M.Sc. Thesis

Meteorology

Feasibility of assimilating ASCAT surface winds into a Limited Area Model

Pirkka Ollinaho

2010

Supervisors: M.Sc Pirkko Pylkkö (FMI)

M.Sc Kalle Eerola (FMI)

Reviewers: Prof. Hannu Savijärvi

M.Sc Pirkko Pylkkö (FMI)

University of Helsinki

Department of Physics

FIN-00014 University of Helsinki

Helsingin yliopisto — Helsingfors universitet — University of Helsinki

Tiedekunta/Osasto — Fakultet/Sektion — Faculty	Laitos — Institution — Department
Matemaattis-luonnontieteellinen tdk.	Fysiikan laitos

Tekijä — Författare — Author
Pirkka Ollinaho

Työn nimi — Arbetets titel — Title

ASCAT pintatuulihavaintojen käyttökelpoisuus rajatun alueen sääennustemallissa

Oppiaine — Läroämne — Subject
Meteorologia

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Pro gradu	Maaliskuu 2010	81

Tiivistelmä — Referat — Abstract

Edellisen sukupolven skatterometrien merenpinnan tuulihavaintoja on tuloksellisesti assimiloitu numeerisiin sääennustemalleihin. Näistä assimilaatiototeutuksista tehdyt vaikutustutkimukset ovat osoittaneet selviä parannuksia mallien analyysi- ja ennustetarkkuuksiin. Metop-A:n hyötykuormana oleva kehittynyt skatterometri (ASCAT) tarjoaa tarkempia havaintoja merenpintatuulista sekä kattaa mittauksillaan suuremman alueen edellisiin skatterometreihin verrattuna.

ASCAT:in havaintojen assimilaatiota korkean resoluution rajatun alueen malliin (HIRLAM 4D-Var) tutkitaan viidessä erillisessä tapaustutkimuksessa. Vaikutuksen huomataan olevan neutraali-positiivinen tapauksille, joissa yleinen virtaussuunta on Atlantilta, mutta pohjoisille polaarivirtauksille vaikutuksen havaitaan olevan negatiivinen. Tämän päätellään johtuvan ongelmista pohjoisten virtausten mallintamisessa, mutta myös sopivien verifikaatiometodien puutteellisuudesta. Ehdotuksia sekä esimerkki verifikaatiometodien kehittämisestä esitellään myöhemmässä vaiheessa.

Työssä tarkastellaan myös lähemmin polaarimatalatapausta. Tarkastelusta selviää, että ASCAT:in assimilaatio analyysijärjestelmään parantaa polaarimatalan alkukehityksen ennustetta, mutta mallin dynamiikka/fysiikka advektoi tämän voimakkaan matalapainekeskuksen liian nopeasti itään. Assimilaation toteutuksen viat havaitaan kuitenkin pieniksi ja ASCAT-havaintojen assimilaation operatiiviseen HIRLAM sääennustemalliin nähdään olevan toteuttamiskelpoinen, tosin pidemmän aikavälin validoinnin tarve nousee vielä esiin.

Avainsanat — Nyckelord — Keywords
ASCAT, skatterometri, data-assimilaatio, vaikutustutkimus, HIRLAM

Säilytyspaikka — Förvaringsställe — Where deposited
Kumpulan tiedekirjasto, Helsingin yliopisto

Muita tietoja — Övriga uppgifter — Additional information

Helsingin yliopisto — Helsingfors universitet — University of Helsinki

Tiedekunta/Osasto — Fakultet/Sektion — Faculty	Laitos — Institution — Department
Faculty of Science	Department of Physics

Tekijä — Författare — Author
Pirkka Ollinaho

Työn nimi — Arbetets titel — Title

Feasibility of assimilating ASCAT surface winds into Limited Area Model

Oppiaine — Läroämne — Subject
Meteorology

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
M.Sc Thesis	March 2010	81

Tiivistelmä — Referat — Abstract

Sea-surface wind observations of previous generation scatterometers have been successfully assimilated into Numerical Weather Prediction (NWP) models. Impact studies conducted with these assimilation implementations have shown a distinct improvement to model analysis and forecast accuracies. The Advanced Scatterometer (ASCAT), flown on Metop-A, offers an improved sea-surface wind accuracy and better data coverage when compared to the previous generation scatterometers.

Five individual case studies are carried out. The effect of including ASCAT data into High Resolution Limited Area Model (HIRLAM) assimilation system (4D-Var) is tested to be neutral-positive for situations with general flow direction from the Atlantic Ocean. For northerly flow regimes the effect is negative. This is later discussed to be caused by problems involving modelling northern flows, and also due to the lack of a suitable verification method. Suggestions and an example of an improved verification method is presented later on.

A closer examination of a polar low evolution is also shown. It is found that the ASCAT assimilation scheme improves forecast of the initial evolution of the polar low, but the model advects the strong low pressure centre too fast eastward. Finally, the flaws of the implementation are found small and implementing the ASCAT assimilation scheme into the operational HIRLAM suite is feasible, but longer time period validation is still required.

Avainsanat — Nyckelord — Keywords
ASCAT, scatterometer, data-assimilation, impact study, HIRLAM

Säilytyspaikka — Förvaringsställe — Where deposited
Kumpula Science Library, University of Helsinki

Muita tietoja — Övriga uppgifter — Additional information

Contents

1 Introduction

The weather affects us all every day of our lives. Sociological or recreational impacts can be seen for example on a sunny summer day making the beaches crowded and getting people escape to their summer cottages (view from a lakeside cottage shown in Fig 1), or on a rainy one keeping everyone huddled inside. From economic viewpoint, the weather has a major impact on business (selling ice creams on a sunny day), and more importantly on transportation, whether it is on land, seas or in air. Since the impacts of the weather are so wide and profound, knowing what the weather is going to be in advance has become an essential aid in many aspects. Predicting the weather successfully some days ahead can help ships choose a more optimal route or inform the ice cream salesman to stock up! Farmers are also able to determine whether to wait for the rain or turn on the irrigation system. Getting a warning of an extreme weather situation (heavy precipitation, high wind conditions etc.) beforehand also helps saving lives and preserving infrastructure from imminent destruction.

Figure 1: Lake Ojajärvi and sunny weather.

For these, and many other reasons, the science of numerical weather prediction (NWP) has been one of the focal points of meteorological research in latter part of 20th century. An important milestone in the NWP history was in 1904 when V. Bjerknes recognized that weather forecasting is fundamentally an initial-value problem in mathematical physics. Moreover, he identified that the basic system of equations to be solved was already known (Haltiner & Williams, 1980). However the system of equations was highly non-linear and did not possess any closed solutions. Though attempts to solve the equations numerically were made by L.F. Richardson during and after World War I using a desk calculator (and failed due to poor initial values after months of calculations), it was not until late 1940s and the inventing of electronic computer that stimulated the next milestone (Haltiner & Williams, 1980). In 1950 J.G. Charney, R. Fjørtoft and J. von Neumann produced the first successful dynamical-numerical forecast at 500 hPa using the newly invented computer (Kalnay, 2003). Since then the development of NWP models has been a rapid one, and concomitant with the remarkable growth of computation power.

Around 1990 the NWP forecast accuracies reached a level after which the accuracy improvement has been less rapid. This was a consequence of model stage of development reaching a level after which the accuracy improvement was mainly dominated by improvements of model physics and initial conditions rather than the model dynamics. Therefore the impacts of various improvement methods do not any more have as large an impact in the forecast accuracies. Naturally this does not mean that the improvements are unnecessary. On the contrary, all the improvements help making the NWP systems closer to simulating the behaviour of the real atmosphere without errors. Figure 2 illustrates the improvements in forecast accuracy of the Finnish operational High Resolution Limited Area Model (HIRLAM). The two gateways into affecting the forecast accuracy are improving the model initial state, and adding and tuning model physics. The former is done through procedure called data-assimilation. The aim of data-assimilation is to combine prior information of the atmospheric state, usually taken to be previously made forecast, with observation data in the best possible way. Data-assimilation has grown to one of the heaviest investments in NWP.

Variational data-assimilation methods enable the use of many sorts of observation data in finding out the most accurate initial-condition for NWP. Quite useful of these

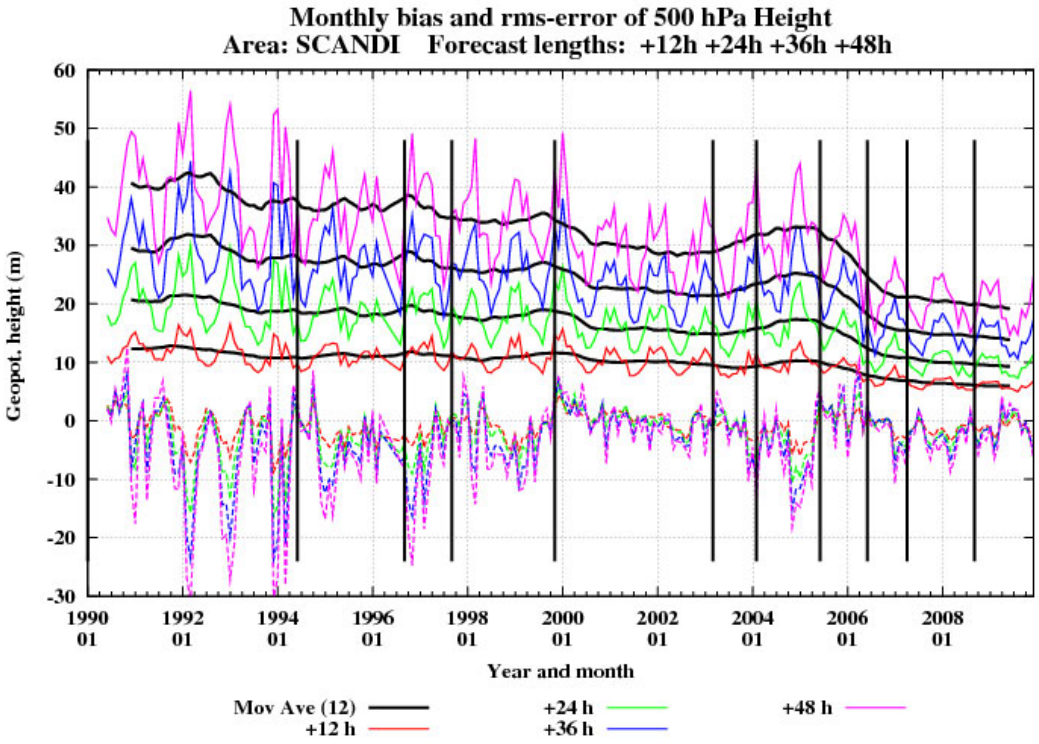


Figure 2: Development of operational (RCR) HIRLAM forecast accuracy from 1990 onwards in Scandinavia area. Solid lines indicate rmse and dashed lines bias for the different forecast lengths. Vertical black lines mark introduction of different HIRLAM version. (From Eerola, Kalle)

data sources are potentially the satellite observations. Space-borne instruments provide global coverage of radiance measurements and thus offer invaluable data from areas where no other observations are normally available, like oceans.

During the 1960s an extensive series of observations using aircraft-mounted radars indicated that there is a relationship between surface wind and radar return (Kidder & Vonder Haar, 1995). This led to a specialized radar know as a *scatterometer*. Though the microwave part of the electromagnetic spectrum was first explored from space by the Soviet satellite Kosmos 243 already in 1968 (Kidder & Vonder Haar, 1995), it was not until 1978 that the first scatterometer was flown as a satellite payload. The satellite, Seasat, unfortunately suffered a power supply failure during the same year. Thus the scientific gain from the scatterometer on board was small. Since then there have been many successful scatterometer missions (e.g. ERS-1 1991-2000, ERS-2 1995 to present and QuickSCAT 1999 to present) providing the much needed wind

information from sea areas.

In addition to surface wind information scatterometer provides ice cover measurements from water surfaces, and soil moisture and snow cover information from land areas. The retrieval of all the quantities is basically done by comparing received backscatter values from the surface with values the surface could theoretically backscatter in various surface conditions. Since microwaves penetrate cloud and rain effectively, scatterometer is able to provide surface information regardless of the weather situation.

The idea of conducting short single case studies in determining scatterometer assimilation impact in a NWP model is not a new one. The usefulness comes from avoiding computationally demanding long period model runs and still producing vital information of the assimilation. Previous studies (see e.g. Isaksen & Stoffelen (2000); Isaksen & Janssen (2000); Leidner *et al.* (2002)) have, however, mainly concentrated on tropical cyclone verification. In this thesis case studies are made for situations more important for European weather, e.g. a polar low. The motivation is to find out whether the ASCAT assimilation effect is noticeable in HIRLAM forecasts, and to determine to which direction this effect is. Since assimilation of ASCAT wind data has proven to have a positive impact on forecast accuracies in global modelling (see Hersbach & Janssen (2007)), it is hypothesised that this will hold for Limited Area Model (LAM) also.

The aim of this study is to find out how feasible implementing an ASCAT assimilation scheme into the operational HIRLAM NWP system is, and recognizing possible flaws in the implementation.

2 Scatterometry

2.1 Microwave remote sensing

The microwave and millimetre region of the electromagnetic spectrum is usually considered to extend from 0.3 to 300GHz (1 m to 1 mm in wavelength). The atmosphere is generally transparent in the low frequency part of the regime with few absorption bands but becomes opaque with higher frequencies (Figure 2.1) . The range is divided into a set of bands denoted with letters P, L, S, C, X, K_u, K and K_o. For *weather radars*, three frequency bands are generally used: X, C and S (WMO, 2006). For example, the previous generation of weather radars in Finland operated in the X-band, but after the upgrade of the radar network, in the C-band (Saltikoff *et al.* , n.d.). Space-borne scatterometers have been using varyingly C- and K_u-bands throughout their short history.

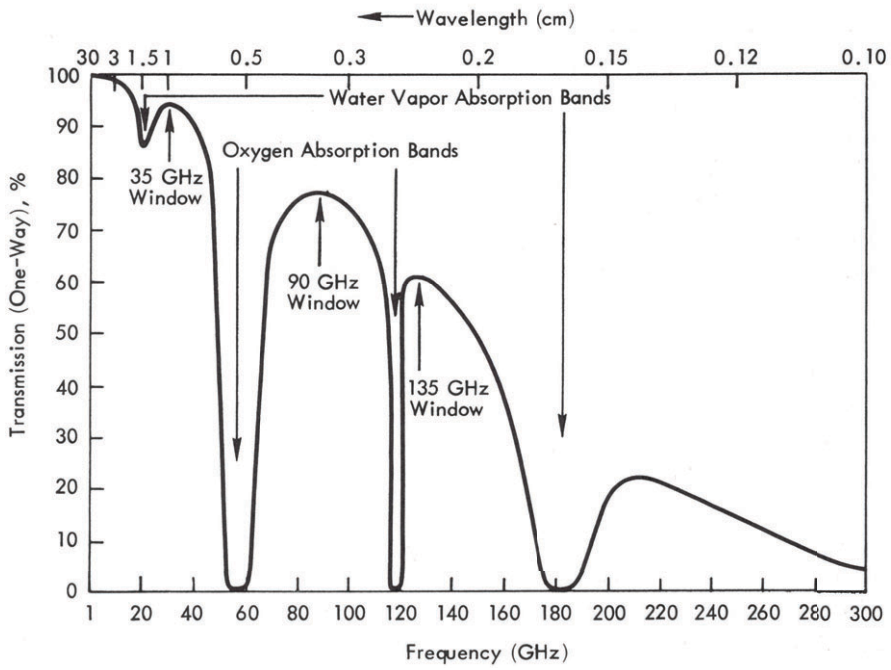
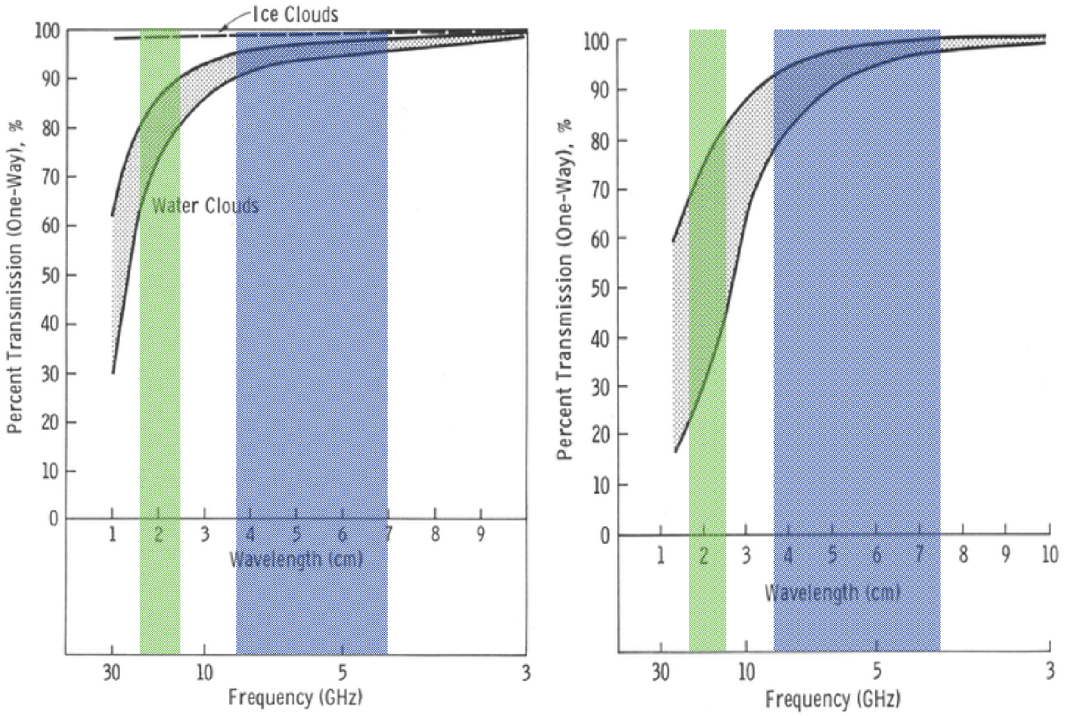


Figure 3: One-way transmission of the atmosphere in the microwave and millimetre region. (After Ulaby *et al.* (1981); Stephens (1994))

The advantage of using microwaves comes from their capability to penetrate into and trough clouds and, depending on the used wavelength, also rain. Figure 4 illustrates the microwave cloud and rain penetration capabilities as a function of frequency.

The blue area represents C-band radar frequency range (4-8 GHz) and the green area the frequency range of K_u-band radar (12-18 GHz). Even though both bands have good cloud penetration capabilities, it is easy to notice that C-band is far less affected by rain than K_u-band.



(a) (b)

Figure 4: a) Effect of clouds and b) rain on one-way transmission in the frequency range 3-30 GHz. Blue area represents the frequency range of C-band radars (4-8 GHz) and green area that of K_u-band radars (12-18 GHz). (Adapted from Ulaby *et al.* (1981))

In addition to the ability to penetrate cloud and rain, microwaves also penetrate more deeply into vegetation when compared to optical waves. The penetration depth or efficiency depends on the frequency used and dielectric properties of the target. By using many frequencies information about the vertical structure of the vegetation can be gained. Moreover, the penetration also applies to the ground itself, which makes microwaves applicable to e.g. geological studies.

2.1.1 Active microwave remote sensing of the ocean

Ocean waves have a very complex topography. The energy transfer between the waves works like reversed turbulence: whereas in turbulence bigger whorls give (or lose)

energy to smaller ones, in the ocean smaller waves give (or lose) energy to bigger ones.

