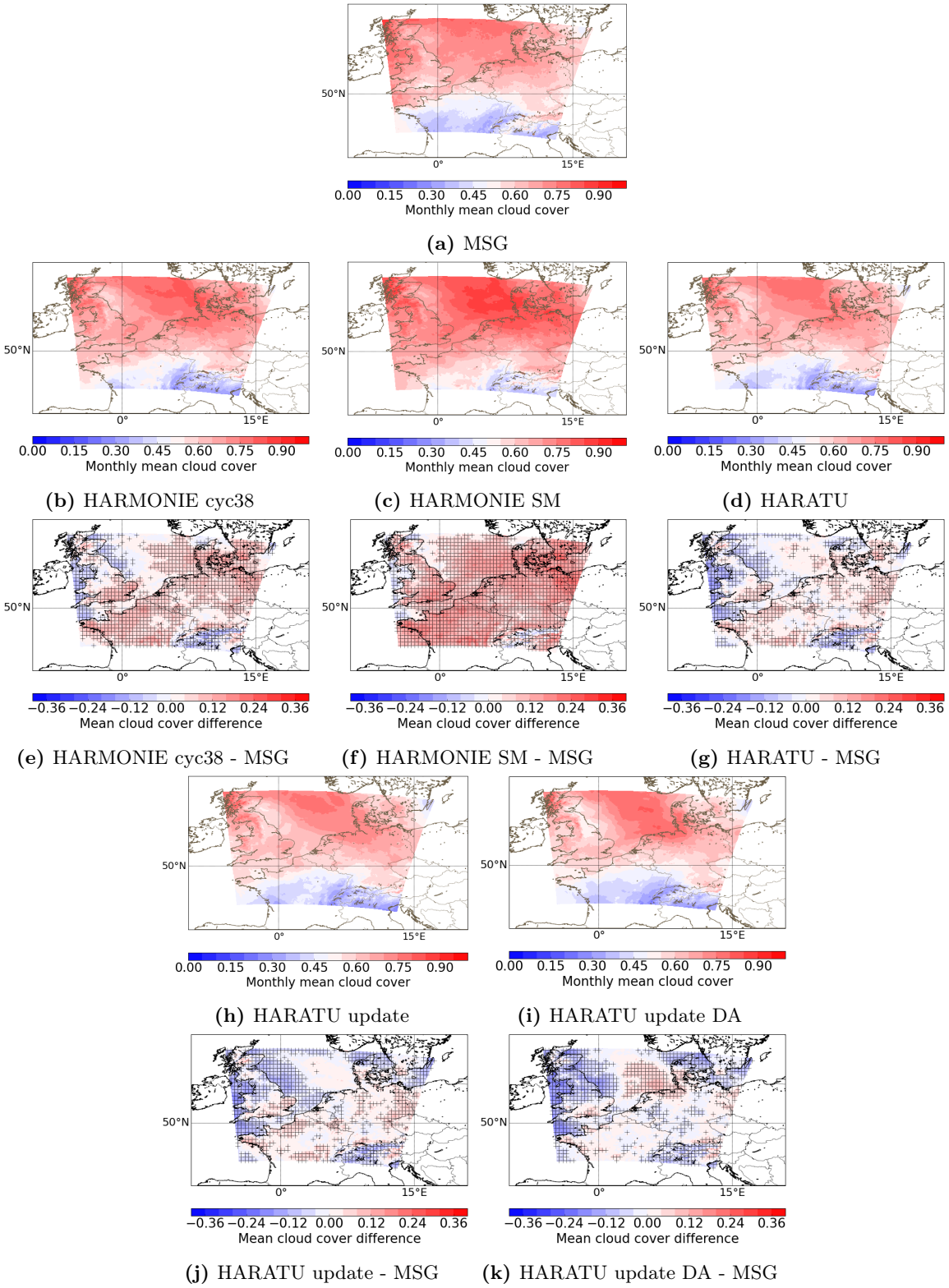


Figure 5.2: True color satellite image taken by MODIS on 25 March 2012, 13 UTC (**a**). The orange box indicates a region of overestimation of MSG. A red box indicates a fog field. The cloud cover field for the same moment according to MSG (**b**), HARATU update (**c**), HARATU update DA (**d**), HARMONIE cyc38 (**e**).

Figure 5.2a shows the true color satellite image taken by MODIS for 25 March 2012, 13 UTC. The total cloud cover according MSG, HARATU update, HARATU update DA and HARMONIE cyc38 are respectively shown in figures 5.2b - 5.2e. The general distribution of clouds of HARATU update is similar to MSG. However, some regions are significantly different, especially the North Sea. As discussed, the MSG data from figure 5.2b shows almost no fractional cloud cover areas, suggesting that many regions have $N = 0.0$ or $N = 1.0$. MSG seems to overestimate the cloud cover in some regions when compared to MODIS, especially in the orange box over the North Sea. Note that cloud cover is mostly overestimated above the sea due to the large albedo difference between surface and cloud (see chapter 2). Cirrus clouds can be present and interpreted as overcast conditions. For this moment, HARMONIE performs relatively bad, with too much cloud cover especially above the North Sea for all three considered versions. The data assimilation run of HARATU update (figure 5.2d) predicts complete overcast above the North Sea. The red box indicates a fog field above the North Sea, which extents beyond the red box into the English channel. All model versions produce a too small fog field in the red box, with HARATU update DA only producing fog outside the box, in a region where no fog is expected. This indicates a possible weakness of HARMONIE regarding fog and/or low clouds.

5.2 Mean cloud cover and observational differences

Mean cloud cover for the considered domain is an important measure of model performance. The average in time has been taken for every grid point. Figure 5.3a shows the mean cloud cover for MSG for March 2012. Figure 5.3 also shows the mean cloud cover for March 2012 for the different versions of HARMONIE in the upper figures. The versions of HARMONIE show approximately the same cloud cover pattern. In the south we observe a relatively low cloud cover, but increases with latitude. The cloud cover is generally higher above sea than land, especially near the North-East coast of England. Some differences between the model versions can be observed. HARMONIE SM shows high cloud cover in the domain, while HARATU update is less cloudy. Note that the land sea difference of MSG is much more distinct than the models. To further investigate the differences between MSG and HARMONIE, we show the mean cloud cover differences between MSG and HARMONIE in the lower figures of figure 5.3. In this figure, bootstrapping is used to determine any significant discrepancy (see section 4.2) and is indicated with a plus. The figure shows what regions and which models deviate significantly from MSG. HARMONIE cyc38 shows an overestimation of the cloud cover, which has also been observed by De Rooy et al. (2014). The overestimation mostly occurs above land. HARMONIE SM severely overestimates the cloud cover. HARATU, HARATU update and HARATU update DA show a similar pattern. Note that HARATU update DA significantly overestimates a region of the North Sea that HARATU update does not. The differences are often larger than the assumed 4% overestimation margin by MSG (see chapter 2), hence land sea differences have to be investigated further.



(a) MSG

(b) HARMONIE cyc38 **(c)** HARMONIE SM **(d)** HARATU

(e) HARMONIE cyc38 - MSG **(f)** HARMONIE SM - MSG **(g)** HARATU - MSG

(h) HARATU update **(i)** HARATU update DA

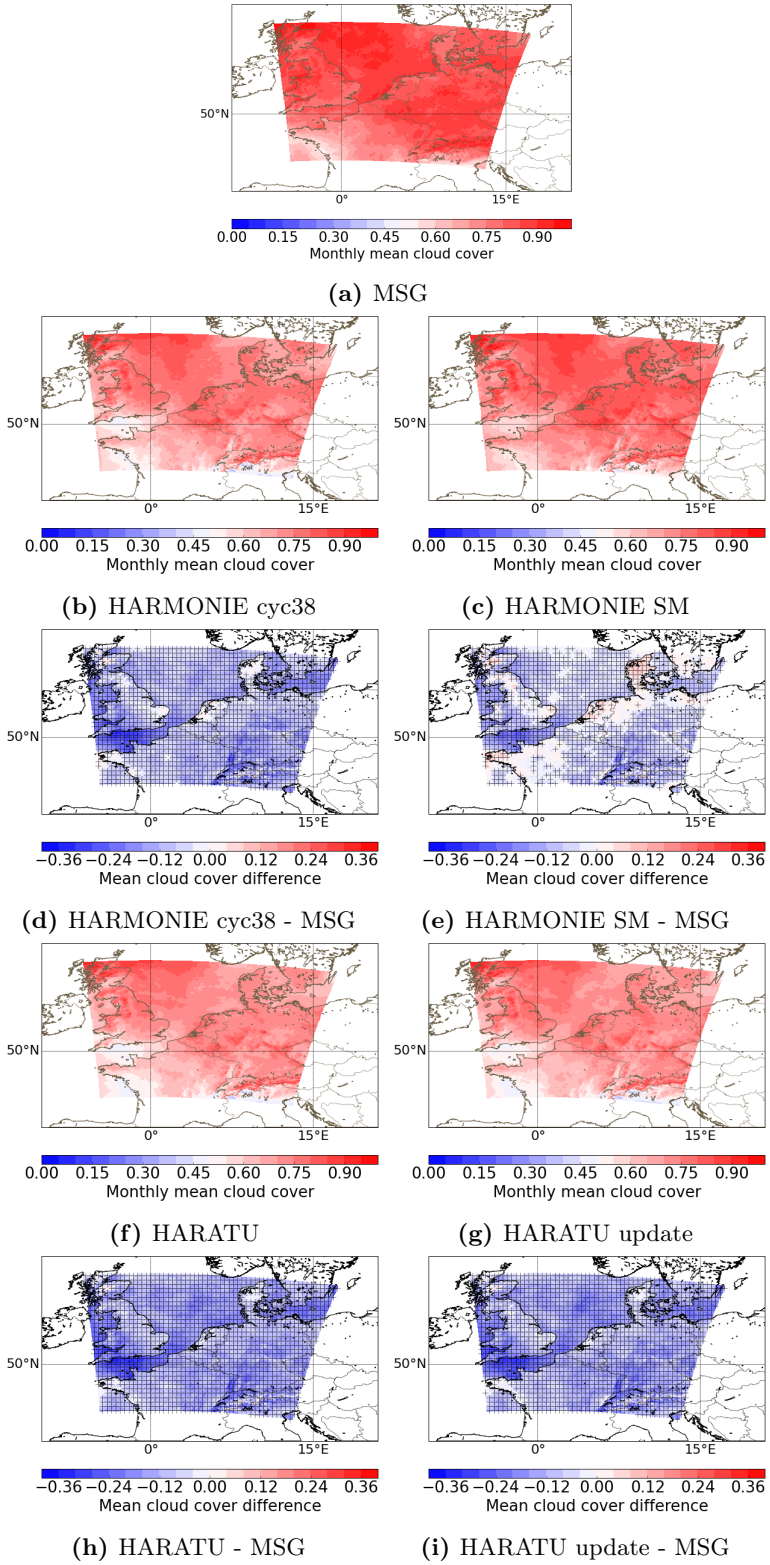
(j) HARATU update - MSG **(k)** HARATU update DA - MSG

Figure 5.3: Mean cloud cover as observed by MSG for March 2012 (**a**). Mean cloud cover for HARMONIE (upper figure) and cloud cover difference between HARMONIE and MSG for March 2012 (lower figure). (**b**) and (**e**) show HARMONIE cyc38, (**c**) and (**f**) show HARMONIE SM, (**d**) and (**g**) show HARATU, (**h**) and (**j**) show HARATU update, and (**j**) and (**k**) show HARATU update DA.

Figure 5.4a shows the mean total cloud cover for August 2006 for MSG. August 2006 is characterized by cloud conditions. The mean cloud cover of HARMONIE and cloud cover difference between MSG and HARMONIE for August 2006 are also shown in figure 5.4. The upper figures represent total cloud cover and the lower figures the cloud cover difference. For this cloudy month, all versions of HARMONIE significantly underestimate the cloud cover when compared with MSG. The differences of figure 5.4 are much larger than the uncertainty of MSG against ground observations. HARMONIE SM shows the best results due to its tendency to produce a high cloud cover, but the total cloud cover is still underestimated.

The total cloud cover for May 2008 according to MSG is shown in figure 5.5a and for HARMONIE cyc38 and HARATU update in figure 5.5b and 5.5c. The differences of HARMONIE cyc38 and HARATU update with MSG are shown in figure 5.5d and 5.5e. This month is characterized by low cloud cover in the North-Eastern corner of the domain. Figure 5.5 shows that the considered HARMONIE versions significantly underestimate the cloud cover for the sea as well as most land regions.

The results of this section show that the HARMONIE versions generally underestimated the cloud cover against MSG above sea. Some regions above land are also underestimated. The performance of the models is dependent on the weather conditions. We also observe that HARATU update performs better than the other versions for March 2012.



(a) MSG

(b) HARMONIE cyc38

(c) HARMONIE SM

(d) HARMONIE cyc38 - MSG

(e) HARMONIE SM - MSG

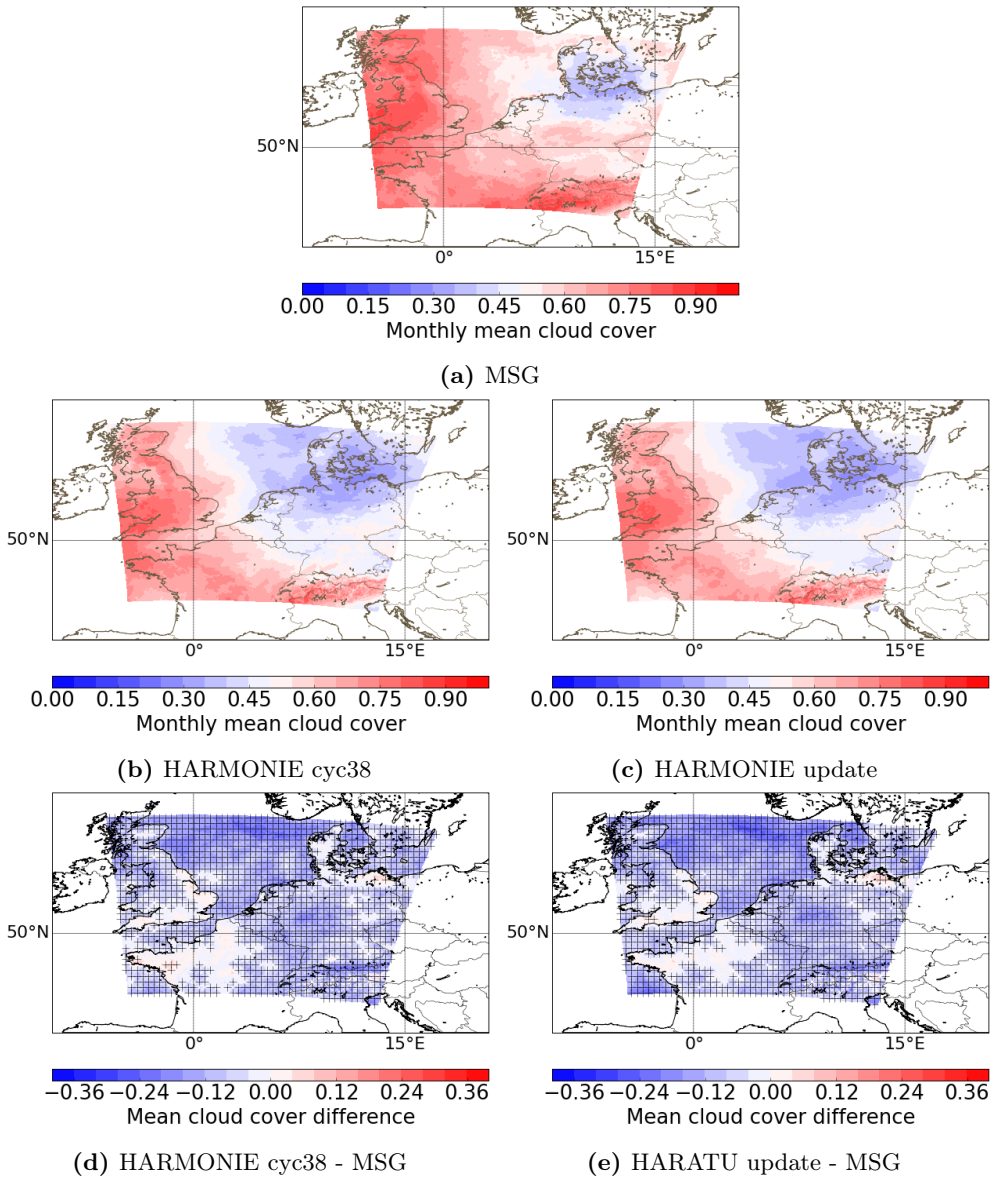
(f) HARATU

(g) HARATU update

(h) HARATU - MSG

(i) HARATU update - MSG

Figure 5.4: Mean cloud cover as observed by MSG for August 2006 (**a**). Mean cloud cover for HARMONIE (upper figure) and cloud cover difference between HARMONIE and MSG for August 2006 (lower figure). (**b**) and (**d**) show HARMONIE cyc38, (**c**) and (**e**) show HARMONIE SM, (**f**) and (**h**) show HARATU, and (**g**) and (**i**) show HARATU update.



(a) MSG

(b) HARMONIE cyc38

(c) HARMONIE update

(d) HARMONIE cyc38 - MSG

(e) HARATU update - MSG

Figure 5.5: Mean cloud cover as observed by MSG for May 2008 (**a**). Mean cloud cover for HARMONIE (upper figure) and cloud cover difference between HARMONIE and MSG for May 2008 (lower figure). (**b**) and (**d**) show HARMONIE cyc38, and (**c**) and (**e**) show HARATU update.

5.3 Absolute mean bias difference and histograms

To be able to assess the performance of the models against each other, we consider the absolute mean bias difference (see section 4.2). Due to the difference between model and observations, which can be positive and negative, the calculated bias (equation 4.1) may result in a wrong impression, as positive and negative differences might compensate. It is therefore convenient to take the absolute values of the bias. This is a measure of the quality of the model. When the absolute mean bias is close to zero, it means that the model output is close to the observations. The same process can be done for another model version. When the absolute mean bias of the model versions are extracted, it

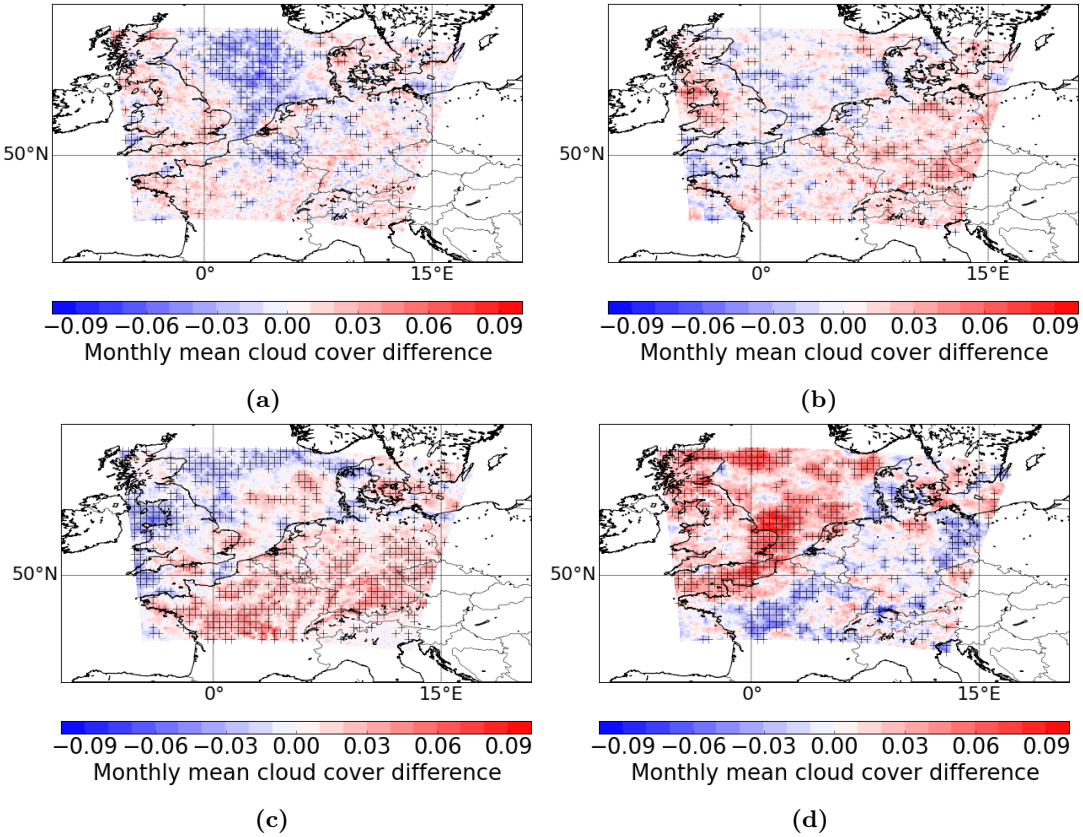


Figure 5.6: The difference between the absolute mean bias of HARMONIE cyc38 and MSG, and HARATU update and MSG for (**a**) August 2006, (**b**) May 2008 and (**c**) March 2012. Figure (**d**) shows the difference between the absolute mean bias of HARMONIE update DA and MSG, and HARATU update and MSG for March 2012. A plus resembles a significant change as determined by bootstrapping. Red shows that HARATU update performs better than HARMONIE cyc38 for (**a**), (**b**) and (**c**), and better than HARATU update DA for (**d**).

shows which version performs better. In figure 5.6, the absolute mean bias difference between HARMONIE cyc38 and MSG, and HARATU update and MSG, now called |HARMONIE CYC38 − MSG| − |HARATU UPDATE − MSG|, is shown for August 2006, May 2008 and March 2012. Figure 5.6d shows |HARATU UPDATE DA − MSG| − |HARATU UPDATE − MSG| for March 2012. To be more precise: HARATU update performs better (worse) with respect to HARMONIE cyc38 for red (blue) areas and for figure 5.6d than HARATU update DA. A plus represents significant change as is determined by bootstrapping. When using the absolute mean bias we lose information about the sign of the deviation of the model with respect to MSG. This information is provided in section 5.2.

For August 2006 we see a worsening of the model for the North Sea. This corresponds with the unusual high cloud cover for this region. The cloud cover in HARATU update is generally lower, so during extensive cloud cover periods it performs worse than HARMONIE cyc38. However, figures 5.6a - 5.6c show an increased performance of HARATU update above land when compared to the reference version HARMONIE cyc38. Figures 5.3 - 5.6 reveal that the performance of models depends on the weather conditions, land sea contrast and possibly seasons. It is remarkable that the data assimilation mode performs significantly worse, mostly above sea. The conclusion that HARATU update has difficulties above sea also hold in data assimilation mode.

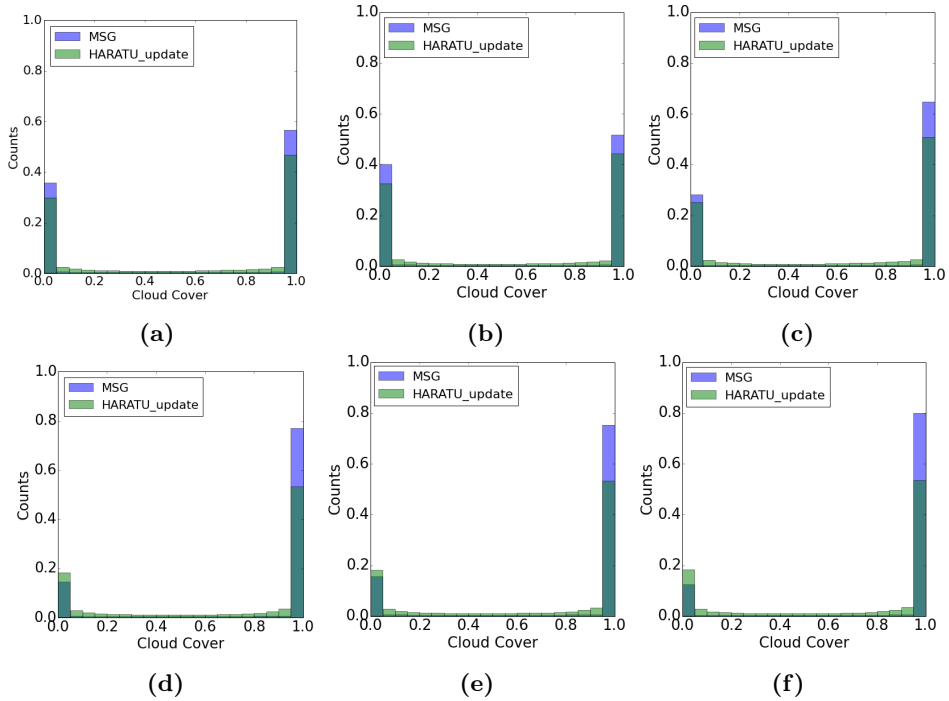


Figure 5.7: Fractional cloud cover occurrence for MSG and HARATU update for March 2012: (**a**) both land and sea, (**b**) land, (**c**) sea, and for August 2006: (**d**) both land and sea, (**e**) land, (**f**) sea. Blue shows the MSG data, green HARATU update and dark green when both data sets overlap.

The impact of the binary cloud scheme of MSG has to be investigated. Due to regridding of the MSG data, some fractional cloud cover can occur. Comparing a mostly binary data set like MSG with a fractional data set like HARMONIE causes uncertainties. It is therefore necessary to investigate this further, especially differences between land and sea. Figure 5.7 shows histograms of the fractional cloud cover occurrence for MSG and HARATU update for March 2012 (upper row) and August 2006 (lower row) for both land and sea, only land and only sea. In blue is MSG, green is HARATU update, and dark green is when both MSG and HARATU update overlap. Every grid point is considered, including all time steps. Note that only a small part of the MSG data has a fractional cloud cover, and is much smaller than the fractional cloud cover of HARMONIE. If only land is considered for March 2012 (figure 5.7b), MSG is both for $N = 0.0$ and $N = 1.0$ equally larger than HARATU update. This is caused by the lack of fractional cloud cover in MSG. Overcast conditions are dominant above sea, but here the difference between MSG and HARATU update is larger than for land (figure 5.7). The difference is beyond the assumed MSG uncertainty range of 4% and can be considered as a significant signal. For August 2006, the figure shows the abundance of clouds, as was observed before. The model slightly overestimates the clear sky conditions and severely underestimates overcast conditions. As expected, HARATU update performs better above land (figure 5.7e) than sea (figure 5.7f). Figure 5.7 supports the hypothesis that the performance of HARMONIE is dependent on the weather conditions (cloudy conditions perform worse) and surface type (land or sea). A further analysis on the land sea difference is done in section 5.8.

5.4 Root-Mean-Square Error

To improve our quality analysis, more advanced statistical tools will be used. The root-mean-square error (RMSE) is an efficient method to measure the accuracy of the data (defined in section 4.2). A

large difference on average between model and observations results in a high RMSE. HARMONIE can produce all possible cloud fractions, while MSG mostly provides a binary scheme, as is seen in the previous sections. The grid point based differences in cloud cover N between MSG and HARMONIE are often large. For example, MSG produces a cloud cover of $N = 1.0$, while HARMONIE produces $N = 0.6$. MSG retrieval algorithms do not allow a cloud fraction of $N = 0.6$ and produces $N = 1.0$ instead (due to regridding, some regions can have a fractional cloud cover with MSG, see section 4.1), hence a large RMSE can be expected. Equation 4.2 shows that large deviations contribute strongly to an increased RMSE.

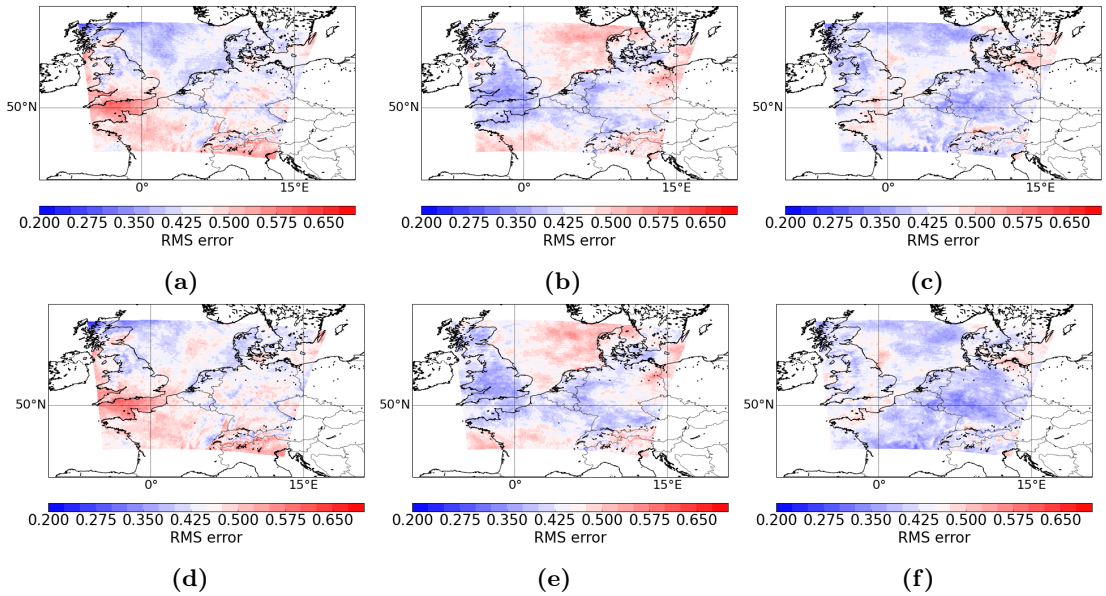


Figure 5.8: RMSE between MSG and HARMONIE cyc38 for (**a**) August 2006, (**b**) May 2008 and (**c**) March 2012. The RMSE between MSG and HARATU update are shown for (**d**) August 2006, (**e**) May 2008 and (**f**) March 2012.

Figure 5.8 shows the RMSE for August 2006, May 2008 and March 2012 between MSG and HARMONIE cyc38, and MSG and HARATU update. Again, the model performs better for March 2012 than for August 2006 and generally better above land. Note that the RMSE is large, especially for August 2006 and May 2008. Cloud fractions are only between 0.0 and 1.0, hence a RMSE of > 0.5 is substantial. As is explained before, this is mostly due to the binary cloud scheme of MSG. However, one could still argue about the quality of HARMONIE. Note that the differences between HARMONIE cyc38 and HARATU update are relatively small. Nonetheless, RMSE is still a valid method to assess the different models. Models with a low RMSE produce forecasts closer to the observations. To investigate the effect of the binary cloud scheme of MSG, let us consider the mean RMSE in time and space on the domain while changing the time steps taken to perform this mean. More time steps are taken to calculate the total cloud cover. This results in the data being transformed from hourly data to daily averaged. The binary scheme of MSG is averaged, and more fractional cloud cover appears if more time steps are considered. It therefore becomes more similar to the model output. As a result, the RMSE is expected to decrease with increasing time steps used to compute an average. Figure 5.9 shows the result.

The figure shows how the RMSE decreases for HARMONIE compared to MSG for March 2012 when the time taken to calculate the average is increased. Thus on the left we have the initial model output, using every hour for the calculation of RMSE. We observe the expected pattern, with the RMSE decreasing with increasing time steps taken to compute the average. We conclude that the large RMSE error is caused mostly by the binary cloud scheme of MSG.

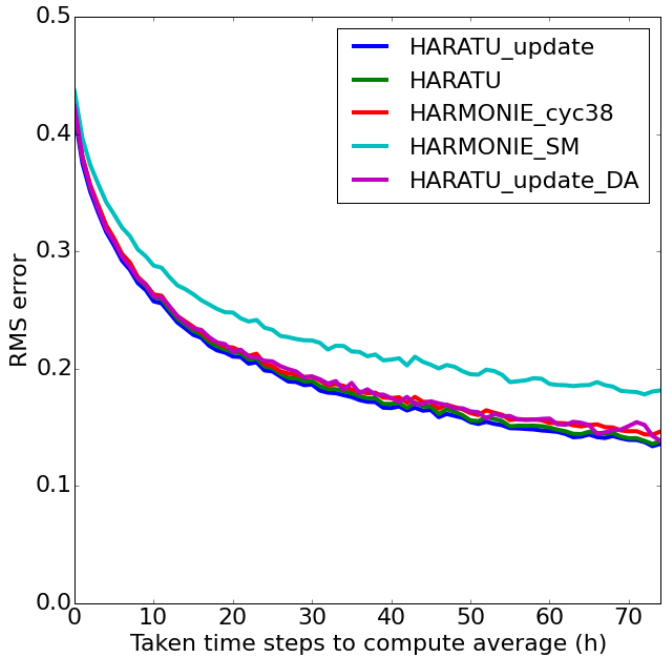


Figure 5.9: RMSE between cloud cover of MSG and HARMONIE as a function of time taken to compute the average for March 2012.

5.5 Correlation coefficient

The quality of model output compared to measurements can be estimated using the correlation coefficient, as is defined in equation 4.6. Remember that $r = 1$ corresponds to a perfect correlation. The correlation coefficient between the different HARMONIE versions and MSG data have been tested as a function of the grid size. With every step, more points around the initial grid point are used for interpolation to a new grid (see section 4.1). The grid then becomes increasingly coarse. Errors in both MSG and HARMONIE average out when interpolated to a new grid. The correlation coefficient is therefore expected to increase with increasing grid size. Figure 5.10 shows the correlation coefficient of different HARMONIE versions against MSG data as a function of grid size. In the figure, "7x7" represents the averaging of 7 by 7 original grid points, to form a new grid point (see figure 4.1). The correlation coefficients for March 2012 are shown in figure 5.10a, May 2008 in figure 5.10b and August 2006 in 5.10c.

The results are comparable with Haiden et al. (2015). According to Haiden et al. (2015), increasing correlation corresponds with increasing grid size. The correlation coefficient keeps improving with grid size due to the continuation of averaging out errors. An important feature of figure 5.10, is the position of the different model versions of HARMONIE, which is an indication of the performance of the model. The higher the model is positioned in figure 5.10, the better it correlates with MSG. The differences between the models can be very small and statistical significance has to be tested. Using the theory of significance between dependent variables (see appendix B.2), p-values can be calculated between models. The differences between the models prove to be statistically significant, due to the large amount of data points, which is in the order of hundred thousands or more. The corresponding p-value (< 0.01) is below the threshold p-value of 0.05. One might consider that such a small correlation

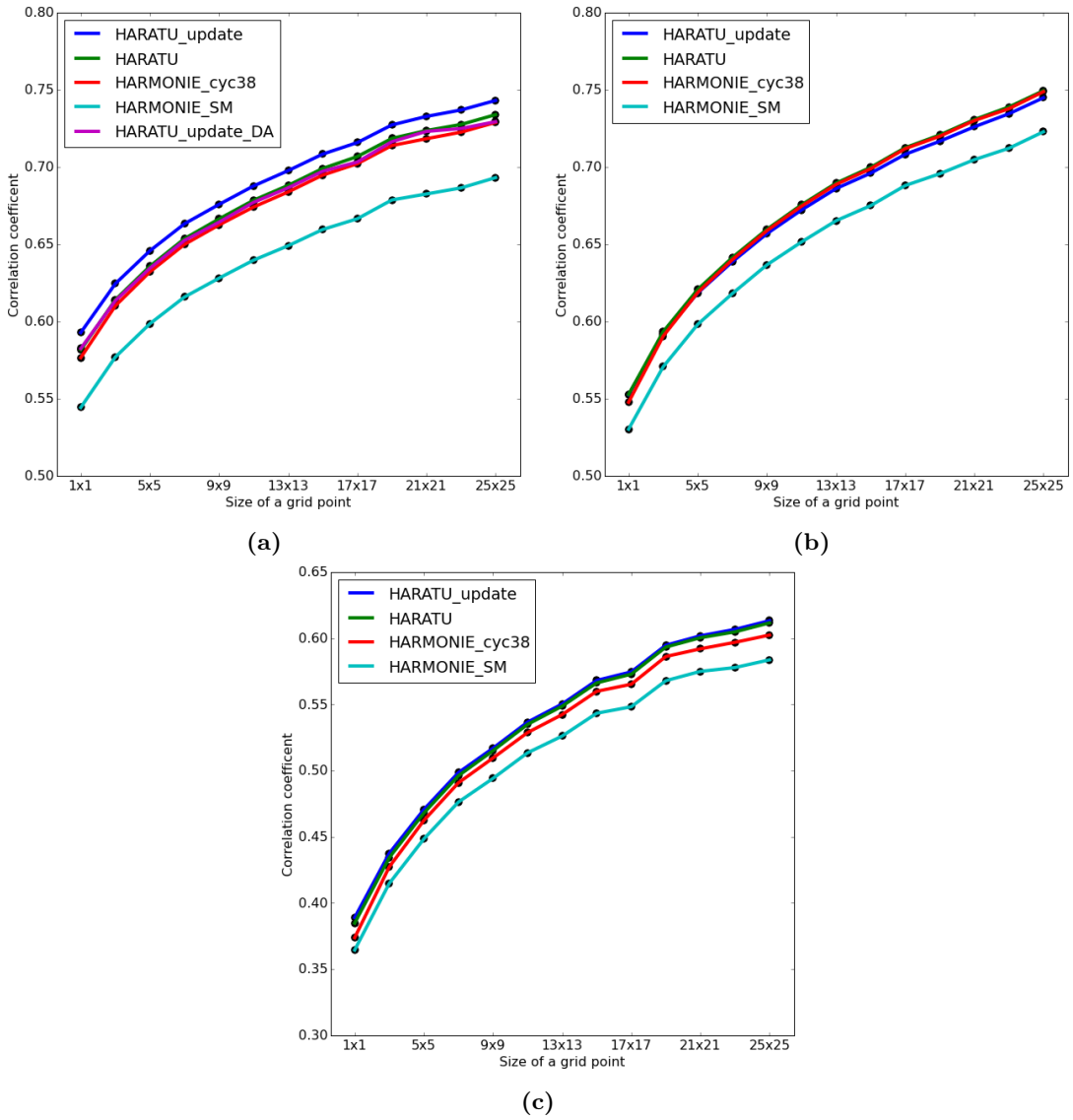


Figure 5.10: Correlation coefficient for different HARMONIE versions as a function of grid size, for March 2012 (**a**), May 2008 (**b**) and August 2006 (**c**).