For turbulence in air flow:

Big whorls have little whorls,
Which feed on their velocity;
And little whorls have lesser whorls,
And so on to viscosity
(in the molecular sense)
Richardson, 1922

When wind starts to disturb a calm sea, short waves form due to frictional drag between the surface and the wind. These waves build up and transfer their energy through non-linear interactions to waves with larger amplitudes and longer wavelengths (Ulaby *et al.* , 1986). When the wind is continuous, longer and longer waves are generated until the surface reaches an equilibrium point. On this equilibrium point dissipation mechanisms balance the tendency for wave growth. The equilibrium depends on the strength of the wind: the stronger wind the longer and higher waves are generated. Nevertheless, the primary transfer of energy from the atmosphere to the sea is by the very shortest waves. If the wind was to die out suddenly, the short waves would decay rapidly, whereas the longer waves could exist as long as several days (Ulaby *et al.* , 1986). The longer waves can thus propagate great distances due to their long life time, whereas the shorter waves attenuate more quickly and are hence more localized. Because of this, the waves at a given point of the sea surface are a complex sum of the locally generated wind waves and the waves that have propagated in from other areas. Adding to the complexity, the propagating waves come from different directions. Ocean waves are also dispersive, hence the wave spectra have single broad peaks (Figure 5).

Radar backscattering from the sea surface from angles of incidence beyond approximately 20° is governed by Bragg scattering (Ulaby *et al.* , 1982). With these incidence angles the backscattering response often arises primarily from resonant components. For scatterometer frequencies these components are tiny capillary (surface-tension) and gravity waves, often referred to as wind-induced gravity-capillary waves.

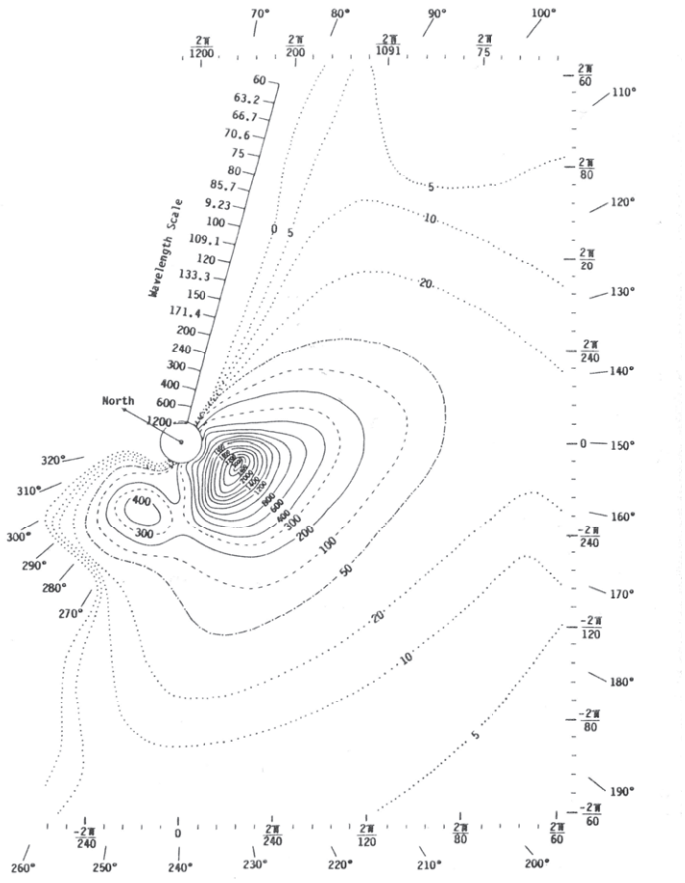


Figure 5: Example of the directional spectrum of a wind-generated sea (determined from stereo photography). Contours of constant amplitude are plotted on the (K, ϕ) plane. (After Ulaby *et al.* (1986))

Even though these waves have a height measured in tenths of a millimetre they have larger backscatter response than the meters high waves they ride upon. Figure 6 illustrates the mechanism: the incoming plane wave has an angle of incidence θ with wavelength λ and the wavelength of the surface wave is L. Now if the excess distance from the source to each successive wave crest (ΔR) is $\lambda/2$, the round-trip phase difference between the return signals from successive crests is 360°, so the signals add in phase (Ulaby *et al.* , 1982). This means that the backscattering power from this resonance is proportional to the square of the number of scatterers ($(N+1)^2$), whereas the average power of non resonant scatterers is only proportional to their number $(N+1)$ (Ulaby *et al.* , 1982). Thus the effect of these very small resonant scatterers in large numbers can exceed that of the much larger ones with no resonance. It should be noted that the resonant wavelength depends on the local angle of incidence, not the

angle relative to the vertical. This means that the Bragg-resonance backscattering from the front to the back face of a larger wave varies as the according Bragg resonant wavelength changes.

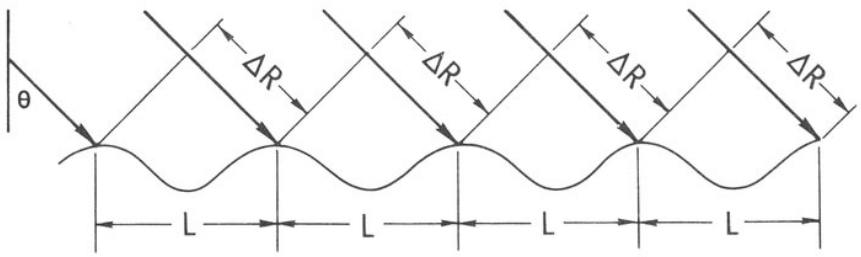


Figure 6: Bragg scattering mechanism: in-phase addition when $\Delta R = n\pi/2$. (From Ulaby *et al.* (1982))

The average of the radar backscattering signal is therefore governed by the local wind speed, but its distribution depends on the slopes of the longer waves that may have propagated from other areas.

At near-vertical angles of incidence the primary backscattering mechanism changes to a quasi specular one, in which geometric- or physical-optics dominate over the Bragg scattering. At (and near) vertical incidence the strongest backscattering would occur with a perfectly flat surface or one that is undulating very gently. When the surface gets rougher, more of incident energy is scattered away from the incidence angle, i.e. the received signal at the radar gets weaker. Hence the radar backscattering decreases with increasing wind speed and wave height near vertical incidence angles, whereas it increases for incidence angles beyond about 12° (Ulaby *et al.* , 1982).

2.1.2 Radar footprint

Consider a point source emitting a microwave pulse of power P_T uniformly into all directions of space. At distance R the power flux will then be $\Phi_T = P_T/(4\pi R^2)$. If Ω is the beam width in steradians (Fig 2.1.2) in which the emitted pulse is contained, the antenna gain can be defined as $G = 4\pi/\Omega$. The gain equals 1 for isotropic emission, and is greater for more contained, i.e. narrower, beams. Thus the power flux can be written as $\Phi_T = GP_T/(4\pi R^2)$ (Ulaby *et al.* , 1982).

Now suppose that microwave radiation in the beam hits a scatterer. Radar cross section of the scatterer is defined as

$$\sigma = \frac{P_S}{\Phi_T} \tag{1}$$

where P_S is the power scattered towards the radar (backscattered power). If the receiving antenna is of same size and geometry as the source antenna, the received backscattered power is

$$P_R = \frac{P_S A}{4\pi R^2} = \frac{G P_T \sigma A}{(4\pi R^2)^2} \tag{2}$$

Assuming an antenna of dimensions L_x by L_y, and a system without power losses, the antenna gain G and surface area A have a relationship since $\Omega = (\lambda/L_x)(\lambda/L_y) = \lambda^2/A$. Keeping in mind the definition for gain, the relationship then becomes $A = \lambda^2 G/4\pi$.

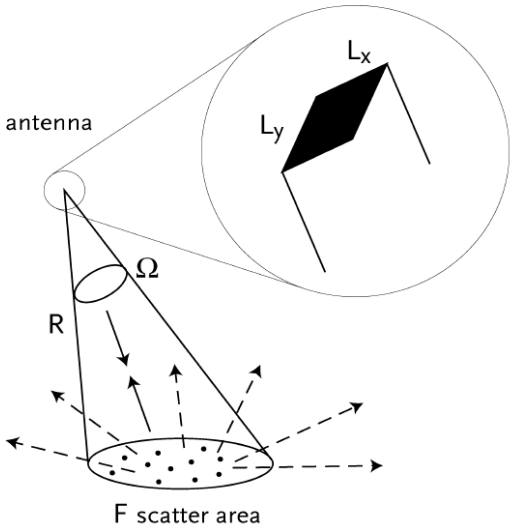


Figure 7: Beam width, antenna geometry and area of backscattering involved in radar footprint. (After Stoffelen (1998))

When considering scatterometer, the received backscatter signal is from an area of tens of kilometres in diameter with generally a large number of various scattering elements in it. For this footprint, F, one can derive a dimensionless microwave cross section per unit surface, often referred to as "Normalized Radar Cross Section",

NRCS, and written by convention as σ^0 (Stoffelen, 1998). The received power can now be written as:

$$P_R = \frac{\lambda^2}{(4\pi)^3} \int_F \frac{P_T \sigma^0 G^2}{R^4} dF \tag{3}$$

If one assumes that σ^0 does not vary over the area in consideration the integral can be solved such that:

$$\sigma^0 = \frac{(4\pi)^3 R^4}{\lambda^2 G^2 F} \frac{P_R}{P_T} \tag{4}$$

Of course in reality the roughness elements on the ocean surface vary greatly depending on the local wind conditions, and the illuminated area encompasses various sized waves. Thus the average spectra covers the effects of ripples on the crests and troughs of waves and those between them (Ulaby *et al.* , 1986), and induces large variability over a scatterometer footprint. As the scattering mechanism does not have a linear dependency on the geophysical condition, the geophysical variability will contribute to σ^0, particularly at low winds (Stoffelen, 1998). Although theoretically not obvious, it is empirically found that σ_0 keeps increasing for increasing wind speed from 25 m/s to 40 m/s, and that a useful wind direction dependency remains (Donnelly *et al.* , 1999), albeit gradually weakening (O&SI-SAF, 2010).

Usually σ^0 is expressed in dB, i.e. as the value 10 log(σ^0).

2.2 Scatterometer

Scatterometer is a monostatic non-nadir looking Real Aperture Radar (RAR). In terms of antenna geometry scatterometers can be classified as (1) side-looking and (2) rotating (Portabella, 2002). Of these the former is discussed in more detail.

The side-looking scatterometers consists of multiple fan-beam antennae with a fixed orientation. All of the antennae point to one or both sides of the satellite flight path. The incidence angles with these configurations are between $15° < \theta < 70°$ thus Bragg scattering and specular reflection determine the radar backscattering.

The idea behind using multiple antennae is to get σ^0 values from an area with different viewing angles. The return signal largely depends on just two geophysical

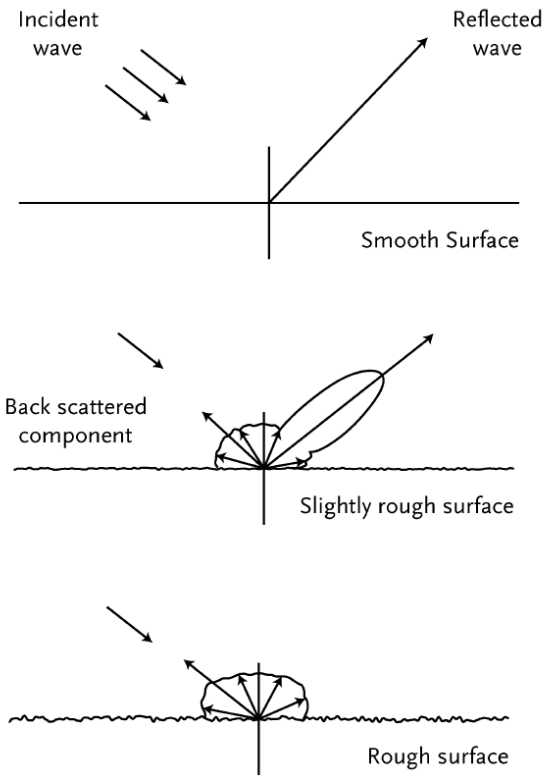


Figure 8: Effect of surface roughness on scattering from a surface. (After Stoffelen (1998))

parameters, which can be taken to be wind speed and direction. Retrieving wind from this dataset is then an *inverse problem*: what kind of wind would generate such sea surface conditions that the backscattering footprint would match the observed one.

To describe the wind-to-backscatter relationship a so called geophysical model function (GMF) is used. Results in finding a theoretical solution to the GMF have not been satisfactory so far due to poor understanding the ocean topography (Stoffelen, 1998). Thus a number of empirical GMFs have been developed and widely used for the wind-to-backscatter retrieval. A general empirically determined forward model function (GMF) which relates wind speed and direction to measured radar backscatter, is defined as (Portabella, 2002):

$$\sigma^0 = B_0(1 + B_1 cos(\phi) + B_2 cos(2\phi))^Z \tag{5}$$

where ϕ is the wind direction, and the coefficients B_0, B_1 and B_2 depend on the wind speed, the local incidence angle, and the polarization and frequency of the radar beam. The value of Z (and the number of possible additional harmonics) depends on the tuning done for each GMF.

Although the energy density of the Bragg waves is actually related to the surface wind stress $\overline{\tau}$ (i.e. the impact the wind has on the sea surface), the reference wind used by the GMFs is the wind at 10-meters height ($\mathbf{U}_{10}$). Granted that there exists a relationship between the two, $\overline{\tau} = -\langle \mathbf{u}' w' \rangle = C_D U_{10} \mathbf{U}_{10}$, the drag coefficient C_D depends on wind speed and its determination is still uncertain. Thus by directly estimating the $\mathbf{U}_{10}$-to-σ^0, the mean behaviour of $\overline{\tau}$ is taken into account implicitly (Portabella, 2002). Moreover $\mathbf{U}_{10}$-observations for validation are much easier to get and more available than $\overline{\tau}$ observations. There also arises another problem from the uncertainty in determining C_D, since atmospheric stability is known to effect C_D (Savijärvi & Vihma, 2001). This means that the stratification of the lowest 10 meters should be taken into account when using real winds in the estimation of GMF. An alternative is to correct the measured winds ($\mathbf{U}_{10}$) into equivalent neutral winds ($\mathbf{U}_{10N}$) when determining the GMF. This is done by using drag coefficient C_{DN} corresponding to neutral stratification. Although this transformation is useful, neutral winds do not have meteorological usage as such, and buoy or NWP model data is required for correcting the neutral winds into real winds.

For the inversion to be successful and more accurate the number of independent σ^0-values from the same area (usually referred to as Wind Vector Cell or WVC) is of particular importance. In the case of three (or more) views and a good azimuth diversity (spread of views) the inversion problem is overdetermined. 90° angle between fore and aft views with mid view precisely in the middle fulfils the azimuth diversity requirement (Portabella, 2002). Figure 9 presents an ideal case for three σ^0-values with good azimuth diversity. The unique intersection (right circle) of the three solution curves potentially denotes a unique solution, however there is another location where the lines nearly intersect (left circle), denoting a secondary solution.

In reality, measurement noise will almost always prevent any triple intersection, thus two solutions with similar minimum curve-distance values are produced. Furthermore the σ^0 triplets place themselves around a conical surface in [σ^0_{fore}, σ^0_{aft}, σ^0_{mid}]-space. The minimum curve-distance of these points from a conical surface, defined by the GMF in use, is then derived by using e.g. the Maximum Likelihood Method. As a result, the inversion gives two equally likely ambiguous wind solutions, and usually additional less likely solutions. To select one of the ambiguous wind solutions addi-

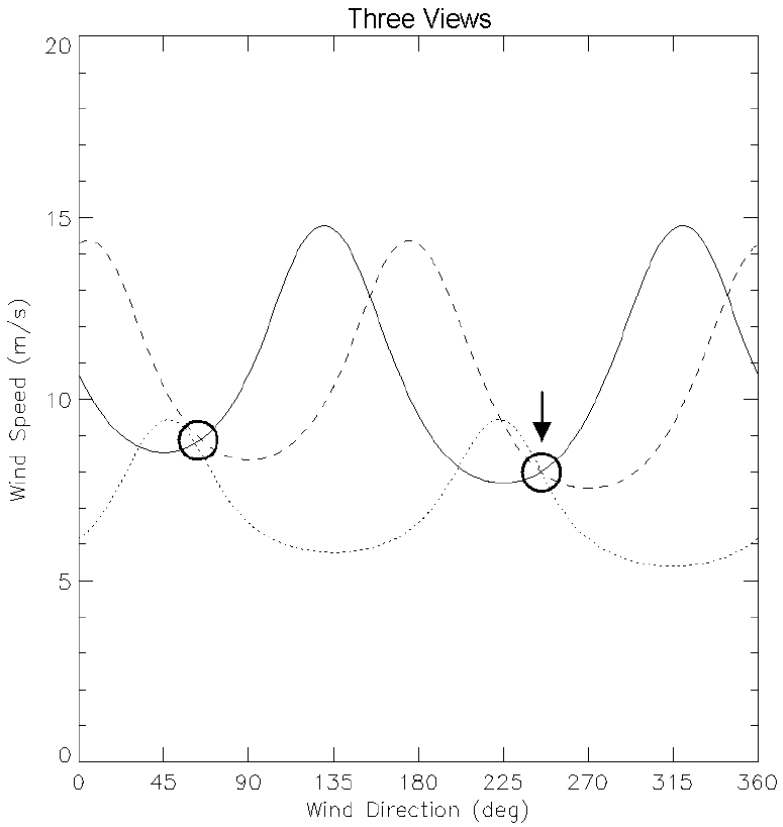


Figure 9: Curves representing set of wind speed and direction values, which satisfy the GMF for a single σ^0 measurement, produced by a wind of 8 m/s and 245° (arbitrary reference). The incidence angle of the views is 54°. Solid curve represents 45°, dashed 90° and dotted 135° view. The arrow points the "truth" and circles show possible wind solutions. (Adapted from Portabella (2002))

tional information from NWP models or buoys and spatial consistency constraints are therefore needed. In addition to this ambiguity removal, the output received from the inversion goes through a comprehensive quality control. The goal of quality control is to detect and reject poor WVCs (Portabella, 2002). Geophysical phenomena, such as sea ice or confused sea state (also rain effects for K_u-band radars), can contaminate the radar observations and in turn decrease the quality of the retrieved winds. A schematic illustration of the processes that the scatterometer data goes through can be seen in Figure 10.

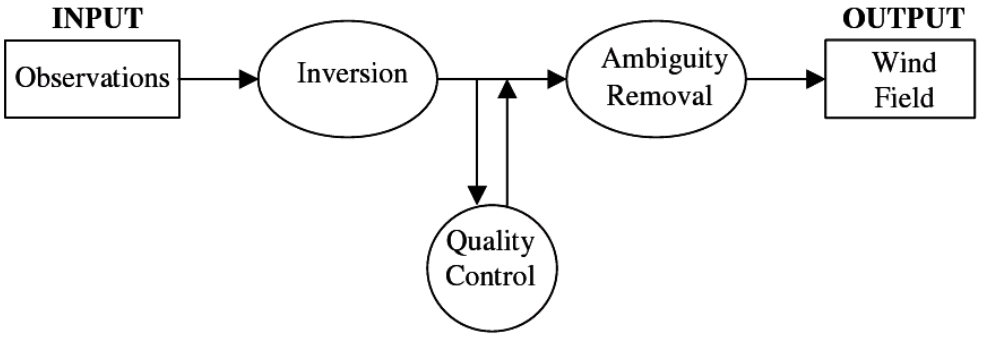


Figure 10: Scatterometer wind retrieval process chart. (After Portabella (2002))

2.3 ASCAT

The Advanced Scatterometer, ASCAT, is part of the payloads of the Meteorological Operational (Metop) polar satellites launched by the European Space Agency (ESA) and operated by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT), and constitute the space segment of EUMETSAT s Polar System (EPS). Metop-A, the first in the series of three satellites, was launched on 19th of October 2006 and two successors are planned to be launched approximately 4-5 years after each other (O&SI-SAF, 2010). Metop orbit achieves global coverage in a period of five days (Gelsthorpe *et al.* , 2000).

ASCAT is a real aperture radar operating at 5.255 GHz (C-band) with system geometry based on the use of fan-beam antennas. The system covers two 550 km swaths separated by about 700 km. The antennae are oriented 45°, 90° and 135° with respect to the satellite track. The system geometry is illustrated in Figure 11.

The GMF used for calculating surface winds is currently CMOD5.n. This GMF solves equivalent neutral winds instead of real winds and can provide up to four wind solutions in each WVC. These ambiguities are removed by applying constraints on the spatial characteristics of the output wind field, such as on rotation and divergence. Furthermore European Centre for Medium-Range Weather Forecasts (ECMWF) wind forecasts are used for final ambiguity removal. The wind product also undergoes validation and quality control through an ad hoc visual examination of the graphical products and the automatic production of control parameters.

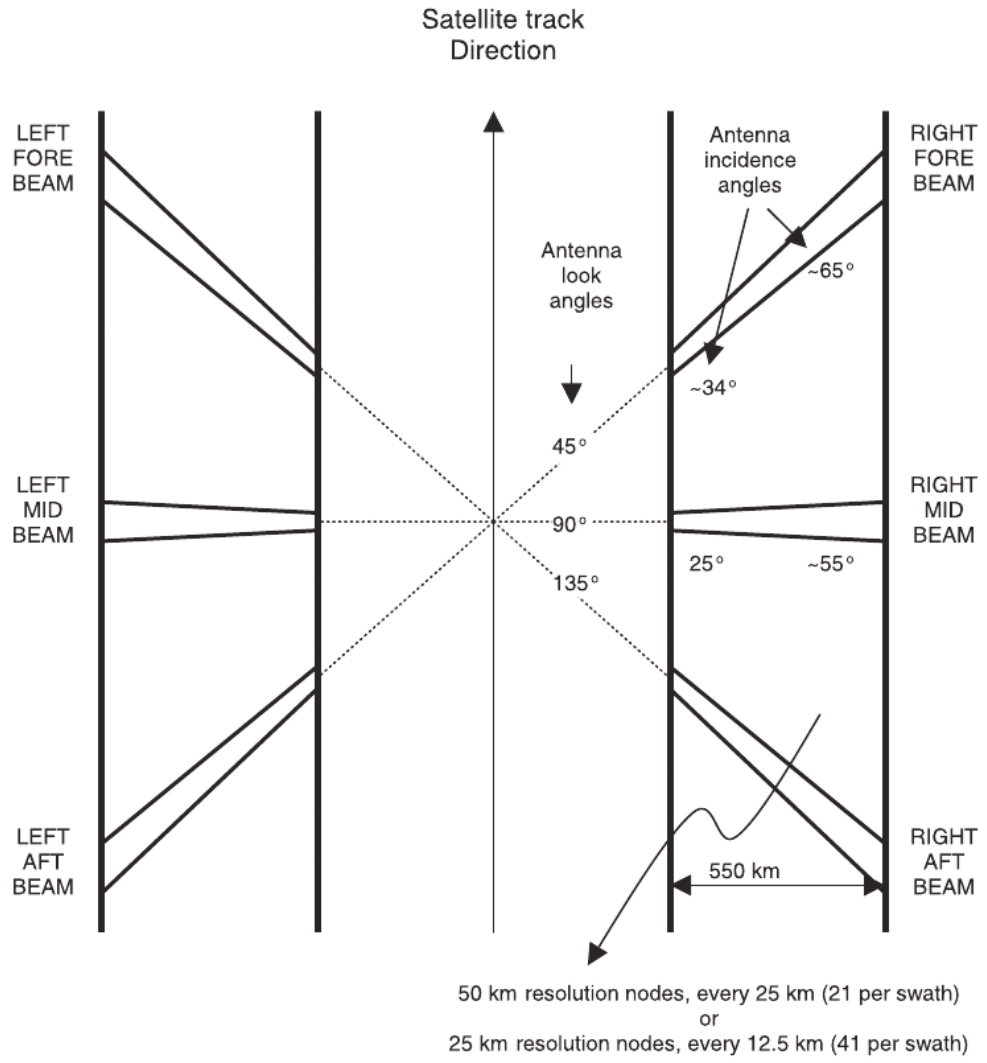
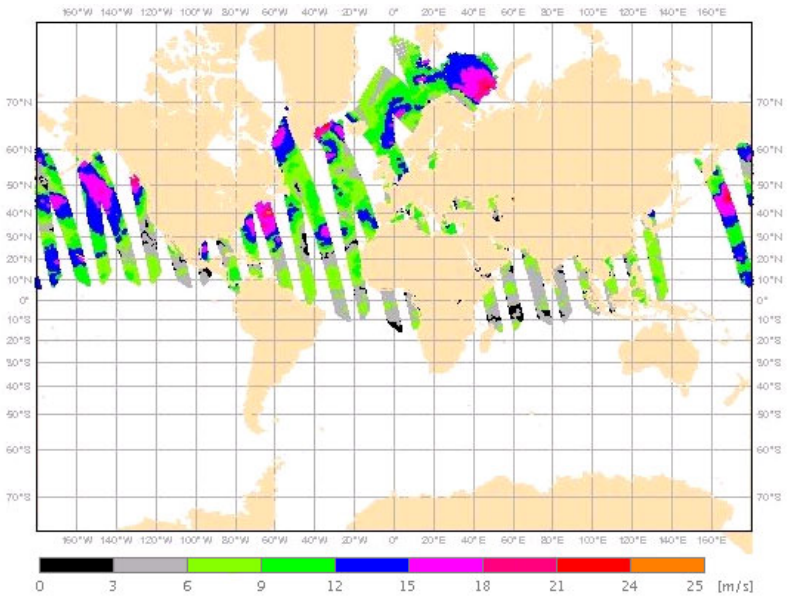
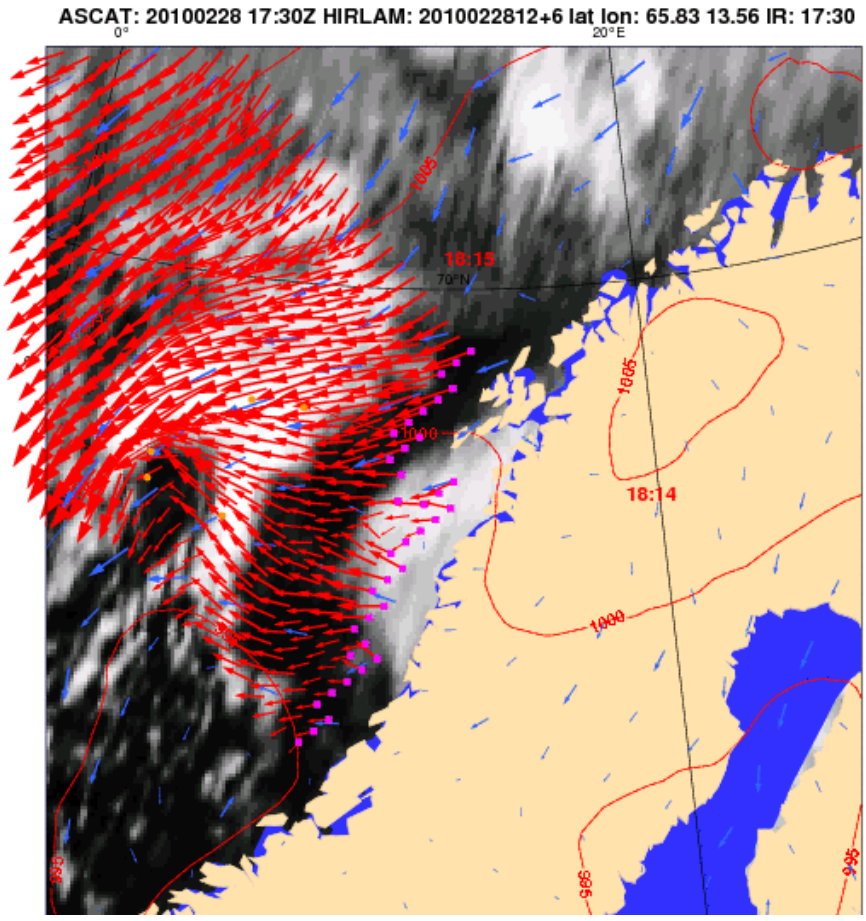


Figure 11: ASCAT system geometry . (After Figa-Saldaña *et al.* (2002))

The final wind products are available in two resolutions: a 50-km resolution product with 25-km cell spacing and a 25-km resolution product with 12.5-km cell spacing. Both resolutions are available as a regional EUMETSAT Advanced Retransmission Service (EARS) product with a timeliness of approximately 30 minutes from sensing time (O&SI-SAF, 2010). Figure 12a illustrates an arbitrary 22-hour coverage from EARS network area and Figure 12b shows finalized product of ASCAT surface winds.



(a)



(b)

Figure 12: a) 22h ascending passes over EARS network coverage at 11UTC 01.03.2010. b) Close up of 10 m wind vectors from Norwegian west coast. (Adapted from KNMI scatterometer webpage)

3 Numerical weather prediction

3.1 Numerical modeling

Numerical meteorology can be described being computational flow dynamics (Savijärvi, 2007). The chaotic behaviour of the atmosphere can be solved to some extent by approximating laws of nature with a system of highly non-linear partial differential equations. In order to predict (forecast) the behaviour of the atmosphere from a given starting time, the initial state of the atmosphere needs to be known. This initial state is presented in a 3-dimensional grid, each grid point containing values of air pressure (p), temperature (T), water vapour mixing ratio (q) and wind components (u, v) at the location represented by the grid point, or rather as an areal average of the grid point location. By applying our system of equations to this initial state local tendencies (rates of change) of the variables can be solved at each grid point. Furthermore the tendencies can then be used for calculating new prognostic values for the given variables at each grid point by advancing in time with a given time step. By replacing the initial values with the prognostic ones new tendencies can be calculated for each grid point. By repeating this through predefined time steps a forecast of the future state of the atmosphere can be made. An example of the final model output is shown in Figure 13.

As this thesis is inclined towards weather forecasting, I will discuss numerical modelling from the perspective of numerical weather prediction (NWP).

Given the chaotic nature of the atmosphere and the approximative nature of our computational methods (truncation error), the predictability of the weather gets worse the more ahead in time the forecast reaches. Due to the chaotically behaving atmosphere small differences between the real state of the atmosphere and the initial estimate formulated from non uniformly spaced observations introduce a growing error to our forecast (initial-value problem). J. G. Charney suggested in 1951 that there was an upper limit to weather predictability, which E. N. Lorentz later in 1965 estimated to be about two weeks (Kalnay, 2003). Since that time numerical forecasting has taken huge steps forward, but the 14-day rule holds. Only in special cases, when the weather is highly predictable, the limit can occasionally be reached and even exceeded through

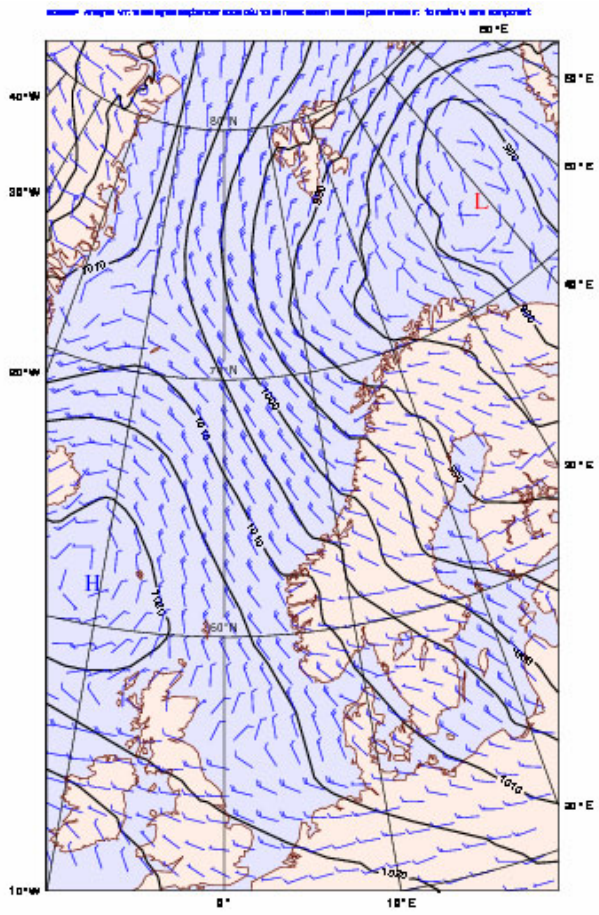


Figure 13: Output from a numerical model. ECMWF mean sea level pressure and 10m wind field over Scandinavia.

techniques like ensemble forecasting (Kalnay, 2003).

The amount of detail a numerical model is able to present is determined by the spacing between its grid points. Naturally the smaller grid increment a model has the more grid points it also has. The tendencies must be calculated for each individual grid point, thus making the grid spacing smaller increases the computation time the model needs. As computing power is by no means limitless, this is a problem. In order to achieve better horizontal resolution, a model with smaller grid increment but also with smaller areal coverage (limited area model, LAM) is *nested* into a model covering larger area but having larger grid increment. Operative forecast from the bigger areal coverage model is used for updating the boundary conditions of the limited area model. Thus information of the synoptic evolution outside the LAM area is provided for the regional model. This information feeding leaves the regional model somewhat a slave to its boundary conditions, and any errors induced to the regional model at

the boundary are subject to growth and might advect further inside the regional model domain with suitable wind conditions (Savijärvi, 2007). Similarly to the initial-value problem, these errors i.e. differences from the real state of the atmosphere in the boundary conditions must be taken into account in regional models (boundary value problem).

Since the atmosphere is three-dimensional, the numerical models trying to simulate it need to have horizontal dimension as well. Basically this is done by just overlaying grid points vertically. It is conventional for meteorological usage to employ constant pressure levels (e.g. in radio soundings) The problem with pressure for models is that surface pressure varies and especially in the mountainous areas one or several low model levels would always be inside the ground. To overcome this complication, Phillips (1959) introduced a "normalized pressure" or "sigma" coordinate system, where $\sigma = p/p_s$ and $p_s(x, y, t)$ is the surface pressure (Kalnay, 2003). It follows from these that at the surface, $\sigma = 1$, and at $p = 0$, $\sigma = 0$, so that at the top and bottom boundary of the atmosphere $d\sigma/dt = 0$. These simple boundary conditions, and the fact that the sigma-levels near the ground follow the shape of the ground, have made the sigma-coordinate system useful. Many models use hybrid vertical coordinates which follow the sigma surfaces near thhe ground but gradually change into constant pressure levels when nearing the stratosphere.

The model grid is usually staggered in horizontal and vertical directions. For example the geopotential height ϕ can be calculated at the u, v, T, q-levels (Lorentz grid) or between them (Charney-Phillips grid). Horizontal staggering is done e.g. according to the Arakawa C-grid, which defines wind components u and v between the $T, q, \phi, d\sigma/dt$-points. Applying such a grid staggering to the model improves accuracy and reduces computational "noise" (Savijärvi, 2007).

3.1.1 Model dynamics

V. Bjerknes recognized in 1904 that 5 fundamental equations formed a determinate system which, in principle, could be solved to forecast the subsequent state of the atmosphere from a known initial state. These equations are:

- Newton s second law of motion

- The first law of thermodynamics
- The law of conservation of mass or continuity equation
- The equation of state
- The conservation equation for the water substance

When applied to a parcel of air these equations form a system of conservation laws which the parcel must abide: Newton s second law of motion represents the conservation of (three-dimensional) momentum, the first law of thermodynamics the conservation of energy and continuity equation the conservation of dry air mass. As its name states conservation of water substance controls the conservation of moisture in all its phases. For a large range of scales of atmospheric motion the atmosphere may be treated as if it were a perfect gas, hence the equation of state follows that for a perfect gas. See e.g Haltiner (1971) or Kalnay (2003) for the presentation in mathematical form and throughout deduction of the equations.

3.1.2 Numerical methods

The utilization of the primitive equations is done through using numerical integration techniques since analytical solutions are not generally obtainable. The most common numerical integration procedure for weather prediction has been the finite-difference method in which the derivatives in the differential equations of motion are replaced by finite difference approximations at a discrete set of points in space and time (Haltiner & Williams, 1980). With appropriate restrictions the resulting set of equations can be then solved by algebraic methods.

Consider an arbitrary function $f(x)$ and its (positive) Taylor series expansion:

$$f(x+\delta x) = f(x) + f'(x)\Delta x + f''(x)\frac{\Delta x^2}{2!} + f'''(x)\frac{\Delta x^3}{3!} + ... \tag{6}$$

Now solving for $f'(x)$ gives

$$f'(x) = \frac{f(x+\Delta x) - f(x)}{\Delta x} + R \tag{7}$$

where R represents the remaining terms of the series and is referred to as *truncation error*. The truncation error plays a major role in model output error, as the methods are applied for calculating tendencies at grid points.

Carrying out the tendency calculations through model time steps is not usually a simple computational issue. Though one can use a simple straightforward extrapolation (Euler-method), $y_{n+1} = y_n + \Delta t \frac{dy_n}{dt}$, for advancing through time, the truncation error involved in this is big. The more advanced methods used these days use implicit methods, meaning that information from the future time step is needed for the current time step, e.g. Crank-Nicholson method: $y_{n+1} = y_n + \frac{\Delta t}{2}(\frac{dy_n}{dt} + \frac{dy_{n+1}}{dt})$. Though these methods improve numerical accuracy, and allow usage of longer time steps, they are expensive for computation time. To find a balance between the accuracy and computation time, semi-implicit method treats implicitly terms in the equation of motion most susceptible to numerical noise. For a more complete presentation see e.g. Haltiner & Williams (1980).

In addition to error due to discretization, round off during numerical calculations contribute to the total error caused by the approximative nature of the numerical methods to the NWP models.

3.1.3 Physical processes

Phenomena with scales below that of the model resolution (four times the grid spacing at minimum (Grasso, 2000)) can not be simulated with model dynamics. On the other hand these sub grid-scale processes can not be neglected, for they can have enormous impact on local weather. To tackle this problem these phenomena are presented with special *parametrization* in the model. Parametrizations aim to present the subgrid-scale processes with the help of the known model variables and capture the impact of the physical process on average. These diabatic sink and source processes of heat and momentum can be categorized into three main types (Savijärvi, 2007):

- Turbulence and surface friction, and other atmosphere-ground interaction
- Cloud physics; condensation, evaporation and transpiration of water vapour, rain
- Radiation; Solar and thermal radiation and radiative transfer in the atmosphere

The parametrizations have especially large impact near the ground and sea surface, effecting surface wind, temperature at 2 metres, humidity and rain output of the

model. In addition to the model dynamics, the processes affect each other and are also subject to various feedback effects (shown in Figure 14). This makes the whole parametrization process very complex and subject to very intense research (Kalnay, 2003). The parametrizations work usually reasonably well in ordinary weather situations, for which they are made in the first hand, but may leave the model "vulnerable" to extraordinary weather situations (Savijärvi, 2007).