coefficient difference (∼0.01) is physically unimportant. However, the relative position of each model does provide us information. The relative difference between the models remain approximately the same for small grid sizes. The relative difference between the model versions mostly deviate beyond 10 by 10. Consequently, we can use a 7 by 7 grid for computational intensive calculations. The results can be compared easily with coarser resolution models like the EC-Earth model of ECMWF, because the conclusions still hold for the resolution of HARMONIE, which is between 5 by 5 and 9 by 9. In figure 5.10, HARMONIE SM (cyan) performs the worst for all months. The other versions of HARMONIE are relatively close to each other for all considered months, especially for May 2008, but are much higher than HARMONIE SM. This suggests that the changes made in the microphysics scheme of HARMONIE contribute the most to a better performance of the cloud cover forecast. For March 2012 and August 2006, a gradual increase in performance to newer versions of HARMONIE is observed. HARATU update therefore performs best for March 2012 and August 2006. Consequently, we conclude

that HARATU update often provides the best results or at least comparable with the other versions of HARMONIE. Figure 5.10c shows that all model versions perform worse during August 2006 than the other months. The correlation coefficient is in the order of 0.2 lower than the correlation coefficient in March 2012.

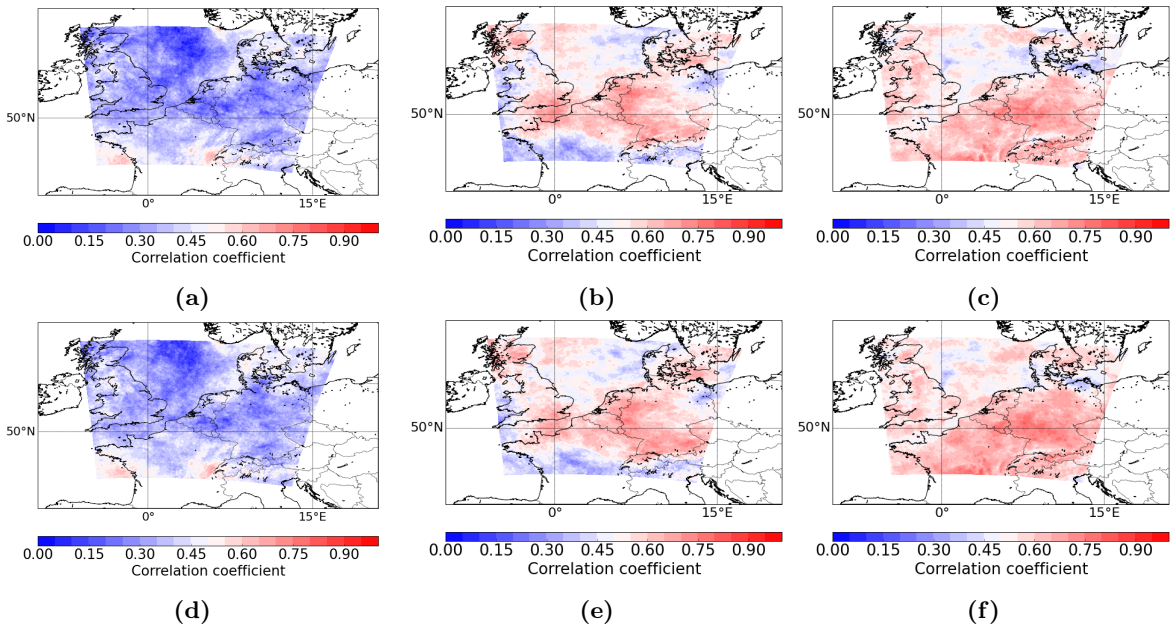


Figure 5.11: The correlation coefficient between MSG and HARMONIE cyc38 for (**a**) August 2006, (**b**) May 2008 and (**c**) March 2012. The correlation coefficient between MSG and HARATU update are shown for (**d**) August 2006, (**e**) May 2008 and (**f**) March 2012.

Figure 5.11 provides an overview of correlation coefficients on the domain. In the upper row the correlation coefficient between MSG and HARMONIE cyc38 are shown and in the lower row for MSG and HARATU update. The left column represents August 2006, followed by May 2008 and March 2012. The model generally has a higher correlation coefficient above land than above sea. However, the results of August 2006 shows a low correlation coefficient, suggesting the correlation coefficient is dependent on the weather conditions. The spatial variability and temporal or seasonal variability can be large. The differences between HARMONIE cyc38 and HARATU update are small.

5.6 Taylor diagram

The Taylor diagram (Taylor, 2001) is a diagram used for comparing the same variable for different models. The Taylor diagram combines multiple parameters that can be used for a quality check (see section 4.2). The total cloud cover for the different months and models are verified against MSG. The Taylor diagram consists of the standard deviation on the x and y axis. The MSG standard deviation is given as a star on the x axis and follows the quarter circle toward the y axis. The standard deviation is therefore the radial distance to the origin. If the spread in the model is estimated correctly, it should be on the dashed line. Circles from the MSG point on the x axis are drawn, representing the RMSE. The correlation coefficient is related to the angle, with highest correlation coefficients closest to the x axis. Figure 5.12 shows the results for the considered HARMONIE versions against MSG for August 2006, May 2008 and March 2012. For the considered months, the standard deviation of the models is relatively close to the prescribed standard deviation by MSG. The order of magnitude of the spread is therefore approximately correct for the models. Note that during the extensive cloud cover

month August 2006, the standard deviation of the models and MSG is lower than during May 2008 and March 2012. Due to the high cloud cover, measurements are more often close to the expected value, thus the standard deviation is reduced. The RMSE is always large, but almost the same for the different months and models and is already described in section 5.4. The correlation coefficient does vary per month and model. When the correlation coefficient is considered, HARMONIE SM performs the worst during most circumstances, and HARATU update performs the best. During the high cloud cover month August 2006, HARMONIE performs significantly worse when compared with the other months, resulting in the model points close to the y axis. The spatial variability is also slightly overestimated, resulting in an overestimation of the standard deviation. These results are in accordance with the results of section 5.5. After all, we see that the model versions do not differ much, but that HARATU update performs slightly better in general.

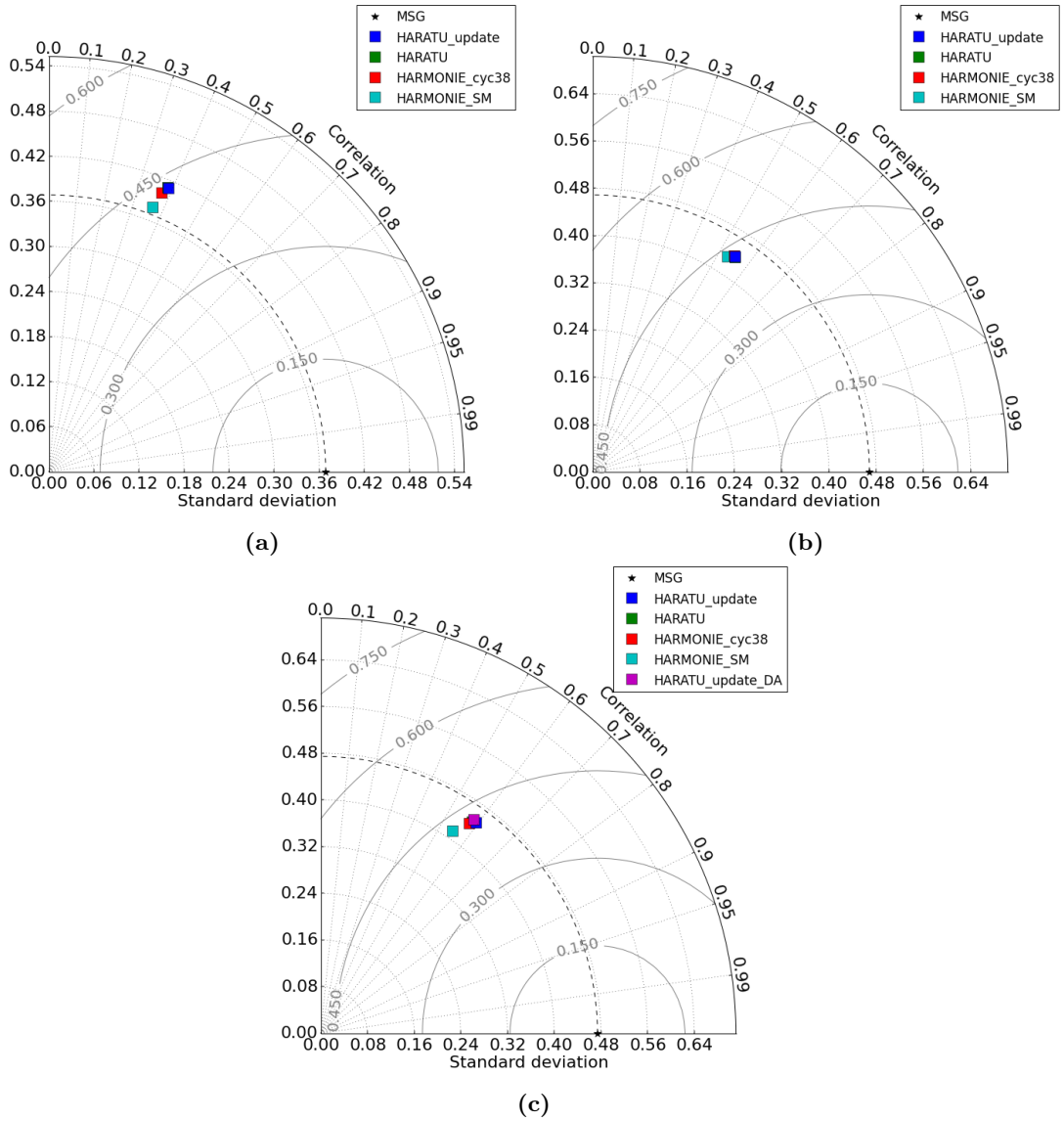


Figure 5.12: Taylor diagram (Taylor, 2001) for August 2006 (**a**), May 2008 (**b**), March 2012 (**c**). The star on the x axis represents the MSG data set.

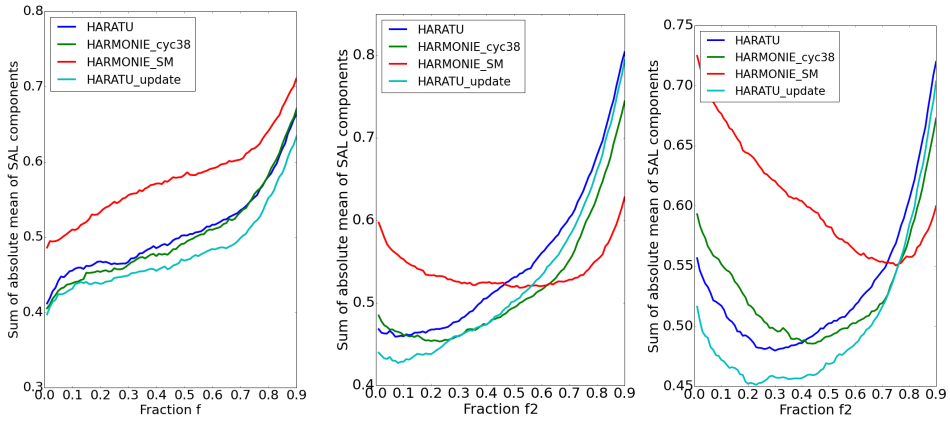
5.7 SAL analysis

In this section we discuss the application of SAL on the data set. The SAL analysis is discussed in section 4.3. We use the SAL analysis to obtain information about the performance of HARMONIE with respect to MSG, regarding the location and structure of clouds.

5.7.1 Threshold analysis

Before the SAL method can be applied, a threshold analysis on our data set has to be done. Finding a minimum in the sum of the structure, amplitude and location components can lead to an optimum value for the threshold. Figure 5.13a shows $|S| + |A| + |L|$ as a function of the threshold f for March 2012. The different versions of HARMONIE are indicated in colours. The quantity $|S| + |A| + |L|$ is a measurement of the correctness of the model compared to the observations. The lower it is, the closer the model resembles the observations. One would expect the sum of the absolute mean of the SAL components to increase with increasing f. An increasing f results in smaller objects. Smaller objects are harder to predict in terms of the structure and location. For example, the location of a large cloud is easily forecasted, while the location of a small cloud is not. A small increase of the volume of a cloud has a relatively large effect on a small cloud. This leads to higher SAL components. Observational uncertainties also contribute to the instability of the threshold, as is described by Weniger and Friederichs (2016). Therefore one would expect $|S| + |A| + |L|$ to be higher close to $f = 0.0$. An optimum is thus predicted. Figure 5.13a shows the expected behaviour for increasing f. No increase of $|S| + |A| + |L|$ is observed close to $f = 0.0$. From this figure one could not extract an optimum for f to use in further analysis. The figure does illustrate the improvement of the models. HARMONIE SM performs the worst with respect to MSG, while HARATU update performs the best.

The model and the observations both provide a different uncertainty. It is therefore convenient to split the threshold f in two parts, f_1 for the observations and f_2 for the model. An estimation of the uncertainties in the observations is given in chapter 2. Estimating the uncertainty range of HARMONIE a-priori is more difficult. Therefore f_2 is set to vary and f_1 is taken constant.

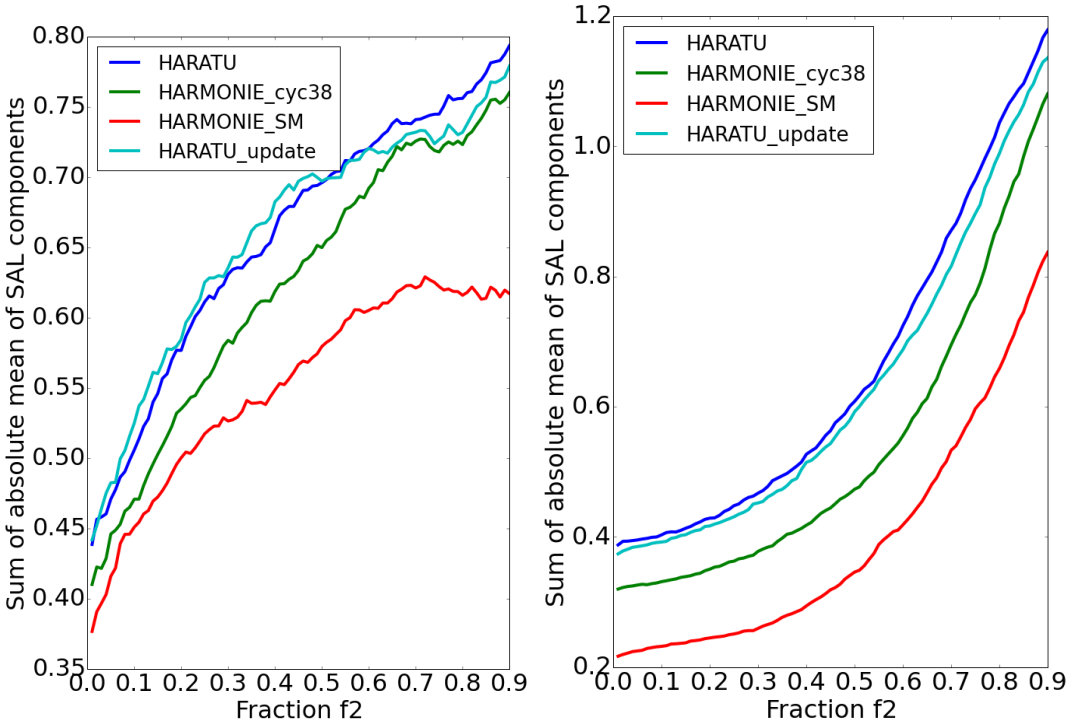


(a) **(b)** **(c)**

Figure 5.13: (**a**) $|S| + |A| + |L|$ as a function of f, (**b**) as a function of f_2 with $f_1 = 0.2$ and (**c**) as a function of f_2 with $f_1 = 0.5$ for March 2012, using MSG data as observations. Colours indicate different versions of HARMONIE: blue HARATU, green HARMONIE cyc38, red HARMONIE SM, cyan HARATU update.