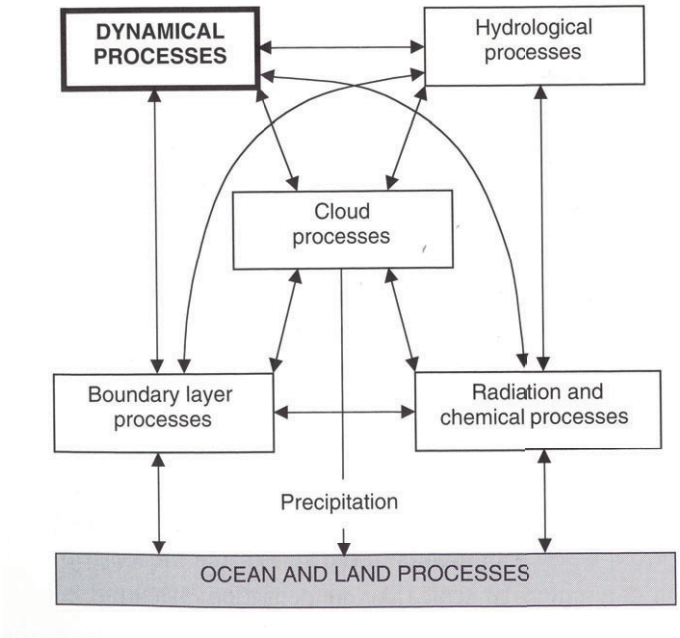


Figure 14: Physical processes in the atmosphere and their interactions. (After Kalnay (2003))

3.2 HIRLAM

The HIRLAM project was established in order to provide the best available operational short-range forecasting system in the member institutes, consisting of National Meteorological Services in Denmark, Finland, Iceland, Ireland, Netherlands, Norway, Spain and Sweden. In addition Météo-France has a research cooperation agreement with HIRLAM consortium. The HIRLAM system is a complete NWP system including Data Assimilation (see next Chapter) with analysis of conventional and non-conventional observations and a limited area forecasting model with a comprehensive set of physical parametrization (Únden *et al.* , 2002).

The HIRLAM model is a hydrostatic grid-point model, with dynamical core based

on a semi-implicit semi-Lagrangian discretization of the multi-level primitive equations, using a hybrid coordinate in the vertical. The prognostic variables u, v, T, q and linearised geopotential height G are defined at full model levels. Pressure p, geopotential height ϕ and vertical wind velocity are calculated at half levels. The Arakawa C-grid is used for horizontal discretization (HIRLAM webpage). The ECMWF global model is used for providing horizontal boundary conditions, but also larger area HIRLAM models can be used for providing these.

The forecast model area of the "reference" HIRLAM (HIRLAM RCR) and area of the meso-beta scale HIRLAM suite (HIRLAM MB71) used at FMI are shown in Figure 15. The reference HIRLAM i.e. the current operational suite, denoted as 7.2., has a horizontal grid increment of 0.15° (about 16 km) with time step of 6 minutes (FMI), and has 60 vertical levels.

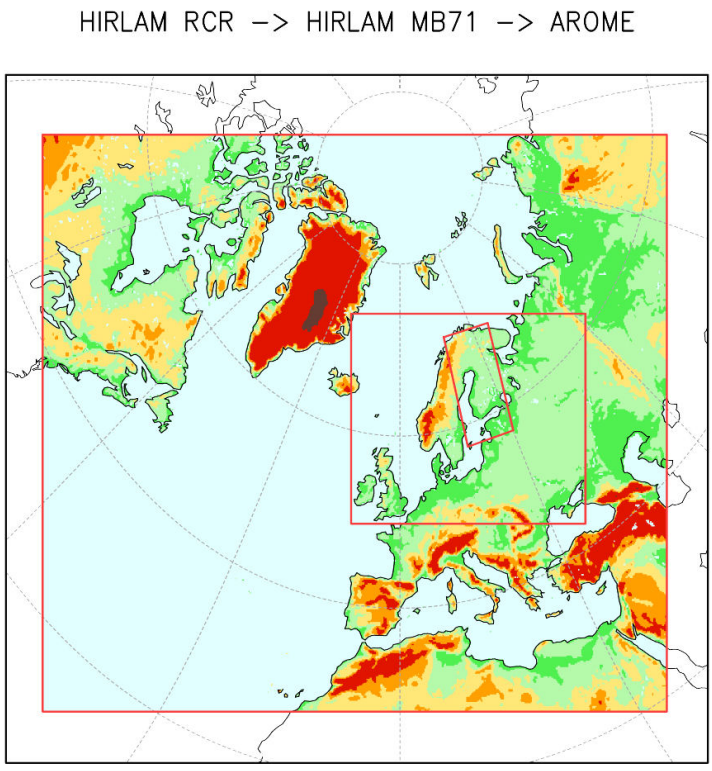


Figure 15: Forecast areas of HIRLAM RCR, MB71 and mesoscale model AROME. (Adapted from FMI HIRLAM webpage)

4 Data assimilation

To conclude from the previous Chapter, NWP is an initial/boundary value problem: given an estimate of the present state of the atmosphere (initial conditions) and appropriate surface and lateral boundary conditions the model simulates (forecasts) the atmospheric evolution (Kalnay, 2003). Due to the chaotic nature of the atmosphere small errors in the initial condition may lead to major errors in the forecast. Thus by improving the accuracy of the initial state the quality of the forecasts is enhanced. The approach of a statistical combination of observations and short-range forecasts, known as data-assimilation, has proven to be practical. The purpose of data-assimilation is defined by Talagrand (1997) as "using all the available information, to determine as accurately as possible the state of the atmospheric (or oceanic) flow." Figure 16 illustrates how the data flow typically proceeds.

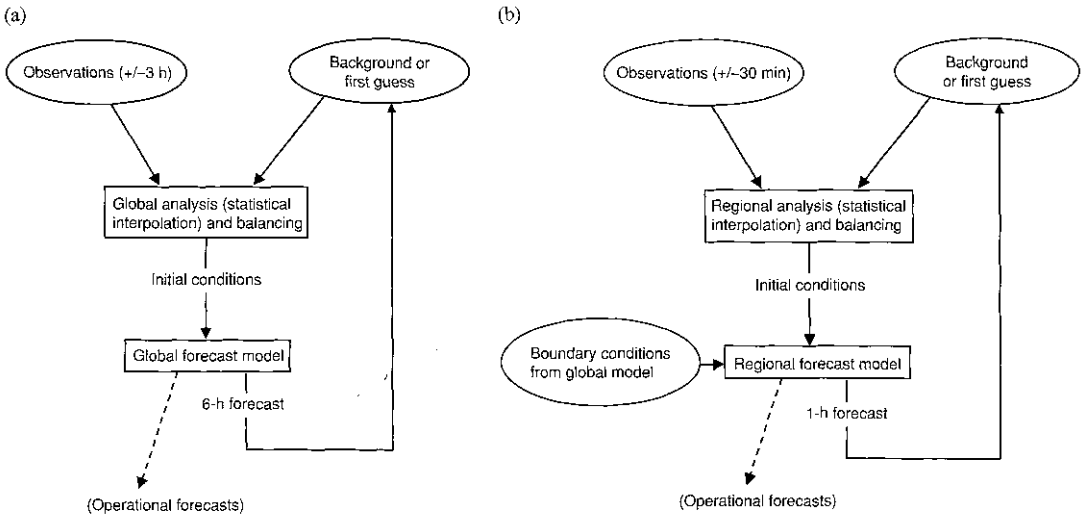


Figure 16: Flow chart of a) global and b) limited area NWP model data assimilation cycle. (After Kalnay (2003))

The task of determining the three-dimensional initial conditions is not an easy task, mostly due to our network of observation stations being extremely heterogeneous: for example land areas of Europe and North-America are well covered with surface synoptic observation (SYNOP) stations, but all ocean areas and the whole southern hemisphere is poorly covered. Satellite data is extremely helpful in filling these observation station gaps, but implementing the data to the data-assimilation system is laborious.

During the previous decade or so the focus of research done with numerical models has shifted from near-total model development towards a situation where data-assimilation is as important as model development. This has been made feasible by the rapid growth of observations available provided by the space-borne instruments. Data-assimilation is quite heavy computationally and in modern numerical models the data-assimilation system can take up to 90% of the computation time and the actual deterministic forecast only 10% (Järvinen, 2003). This kind of separate treatment of the forecast and analysis systems is, however, less valid in the modern NWP world where the two form an inseparable wholeness.

4.1 Basics concepts

4.1.1 Analysis

The aim of analysis is to describe the state of the atmosphere using all available information (Järvinen, 2003). Usually a set of numbers retaining information of the variables in a numerical model grid is used for representing the objective analysis. Subjective weather analysis made by a meteorologist is also an analysis, but it is more of a "diagnosis" of the present state of the atmosphere.

Observations of the true state of the atmosphere act as the basic objective source of information for the analysis. As the number of these observations is very limited, the analysis problem is, local exceptions taken aside, always undetermined, i.e. the number of observations does not fulfil the degrees of freedom of the model. This problem can be evaded by using a *background eld* or *prior information* of the model variables. Normally this a priori knowledge is taken to be a short-range forecast.

4.1.2 Least squares estimation

Suppose there exists a scalar entitity s with N independent observations that are $s_1, s_2, ..., s_N$, and that the observational error is given by $\varepsilon_n = s_n - s$. Assume that the observational errors are random, unbiased and normally distributed. The probability that the error of the nth observation lies between $[\varepsilon_n, \varepsilon_n + d\varepsilon_n]$ is then (Daley, 1991)

$$p(\varepsilon_n) = \frac{1}{\sigma_n \sqrt{2\pi}} \exp\left[-\frac{1}{2}\varepsilon_n^2 \sigma_n^{-2}\right] \tag{8}$$

The function $p(\varepsilon_n)$ is the so-called *probability density function* of the normal distribution and σ_n the standard deviation of the probability distribution.

If $N = 1$, i.e. there is only one observation, the most probable value of s is that value which maximizes the probability $p(\varepsilon_1)$. This occurs at $\varepsilon_1 = 0$, or $s = s_1$. With N observations the joint probability that, $\varepsilon_n \in [\varepsilon_n, \varepsilon_n + d\varepsilon_n]$ for each $n = 1, 2, ..., N$, is the product of all individual probabilities

$$\begin{aligned} p(\varepsilon_1, \varepsilon_2, ...\varepsilon_N) &= p(\varepsilon_1)p(\varepsilon_2)...p(\varepsilon_N) \\ &= \prod_{n=1}^{N} \frac{1}{\sigma_n\sqrt{2\pi}} \exp\left[-\frac{1}{2}\varepsilon_n^2\sigma_n^{-2}\right] \\ &= \left[\prod_{n=1}^{N} \frac{1}{\sigma_n\sqrt{2\pi}}\right] \exp\left[-\frac{1}{2}\sum_{n=1}^{N}(s_n - s)^2\sigma_n^{-2}\right] \end{aligned} \tag{9}$$

where $\prod$ is the product and $\sum$ the summation operator. The most probable value of s is found from the maximum of the total probability distribution $p(\varepsilon_1, \varepsilon_2, ...\varepsilon_N)$. The most probable value for s, i.e. the *maximum likelihood estimate* s_a, can be found by minimizing the sum:

$$J = \frac{1}{2}\sum_{n=1}^{N}(s_a - s_n)^2\sigma_n^{-2} = \frac{1}{2}(s_a - s_1)^2\sigma_1^{-2} + \frac{1}{2}(s_a - s_2)^2\sigma_2^{-2} + ... + \frac{1}{2}(s_a - s_N)^2\sigma_N^{-2} \tag{10}$$

where the quantity J is the so-called *cost function*.

The value of J obviously changes with varying s_a, thus to minimize J a suitable value for s_a needs to be found. By differentiating J with respect to s_a and by setting the differential to zero,

$$\frac{\partial J}{\partial s_a} = \sum_{n=1}^{N}(s_a - s_n)\sigma_n^{-2} = (s_a - s_1)\sigma_1^2 + (s_a - s_2)\sigma_2^2 + ... + (s_a - s_N)\sigma_N^2 = 0 \tag{11}$$

Now s_a can be solved as

$$s_a = \frac{\sum_{n=1}^{N}\sigma_n^{-2}s_n}{\sum_{n=1}^{N}\sigma_n^{-2}} \tag{12}$$

which is the *least squares method estimate* of s_a. The most probable value for s is therefore obtained by formulating a weighted average of the observations, where the weights are inversely proportional to the expected error variance of each observation.

Let $\varepsilon_a = s_a - s$ be the error of the estimate s_a. The variance of s_a is then

$$\langle\varepsilon_a^2\rangle = \langle\sigma_a^2\rangle = \langle(s_a - s)^2\rangle = \left\langle\left[\frac{\sum_{n=1}^{N}\sigma_n^{-2}(s_n - s)}{\sum_{n=1}^{N}\sigma_n^{-2}}\right]^2\right\rangle = \left[\sum_{n=1}^{N}\sigma_n^{-2}\right]^{-1} \tag{13}$$

over vast enough sample. It would appear that all additional information reduces the error variance of the estimate, i.e. makes the estimate more accurate. The accuracy improvement is the larger the smaller error variance the additional information has (Järvinen, 2003).

The scalar, or 0-dimensional, result can naturally be expanded to cover vectors, i.e. quantities having time or space dependencies. Consider $f(\mathbf{r})$ to be a function of space $\mathbf{r} = \mathbf{r}(x, y, z)$ with two observations at each point, i.e. $N = 2$ (observation and background). Assume there is prior knowledge of the estimate f at the observation points, $f_b(\mathbf{r}_k)$. Now for each K observation points let the errors of both the observations and background be random, unbiased, normally distributed and correlated in space but not correlated with each other. Let also the maximum likelihood estimate for $f(\mathbf{r})$ be $f_a(\mathbf{r})$. Now the expression to be minimized can be written in vector form as (for a more detailed approach and summation form see e.g. Daley (1991))

$$J = \frac{1}{2}[\mathbf{f}_o - \mathbf{f}_a]^T \mathbf{O}^{-1}[\mathbf{f}_o - \mathbf{f}_a] + \frac{1}{2}[\mathbf{f}_b - \mathbf{f}_a]^T \mathbf{B}^{-1}[\mathbf{f}_b - \mathbf{f}_a] \tag{14}$$

where column vectors $\mathbf{f}_a$, $\mathbf{f}_o$ and $\mathbf{f}_b$, each of length K, contain estimated, observed and background values at the observation points respectively. $\mathbf{O}$ is a $K \times K$ covariance matrix of the observation error containing variances of the observational error $\langle \varepsilon_o(\mathbf{r}_k)\varepsilon_o(\mathbf{r}_l)\rangle$, which reduces to $\langle \varepsilon_o^2(\mathbf{r}_k)\rangle$ when the observational errors are uncorrelated with each other. This observation error is caused by instrumental errors as well as representativeness error. $\mathbf{B}$ is the error covariance matrix of a priori information with $\langle \varepsilon_b(\mathbf{r}_k)\varepsilon_b(\mathbf{r}_l)\rangle$ as its factors.

Again in order to find the minimum of J, differentiate Equation 14 with respect to $f_a(\mathbf{r}_k)$ and set the result to zero

$$\frac{\partial J}{\partial f_a(\mathbf{r}_k)} = \mathbf{B}^{-1}[\mathbf{f}_a - \mathbf{f}_b] + \mathbf{O}^{-1}[\mathbf{f}_a - \mathbf{f}_o] = 0 \tag{15}$$

Keeping in mind the matrix calculation rules, $\mathbf{f}_a$ can now be solved

$$\begin{aligned} \mathbf{f}_a &= \left[\mathbf{B}^{-1} + \mathbf{O}^{-1}\right]^{-1} \left[\mathbf{B}^{-1}\mathbf{f}_b + \mathbf{O}^{-1}\mathbf{f}_o\right] \\ &= \mathbf{f}_b + \mathbf{B}[\mathbf{B} + \mathbf{O}]^{-1}[\mathbf{f}_o - \mathbf{f}_b] \end{aligned} \tag{16}$$

$\mathbf{f}_a$ is now a column vector with the maximum likelihood estimate of f as its factors. Note that the estimate $\mathbf{f}_a$ found this way is only valid for locations with both observation and background information.

4.1.3 Bayesian estimation

To extend the maximum likelihood method to be more convenient with numerical models in mind, i.e. to expand the estimate $\mathbf{f}_a$ to cover model grid points, Bayesian estimation needs to be presented. So-called *observation operator* is used for producing the equivalent of an observation from the model output.

Let column vector $\mathbf{x}$, of length N, be the model state vector to be estimated, and column vector $\mathbf{y}$, of length K, be the observation vector. No direct calculations between the two quantities can be carried out since the vectors are of different length and may even represent totally different physical variables. The former may be referred to as representative of model space and the latter as representative of observation space. Let the quantity representing the observations be unambiguously described from the model variables with observation operator H,

$$\mathbf{y} = H\mathbf{x}$$

The observation operator may be thought to be a mapping from model space to observation space, i.e. it makes the vectors $\mathbf{x}$ and $\mathbf{y}$ equally sized. In a simplified case, with both vectors $\mathbf{x}$ and $\mathbf{y}$ representing same physical quantity, the observation operator merely interpolates the model variable to the observation point. In this case the error in the observation operator is just the error of interpolation, but in a more general case the error coming from any other numerical operation on the variables is also added to the total error (Järvinen, 2003). The observation operator also enables direct use of satellite measured radiances. This, however, requires existence of a mapping tool for changing the model quantities x into radiances.

Denote the probability density function of a priori state of the model with $P(\mathbf{x})$ before any observations are taken into consideration. Let $P(\mathbf{y})$ then be the probability density function of the observations. The aim is to find an analysis representing the maximum of conditional probability $P(\mathbf{x}|\mathbf{y})$, which is the combined probability for events $\mathbf{x}$ and $\mathbf{y}$, i.e. the probability that $\mathbf{x}$ ja $\mathbf{y}$ occur simultaneously, is (Järvinen, 2003)

$$P(\mathbf{x} \cap \mathbf{y}) = P(\mathbf{x}|\mathbf{y})P(\mathbf{y}) = P(\mathbf{y}|\mathbf{x})P(\mathbf{x}) \tag{17}$$

This is the Bayes theorem.

Since the observations have been made at the time of the analysis $P(\mathbf{y})$ is a constant, thus

$$P(\mathbf{x}|\mathbf{y}) \propto P(\mathbf{y}|\mathbf{x})P(\mathbf{x}) \tag{18}$$

$P(\mathbf{x}|\mathbf{y})$ is the probability that analysis gets value $\mathbf{x}$ when the observations get value $\mathbf{y}$. Respectively $P(\mathbf{y}|\mathbf{x})$ is the probability that the observations get value $\mathbf{y}$ when the state of the atmosphere is $\mathbf{x}$.

Assume that the errors of observations and the prior field are normally distributed. Then the a priori probability distribution function $P(\mathbf{x})$ of the atmospheric state $\mathbf{x}$ is

$$P(\mathbf{x}) \propto \exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b)\right] \tag{19}$$

This is a multivariate normal distribution which states that at grid point i the best a priori estimate of the atmospheric state $\mathbf{x}_i$ is the prior field $\mathbf{x}_{b,i}$.