Figures 5.13b and 5.13c show $|S| + |A| + |L|$ as a function of f_2, while taking $f_1 = 0.2$ for figure 5.13b and $f_2 = 0.5$ for figure 5.13c. The results are vastly different, compared to figure 5.13a. Now

an optimum of f_2 is observed. However, the location of the optimum of f_2 strongly depends on the chosen HARMONIE version. The location of the optimum of f_2 is dependent on the cloud cover of the model version. A low cloud fraction leads to a low optimum of f_2. When the cloud cover is low, a low threshold performs the best due to the relative abundance of small clouds. The difference between figure 5.13b and figure 5.13c is also large, especially the location of the optimum of f_2 is different. For these results, it is inconclusive what value to use for f_2 that is valid for all model versions. The choice of f_1 is also uncertain.



(a) **(b)**

Figure 5.14: Calculating $|S| + |A| + |L|$ as a function of f_2, for $f_1 = 0.5$, during May 2008 (**a**) and August 2006 (**b**), using MSG data as observations. Colours are the same as figure 5.13a.

Figure 5.14 shows the result when $f_1 = 0.5$ for May 2008 and August 2006. It shows a completely different result compared to figure 5.13. No optimum is visible. Note that HARATU and HARATU update perform worse than HARMONIE SM and HARMONIE cyc38 in terms of $|S|+|A|+|L|$. This is likely caused by the weather conditions during these months. These months have a high cloud cover, so the model versions that also produce a high cloud cover performs the best (HARMONIE SM). Figures 5.13 and 5.14 show a large variability per month. Considering figures 5.13 and 5.14, we are unable to choose an optimal threshold from this method that is valid for all months and model versions. For this study we have chosen a single threshold of $f = 0.5$, which is approximately the optimum of HARMONIE cyc38 in figure 5.13c. We choose f to be sufficiently high enough to incorporate the uncertainties in MSG and HARMONIE. The smoothing parameter and minimum size method of section 4.3.6 are tested as well, but prove to perform worse than the threshold parameter method for these months and model versions.

5.7.2 SAL analysis on cloud cover of HARMONIE and MSG

In this research we use the SAL analysis to study the strength and weakness of HARMONIE. The sum of $|S|$, $|A|$ and $|L|$ can be used, which is an estimate of the performance and spread. The strength of the SAL method is to distinguish model output performance considering the structure, amplitude and location. The SAL diagram, as was first produced by Wernli et al. (2008), shows the result for total cloud cover. The value of each of the SAL components is calculated hourly for the whole domain, and is relative to the MSG satellite data. Figure 5.15 shows a SAL diagram for March 2012 ($f = 0.5$). The figure shows the S component against the A component, coloured by the L component. Median values of S and A are the non-dashed lines. The 25 to 75 percent of the distribution of A and S are represented by the grey box. In the upper right corner is the sum of the absolute values of S, A and L.

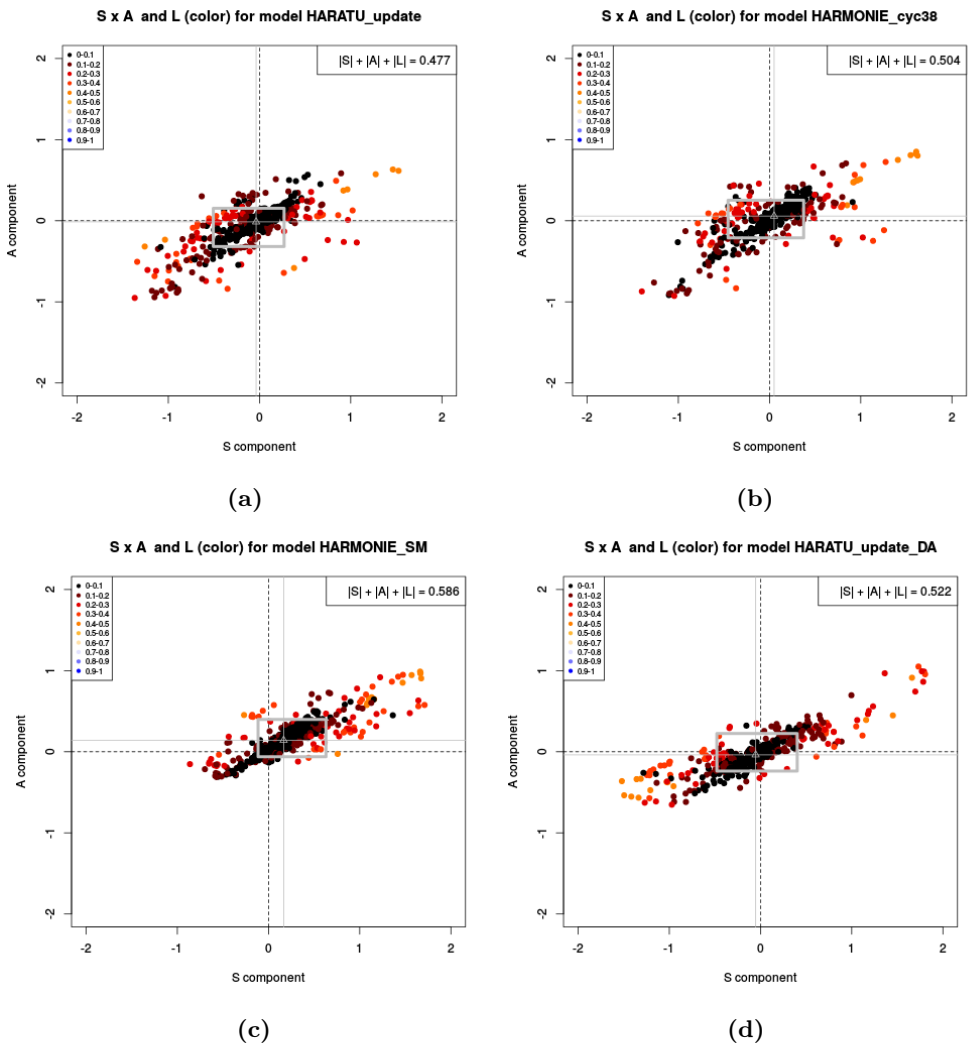


Figure 5.15: SAL diagrams for hourly cloud cover forecast compared to MSG for March 2012 of (**a**) HARATU update, (**b**) HARMONIE cyc38, (**c**) HARMONIE SM and (**d**) HARATU update DA. On the x (y) axis the S (A) component. Colour indicates the L component. The median values of S and A are the non-dashed lines. The grey box is 25 to 75 percent of the distribution of A and S. In the upper right corner is the sum of the absolute values of S, A and L. The threshold $f = 0.5$ is used.

For HARATU update (figure 5.15a), the median of the SAL components is around 0.0 for both A and S. For HARMONIE cyc38 (figure 5.15b), the median of both A and S are slightly positive, but very similar. An overestimation of the cloud cover for HARMONIE SM is observed, especially the structure (figure 5.15c). The result for HARMONIE update DA is similar to HARATU update (figure 5.15d). The spread of the data points is still large, but most points are close to the main diagonal through the top right quadrant and the lower left quadrant. The spread is smaller for HARATU update, indicating that it performs better than HARMONIE cyc38 ($|S| + |A| + |L|$ is smaller for HARATU update than HARMONIE cyc38). The high density of calculated points around the diagonal indicates a correlation between the amplitude and the structure of the cloud cover. This is expected, due to the often occurring $N = 1.0$ or $N = 0.0$. An increase in the cloud cover volume is often linked to a higher mean cloud fraction. Most points are therefore found in the top right quadrant (overestimation of A and S) and the lower left quadrant (underestimation of S and A). However, the entities do not follow exactly the diagonal, but follow a line with a lower slope. This reveals that HARMONIE struggles with the structure component more often than the amplitude component. The information acquired from the SAL analysis provides us new spatial information, which we were unable to acquire with the other methods. Other months have been considered as well, which are shown in figure 5.16.

Figure 5.16 shows the SAL diagrams for August 2006 (5.16a and 5.16b) and May 2008 (5.16c and 5.16d) for HARATU update and HARMONIE cyc38. The results for both August 2006 and May 2008 show us a different distribution than figure 5.15. More entities are found in the lower left quadrant. Consequently, the structure and amplitude component are systematically underestimated. Especially the structure component is strongly underestimated, which can be as large as -1.5 (remember that a factor of -1 corresponds with an underestimation of the domain averaged cloud cover by a factor 3). The amplitude component is mildly underestimated, to -0.5, what corresponds with an underestimation of approximately 1.7 with respect to the observations. The spread of points in May 2008 is larger than for August 2006. This is likely caused by the variable weather during May 2008. August 2006 was characterized by high precipitation and cloud cover, suggesting less variability. The differences between HARATU update and HARMONIE cyc38 are small for the SAL method for May 2008. For August 2006, the differences between HARATU update and HARMONIE cyc38 are larger. Regarding $|S| + |A| + |L|$ for May 2008 and August 2006, we observe that HARMONIE cyc38 performs better than HARATU update. After all, we conclude that especially the structure (volume) of the cloud cover is often underestimated by the models during high cloud cover conditions.

The results of the SAL analysis of this study can be compared to other studies. Crocker and Mittermaier (2013) used the threshold method to apply the SAL analysis on cloud cover forecasts of the United Kingdom. The observational data set of Crocker and Mittermaier (2013) is also MSG. The used threshold was 0.4 and 0.2, although Crocker and Mittermaier (2013) concluded that their results are dependent of the choice of the threshold, which is in accordance with Weniger and Friederichs (2016). The results of Crocker and Mittermaier (2013) are comparable with figures 5.15 and 5.16. Most of their results also follow the main diagonal to some extent and are positioned in the upper right and lower left quadrant, concluding that S and A are correlated. The results of Morales and Calvo (2013) are similar, with most points following the main diagonal. They conclude that SAL can be used for ranking models and to acquire new spatial information.

5.7.3 Time analysis of SAL components

Using the SAL method to consider a whole month can be used for a statistical analysis, but one loses information about the performance of the models during different weather conditions. Let us look at the evolution of SAL during March 2012 for the different model versions.

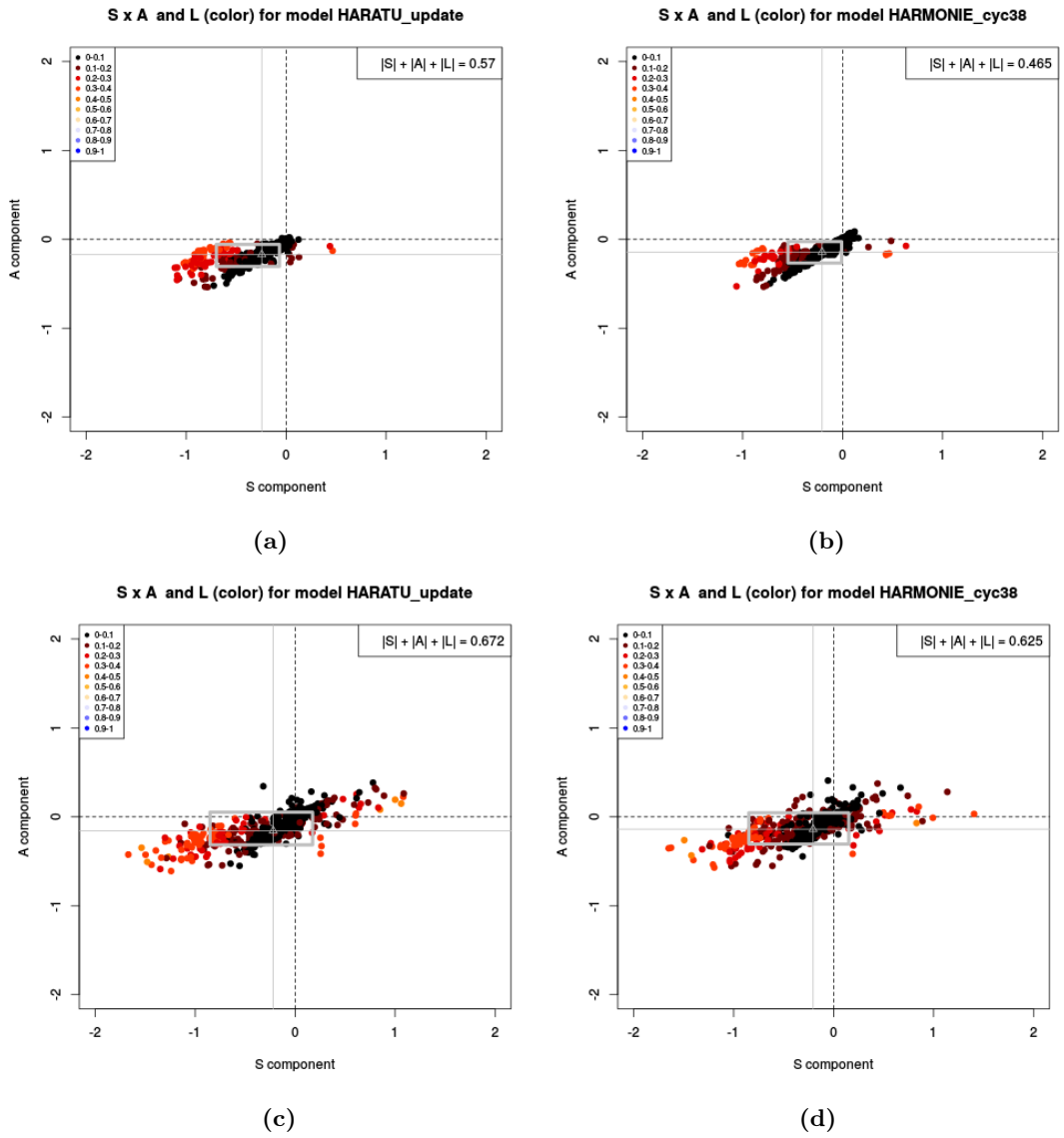
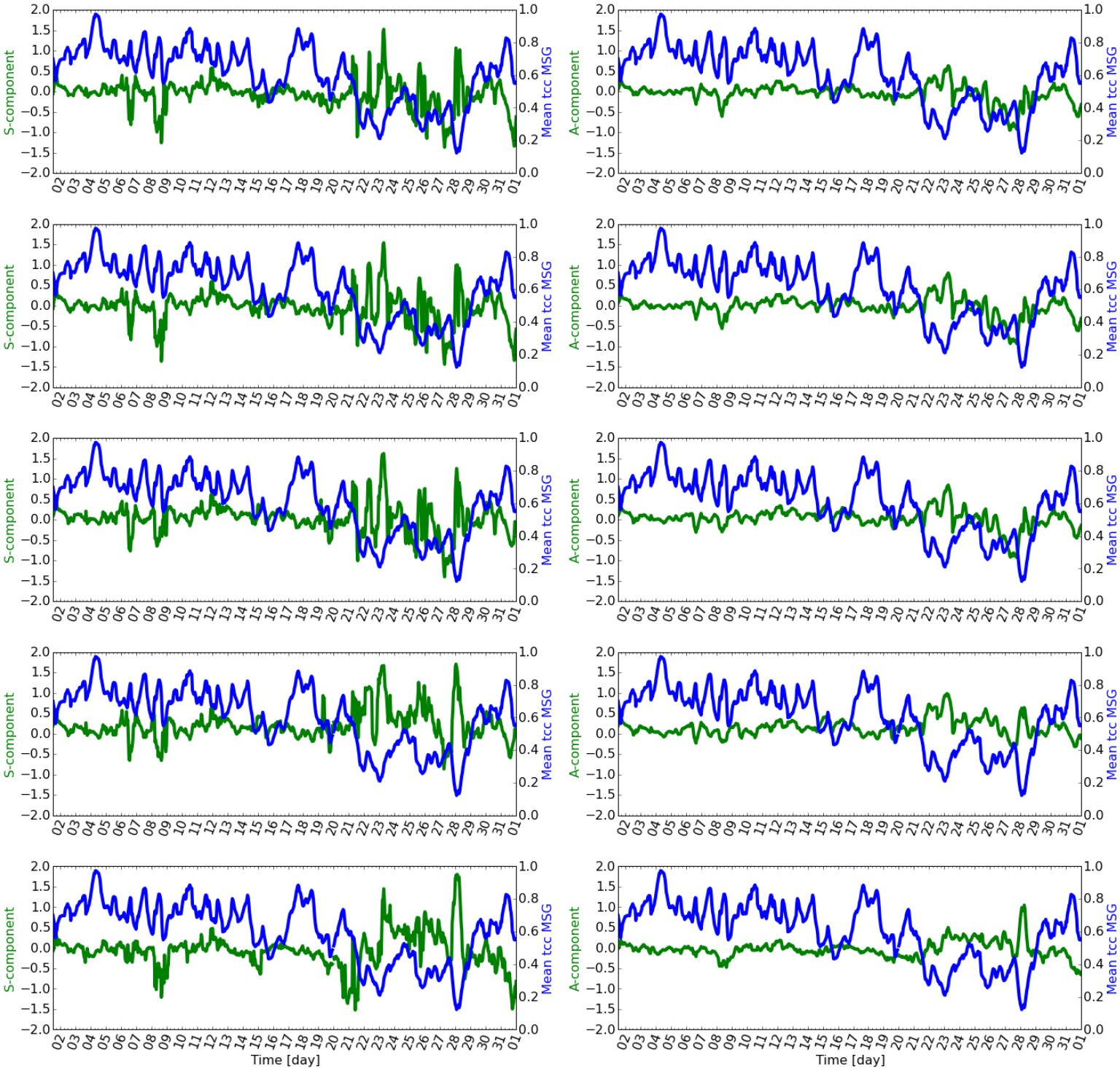
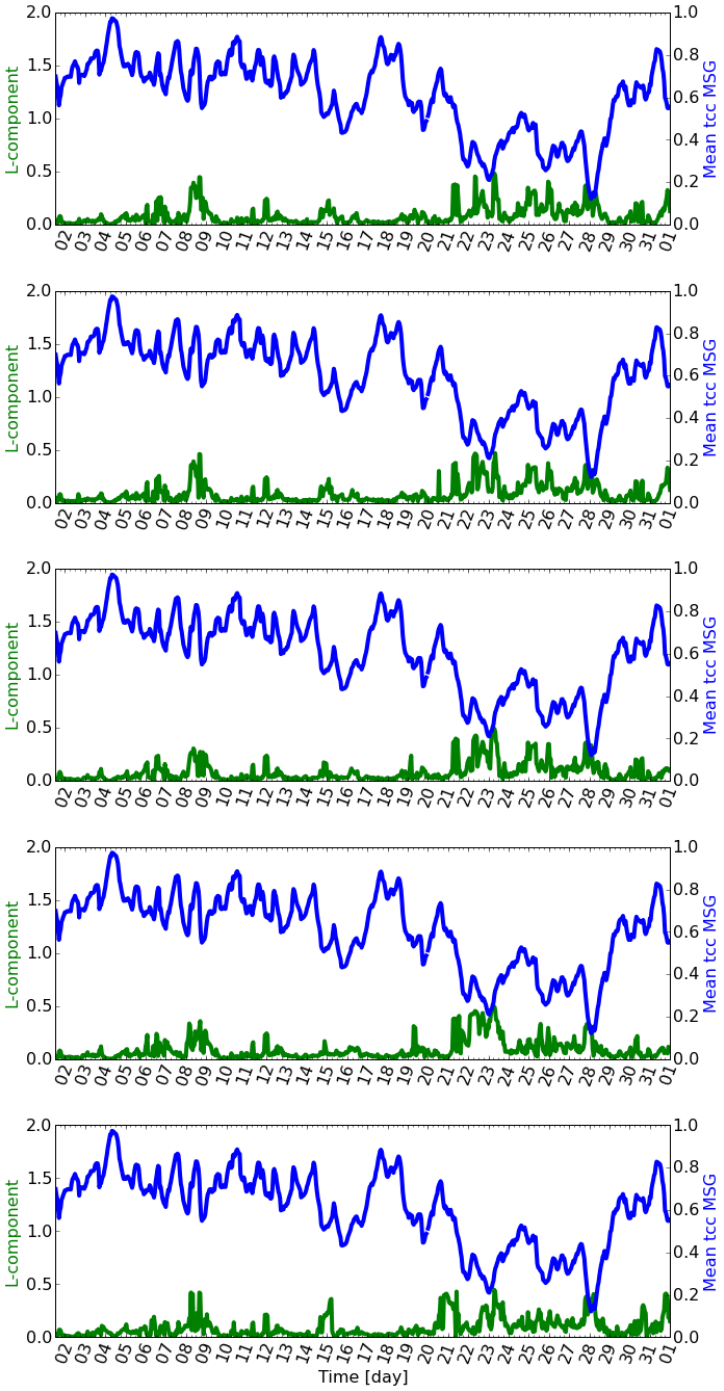


Figure 5.16: SAL diagrams for hourly cloud cover forecast for August 2006 of HARATU update (**a**) and HARMONIE cyc38 (**b**) and for May 2008 HARATU update (**c**) and HARMONIE cyc38 (**d**), compared to MSG. On the x (y) axis the S (A) component. Colour indicates the L component. The median values of S and A are the non-dashed lines. The grey box is 25 to 75 percent of the distribution of A and S. In the upper right corner is the sum of the absolute values of S, A and L. The threshold method is used, with $f = 0.5$.