Likewise if the state of the atmosphere is $\mathbf{x}$ the conditional distribution function of the observations is

$$P(\mathbf{y}|\mathbf{x}) \propto \exp\left[-\frac{1}{2}(\mathbf{y}-H[\mathbf{x}])^T\mathbf{R}^{-1}(\mathbf{y}-H[\mathbf{x}])\right] \tag{20}$$

where $\mathbf{R}$ is observation error covariance matrix retaining the error caused by the limited nature of the observation operator. It also includes a *representativeness* element: observation and grid box represent different volumes and are not compatible as such. The observation can be almost point-like or represent some volume (e.g. radiosonde measurement), whereas the model volume is dependent on the model dimensions $\Delta x, \Delta y, \Delta z$.

The conditional distribution function $P(\mathbf{x}|\mathbf{y})$ is now

$$P(\mathbf{x}|\mathbf{y}) \propto \exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) - \frac{1}{2}(\mathbf{y}-H[\mathbf{x}])^T\mathbf{R}^{-1}(\mathbf{y}-H[\mathbf{x}])\right] \tag{21}$$

The maximum of $P(\mathbf{x}|\mathbf{y})$ is then found by minimizing the exponent

$$J = \frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + \frac{1}{2}(\mathbf{y}-H[\mathbf{x}])^T\mathbf{R}^{-1}(\mathbf{y}-H[\mathbf{x}]) \tag{22}$$

The minimum of the penalty function again represents the most probable state for the vector $\mathbf{x}$, and is thus the searched analysis $\mathbf{x} = \mathbf{x}_a$.

In order to solve for the minimum of J the observation operator H needs to be linear enough. Assuming the analysis is a small deviation from the background field,

and that the observation operator is reasonably linearisable, it holds that

$$\mathbf{y} - H(\mathbf{x}) \simeq \mathbf{y} - \mathbf{H}(\mathbf{x} - \mathbf{x}_b) - H(\mathbf{x}_b) \tag{23}$$

where $\mathbf{H}$ is *tangent-linear observation operator*.

The cost function can be now minimized with help of this tangent-linear hypothesis. By differentiating the cost function with respect to the analysis column vector $\mathbf{x}_a$ and setting the differential to zero

$$\frac{\partial J}{\partial \mathbf{x}_a} = \mathbf{B}^{-1}(\mathbf{x}_a - \mathbf{x}_b) - \mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y} - H[\mathbf{x}_a]) = 0 \tag{24}$$

By applying the tangent-linear hypothesis again this can be solved for $\mathbf{x}_a$,

$$\mathbf{x}_a = \mathbf{x}_b + (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y} - H[\mathbf{x}_b]) \tag{25}$$

which is usually written as

$$\begin{aligned} \mathbf{x}_a &= \mathbf{x}_b + \mathbf{K}(\mathbf{y} - H[\mathbf{x}_b]) \\ \mathbf{K} &= \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \end{aligned} \tag{26}$$

The matrix $\mathbf{K}$ is the *gain matrix* or the *weight matrix*. It stresses the deviation between the observations y and the background field at the observation points. The quantity $(\mathbf{y} - H[\mathbf{x}_b])$ is called the *innovation*, and it acts as a source of new information.

In order to solve $\mathbf{K}$ the covariance matrices of the observation error and the background error are needed. The gain matrix has two functions: firstly it gives the innovation an optimal weight, i.e. it gives an optimal weighting for the observations *with respect to the prior information*. Secondly $\mathbf{K}$ spreads the observation information to the model space.

4.2 Variational assimilation

4.2.1 3D-Var

The task of solving the weight matrix $\mathbf{K}$ directly is a burdensome task, though it can be done with some simplifying assumptions (Optimal Interpolation). To avoid this problem the variational assimilation method (three-dimensional variational assimilation, 3D-Var) minimizes the cost function J by "guessing" values for the argument

$\mathbf{x}$. The minimum cost function J found then represents the searched analysis $\mathbf{x}_a$. In practice this is always an approximation of the "real" $\mathbf{x}_a$. The minimizing is done iteratively by

1. evaluating J i.e. solving the equation

$$J = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + \frac{1}{2}(\mathbf{y} - H[\mathbf{x}])^T \mathbf{R}^{-1}(\mathbf{y} - H[\mathbf{x}]) \tag{27}$$

   for an initial guess $x = x_0$

2. calculating the gradient

$$\nabla J(\mathbf{x}) = \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) - \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{y} - H[\mathbf{x}]) \tag{28}$$

3. adjusting the iterand $\mathbf{x}$ to the direction $-\nabla J(\mathbf{x})$ by a step $-\alpha \nabla J(\mathbf{x})$, where α is some predefined constant.

So each coefficient of the iterand $\mathbf{x}$ is optimally perturbed in a way that the next evaluation gives the cost function J a smaller value. It usually takes many iteration rounds (i.e. repetition of steps 1 to 3) to find the minimum for J. The practicality behind the 3D-Var method is that the number of iteration rounds is still relatively small. Since it is usually sufficient to find the approximate minimum point, the iteration can be terminated after 70-100 iteration rounds even though the iteration has not converged i.e. the absolute minimum for J has not been found. A demonstration of 3D-Var is illustrated in Figure 17.

The first guess x_0 of the iteration is usually the background field $\mathbf{x}_b$, though it does not need to be so. The closer the first guess is to the value $\mathbf{x}$ to minimize J the faster the iteration converges. Also a first guess too far from the minimum of J may lead to wrong result. Thus the convention of using the background field is a practical one as the $\mathbf{x}_b$ is the best a priori information of the state of the atmosphere.

The 3D-Var requires that the error covariance matrices for the background $\mathbf{B}$ and observations $\mathbf{R}$ are formulated. Defining $\mathbf{B}$ is the most challenging task in the method. Usually the background field error covariances are presumed to be isotropic and homogeneous. In addition, a fitting variable change is generally required. In the best scenario the variable change makes the covariances between the variables to be analysed negligible.

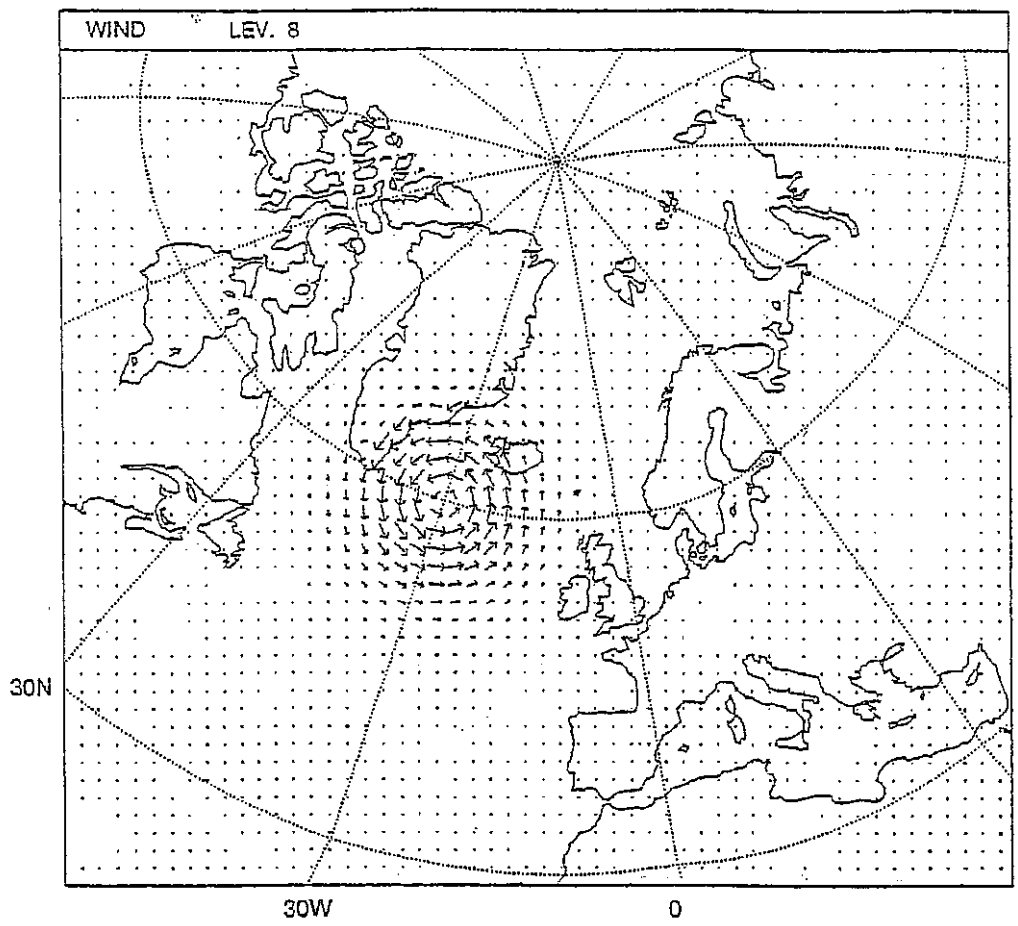


Figure 17: Wind field at HIRLAM model level 8. Example of the impact of one observation in the HIRLAM 3D-Var analysis system. A single temperature observation at the 500hPa level not only effects the temperature distribution but has also an enormous effect on the wind field. (After Järvinen (2003))

The observation error covariance matrix $\mathbf{R}$ is much easier to construct. It is usually a diagonal or block-diagonal matrix, depending whether the observations are correlated with each other or not. Since the observation errors of different instruments, observing stations or platforms are generally uncorrelated, the diagonal assumption is predominant. On the other hand, e.g. radiosonde measurements from one balloon are correlated which must be taken into account when constructing $\mathbf{R}$. Flexibility in defining the cost function ($J = J_b + J_o$, where J_b is the contribution of the background field and J_o that of the observations) allows that the correlated observation errors can be treated separately from the uncorrelated ones. When calculating the impact of observations on the cost function, the fraction J_o can be constructed from the different

observations as (Järvinen, 2003)

$$J_o(x) = \sum_{i=1}^{P} J_{o,i}(x) \tag{29}$$

where

$$J_{o,i} = (\mathbf{y}_i - H_i[\mathbf{x}])^T \mathbf{R}_i^{-1} (\mathbf{y}_i - H_i[\mathbf{x}]) \tag{30}$$

and $\mathbf{R}_i$ is constructed for each observation and given information whether the observation is correlated or not. The same fractioning can be applied for the gradient of the cost function, ∇J_o.

The additive nature of the cost function can be used as a powerful diagnostic tool of the 3D-Var analysis. It is also a window for observing the the impact of different observation system to the cost function.

All in all, the strengths of 3D-Var can be taken to be that it is (1) conceptually simple, (2) use of indirect observations is straightforward: In order to calculate the cost function only the observation operator H is needed and for calculating the gradient of the cost function the transpose of the tangent-linear observation operator $\mathbf{H}^T$. Furthermore (3) error covariances of the background field are only required between the model variables and (4) implementation of additional (weak) constraints is easy. This means that filtering of unwanted physical behaviour can be done by adding a term J_c to the cost function if the minimum is representing an unwanted physical phenomenon.

On the other hand developing the observation operator H can be laborious and requires heavy investments when implementing new observation types. Also programming and developing the tangent-linear observation operator $\mathbf{H}^T$ is not simple. Defining the error covariance matrix of the background field $\mathbf{B}$ is the biggest problem. Usually $\mathbf{B}$ is made as diagonal as possible through complex variable changes. In addition 3D-Var requires a lot computation resources.

4.2.2 4D-Var

The four-dimensional variational assimilation (4D-Var) is a generalization of the 3D-Var method. Whereas 3D-Var is a static analysis method, i.e. all the observations are taken to be from the same time, 4D-Var method takes the asynoptic nature (satellites, automatic weather stations) of the observations into account. The cost function and its gradient are formally the same as in 3D-Var.

Consider a time interval, *time window*, into which the observations have been spread to n steps. Analysis is defined at the starting point of the time window and the model state vector for the next time step is obtained directly from this initial condition by using a prediction model. Let i be some observation time and $\mathbf{y}_i$ observations from this time. The corresponding state of the model is $\mathbf{x}_i$ and the true state of the atmosphere $\mathbf{x}_{ti}$. The observation error for the time i is defined as $\varepsilon_{oi} = \mathbf{y}_i - H_i[\mathbf{x}_i]$ and the observation error covariance matrix as $\mathbf{R}_i$. The error covariance matrix of the background field $\mathbf{B}$ and the background field itself $\mathbf{x}_b$ are only defined for the starting point of the time window corresponding to the analysis $\mathbf{x}_a$.

In a general form the 4D-Var analysis problem is a minimization of cost function (Järvinen, 2003)

$$J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + \sum_{i=0}^{n} (\mathbf{y}_i - H_i[\mathbf{x}_i])^T + \mathbf{R}_i^{-1} (\mathbf{y}_i - H_i[\mathbf{x}_i]) \quad (31)$$

The solution of the minimization problem needs to be valid for successive model states $\mathbf{x}_i$ satisfying the model equation

$$\mathbf{x}_i = M_{0 \to i}(\mathbf{x}) \quad (32)$$

for all i. Here $M_{0 \to i}$ is the time integration of the prediction model M from time 0 to step i, i.e. from the beginning of the time window to the time of the observation. 4D-Var is thus a non-linear constrained minimization problem. This is generally an extremely demanding problem to solve, but two hypotheses, *causality of the model states* and *tangent-linearity of the observation operator*, transform the problem into a much simpler *unconstrained quadratic minimization problem*.

Causality of the model states says that the prediction model M can be described as a product of successive model prediction steps. The time integration is started from the initial condition of the model, $\mathbf{x}$, at the starting time. Let time moments i be successive and let $\mathbf{x}_0 = \mathbf{x}$ at the starting time. Denote the prediction step from time $i-1$ to time i as M_i, i.e. $\mathbf{x}_i = M_i \mathbf{x}_{i-1}$. M_i is the *time step operator* of the prediction model. Now

$$\mathbf{x}_1 = M_1 \mathbf{x}_0 = M_1 \mathbf{x}, \quad \mathbf{x}_2 = M_2 \mathbf{x}_1 = M_2 M_1 \mathbf{x}, ... \quad (33)$$

and recursively

$$\mathbf{x}_i = M_i M_{i-1} ... M_1 \mathbf{x} \quad (34)$$

For the cost function J to be unambiguous, it is needed that the observation operator H_i as well as the time step operator of the prediction model are linearisable, i.e.

$$\mathbf{y}_i - H_i M_{0\to i}(\mathbf{x}) \simeq \mathbf{y}_i - H_i M_{0\to i}(\mathbf{x}_b) - \mathbf{H}_i \mathbf{M}_{0\to i}(\mathbf{x} - \mathbf{x}_b) \tag{35}$$

where $\mathbf{M}$ is tangent linear prediction model, a differential of the model M. The tangent linear hypothesis is more realistic the shorter the time window is, but also the quality of the prediction model M affects the realism of the hypothesis.

The first term of the cost function J_b is exactly the same as for 3D-Var, but solving the second term J_o is harder since the time integration of the prediction model is needed. Instead of integrating the prediction model n times from the analysis time to the observation time i the evaluation of J_o (and of the gradient ∇J_o) can be done more efficiently.

Evaluating J_o is straightforward since the prediction model needs to be integrated forward in time from the initial state $\mathbf{x}$ to the state $\mathbf{x}_n$ at time n and compare the prediction with the observations made at the that time step. After the integrations J_o is the sum over the observation times $J_o(\mathbf{x}) = \sum_{i=0}^{n} J_{oi}(\mathbf{x})$.

Evaluation of the gradient ∇J_o is more complicated and requires a complex construction. This requires input from evaluation of J_o (deviations of the observations and prediction $\mathbf{d}_i = \mathbf{R}_i^{-1}(\mathbf{y}_i - H_i[\mathbf{x}_i])$ from each integration time i) and an *adjoint integration* from time n to the time 0. The adjoint integration is a backwards in time operation of the prediction model and is a product of the transposes of the time step operators $\mathbf{M}_i^T$, i.e. $\mathbf{M}_{n\to 0}^T$. For n time moments the formulation of the gradient is

$$\begin{aligned} -\nabla J_o &= -\sum_{i=0}^{n} \nabla J_{oi} \\ &= \sum_{i=0}^{n} \mathbf{M}_1^T ... \mathbf{M}_i^T \mathbf{H}_i^T \mathbf{d}_i \\ &= \mathbf{H}_0^T \mathbf{d}_0 + \mathbf{M}_1^T \mathbf{H}_1^T \mathbf{d}_1 + \mathbf{M}_1^T \mathbf{M}_2^T \mathbf{H}_2^T \mathbf{d}_2 + ... + \mathbf{M}_1^T \mathbf{M}_2^T ... \mathbf{M}_n^T \mathbf{H}_n^T \mathbf{d}_n \end{aligned} \tag{36}$$

This is applied via an algorithm, integrating the adjoint model backwards in time. Essentially the model is forced at each time step i with term $\mathbf{H}_i^T \mathbf{d}_i$ which depends on the normalized distance of the observations and model trajectories.

Assume the background field $\mathbf{x}_b$ and its error covariance matrix $\mathbf{B}$ are available at the beginning of the time window, and let the observations be spread throughout the

time window. The cost function is determined by the difference from the background field (J_b) and the distance of the model trajectories from the available observations (J_b) (illustrated in Figure 18). The aim is to define the analysis for the time responding to the background field in such a way that the analysis is in tune with the background field and on the other hand so that the forecast made from the analysis is well fit with the observations. The gradient of the cost function adjusts the analysis away from the background field, to make the analysis more consistent with the observations, but also towards the background field, to restrict the analysis from getting too far from the prior information. The analysis thus made is the best fit to both the background field and the observations in the least squares estimation sense.

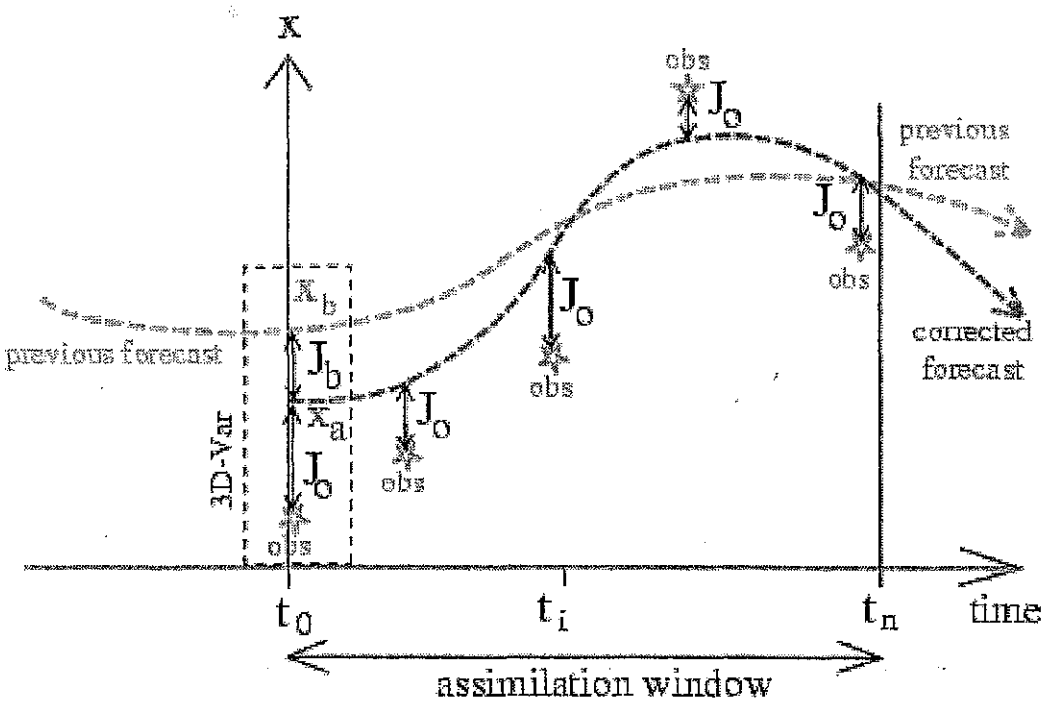


Figure 18: Impact of 4D-Var analysis on the background field (grey dashed line). (After Järvinen (2003))

4D-Var has all the strengths of 3D-Var method and in addition it (1) allows the use of observations in the time they were taken, allowing more powerful utilization of time-rich observations from e.g. satellite and buoy data. (2) The analysis satisfies the primitive equations precisely, which reduces the need for a separate initialization of the analysis. On the other hand the method is extremely expensive in computation time and takes considerable resources to develop. 4D-Var also binds the data-assimilation

and the deterministic forecast model closely together, as the prediction model used for time integration is usually the deterministic model.

4.3 The HIRLAM ASCAT data assimilation scheme

The ASCAT data assimilation in HIRLAM 7.3beta2-version follows closely the HIRLAM QuikScat implementation made by the Norwegian Meteorological Institute (met.no) (Tveter, Frank T.). The input is ASCAT final product of neutral winds after quality control and ambiguity removal. A second ambiguity is created within the NWP system and is always 180° off the primary wind solution (Tveter, Frank T.). The implementation is trying to estimate the "expected state of the atmosphere" rather than use the "most probable state of the atmosphere" approach introduced earlier and used in scatterometer data-assimilation by e.g. Portabella & Stoffelen (2004). The former is the second of the two data-assimilation principles suggested by Lorenc (Tveter, 2006), and is based on minimizing *Bayes risk*. For a quadratic loss function $l[\mathbf{x}, \mathbf{x}_a] = (\mathbf{x}-\mathbf{x}_a)^T(\mathbf{x}-\mathbf{x}_a)$, where $\mathbf{x}$ is representing the true state of the atmosphere, the Bayes risk would be

$$\begin{aligned} B(\mathbf{x}_a) &= E[l[\mathbf{x}, \mathbf{x}_a]] \\ &= E[E[(\mathbf{x}-\mathbf{x}_a)^T(\mathbf{x}-\mathbf{x}_a)|\mathbf{y} \cap \mathbf{x_b}]] \\ &= E\left[\mathsf{Var}[\mathbf{x}|\mathbf{y} \cap \mathbf{x_b}]] + E[(E[\mathbf{x}|\mathbf{y} \cap \mathbf{x_b}] - \mathbf{x}_a)^T(E[\mathbf{x}|\mathbf{y} \cap \mathbf{x}_b] - \mathbf{x}_a)\right] \end{aligned} \tag{37}$$

Function that minimizes this Bayes risk is given by

$$\mathbf{x}_a^{**} = E[\mathbf{x}|\mathbf{y} \cap \mathbf{x}_b], \tag{38}$$

which is the "expected state of the atmosphere" (Tveter, 2006). Bayesian risk theory corresponds to how most NWP systems are verified (via (root-)mean-square-error methods). Since the optimal analysis depends on the verification method this approach is well-grounded (Tveter, Frank T.).

As the scatterometer wind observations are non-Gaussian an extension to the Bayesian estimation is needed. First assume a set of observations with Gaussian error statistics $\mathbf{y}$, and another set of independent non-Gaussian observations $\tilde{\mathbf{y}}$. Now recall Equation

(26) and define "Gaussian analysis", $\breve{\mathbf{x}}_b$, as the analysis with only the first guess $\mathbf{x}_b$ and the Gaussian observations $\mathbf{y}$ assimilated into it

$$\begin{aligned} \breve{\mathbf{x}}_b &= \mathbf{x}_b + \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}_b) \\ \breve{\mathbf{B}}^{-1} &= \mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} \end{aligned} \tag{39}$$

The probability density function for the state of the atmosphere may be written as (for detailed approach see Tveter (2006))

$$\begin{aligned} P[\mathbf{x}|\tilde{\mathbf{y}} \cap \mathbf{y} \cap \mathbf{x}_b] &= \frac{P[\tilde{\mathbf{y}}|\mathbf{x} \cap \mathbf{y} \cap \mathbf{x}_b]P[\mathbf{x}|\mathbf{y} \cap \mathbf{x}_b]}{P[\tilde{\mathbf{y}}|\mathbf{y} \cap \mathbf{x}_b]} \\ &= \frac{P[\tilde{\mathbf{y}}|\mathbf{x}]P[\mathbf{x}|\breve{\mathbf{x}}_b]}{P[\tilde{\mathbf{y}}|\breve{\mathbf{x}}_b]} \\ &= P[\mathbf{x}|\tilde{\mathbf{y}} \cap \breve{\mathbf{x}}_b] \end{aligned} \tag{40}$$

Thus the Gaussian analysis $\breve{\mathbf{x}}_b$ works as the background field for the non-Gaussian observations $\tilde{\mathbf{y}}$.

A scatterometer observation consists of n ambiguities, $\tilde{\mathbf{y}} = \mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n$. For a single ambiguous observation the probability density function can be modelled as

$$\begin{aligned} P[\tilde{\mathbf{y}}|\mathbf{x}] &\propto \hat{\gamma} + \sum_{k=1}^{n} \hat{q}_k P[\mathbf{y}_k|\mathbf{x}] \\ P[\mathbf{y}_k|\mathbf{x}] &\propto e^{-\tilde{J}_k[\mathbf{x}]} \\ \tilde{J}_k[\mathbf{x}] &= \frac{1}{2}(\mathbf{y}_k - \mathbf{H}\mathbf{x})^T\tilde{\mathbf{R}}^{-1}(\mathbf{y}_k - \mathbf{H}\mathbf{x}) \end{aligned} \tag{41}$$

where $\hat{\gamma}$ is the a priori probability for "gross error" (within a given range), $\hat{q}_k$ the a priori probability that ambiguity k is "correct", $\mathbf{y}_k$ the value of the kth wind vector ambiguity and $\tilde{\mathbf{R}}$ represents the observation covariance matrix for the "correct" ambiguity (Tveter, 2006). The a priori probabilities for the ambiguities are based on how well the simulated backscatter (derived for each ambiguity), agrees with the actual backscatter observation, and is available from the retrieval algorithm.

Now assuming the observation error is independent of the first guess error, and using the Gaussian analysis and non-Gaussian scatterometer observation, the probability

for the true atmospheric state may be written as

$$\begin{aligned} P[\mathbf{x}|\tilde{\mathbf{y}} \cap \breve{\mathbf{x}}_b] &= \frac{P[\tilde{\mathbf{y}}|\mathbf{x}]P[\mathbf{x}|\breve{\mathbf{x}}_b]}{P[\tilde{\mathbf{y}}|\breve{\mathbf{x}}_b]} \\ P[\tilde{\mathbf{y}}|\breve{\mathbf{x}}_b] &= \int_{-\infty}^{\infty} P[\tilde{\mathbf{y}}|\mathbf{x}]P[\mathbf{x}|\breve{\mathbf{x}}_b]d\mathbf{x} \\ &= \hat{\gamma} + \sum_{k=1}^{n} \hat{q}_k P[\mathbf{y}_k|\breve{\mathbf{x}}_b] \\ P[\mathbf{y}_k|\breve{\mathbf{x}}_b] &\propto e^{-\tilde{J}_k} \\ \tilde{J}_k &= \frac{1}{2}(\mathbf{y}_k - \mathbf{H}\breve{\mathbf{x}}_b)^T(\mathbf{H}\breve{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}})^{-1}(\mathbf{y}_k - \mathbf{H}\breve{\mathbf{x}}_b) \end{aligned} \tag{42}$$

And further

$$\begin{aligned} P[\mathbf{x}|\tilde{\mathbf{y}} \cap \breve{\mathbf{x}}_b] &= \gamma[\breve{\mathbf{x}}_b]P[\mathbf{x}|\breve{\mathbf{x}}_b] + \sum_{k=1}^{n} q_k[\breve{\mathbf{x}}_b]\frac{P[\mathbf{y}_k|\mathbf{x}]P[\mathbf{x}|\breve{\mathbf{x}}_b]}{P[\mathbf{y}_k|\breve{\mathbf{x}}_b]} \\ \gamma[\breve{\mathbf{x}}_b] &= \frac{\hat{\gamma}}{\hat{\gamma} + \sum_{i=1}^{n} \hat{q}_i P[\mathbf{y}_i|\breve{\mathbf{x}}_b]} \\ q_k[\breve{\mathbf{x}}_b] &= \hat{q}_k \frac{P[\mathbf{y}_k|\breve{\mathbf{x}}_b]}{\hat{\gamma} + \sum_{i=1}^{n} \hat{q}_i P[\mathbf{y}_i|\breve{\mathbf{x}}_b]} \end{aligned} \tag{43}$$

Now applying this to Eq 38 gives

$$\begin{aligned} \mathbf{x}_a^{**} &= E[\mathbf{x}|\mathbf{y} \cap \breve{\mathbf{x}}_b] \\ &= \int_{-\infty}^{\infty} \mathbf{x}P[\mathbf{x}|\tilde{\mathbf{y}} \cap \breve{\mathbf{x}}_b]d\mathbf{x} \\ &= \gamma[\breve{\mathbf{x}}_b]\breve{\mathbf{x}}_b + \sum_{k=1}^{n} q_k[\breve{\mathbf{x}}_b]\mathbf{x}_k \\ \mathbf{x}_k &= \breve{\mathbf{x}}_b + \breve{\mathbf{B}}\mathbf{H}^T(\mathbf{H}\breve{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}})^{-1}(\mathbf{y}_k - \mathbf{H}\breve{\mathbf{x}}_b) \end{aligned} \tag{44}$$

This is the "expected state of the atmosphere" analysis when assimilating one ambiguous scatterometer observation.

Assume that the a posteriori probabilities for each ambiguity to be correct, $q_k[\breve{\mathbf{x}}_b]$, is known. A spatial average of raw measurements, so-called *superobservation*, may be defined according to

$$\tilde{\mathbf{y}}_s[\breve{\mathbf{x}}_b] = \gamma[\breve{\mathbf{x}}_b]\mathbf{H}\breve{\mathbf{x}}_b + \sum_{k=1}^{n} q_k[\breve{\mathbf{x}}_b]\mathbf{y}_k \tag{45}$$

If this superobservation is assimilated as a conventional wind observation with Gaussian error characteristics ($\tilde{\mathbf{R}}$) and in accordance to Eq. 26 the result is identical that of Eq. 44, i.e. the "expected state of the atmosphere" analysis when assimilating one ambiguous scatterometer observation (Tveter, 2006).

Since scatterometer observations have correlated errors a correct and complex analysis would need to ensure that observations with correlated errors received lesser weight than similar observations with uncorrelated errors. This error correlation effect is approximated by reducing the weight of the superobservations by increasing $\tilde{\mathbf{R}}$, and otherwise treating the observations as having independent errors. In HIRLAM 3D-Var the cost function contribution from each scatterometer observation is

$$J_{o,k}[\mathbf{x}] = \frac{1}{2}(\tilde{\mathbf{y}}_{s,k}[\breve{\mathbf{x}}_b] - \mathbf{H}_k\mathbf{x})^T\tilde{\mathbf{R}}^{-1}(\tilde{\mathbf{y}}_{s,k}[\breve{\mathbf{x}}_b] - \mathbf{H}_k\mathbf{x}) \tag{46}$$

where the index k refers to superobservation number k (Tveter, 2006).

There are fixed number of iterations in the minimization algorithm, usually 60-100 iterations are reasonable to assume. The scatterometer observations are ignored for the first 10 iteration rounds after which it is assumed that the Gaussian analysis is reached. The contribution of the superobservations to the cost function are then calculated according to Eq. 46. The ambiguities of the superobservations are allowed to fluctuate freely before iteration round 30. After this one of the ambiguities is chosen as the correct one. Due to the larger grid increment of the assimilation calculations the scatterometer data is thinned to 100 km before any iterations. The observation window for both the 3D-Var and 4D-Var analyses is forecast base time -4 h to +2 h. This allows more data inflow from sources with longer transmission times (as well as more observations from asynoptic sources).

5 Impact study

5.1 Background

The general circulation in the atmosphere is driven by the uneven distribution of solar radiation on the surface. This leads to an uneven temperature distribution and furthermore density and pressure differences which in turn generate circulation into the atmosphere (Holopainen *et al.* , 2008). Strong disturbances, transferring heat and momentum effectively from one latitude to another, form especially to the mid-latitude troposphere. The ocean-land distribution and existence of mountains further affect the circulation in mainly east-west direction by inducing stationary waves (for more detailed approach to general circulation see e.g. Holton (2004)). Figure 19 illustrates the mean 500-hPa pressure level in January in the Northern Hemisphere. It is easy to

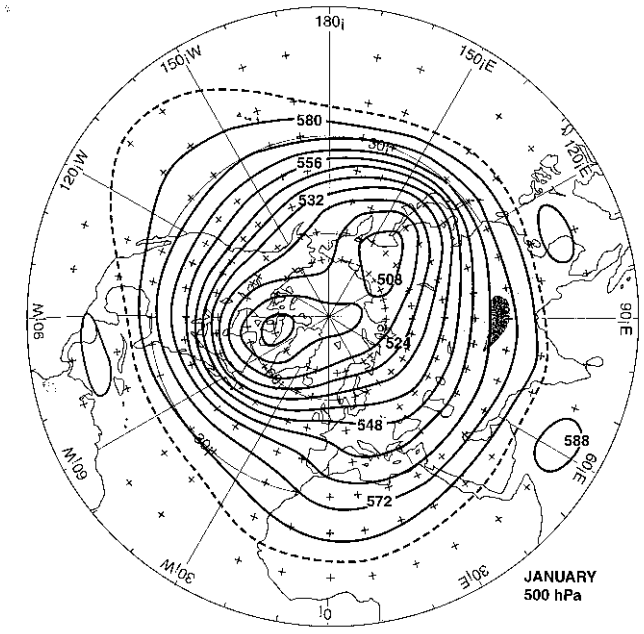


Figure 19: Northern Hemispheric mean 500-hPa contours in January. Heights shown in tens of meters. (After Holton (2004))

notice that Europe is located downstream of the westerly geostrophic flow pattern induced by the strong pressure anomaly on top of Canada and Greenland. This flow also crosses the Atlantic Ocean being thus subject to massive evaporation from the ocean. The flow pattern is known as the North Atlantic storm track and it makes especially the North Europe and Ireland-England-region subject to constant storm activity.

The storm track dominates the weather in the long term. To discern what kind of impact assimilating ASCAT winds into HIRLAM has, this study focuses on several subjectively picked "interesting" weather situations lasting from 2 to 5 days in length. In addition to cases with the prevailing westerly flow (Atlantic) also few less frequent weather patterns, where the prevailing flow direction is from the polar areas, are taken into closer examination. The categorization of the cases is shown in Table 1.

Table 1: The weather situations used for this study. The table defines the prevailing flow direction, interesting effects associated to the flow pattern and how long the situation can be seen lasting. Also the run time for HIRLAM model and type of data assimilation method are presented.

Type	Special	Days of interest	HIRLAM run for	Assimilated
Atlantic		08.10.2009 to 13.10.2009	04.10.2009 to 13.10.2009	4D-Var
Atlantic	Storm in Finland	21.01.2009 to 25.01.2009	17.01.2009 to 26.01.2009	4D-Var
Atlantic	Comma	27.11.2009 to 02.12.2009	23.11.2009 to 02.12.2009	4D-Var
Northern	Storm in Finland	27.09.2009 to 29.09.2009	22.09.2009 to 30.09.2009	4D-Var
Northern	Polar low Storm in Finland	28.12.2008 to 01.01.2009	24.12.2008 to 01.01.2009	4D-Var

The attribute "Storm in Finland" indicates that there was a warning of stormy sea conditions in the Baltic Sea area given by Finnish Meteorological Institute (FMI). The warnings were found from a storm archive maintained by the operational forecasters of FMI. "Comma" means that a substantial comma cloud formed and could have effected the weather on land areas. The "polar low" case indicates a polar low formed and played important role in the weather. For a description of the physical background and detailed impact on weather of the two phenomena see e.g. Bader *et al.* (1995) or SatManu webpage.

In this study I conducted the HIRLAM runs on ECMWF supercomputer cluster "c1a" via ECaccess gateway. The simulations were made with tag "HIRLAM7.3beta2" of the HIRLAM trunk. The HIRLAM switch for using host model upper air analysis for redoing background (LSMIX) was turned off. Since LSMIX combines analyses from host and nested model (see Yang (2009) for principles behind the technique), omitting the host model analysis prevents "contamination" of the background field (due to ASCAT data already being assimilated into ECMWF model). Rest of the

switches were left as default. All the cases were run first without ASCAT data and afterwards with ASCAT data, taking boundary and climate files from the runs conducted without ASCAT. HIRLAM runs were initiated 4 days prior the days of interest to let the model dynamics "warm up". A 48 h forecast was made at 00UTC and 12UTC and a 6 h forecast for background field at 06UTC and 18UTC for each day HIRLAM was run.

For practicality, the HIRLAM runs conducted *without* assimilating ASCAT data shall be referred to as "control (HIRLAM)" and the runs conducted *with* ASCAT assimilation included to as "ASCAT (HIRLAM)". Not to confuse the reader, these two are also named "NO_ASCAT_4D" and "ASCAT_4D", respectively, in some of the graphs.

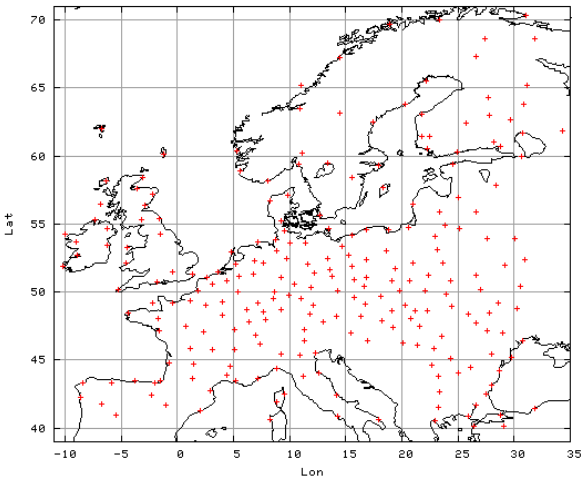
Verification was done by comparing observations to corresponding forecast values for the same location, and averaging them over all stations and time moments. The mean systematic error or bias $E(\hat{X} - X)$ is based on estimating the sample statistics by

$$b = \frac{1}{k}\sum_{t=1}^{k}\frac{1}{n}\sum_{i=1}^{n}(\hat{x}_{k,i} - x_{k,i}) \tag{47}$$

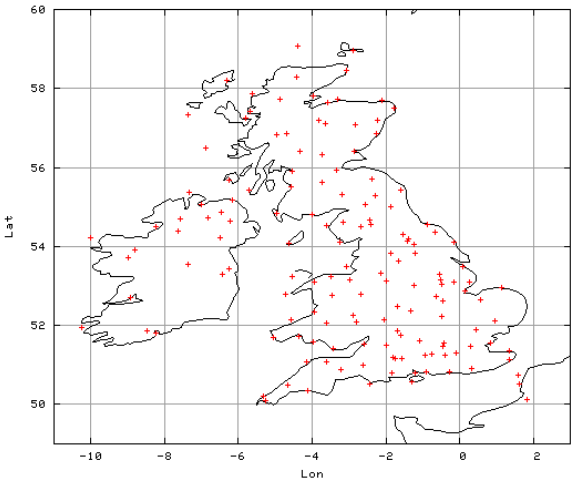
where $\hat{x}_{k,i}$ is the forecast value and $x_{k,i}$ is the observed value at time k in location i (after Jolliffe & Stephenson (2003)). The other verification quantity used, root mean squared error (rmse), is the square root of the mean squared error $E[(\hat{X} - X)^2]$. The time and spatial average can be presented by

$$\mathsf{rmse} = \sqrt{\frac{1}{k}\sum_{t=1}^{k}\frac{1}{n}\sum_{i=1}^{n}(\hat{x}_{k,i} - x_{k,i})^2} \tag{48}$$

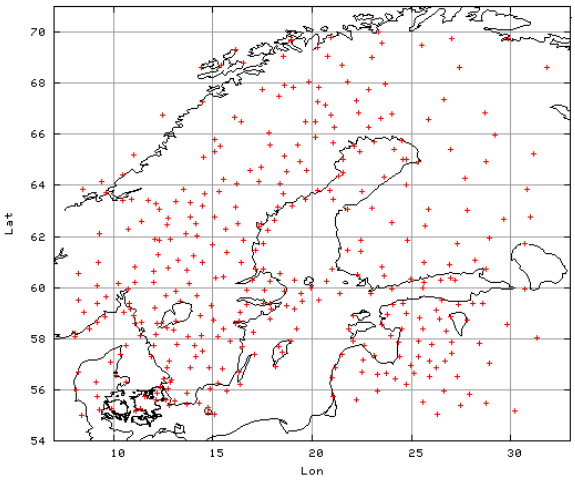
Figure 20 presents three different areas from which observation station data was used for making verification. EWGLAM (Fig 20a) is representative of central European continent, Ireland and England (Fig 20b) represent the observations from the two islands and "Scandinavia" (Fig 20c) from the Nordic and Baltic countries.