S-component
Mean tcc MSG
A-component
Mean tcc MSG
Time [day]
Time [day]

(a) S **(b)** A



(c) L

Figure 5.17: Time evolution of the SAL components for March 2012, with MSG as observations. On the x axis is the day of the month. From top to bottom: HARATU update, HARATU, HARMONIE cyc38, HARMONIE SM and HARATU update DA. Green represents the component of SAL on the left axis, blue the mean cloud cover according to MSG on the right axis. (**a**) S, (**b**) A and (**c**) L.

Figure 5.17 shows the evolution of S (figure 5.17a), A (figure 5.17b) and L (figure 5.17c) during March 2012 from top to bottom for HARATU update, HARATU, HARMONIE cyc38, HARMONIE SM and HARATU update DA in green on the left axis. In blue, on the right axis, the mean cloud cover according to MSG is shown. The difference between models is marginal. Whenever one model is off, than so is the other. The month is characterized by distinct periods of high and low model performance, especially the S component. These periods are often, but not always, associated with a low average total cloud cover. This supports the results of figures 5.15 and 5.16, in which it was shown that especially the S component is often not predicted correctly by the models. Figure 5.17 shows that overcast conditions are predicted better than periods with a low cloud cover.

The diurnal cycle in figure 5.18 is relatively large for a domain averaged mean. The diurnal cycle is likely caused by the reduced turbulence and convection during night, producing a stable boundary layer. It is also caused by the measuring method of MSG. MSG uses only infrared light to measure cloud cover during night, while it uses both infrared and visible light to measure cloud cover during day (see chapter 2). To further investigate the diurnal cycle, we consider the SAL components for every hour of the day. The results are an average of the total cloud cover per hour of the day.

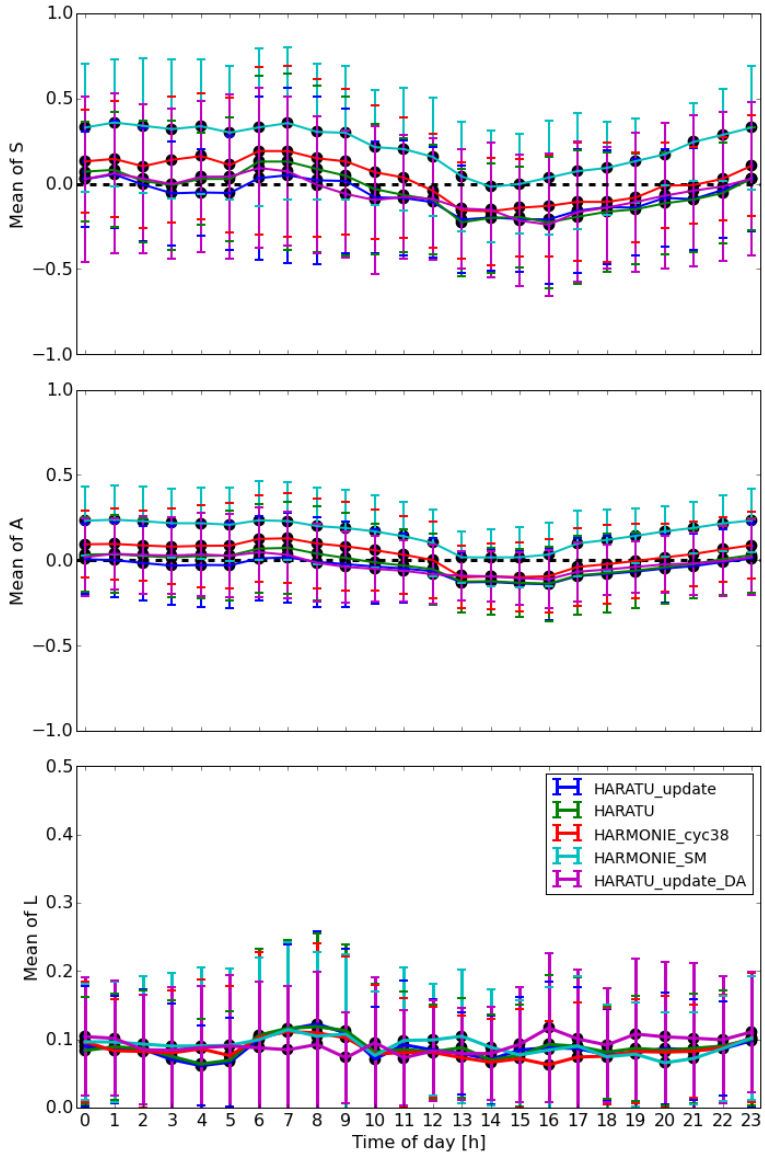
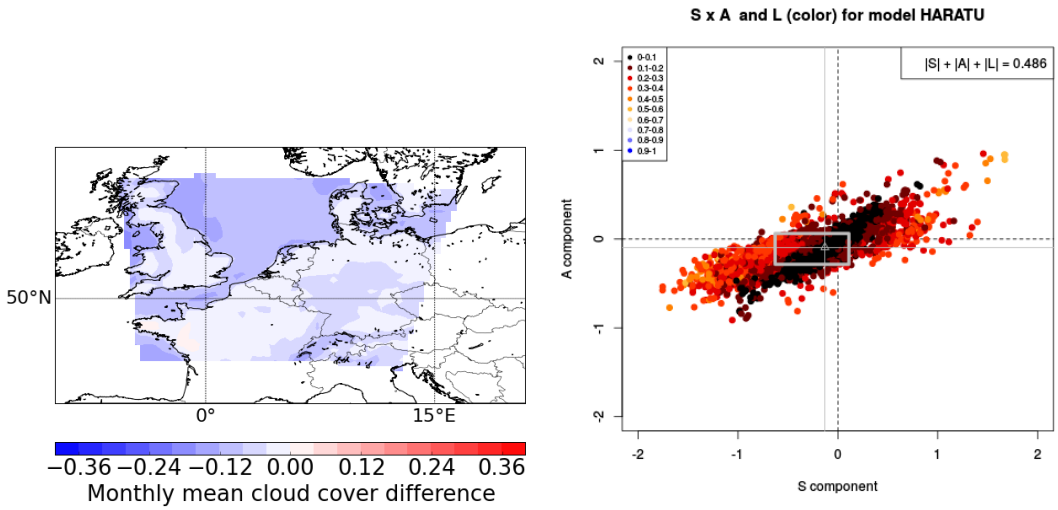


Figure 5.18: The SAL components per hour for March 2012, using MSG data as observations and $f = 0.5$ as a threshold. Colours indicate different versions of HARMONIE. The bars indicate the standard deviation. The horizontal dashed line represents the ideal value for the mean of S and A, which is the 0.0 line.

Figure 5.18 shows the result. The mean of S, A and L are shown as a function of the day in March 2012. The performance of all model versions are almost equal for the location component per hour. However, for S and A, HARMONIE SM performs systematically worse, followed by HARMONIE cyc38, HARATU, HARATU update DA and HARATU update. During the night, S and A are predicted well for HARATU update. The older HARMONIE versions perform worse during the night. The convection and turbulence are then reduced, often leading to a stable boundary layer. During day time, turbulence and convection are increased, which is more difficult to calculate for the model. Not enough convection and turbulence leads to a reduced average cloud cover. This results in a decrease of the mean of S and A, which is observed in figure 5.18. The model performance converge during the day, underestimating S and A. This suggests that some improvements can be made during day



(a) **(b)**

Figure 5.19: The cloud cover difference between HARATU and MSG for the year 2012 (**a**) and a SAL analysis for HARATU, using a 7x7 grid (**b**).

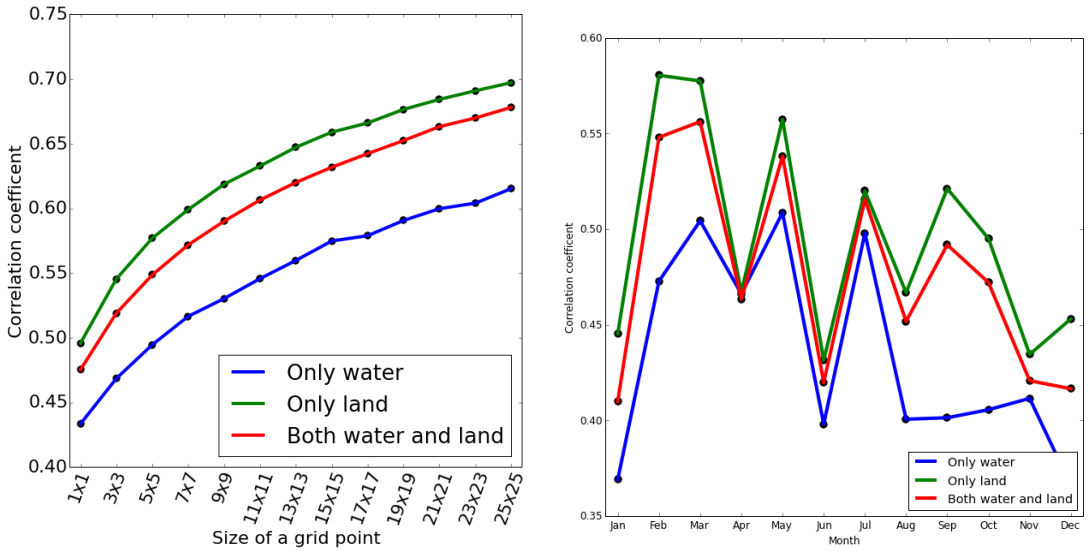
time. The diurnal cycle of HARMONIE for a domain around Spain has been investigated by Morales and Calvo (2013). They observe an overestimation of S and A during the night, while they observe an underestimation during daytime. The results are similar to the results of figure 5.18. Note that Morales and Calvo (2013) do not use HARATU update, but use HARMONIE cycle 36.

5.8 Whole year and land sea analysis

Up to now we have considered individual months. Some of the considered months are rather exceptional (August 2006) and thus one could question the generality of the results. It is therefore convenient to extend the considered time period to a whole year. The year 2012 is examined.

Figure 5.19a shows the cloud cover difference for the year 2012 between HARATU and MSG. Note that the year 2012 is only run for the regular HARATU version. The model and MSG difference is calculated first and then the annual average is taken. An underestimation of the total cloud cover is evident for almost all regions. Above land, the difference is often smaller than 4%, so the model performs relatively well and is still in the assumed uncertainty range of MSG. The model performs worse above sea. The difference between land and sea is very distinct. One possibility for this difference is the data retrieval of MSG above sea. Due to the lower albedo of water, thin cirrus clouds are more easily observed and interpreted as overcast (see chapter 2). A part of the deviations may be caused by retrieval error, but does not fully explain the relative large land sea difference of more than 0.1. The conclusion that the cloud cover is underestimated above sea by the model is still valid when considering a whole year. Figure 5.19b shows the SAL analysis for the whole year 2012. Most model output is in the upper right or lower left quadrant. Most entities follow a line with a lower slope than the main diagonal. A relatively large amount of data points underestimate the structure and in a lesser extent the amplitude. To conclude, the structure component is often underestimated for 2012 and HARATU performs worse above the sea.

To investigate this land sea difference further, let us look at figure 5.20a. Figure 5.20a shows the correlation coefficient with increasing grid size for only land, water and both for 2012. It emphasizes the difference between land and sea. The correlation coefficient differences are large, more than 0.1. Figure 5.20a shows that HARATU performs better above land than water for 2012, supporting the



(a) **(b)**

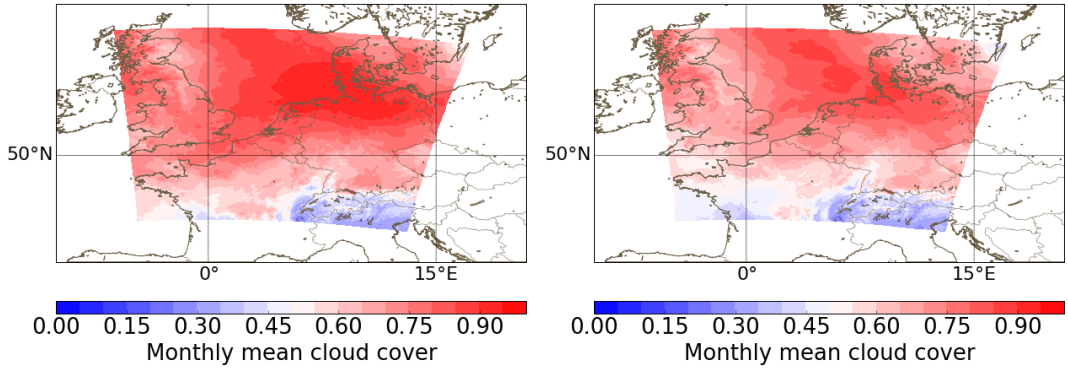
Figure 5.20: Correlation coefficient for 2012 considering land, water and both for HARATU, per grid size (**a**) and per month (**b**).

previous findings. Considering the large amount of data points, the difference is therefore significant (see appendix B.2). To investigate a possible seasonal dependence, we investigate the correlation coefficient per month for land, water and both in figure 5.20b. The fluctuations between months can be large, but no clear signal is visible per season. The relative differences between land and water is for most months relatively large, just like figure 5.20a. From figure 5.20b we conclude that the performance of HARMONIE is dependent on the month, but a clear seasonal signal is not visible. For all months, land performs better than sea.

5.9 Climate mode

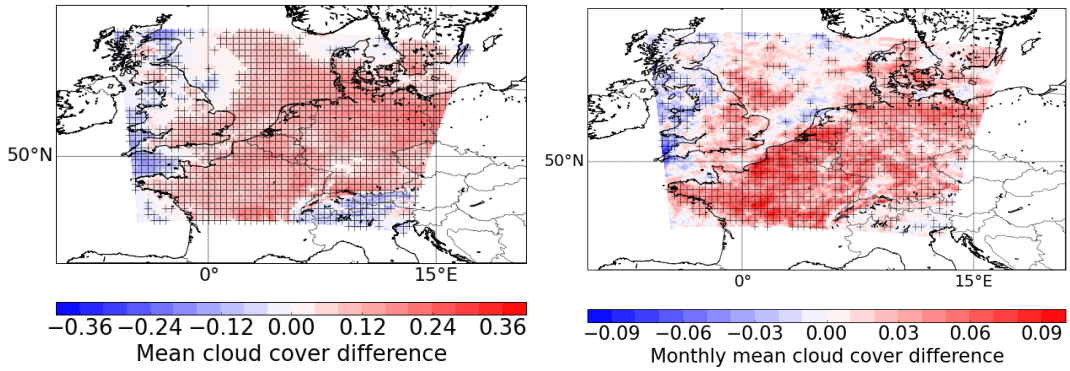
All previous experiments have been done in HC and DA mode. In this section, we test the ability of HARMONIE to simulate total cloud cover in climate mode.

Let us start with the mean cloud cover for March 2012 for HARMONIE cyc38 and HARATU (HARATU update is not run in CL mode) in figure 5.21. If we compare these results with the HC mode of figure 5.3, we observe more cloud cover in the domain. To further show these differences between model and MSG, let us look at figure 5.22a, which shows the monthly mean cloud cover difference between HARATU and MSG for March 2012. A significant overestimation is calculated above land and a large part of the North Sea. Figure 5.22b shows the difference |HARMONIE CYC38 − MSG| − |HARATU − MSG| for March 2012. Again, red indicates a better performance of HARATU than HARMONIE cyc38. The figure illustrates that HARATU performs much better than HARMONIE cyc38, especially above land. This is supported by figure 5.23a, which shows the correlation coefficient with increasing grid size for March 2012. The difference between HARATU and HARMONIE cyc38 is large and very significant (p-values < 0.001). In climate mode, HARMONIE performs worse when compared with the correlation coefficient of figure 5.10a. The correlation coefficients are lower, hence the climate mode does not resemble the observations as good as the HC mode. This is expected due to the lack of guidance of HARMONIE in climate mode. Errors early in the experiment can grow large due to the lack of ERA-Interim input during the run, which is inherent to the climate mode (only input at the boundaries is provided). Comparison with hourly observations is therefore not ideal.



(a) **(b)**

Figure 5.21: Mean cloud cover for March 2012 climate mode, for HARMONIE cyc38 (**a**) and HARATU (**b**).

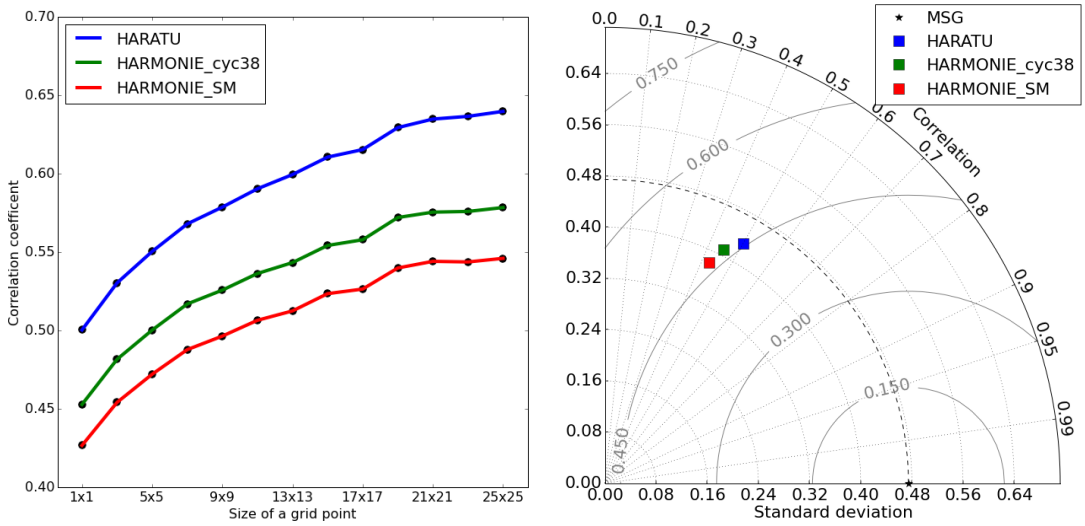


(a) **(b)**

Figure 5.22: Mean cloud cover difference between HARATU and MSG for March 2012 climate mode. A plus signifies significance as is calculated using bootstrapping (**a**). The absolute mean bias difference |HARMONIE CYC38 − MSG| − |HARATU − MSG| for March 2012 climate mode (**b**).

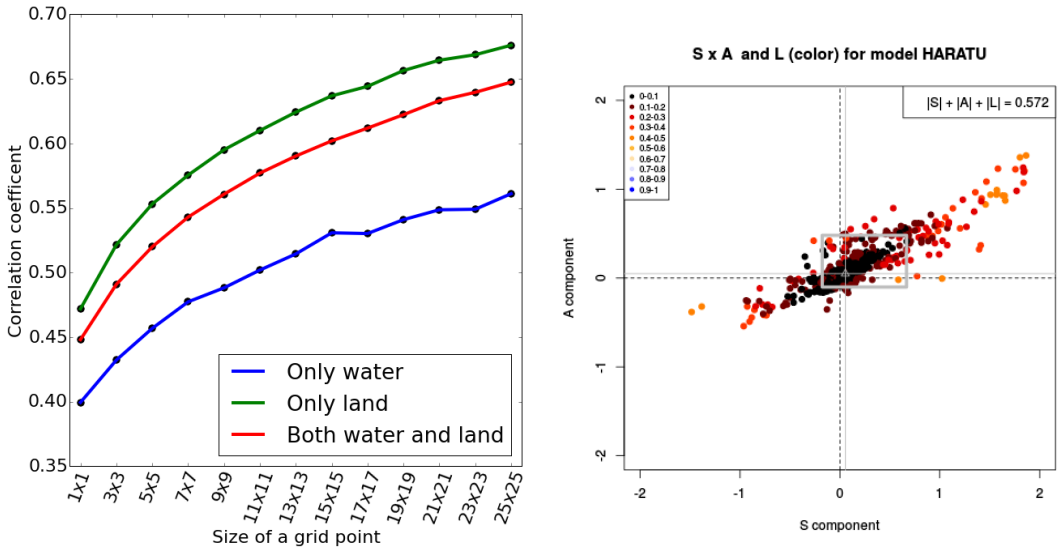
Figure 5.23b shows the Taylor diagram for HARMONIE in the climate mode for March 2012. The differences between the models are much larger than in figure 5.12c, showing that climate mode performs worse than the HC mode. HARATU predicts more spatial variability (higher standard deviation) than the other models. Note that the RMSE is approximately the same as in HC mode of figure 5.12c. This suggests that the high RMSE is caused by comparing the binary cloud cover of MSG with the fractional cloud cover of HARMONIE.

So far we have considered the CL mode only for individual months. Let us consider the results for the CL mode for the year 2012. Figure 5.24a shows the correlation coefficient for land, water and both for 2012 for HARATU. The model performs better above land than water, as we have seen in figure 5.20a. Figure 5.24b shows the SAL analysis when both land and water are considered for 2012 for HARATU. Most data are in the upper right quadrant and mostly overestimate the structure. The sum of the absolute values of the components is higher than in figure 5.19. This shows that the HC mode performs better than he CL mode. Again, we conclude that the cloud cover performs significantly worse above the sea. For the climate mode, the cloud cover is generally overestimated, especially the structure. Other months show approximately the same result. Only August 2006 shows an underestimation of the cloud cover.



(a) **(b)**

Figure 5.23: The correlation coefficient as a function of increasing grid size for March 2012 (**a**), and a Taylor diagram for different HARMONIE versions and MSG for March 2012 in climate mode (**b**).



(a) **(b)**

Figure 5.24: The correlation coefficient for only land, water and both for 2012 for HARATU (**a**). The SAL analysis for 2012 for HARATU in climate mode (**b**).

6 Results part II: Study of cloud schemes

6.1 Statistical and relative humidity cloud schemes

In the second part of this study, we implement a relative humidity scheme and make changes to the statistical cloud scheme. We assess the impact of cloud schemes on the model performance. Section 3.2 and figure 3.1 show the importance of a correct description of subgrid variability. For the statistical cloud scheme experiments we tune the parameters a and b of equation 6.1. This equation is restated here:

$$\sigma_s = \sqrt{a\left(\sigma_{con}^2 + \sigma_{turb}^2\right) + b\sigma_{extra}^2}. \tag{6.1}$$

Table 6.1: Table of the tuning parameters used in the statistical cloud schemes.

	ref	fac1	extravar	novar
a	1	1/2	0	0
b	1	1	1	0

Table 6.1 shows the chosen tuning parameters for the different experiments. The reference statistical cloud scheme is called ref in the table, and uses $a = 1$ and $b = 1$. As a sensitivity experiment, we change a from 1 to 1/2 and call it fac1. This factor will lead to a lower variance. When the cloud cover is overestimated, $a = 1/2$ will lead to a better prediction. Another experiment is considering only the extra variance term of equation 6.1 and is called extravar. The total variance is then reduced. It is likely that it will have the same consequences as described by the experiment with $a = 1/2$. In HARMONIE, the extra variance term is also multiplied with a height dependent factor. Another experiment consists of neglecting all variance terms ($a = 0$ and $b = 0$, called novar), and is the most radical scenario. This results in a binary cloud cover distribution. Cloud fractions in between 1.0 and 0.0 are caused by subgrid variations, which are produced by turbulence. Without variance, this results in cloud formation only for situations with 100% relative humidity. A binary cloud scheme assumes that the subgrid variability does not significantly contribute to cloud cover formation. However, the resolution of HARMONIE is not high enough to calculate the smallest turbulence, so a binary cloud scheme is expected to perform worse than the other experiments.

We also experiment with a relative humidity cloud scheme (as described in section 3.2). All relative humidity experiments are done in HARATU update (now called HARATUUP). The first relative humidity experiment consists of a simple relative humidity scheme, neglecting ice and liquid water in the cloud scheme. This experiment is called HARATUUP RH. The second experiment is a simple relative humidity scheme, including liquid and ice water. This experiment is called HARATUUP RH qt. The last experiment consists of a relative humidity scheme like HARATUUP RH, but with a different vertical profile for RH_{crit}, following Sundqvist et al. (1989); Giorgetta et al. (2013) (which is

Table 6.2: Table of the considered model versions with its description.

Model version	Model description
HARATUUP stat	Reference HARATU update with statistical cloud scheme
HARATUUP stat fac1	HARATUUP stat with $a = 1/2$
HARATUUP stat extravar	HARATUUP stat with only the extra variance term
HARATUUP stat novar	HARATUUP stat with no variance
HARATUUP RH	HARATU update with simple relative humidity scheme
HARATUUP RH qt	HARATUUP RH, with ice and liquid water included
HARATUUP RH sundqvist	HARATUUP RH with RH_{crit} profile following Sundqvist et al. (1989)

called HARATUUP RH sundqvist). Table 6.2 shows an overview of the experiments. The statistical cloud scheme experiments are called HARATUUP stat combined with the corresponding name from table 6.1. The reference statistical cloud scheme is called HARATUUP stat from now on.

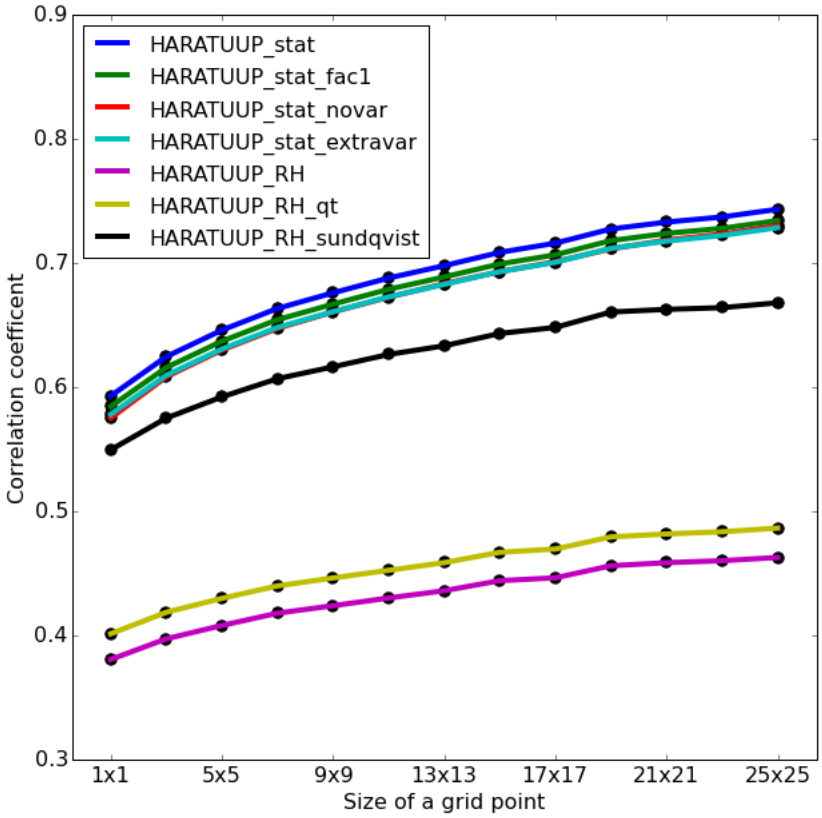


Figure 6.1: Correlation coefficients of different cloud scheme scenarios versus MSG as a function of grid size for March 2012, hindcast.

To assess the performance of the cloud scheme experiments, we calculate the correlation coefficient as a function of the grid resolution for the different cloud schemes with MSG for March 2012 HC mode in figure 6.1. As figure 6.1 shows, all cloud schemes perform worse in terms of correlation with MSG with respect to HARATUUP stat. The statistical cloud schemes perform relatively well and are positioned close to HARATUUP stat. Note that a reduction in variance leads to a lower correlation coefficient. The difference between the novar and extravar experiment is very small due to

a contribution to cloud cover by ice clouds, which is calculated outside the cloud scheme.

The implemented relative humidity schemes perform much worse. The critical relative humidity profile of Quaas (2012), as is used in HARATUUP RH and HARATUUP RH qt, is a global average and causes too much variance for high resolution models. The better performing approximation of Sundqvist et al. (1989) is used in their models and is highly optimized. When we use this critical relative humidity profile, we lose the direct link with measurements. Figure 6.1 does show a significant correlation coefficient increase when the scheme of Sundqvist et al. (1989) is implemented (HARATUUP RH sundqvist). The reduced variance of this scheme produces a much better cloud cover forecast. The low critical relative humidity at higher altitudes causes a relatively large fractional cloud cover for HARATUUP RH and HARATUUP RH qt (see section 3.2). The regions with $N = 0.0$ or $N = 1.0$ are therefore expected to be smaller than HARATUUP stat. The difference between HARATUUP RH and HARATUUP RH qt shows the impact of ice and liquid in the cloud scheme. As figure 6.1 shows, the contribution of ice and liquid results in a better result compared to MSG. The difference between HARATUUP RH and HARATUUP RH sundqvist shows the importance of the choice of the RH_{crit} profile. Figure 6.1 does not show the general pattern of cloud cover of the model versions in the domain. Fractional cloud cover conditions are expected when the variance is high, like in HARATUUP RH and HARATUUP RH qt. To investigate this further, a case study is done.

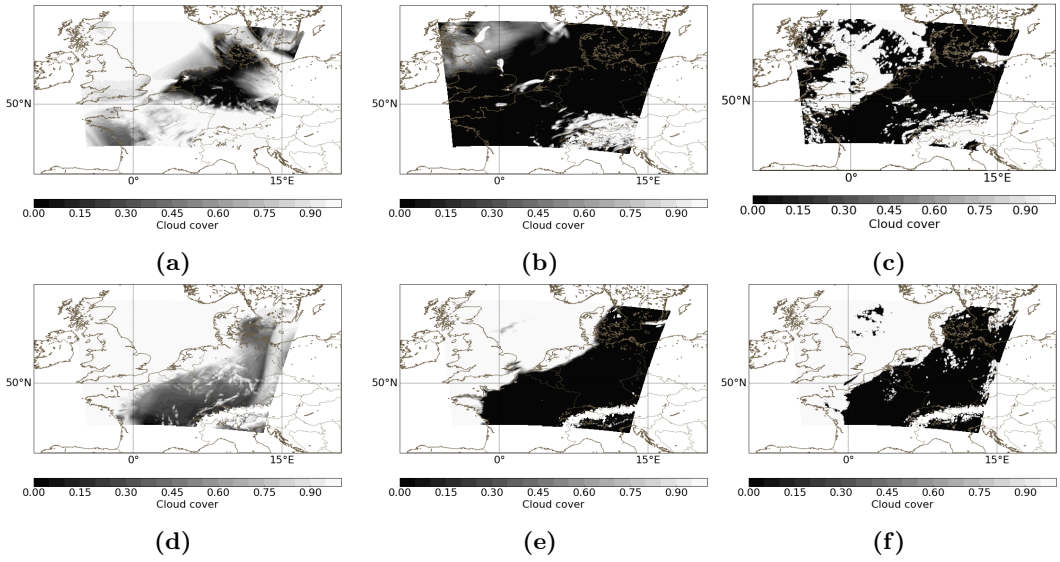
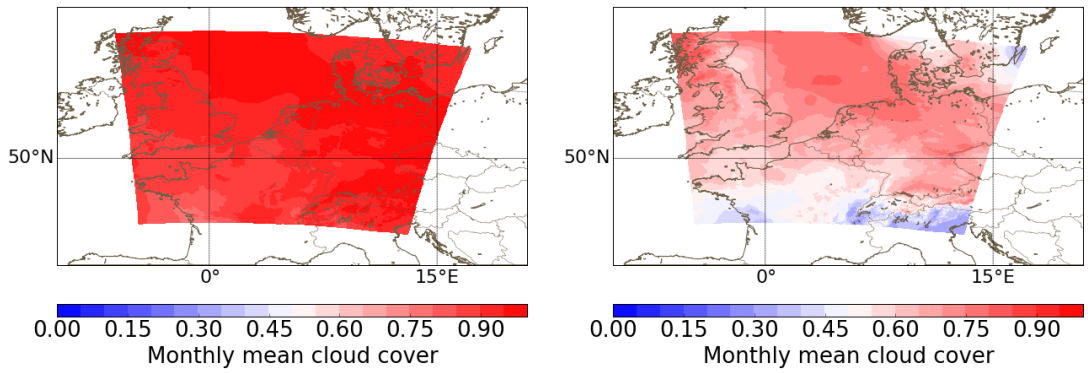


Figure 6.2: The cloud cover according to HARATUUP RH (**a**), HARATUUP RH sundqvist (**b**) and MSG (**c**) during 25 March 2012, 13 UTC. The cloud cover for 16 March 2012, 13 UTC for HARATUUP RH (**d**), HARATUUP RH sundqvist (**e**) and MSG (**f**).