(a) EWGLAM, 243 stations



(b) Ireland and England, 146 stations



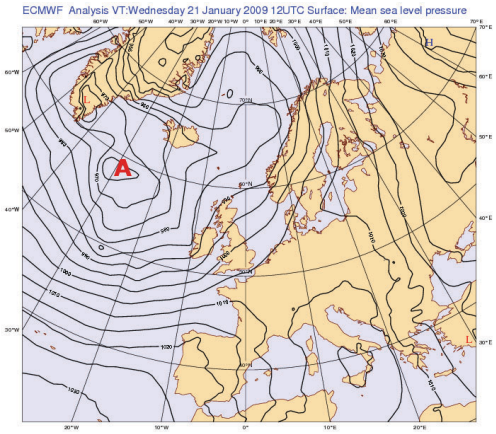
(c) Scandinavia, 270 stations

Figure 20: Areas from which observations are used for verification purposes.

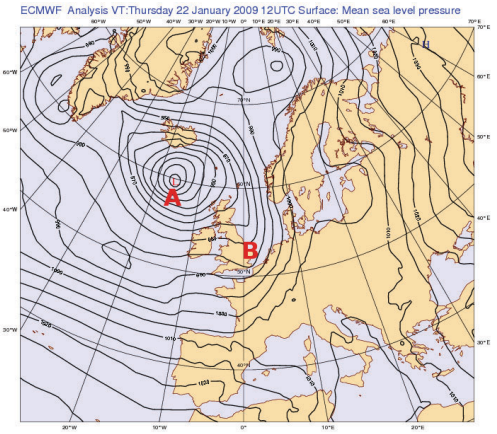
5.2 Atlantic flow

5.2.1 Case 21.01.-25.01.2009 - Storm

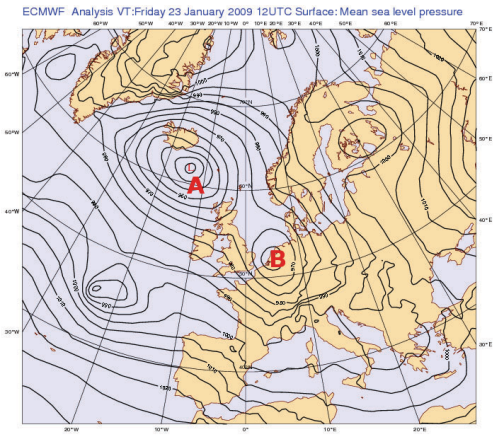
The synoptic regime of time-period 21.1.-25.1.2009 is presented in Figures 21a-21d. The flow regime is dominated by a strong low pressure centre located south of Iceland (marked with red A in Figures 21a-21d). The centre forms and starts to deepen around 00UTC 22nd January and slowly drifts, lessening in strength, to the south-west coast of Iceland by 12UTC 24th. A second weaker low pressure centre starts to develop on top of Britain after 12UTC 22nd (B in fig 21b) and moves with a north-easterly track after reaching the Netherlands area. A storm warning was given by the FMI for 24th with predicted wind speeds of over 17 m/s in the sea areas (Storm archive).



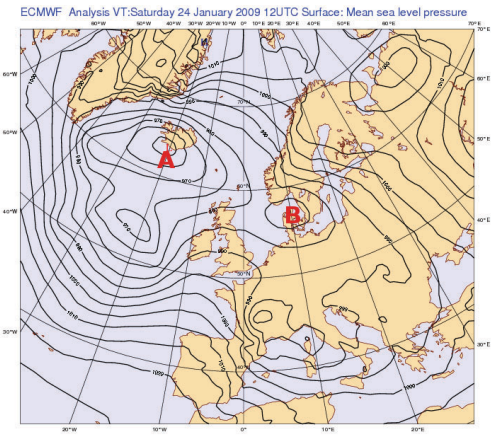
(a) 21.01. 12UTC



(b) 22.01. 12UTC



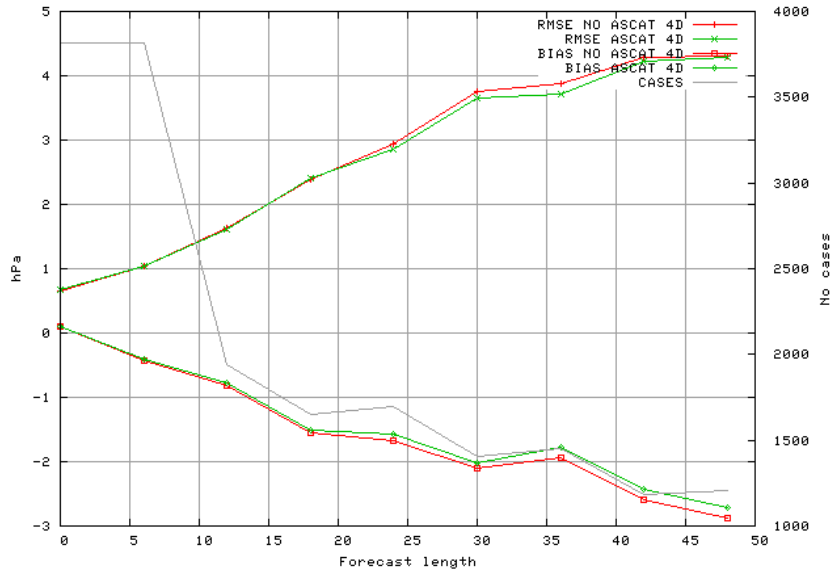
(c) 23.01. 12UTC



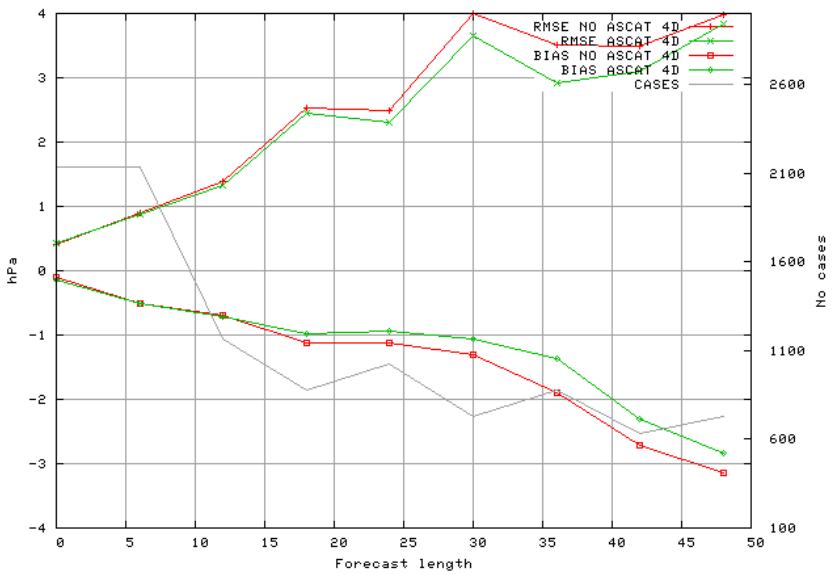
(d) 24.01. 12UTC

Figure 21: Synoptic evolution of mean sea level pressure between 12UTC 21.01. and 12UTC 24.01 from ECMWF analysis. Contours are drawn with 5hPa increments.

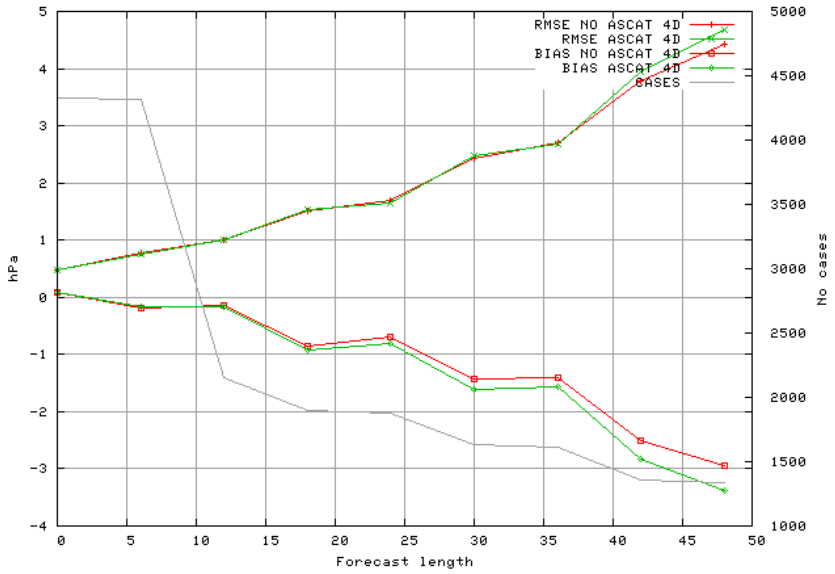
Figures 22a to 22c contain the verification results from EWGLAM, Ireland and England and Scandinavia area respectively. The verification from EWGLAM area (Fig 22a) shows small improvement in both rmse and bias from runs conducted with ASCAT data in forecast lengths longer than 24 h. The improvement is enhanced when looking at the Ireland-England area (Fig 22b) and can be seen from forecast length 18 h onwards. The verification from the Scandinavian area (Fig 22c) shows a decline of the bias with ASCAT data from forecast length 18 h onwards and a worse rmse score for forecast lengths 42 h and 48 h.



(a) EWGLAM



(b) Ireland and England

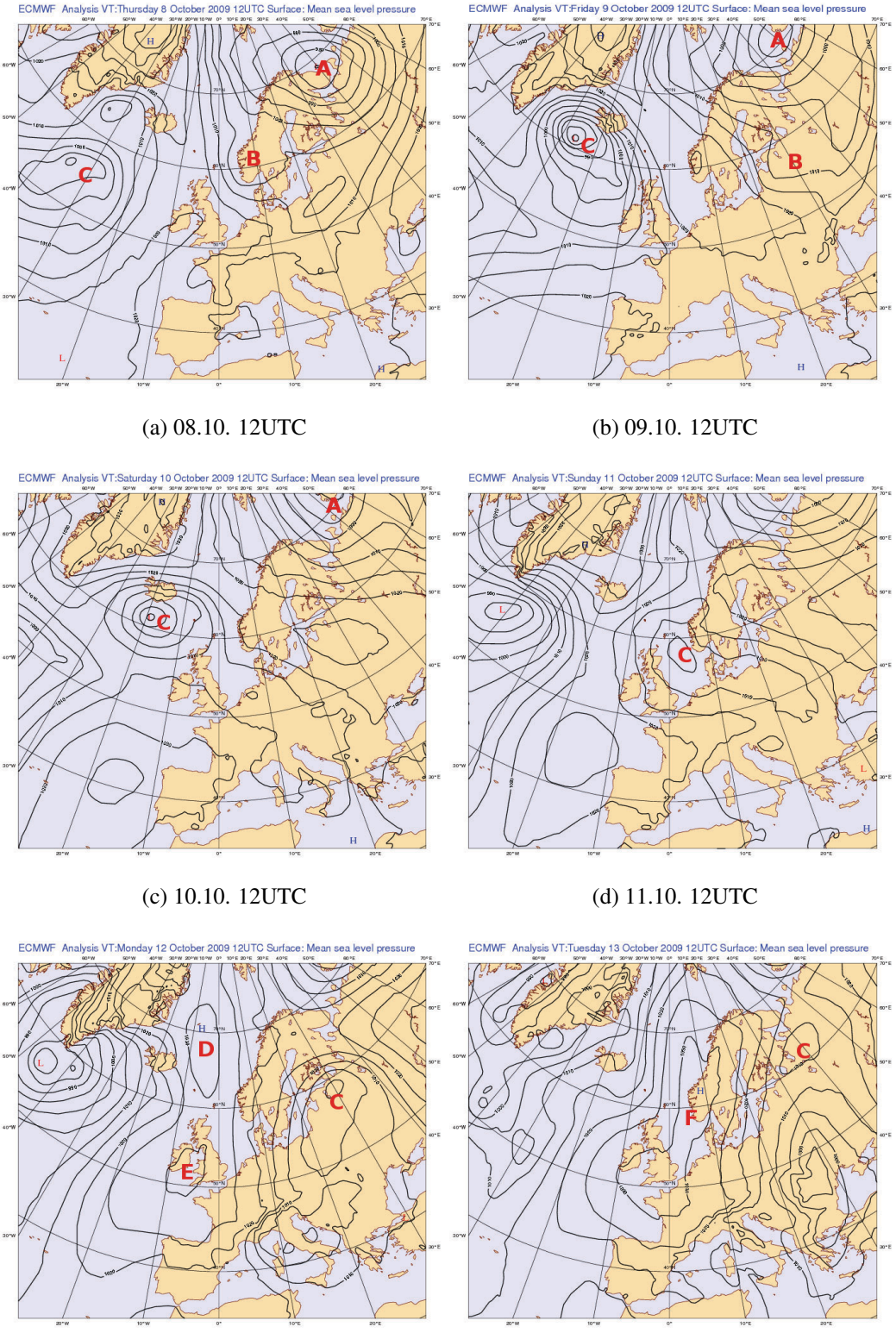


(c) Scandinavia

Figure 22: Verification results for mean sea level pressure from time period of 22.1.-25.1.2009. x-axis shows forecast length in hours and y-axis is deviation from observed pressure value in hPa. Red line with squares represent bias of control HIRLAM runs and green line with circles represents bias of ASCAT HIRLAM runs. Red line with +-signs represent rmse of the control run and green line with x-signs that of the ASCAT run.

5.2.2 Case 08.10.-13.10.2009

Strong low pressure centre on top of northern Finland and Kola peninsula (A in Figure 23) dominates the flow in the Scandinavia region in the beginning of the time period. The pressure centre moves east and lessens in influence by 00UTC 11th. Another weaker low pressure centre forms around 06UTC 8th (B in Fig 23a), and travels through the south parts of Norway, Sweden and reaches Finnish south coast by 00UTC 9th and continues on its eastward track. At 00UTC 9th a strong low pressure centre (C in all figures) starts to form south-west of Iceland. It stays near the Icelandic coast until 00UTC 10th when it starts a slow drift towards Scotland. The centre fills on the way and passes by north of Scotland on 06UTC 11th and reaches Denmark by 18UTC 11th. After reaching the continent the low transforms into a large weak pressure anomaly, and advects north-eastward. The flow pattern at the end of the time period is dominated by a large high pressure area (F in Fig 23f) merging from high pressure centres on top of Britain and western Norway (D and E in Fig 23e).



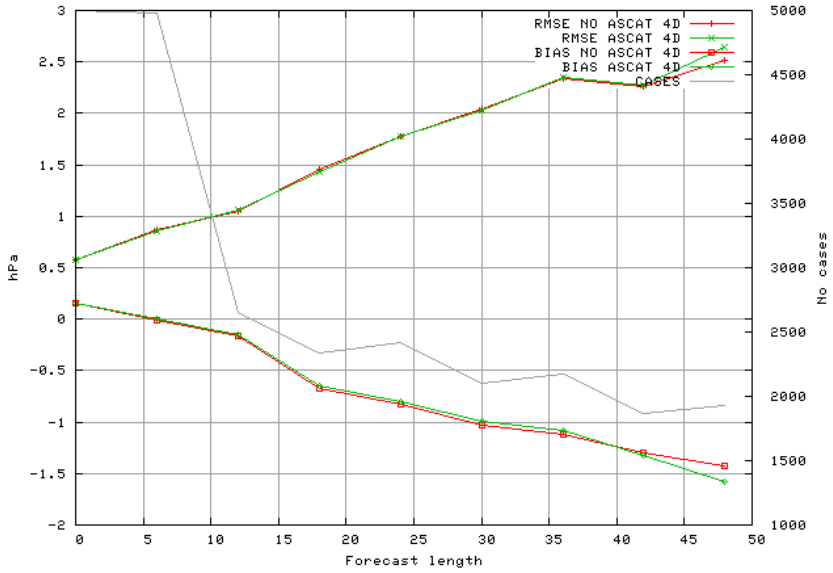
(a) 08.10. 12UTC (b) 09.10. 12UTC

(c) 10.10. 12UTC (d) 11.10. 12UTC

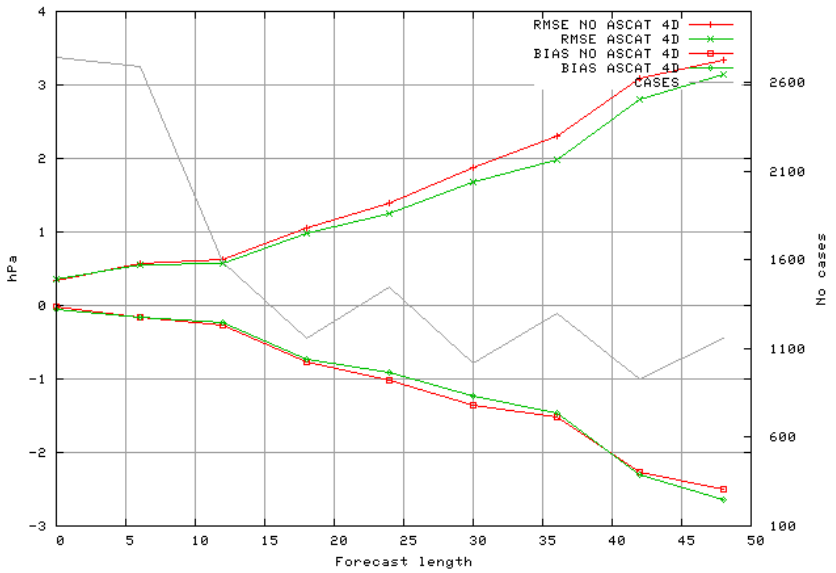
(e) 12.10. 12UTC (f) 13.10. 12UTC

Figure 23: Synoptic evolution of mean sea level pressure between 12UTC 08.10. and 12UTC13.10 from ECMWF analysis. Contours are drawn with 5hPa increments.