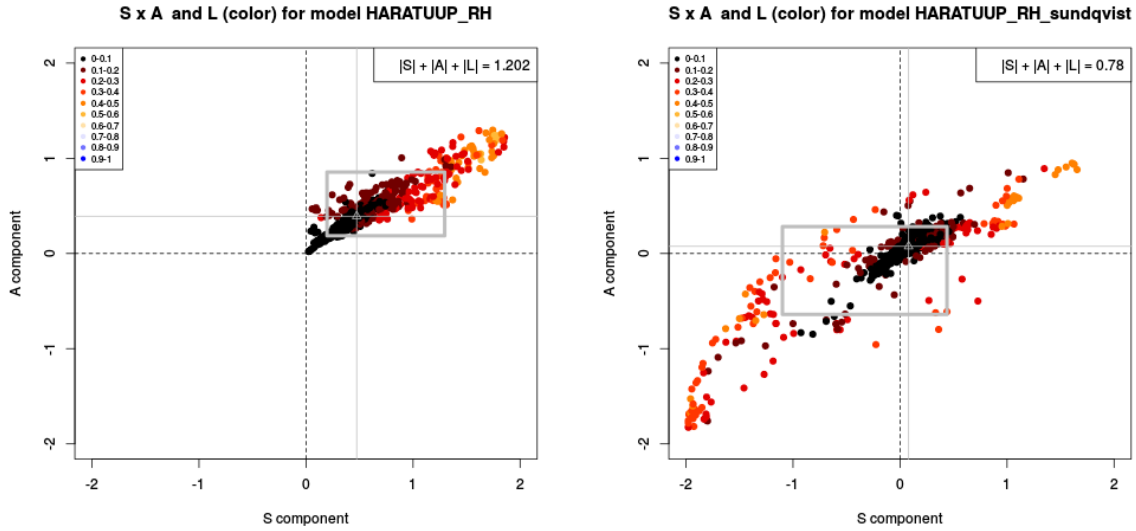
Figures 6.2a - 6.2c show the cloud cover distribution for 25 March 13 UTC in HC mode respectively for HARATUUP RH, HARATUUP RH sundqvist and MSG. Cloudy conditions are more dominant in HARATUUP RH, while clear sky conditions are more dominant in HARATUUP RH sundqvist. A large region with a fractional cloud cover is observed in figure 6.2a, as is expected. Figure 6.2d shows another case. It shows the cloud cover distribution for 16 March 2012 13 UTC according to HARATUUP RH and figure 6.2e according to HARATUUP RH sundqvist. The differences between the models is smaller. Keep in mind that the only difference between the models is the critical relative humidity profile. As expected, HARATUUP RH produces in a large area a fractional cloud cover, while for the same area HARATUUP RH sundqvist produces mostly clear sky conditions. This illustrates the overestimation of variance of HARATUUP RH and shows the better performance of HARATUUP RH sundqvist for these cases, especially when compared with the MSG measurement of figure 6.2f. These results show the large impact of the choice of the RH_{crit} profile.



(a) **(b)**

Figure 6.3: The mean cloud cover according to the HARATUUP RH (**a**) and for HARATUUP RH sundqvist (**b**) during March 2012.

Figure 6.3a shows the mean cloud cover for HARATUUP RH during March 2012 and figure 6.3b for HARATUUP RH sundqvist. The cloud cover of HARATUUP RH is severly overestimated when compared to figure 5.3a. The high variance of the RH_{crit} profile following the measuremetns of Quaas (2012) results in a very strong overestimation of the cloud cover. HARATUUP RH sundqvist is more in line with MSG. Again, both figures suggest that the choice of the critical relative humidity profile is of utmost importance.



(a) **(b)**

Figure 6.4: SAL analysis for HARATUUP RH (**a**) and HARATUUP RH sundqvist (**b**) for March 2012.

To quantify the structure and location performance, we use the SAL analysis. Figure 6.4a shows the SAL analysis for HARATUUP RH and figure 6.4b for HARATUUP RH sundqvist during March 2012. Figure 6.4a shows a strong overestimation of the total cloud for HARATUUP RH, for the S and the A component. This is in agreement with figures 6.2 and 6.3. Figure 6.4b shows that the

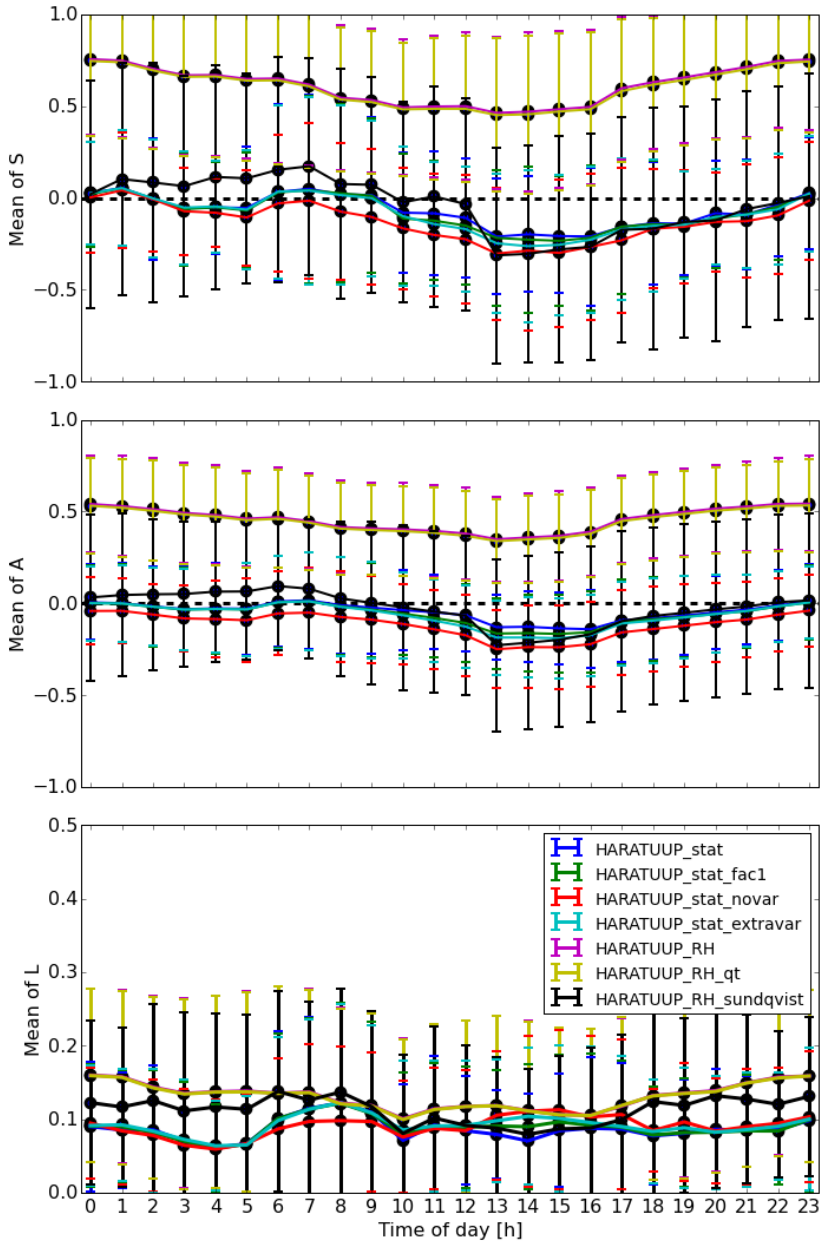


Figure 6.5: The SAL components per hour for March 2012, using MSG data as observations and $f = 0.5$ as a threshold. Colours indicate different versions of HARMONIE. The bars indicate the standard deviation. The horizontal dashed line represents the ideal value for the mean of S and A, and is the 0.0 line.

spread of some of the entities is very large for HARATUUP RH sundqvist. Some entities have extreme values, like $S = -2$ and $A = -2$. A relative humidity scheme may be a too simple representation.

To investigate the SAL per hour of the day, let us consider figure 6.5. It shows the SAL every hour of the day for the different cloud schemes for March 2012. HARATUUP RH qt and HARATUUP RH are far off with respect to the other models. HARATUUP RH sundqvist is relatively close to the statistical cloud scheme versions. The figure shows an important finding. The performance between the statistical cloud schemes is almost the same during night, and starts to deviate during the day. Remember that a value closer to 0.0 for S, A or L is better. The deviation can be explained with the stable conditions during night (low variance) and more turbulent and convective conditions during the day (increased variance). When the statistical cloud schemes are compared with the no variance run, we observe significantly better results during the night. This shows the importance of the use of the statistical cloud scheme during the night. The amplitude and structure are underestimated during the day (except for the simple relative humidity schemes), so the statistical cloud scheme model with the most variance (HARATUUP stat) performs best during the day. We also observe that each tested component of the statistical cloud scheme contributes to the performance of HARATUUP stat during the day. The tested components contribute to the quality of the calculation of cloud cover during convective and turbulent conditions.

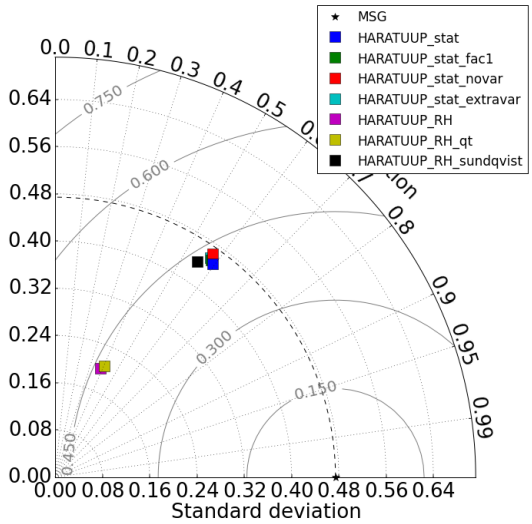


Figure 6.6: Taylor diagram for the different cloud schemes of HARMONIE compared with MSG for March 2012.

Other statistics for the different versions of HARMONIE are shown in figure 6.6 in the form of a Taylor diagram. All versions are again close to each other, except the new simple relative humidity schemes HARATUUP RH and HARATUUP RH qt. For these schemes, the RMSE is comparable with the other versions. The correlation coefficient is worse, as is observed in figure 6.1. The standard deviation is underestimated compared to MSG. This can be explained by the systematic overestimation of cloud cover. The spread in the data is small, what leads to a low standard deviation. It also shows how significantly different the simple relative humidity scheme is compared to MSG. On the other hand, HARATUUP RH sundqvist is close to the statistical cloud scheme versions. The statistical cloud schemes perform relatively similar, although HARATUUP stat correlates the best. The figure shows that small changes in the statistical cloud scheme will not alter the results significantly.

6.2 Variance with height

To be able to link the critical relative humidity scheme of the relative humidity cloud schemes to the statistical cloud scheme, one has to look at the variance of moisture and temperature with height. The critical relative humidity is a measure of the variance. When the variance is high, local turbulence causes the critical relative humidity to drop. If the profile of the standard deviation of the saturation curve σ_s is different than the critical relative humidity profile of Quaas (2012) or Sundqvist et al. (1989), it suggests some improvements can be made. The standard deviation σ_s can be split into a turbulence, convection and an extra variance part (see equation 6.1). Figure 6.7 shows the turbulence, convection and extra variance terms of σ_s, as well as σ_s as a function of height for the HARATUUP stat version. This is a time average over the whole domain for March 2012.

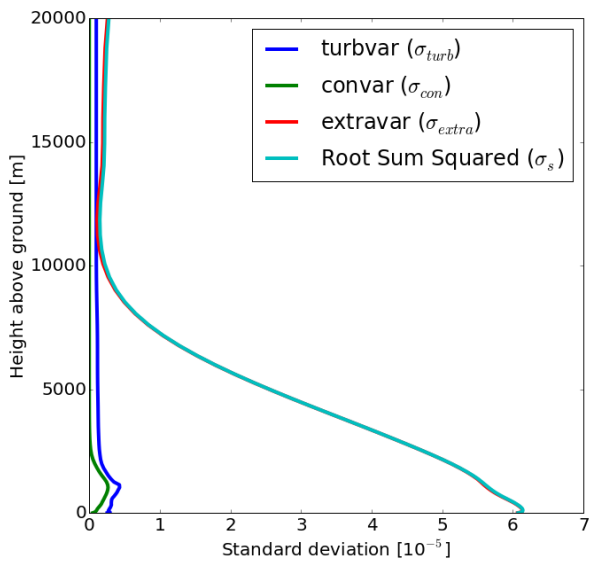


Figure 6.7: The time and domain averaged components of the standard deviation σ_s as a function of height for March 2012 in the HARATUUP stat version.

Figure 6.7 shows that the extra variance term is the dominant term in the atmosphere. On average, the turbulence and convection contribution are only small and limited to the lower atmosphere. The standard deviation σ_s decreases with height, what means that the variance decreases with height. According to Quaas (2012) and Sundqvist et al. (1989), a decrease of RH_{crit} is expected. This profile can be explained by the long life time of clouds higher in the atmosphere. This suggests that some changes in the statistical cloud scheme can be optimized regarding total cloud cover, especially in the extra variance term. In future research, the extra variance term can be converted to an RH_{crit} profile and can be enhanced to follow a different vertical variance profile, either following Quaas (2012) or Sundqvist et al. (1989).

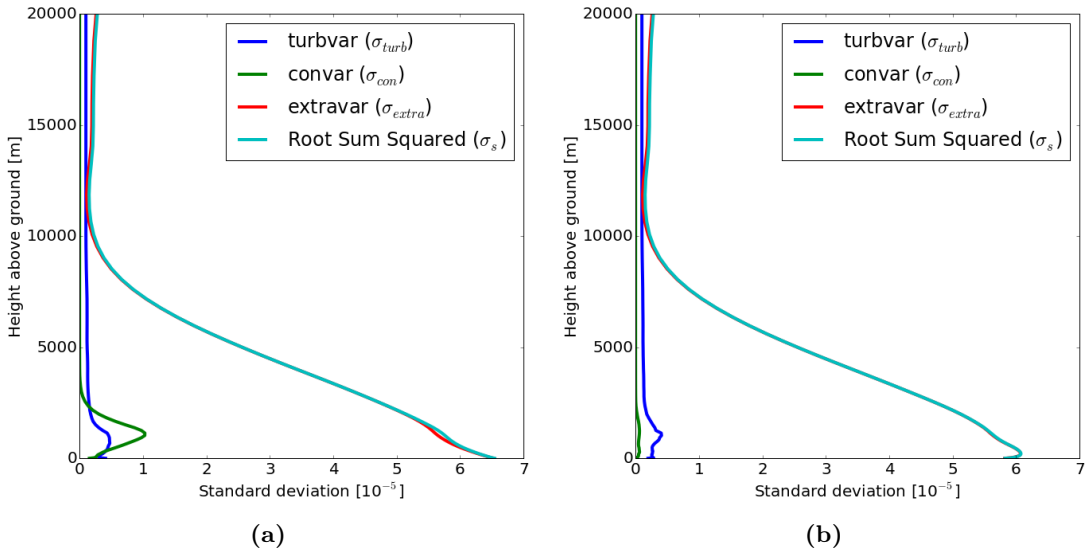


Figure 6.8: The time and domain averaged components of the standard deviation σ_s as a function of height for March 2012 in HARATUUP stat for (**a**) 13 UTC, and (**b**) 00 UTC.

Figure 6.7 does not provide a detailed image about turbulence and convection terms during day and night. Figure 6.8 shows the time and space averaged profile of σ_s and its components during 13 UTC (figure 6.8a) and during 00 UTC (figure 6.8b) for March 2012. It shows that the convection during the day is significantly higher than during the night. Also the turbulence is increased during the day. As expected, during the stable conditions of the night, convection is almost non-existent. Both the turbulence and convection peak correctly in the boundary layer and decrease above it. The convection and turbulence parts of equation 6.1 therefore follow the expected profile. However, we cannot say anything about the quality of the order of magnitude of the convection and turbulence parts.

7 Conclusion and Discussion

In this study we evaluate the cloud representation in HARMONIE. Our main goal is to assess the performance of different HARMONIE versions. We also do experiments with cloud schemes. The impact of various relative humidity and statistical cloud schemes on cloud cover formation is investigated.

Results of this study showed a general underestimation of the total cloud cover above sea when verified against MSG. MSG overestimates the total cloud cover above sea with respect to ground observations. The degree of overestimation is dependent on the measurement angle, time of the day, albedo of the surface and cloud type. In this study, we assume the uncertainty of MSG to be approximately 4%. After converting HARMONIE and MSG to the same grid, we have observed a significant underestimation of the total cloud cover of the newer versions of HARMONIE above the sea. The degree of underestimation is dependent on weather conditions. When we compare the cloud cover of the different HARMONIE versions, we see that the latest version of HARMONIE (HARATU update) performs better than the other versions. Only for May 2008 we have observed a correlation coefficient comparable with the other versions. The conclusion holds when converting to a coarser grid, so the results can be easily compared to other models with a coarse grid. The correlation coefficient with increasing grid size shows that the correlation coefficient increases when converted to a coarser grid. This is due to the averaging of errors. The correlation coefficient also shows a strong difference between land and sea. It shows a better performance of HARATU update above land than sea. No seasonal signal is observed. The absolute mean bias difference reveals an improvement of HARMONIE above land with respect to the other versions. Histograms of the fractional cloud cover for HARMONIE and MSG shows that the binary cloud scheme of MSG contributes to the uncertainty, especially during cloudy months. The Taylor diagram shows that the standard deviation is approximated correctly. It also reveals that the RMSE is very large. This is caused by the binary cloud scheme of MSG. Although the differences are small on the Taylor diagram, HARATU update shows the best forecast when compared to MSG for most months.

The SAL analysis shows a relatively large spread in the data for March 2012 with respect to the cloudy month August 2006. This represents that the model deviates stronger from the observations for months with more fractional cloud cover. For months with low cloud cover, the importance of the threshold choice for the SAL analysis increases. It is therefore difficult to derive firm conclusions from the SAL analysis. The SAL analysis does provide information which is not obtained from other statistical methods. The structure of the cloud field is stronger underestimated than the amplitude. This means that the volume of the cloud field is too low. We have also shown that HARATU update is capable of producing more adequate cloud cover forecasts during night in terms of structure and amplitude than the other versions of HARMONIE. During the day, the structure and amplitude of cloud cover is underestimated, and the difference between model versions becomes smaller. During the day, convection and turbulence is increased. This suggests that the model has difficulties in producing the right cloud cover during convective and turbulent conditions. For the climate mode, we observe an overestimation of the cloud cover for March 2012. This is likely caused by the lack of input of ERA-interim data. Errors are therefore allowed to grow in time in the climate mode. The results for the data assimilation mode perform generally worse than the hindcast mode for March 2012.

The second part of this study shows the performance of multiple cloud schemes. Changes have been made to the statistical cloud scheme and new relative humidity schemes have been implemented. The goal is to find a better cloud scheme and assess the contribution of the components of the cloud schemes. Changes to the statistical cloud scheme did not provide better results. However, it does provide information about the importance of the different components. Drastic changes have been tested, converting the cloud scheme to a binary scheme, as well as only considering the extra variance term. The no variance run did not produce a complete binary cloud scheme due to an ice cloud contribution from outside the cloud scheme. The ice cloud contribution should be considered when comparing the no variance run with the extra variance run. The default statistical cloud scheme provides the best results in the amplitude and structure component, especially during the day. Decreasing the strength of the subgrid variability did deteriorate the results, demonstrating that we need substantial subgrid variability. The subgrid variability added with each experiment results in a better forecast during the day. We have also shown that it is important to use a statistical cloud scheme during the night, because it performs significantly better than the model version with no variance. In general, the subgrid variability is too low during the day, so every term contributing to the subgrid variability is likely to produce better results.

Different relative humidity schemes have been tested, including a scheme with a critical relative humidity profile following measurements (Quaas, 2012). Boldly adding subgrid variability without discriminating between atmospheric conditions deteriorates even further. Following a different critical relative humidity profile (Sundqvist et al., 1989) provides better results. This relative humidity scheme does not perform better than the statistical cloud scheme, but it does provide information about the impact of critical relative humidity. The extra variance term of the statistical cloud scheme can be linked to the critical relative humidity profile. We also see that the extra variance term is dominant when compared to the turbulence and convection terms. The vertical profile of the turbulence and convection is limited to the boundary layer and has a diurnal cycle, which is expected. More research is needed to successfully implement a new profile for the extra variance term. This new profile can be linked to a realistic critical relative humidity profile.

A | Turbulence

In this section we derive the equations of the form of equation 3.1. Let us start with the Navier Stokes equations, including the Coriolis force, but neglecting tidal forces and viscous stresses:

$$\frac{d\mathbf{u}}{dt} = -\frac{1}{\rho}\boldsymbol{\nabla} p + f\mathbf{u} \times \hat{\mathbf{k}} - g\hat{\mathbf{k}}. \tag{A.1}$$

In which $\mathbf{u}$ is the three dimensional wind velocity, consisting of $\mathbf{u} = (u, v, w)$, p pressure, ρ density of air and $\hat{\mathbf{k}}$ the unit vector in the vertical direction. $\frac{d}{dt}$ is the total derivative:

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \mathbf{u} \cdot \boldsymbol{\nabla}. \tag{A.2}$$

Turbulence can be interpreted as a covariance between two quantities. Reynolds decomposition will therefore be useful. A given variable x is then decomposed to:

$$x = \bar{x} + x'. \tag{A.3}$$

In which $\bar{x}$ is the mean of x and x' the perturbations with respect to the mean, with $\bar{x'} = 0$. The covariance then becomes:

$$\overline{(xy)} = \overline{(\bar{x} + x') \cdot (\bar{y} + y')} \tag{A.4}$$

$$= \bar{x}\bar{y} + \overline{x'y'}. \tag{A.5}$$

Using Reynolds decomposition on p, u, v and w and substituting this in equation A.1 (considering only in the x-direction as an example, other directions are analogue):

$$\frac{\partial\overline{\bar{u} + u'}}{\partial t} + \overline{(\bar{u} + u')\frac{\partial(\bar{u} + u')}{\partial x}} + \overline{(\bar{v} + v')\frac{\partial(\bar{u} + u')}{\partial y}} + \overline{(\bar{w} + w')\frac{\partial(\bar{u} + u')}{\partial z}} = -\frac{1}{\rho}\frac{d\overline{(\bar{p} + p')}}{dx} + f\overline{(\bar{v} + v')}. \tag{A.6}$$

This can be simplified:

$$\frac{\partial\bar{u}}{\partial t} + \bar{u}\frac{\partial\bar{u}}{\partial x} + \bar{v}\frac{\partial\bar{u}}{\partial y} + \bar{w}\frac{\partial\bar{u}}{\partial z} + \overline{u'\frac{\partial u'}{\partial x}} + \overline{v'\frac{\partial u'}{\partial y}} + \overline{w'\frac{\partial u'}{\partial z}} = -\frac{1}{\rho}\frac{d\bar{p}}{dx} + f\bar{v} \tag{A.7}$$

$$\frac{d\bar{u}}{dt} + \overline{u'\frac{\partial u'}{\partial x}} + \overline{v'\frac{\partial u'}{\partial y}} + \overline{w'\frac{\partial u'}{\partial z}} = -\frac{1}{\rho}\frac{d\bar{p}}{dx} + f\bar{v}. \tag{A.8}$$

Assuming incompressibility for turbulent wind speed:

$$\boldsymbol{\nabla} \cdot \mathbf{u}' = 0, \tag{A.9}$$

with $\mathbf{u}' = (u', v', w')$ and using the product rule of differentiation before averaging:

$$\frac{\partial u'u'}{\partial x} + \frac{\partial u'v'}{\partial y} + \frac{\partial u'w'}{\partial z} = u'\frac{\partial u'}{\partial x} + v'\frac{\partial u'}{\partial y} + w'\frac{\partial u'}{\partial z} + u'\left(\frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y} + \frac{\partial w'}{\partial z}\right) \tag{A.10}$$

$$= u'\frac{\partial u'}{\partial x} + v'\frac{\partial u'}{\partial y} + w'\frac{\partial u'}{\partial z} + u'\boldsymbol{\nabla}\cdot\mathbf{u}' \tag{A.11}$$

$$= u'\frac{\partial u'}{\partial x} + v'\frac{\partial u'}{\partial y} + w'\frac{\partial u'}{\partial z}. \tag{A.12}$$

Taking the average:

$$\frac{\partial\overline{u'u'}}{\partial x} + \frac{\partial\overline{u'v'}}{\partial y} + \frac{\partial\overline{u'w'}}{\partial z} = \overline{u'\frac{\partial u'}{\partial x}} + \overline{v'\frac{\partial u'}{\partial y}} + \overline{w'\frac{\partial u'}{\partial z}}, \tag{A.13}$$

which can be used in equation A.8 to rewrite some components to derivatives of covariance terms.

$$\frac{d\bar{u}}{dt} + \frac{\partial\overline{u'u'}}{\partial x} + \frac{\partial\overline{u'v'}}{\partial y} + \frac{\partial\overline{u'w'}}{\partial z} = -\frac{1}{\rho}\frac{d\bar{p}}{dx} + f\bar{v} \tag{A.14}$$

Assuming horizontal homogeneity, which is a decent assumption if compared to the scales of change in the vertical direction, equation A.14 becomes:

$$\frac{d\bar{u}}{dt} + \frac{\partial\overline{u'w'}}{\partial z} = -\frac{1}{\rho}\frac{d\bar{p}}{dx} + f\bar{v}. \tag{A.15}$$

This is the final equation for x component of the turbulent flow. Many other variables, like v and θ can be written in the same form:

$$\frac{d\bar{\phi}}{dt} = -\frac{\partial\overline{\phi'w'}}{\partial z} + F, \tag{A.16}$$

in which F represents all other terms and ϕ is the considered quantity.

B Significance of correlation coefficient differences

B.1 Independent variables

Let us consider two independent data sets; x_1, y_1 and x_2, y_2. Correlation coefficients can be calculated between x_1 and y_1, and x_2 and y_2, with x a model data set and y the observational data set. The results may tell us something about the quality of a model. However, no statement can be made about the significance between the two resulting correlation coefficients. Fisher z-transformation (Fisher, 1921) will be used to find the probability that the results are equal or extremer than the observations (p-value). The z-transformation is used for stabilizing the variance and transforming it into an almost Gaussian distribution. When the z-values are found, it can be used to find p-values. The z-value between two different independent correlation coefficients can be calculated as follows (Fisher, 1921):

$$z = \frac{\log\left(\frac{1+r_1}{1-r_1}\right) + \log\left(\frac{1+r_2}{1-r_2}\right)}{2\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}}. \tag{B.1}$$

In which r_1 is correlation coefficient between x_1 and y_1 and r_2 between x_2 and y_2. The length of data sets is n_1 x_1 and y_1, and n_2 is the length of data sets x_2 and y_2. In this study is $n_1 = n_2 = n$. The z-values can be used to calculate the p-values:

$$p = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{z}{\sqrt{2}}\right)\right), \tag{B.2}$$

in which erf is the error function:

$$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2}\,\mathrm{d}t. \tag{B.3}$$

When testing for significance, one should use two tailed statistics. Then the two tailed p-value p_2 becomes:

$$p_2 = 2p \tag{B.4}$$

The used confidence interval is 2σ, or 95 %. For two tailed, this corresponds with measurements in the lowest and highest 2.5% of the probability distribution function to be significantly different.

B.2 Dependent variables

Now let us consider two dependent data sets; x_1, y and x_2, y. Correlation coefficients can be calculated as in section 4.2. For calculating p-values, the theory of section B.1 is insufficient and needs to be expanded. Steiger (1980) broadened the theory of Fisher (1921), to include a significance test for dependent variables. The correlation coefficients between x_1 and y (r_{13}), x_2 and y (r_{23}), and x_1 and x_2 (r_{12}) are needed. The first step is to calculate the z-values using the Fisher z-transformation of section B.1, what will result in:

$$z_1 = \frac{1}{2} \log \left(\frac{1 + r_{13}}{1 - r_{13}} \right) \tag{B.5}$$

$$z_2 = \frac{1}{2} \log \left(\frac{1 + r_{23}}{1 - r_{23}} \right) \tag{B.6}$$

The covariance can be written as:

$$\sigma_{31,32} = \frac{r_{12} \left(1 - 2\bar{r}^2\right) - \frac{1}{2}\bar{r}^2 \left(1 - 2\bar{r}^2 - r_{12}^2\right)}{\left(1 - \bar{r}^2\right)\left(1 - \bar{r}^2\right)}, \tag{B.7}$$

in which $\bar{r} = \frac{1}{2}\left(r_{13} + r_{23}\right)$. With the covariance, the z-score can be determined as follows:

$$z = \frac{\sqrt{n-3}\,(z_1 - z_2)}{\sqrt{2 - 2\sigma_{31,32}}}. \tag{B.8}$$

Following equations B.2 and B.4, the two tailed p-value can be calculated. Therefore the significance between correlation coefficients of two different models can be derived. For a more detailed derivation of the significance test between two dependent variables, see Steiger (1980).

Bibliography

Bengtsson, L., Andrae, U., Aspelien, T., Batrak, Y., Calvo, J., de Rooy, W., Gleeson, E., Hansen-Sass, B., Homleid, M., Hortal, M., Ivarsson, K.-I., Lenderink, G., Niemelä, S., Nielsen, K. P., Onvlee, J., Laura Rontu, P. S., Muñoz, D. S., Subias, A., Tijm, S., Toll, V., Yang, X., and ødegaard Køltzow, M. (2017). The HARMONIE-AROME model configuration in the ALADIN-HIRLAM NWP system. *Monthly Weather Review*.

Crocker, R. and Mittermaier, M. P. (2013). Exploratory use of a satellite cloud mask to verify NWP models. *Meteorological Applications*, 20(2):197–205.

Cuijpers, J. W. M. and Bechtold, P. (1995). A Simple Parameterization of Cloud Water Related Variables for Use in Boundary Layer Models. *Journal of the Atmospheric Sciences*, 52(13):2486–2490.

Cuxart, J., Bougeault, P., and Redelsperger, J.-L. (2000). A turbulence scheme allowing for mesoscale and large-eddy simulations. *Quarterly Journal of the Royal Meteorological Society*, 126(562):1–30.

De Rooy, W., Lenderink, G., Van der Plas, E., and Barkmeijer, J. (2014). The fog above sea problem in Harmonie Part II: Experiences with the RACMO turbulence scheme. *ALADIN-HIRLAM Newsletter*, (3).

De Rooy, W. C. and Pier Siebesma, A. (2010). Analytical expressions for entrainment and detrainment in cumulus convection. *Quarterly Journal of the Royal Meteorological Society*, 136(650):1216–1227.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thèpaut, J.-N., and Vitart, F. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656):553–597.

Derrien, M., Le Gléau, H., and Fernandez, P. (2012). Algorithm Theoretical Basis Document for "Cloud Products" (CMa-PGE01v3.2, CT-PGE02 v2.2 & CTTH-PGE03 v2.2). *NWC SAF Newsletter*, 3(2).

Dundee (2016). NERC Satellite Receiving Station, Dundee University, Scotland.

Ekman, V. (1905). On the influence of the earth's rotation on ocean currents. *Ark. Mat. Astron. Fys.*, 2(11).

Finkensieper, S., Hanschmann, T., Stengel, M., Benas, N., Van Zadelhoff, G., Meirink, J. F., Hollmann, R., and Werscheck, M. (2016). Validation Report SEVIRI cloud products Edition 2 (CLAAS-2). *EUMETSAT SAF on CLIMATE MONITORING*, 1.

Fisher, R. A. (1921). On the Probable Error of a Coefficient of Correlation Deduced from a Small Sample. *Metron*, 1(3-32):205–235.

Giorgetta, M. A., Roeckner, E., Mauritsen, T., Bader, J., Crueger, T., Esch, M., Rast, S., Kornblueh, L., Schmidt, H., Kinne, S., Hohenegger, C., Möbis, B., Krismer, T., Wieners, and Stevens, B. (2013). The atmospheric general circulation model ECHAM6 Model description. Technical report, Max-Planck-Institut für Meteorologie.

Gleeson, E., Pagh Nielsen, K., Toll, V., Rontu, L., and Whelan, E. (2015). Shortwave radiation experiments in HARMONIE. Tests of the cloud inhomogeneity factor and a new cloud liquid optical property scheme compared to observations. *ALADIN-HIRLAM Newsletter*, 5.

Haiden, T., Forbes, R., Ahlgrimm, M., and Bozzo, A. (2015). The skill of ECMWF cloudiness forecasts. *ECMWF Newsletter*, (143):14–19.

Hartmann, D., Klein Tank, A., Rusticucci, M., Alexander, L., Brönnimann, S., Charabi, Y., Dentener, F., Dlugokencky, E., Easterling, D., Kaplan, A., Soden, B., Thorne, P., Wild, M., and Zhai, P. (2013). *Observations: Atmosphere and Surface*, book section 2, pages 159 – 254. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Knight, C., Knight, S., Massey, N., Aina, T., Christensen, C., Frame, D., Kettleborough, J., Martin, A., Pascoe, S., Sanderson, B., Stainforth, D., and Allen, M. (2007). Association of parameter, software, and hardware variation with large-scale behavior across 57,000 climate models. *Proceedings of the National Academy of Sciences of the United States of America*, 104(30):12259–12264.

Legendre, M., Fayard, D., Hunter-Anderson, J., Burns, S., and Pili, P. (2010). From MSG to MTG, cost-effective operations of a complex system. *SpaceOps 2010 Conference*.

Mittermaier, M., Ghelli, A., Ebert, E., Wilson, L., and Nurmi, P. (2012). Recommended Methods for Evaluating Cloud and Related Parameters. *World Meteorological Organization*.

Morales, G. and Calvo, J. (2013). Verification of fog and low cloud simulations using an object oriented method. *EGU General Assembly 2013*.

Neggers, R. A. J., Siebesma, A. P., Lenderink, G., and Holtslag, A. A. M. (2004). An evaluation of mass flux closures for diurnal cycles of shallow cumulus. *Monthly Weather Review*, 132(11):2525–2538.

Nielsen, K. P., Gleeson, E., and Rontu, L. (2014). Radiation sensitivity tests of the harmonie 37h1 nwp model. *Geoscientific Model Development*, 7(4):1433–1449.

Perov, V. and Gollvik, S. (1996). A 1-D test of a nonlacal, E = ϵ boundary layer scheme for a NWP model resolution. *Technical Report*, (25).

Platnick, S., King, M. D., Ackerman, S. A., Menzel, W. P., Baum, B. A., Riedi, J. C., and Frey, R. A. (2003). The MODIS cloud products: algorithms and examples from Terra. *IEEE Transactions on Geoscience and Remote Sensing*, 41(2):459–473.

Quaas, J. (2012). Evaluating the "critical relative humidity" as a measure of subgrid-scale variability of humidity in general circulation model cloud cover parameterizations using satellite data. *J. Geophys. Res.*, (117).

Schmetz, J., Pili, P., Tjemkes, S., Just, D., Kerkmann, J., Rota, S., and Ratier, A. (2002). An introduction to meteosat second generation (msg). *Bulletin of the American Meteorological Society*, 83(7):977–992.

Schulz, J., Thomas, W., Müller, R., Behr, H.-D., Caprion, D., Deneke, H., Dewitte, S., Dürr, B., Fuchs, P., Gratzki, A., Hollmann, R., Karlsson, K.-G., Manninen, T., Reuter, M., Riihelä, A., Roebeling, R., Selbach, N., Tetzlaff, A., Wolters, E., Zelenka, A., and Werscheck, M. (2008). Operational climate monitoring from space: the EUMETSAT satellite application facility on climate monitoring (CM-SAF). *Atmos. Chem. Phys. Discuss.*, 8:8517–8563.

Seity, Y., Brousseu, P., Malardel, S., Hello, G., Benard, P., Bouttier, F., Lac, C., and Masson, V. (2011). The arome-france convective-scale operational model. *Mon. Wea. Rev.*, 139:976–991.

Söhne, N., Chaboureau, J.-P., and Guichard, F. (2008). Verification of cloud cover forecast with sateilite observation over west africa. *Monthly Weather Review*, 136(11):4421–4434.

Sommeria, G. and Deardorff, J. (1977). Subgrid scale condensation in models of nonprecipitating clouds. *J. Atmos. Sci.*, (34).

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87(2):245–251.

Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M. (2013). Climate Change 2013: The Physical Science Basis. *Intergovernmental Panel on Climate Change, Working Group I Contribution to the IPCC Fifth Assessment Report (AR5).* 1535 pp.

Stull, R. B. (1988). *An Introduction to Boundary Layer Meteorology.* Springer Netherlands, Dordrecht, Netherlands.

Sundqvist, H., Berge, E., and Kristjànsson, J. E. (1989). Condensation and cloud parameterization studies with a mesoscale numerical weather prediction model. *Monthly Weather Review*, 117(8):1641–1657.

Susskind, J., Barnet, C. D., and Blaisdell, J. M. (2003). Retrieval of atmospheric and surface parameters from AIRS/AMSU/HSB data in the presence of clouds. *IEEE Transactions on Geoscience and Remote Sensing*, 41(2):390–409.

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, 106(D7):7183–7192.

Weniger, M. and Friederichs, P. (2016). Using the SAL technique for spatial verification of cloud processes: A sensitivity analysis. *ArXiv e-prints.*

Wernli, H., Paulat, M., Hagen, M., and Frei, C. (2008). SAL-A Novel Quality Measure for the Verification of Quantitative Precipitation Forecasts. *Monthly Weather Review*, 136(11):4470–4487.

Wild, M., Folini, D., Schär, C., Loeb, N., Dutton, E., and König-Langlo, G. (2013). The global energy balance from a surface perspective. *Climate Dynamics.*