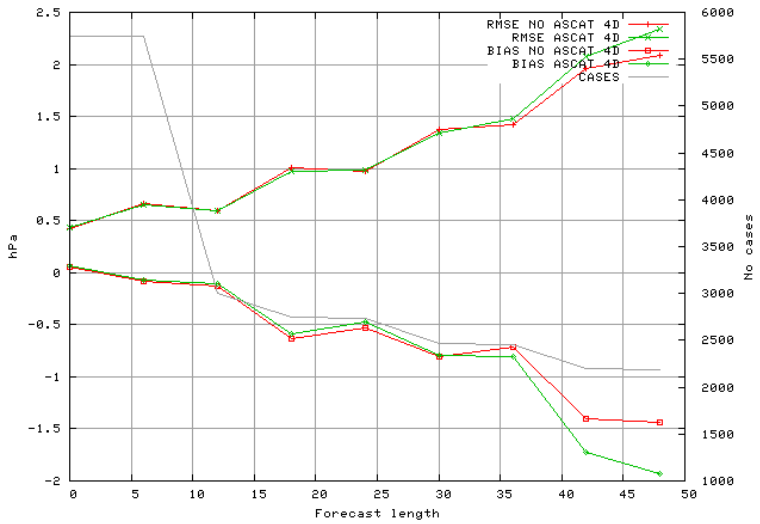
The impact of ASCAT assimilation is neutral in EGLAM area (Fig 24a) with a small decline in both rmse and bias in forecast length 48 h. The Ireland-England area verification (Fig 24b) shows a small positive impact on bias for forecast lengths 24 h and 30 h but a small worsening with 48 h forecasts. The rmse is improved from forecast length 12 h onwards. Scandinavian area verification (Fig 24c) is neutral until forecast length 36 h after which the ASCAT forecasts have a worse bias and rmse than the control runs.



(a) EWGLAM



(b) Ireland and England

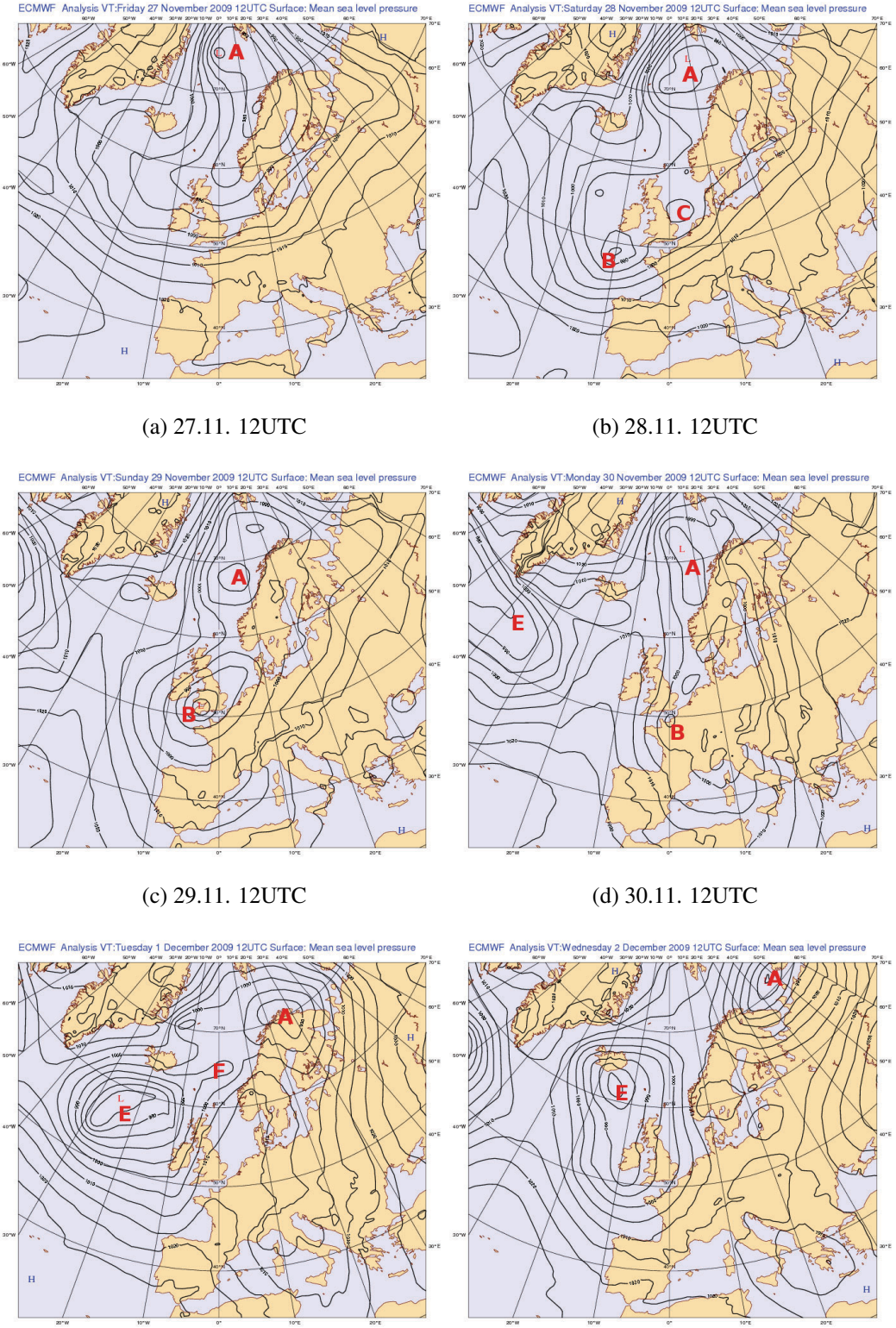


(c) Scandinavia

Figure 24: Verification results for mean sea level pressure from time period of 8.10.-13.10.2009. x-axis shows forecast length in hours and y-axis is deviation from observed pressure value in hPa. Red line with squares represent bias of control HIRLAM runs and green line with circles represents bias of ASCAT HIRLAM runs. Red line with +-signs represent rmse of the control run and green line with x-signs that of the ASCAT run.

5.2.3 Case 27.11.-02.12.2009 - Comma cloud

The period starts with a vast low pressure area dominating the whole European area flow. The main low pressure centre is located west and north-west of the Norwegian coast (A in the synoptic maps). The centre stays close to the Norwegian coast varying in strength until 06UTC 1st when it joins a north-west bound low pressure system (D in Fig 25d) and advects west. Additional low pressure centres form around the British isles starting from 06UTC 28th (B and C in Fig 25b). One of the additional low pressures (B), south-west of Britain, starts to deepen around 12UTC 28th and reaches British island by 06UTC 29th. When reaching land surface the centre starts to fill and dies out by 12UTC 30th. A new strong low pressure centre (E in Fig 25d) starts to deepen far west of Ireland at 18UTC 30th. The system moves north-west and reaches Icelandic south coast by 06UTC 2nd. This pressure centre dominates the flow of western Europe and gradually that of northern Europe by 00UTC 3rd as the original low pressure centre (A) advects further to east. The new pressure system gives birth to a large comma south of Iceland at 18UTC 30th (F in Fig 25e). This comma advects east and dies out near Norwegian west coast around 18UTC 1st.



(a) 27.11. 12UTC

(b) 28.11. 12UTC

(c) 29.11. 12UTC

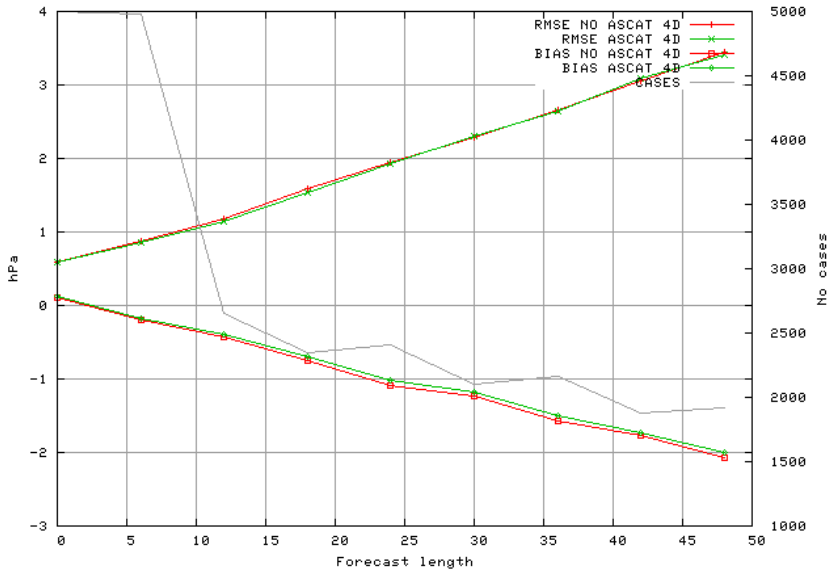
(d) 30.11. 12UTC

(e) 01.12. 12UTC

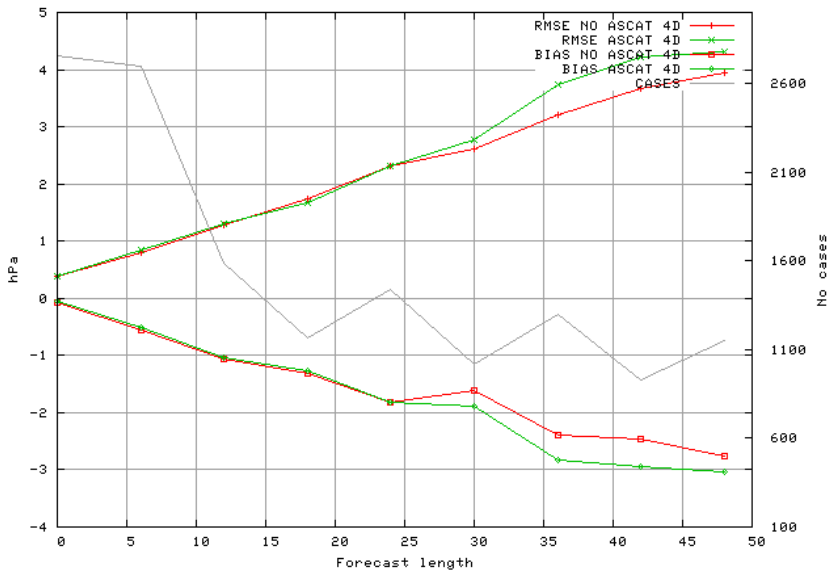
(f) 02.12. 12UTC

Figure 25: Synoptic evolution of mean sea level pressure between 12UTC 27.11. and 12UTC 02.12 from ECMWF analysis. Contours are drawn with 5hPa increments.

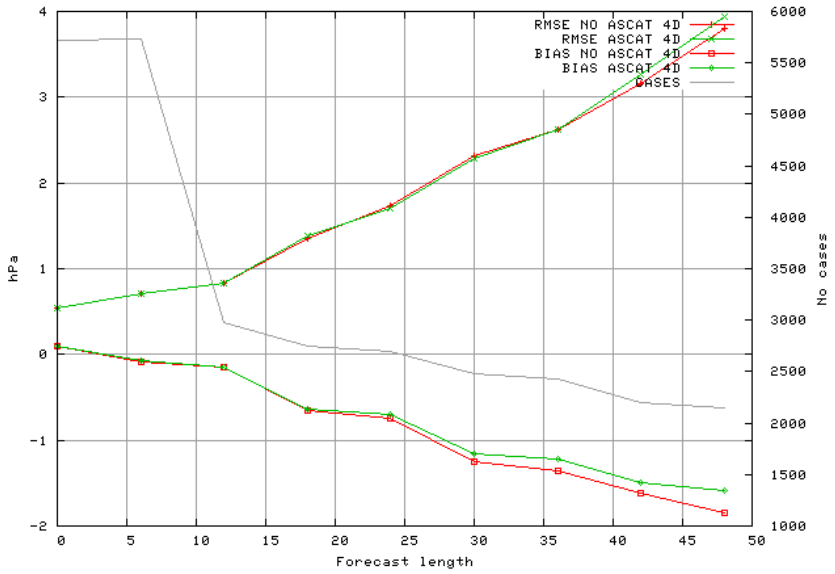
The verification result from EWGLAM area (Fig 26a) shows a neutral-positive impact on bias and a neutral impact on rmse. A negative impact on both bias and rmse can be seen in Ireland-England verification (Fig 26b) from forecast lengths above 30 h. The Scandinavian verification (Fig 26c) indicates a positive bias for forecast lengths longer than 24 h, but a negative rmse score for forecast lengths 42 h and 48 h.



(a) EWGLAM



(b) Ireland and England



(c) Scandinavia

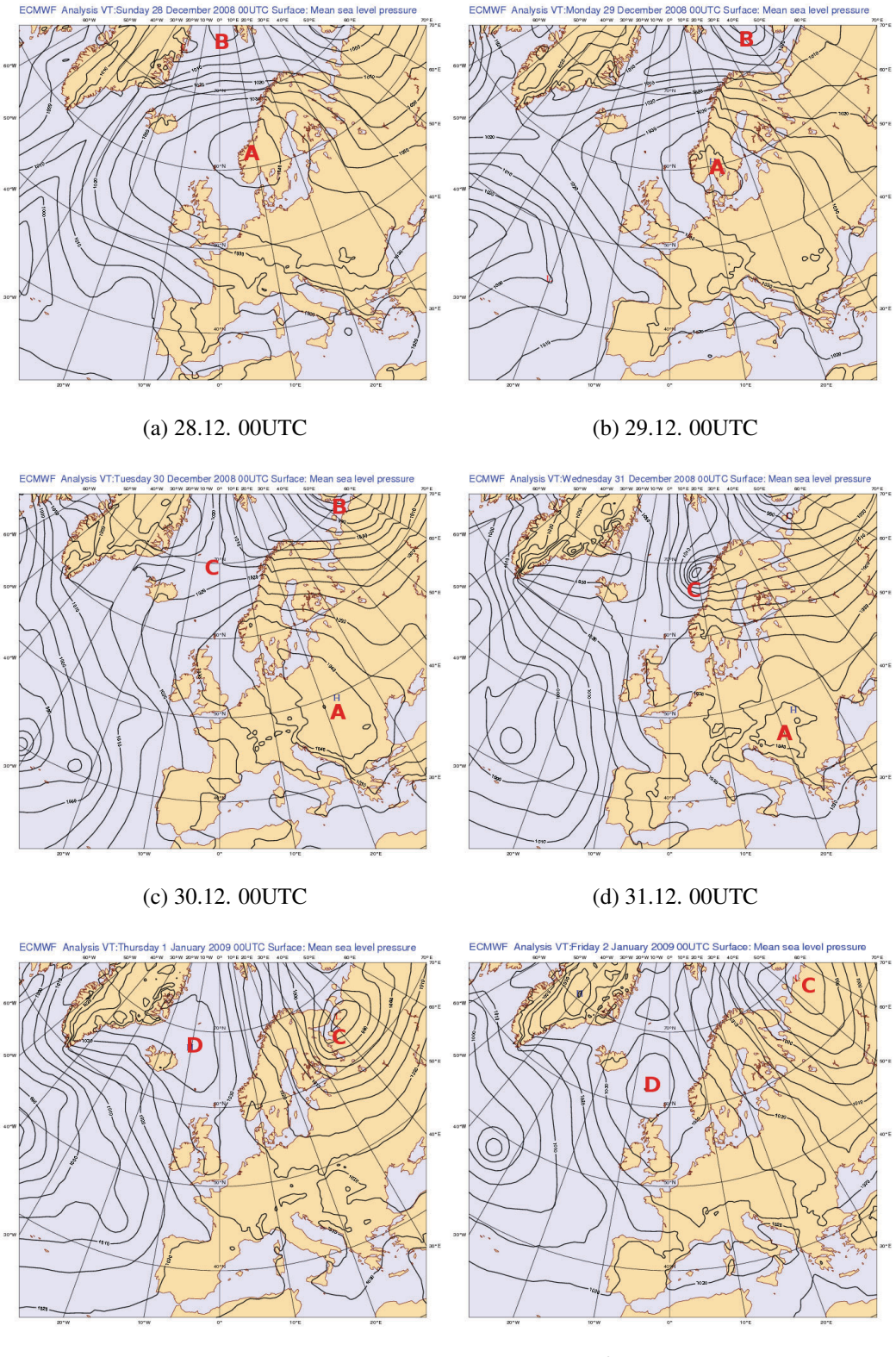
Figure 26: Verification results for mean sea level pressure from time period of 27.11.-2.12.2009. x-axis shows forecast length in hours and y-axis is deviation from observed pressure value in hPa. Red line with squares represent bias of control HIRLAM runs and green line with circles represents bias of ASCAT HIRLAM runs. Red line with +-signs represent rmse of the control run and green line with x-signs that of the ASCAT run.

5.3 Northern flow

5.3.1 Case 28.12.2008-01.01.2009 - Polar low

A strong high pressure located on top of Southern Norway (A in Fig 27) dominates the flow region in the beginning of the time-period. It slowly drifts towards Southern Sweden until at 06UTC 29th it speeds up and moves south-east. The centre stops around Hungary region and stays there until 12UTC 31st after which it dies. Meanwhile a low pressure centre starts eastward advection from Svalbard area at 12UTC 28th (B in the synoptic maps) and reaches south part of Novaya Zemlya at 00UTC 31st. On 12UTC 30th a polar low starts to form between Greenland and Norway (area C in Fig 27c). It deepens quickly and reaches north-west coast of Norway by 00UTC 31st. The low moves eastward and reaches Northern Finland by 12UTC 31st and continues on its eastward track. A weak high pressure forming on the east side of Greenland (D in Fig 27e) further enhances the flow pattern induced by the polar low, and northern air flows into especially the Scandinavia area from 06UTC 31st to 00UTC 2nd. A storm warning was given for 31st, with wind speeds of over 21 m/s in the Gulf of Bothnia

(Storm archive).



(a) 28.12. 00UTC

(b) 29.12. 00UTC

(c) 30.12. 00UTC

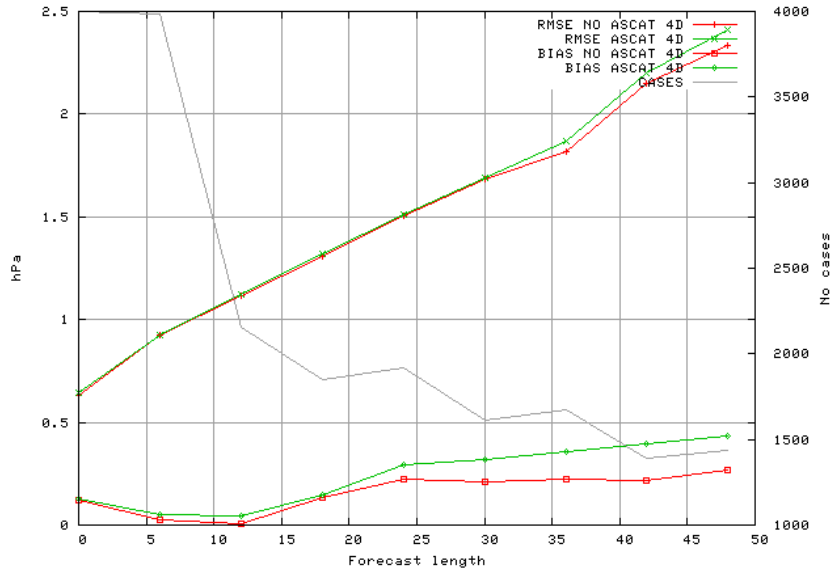
(d) 31.12. 00UTC

(e) 01.01. 00UTC

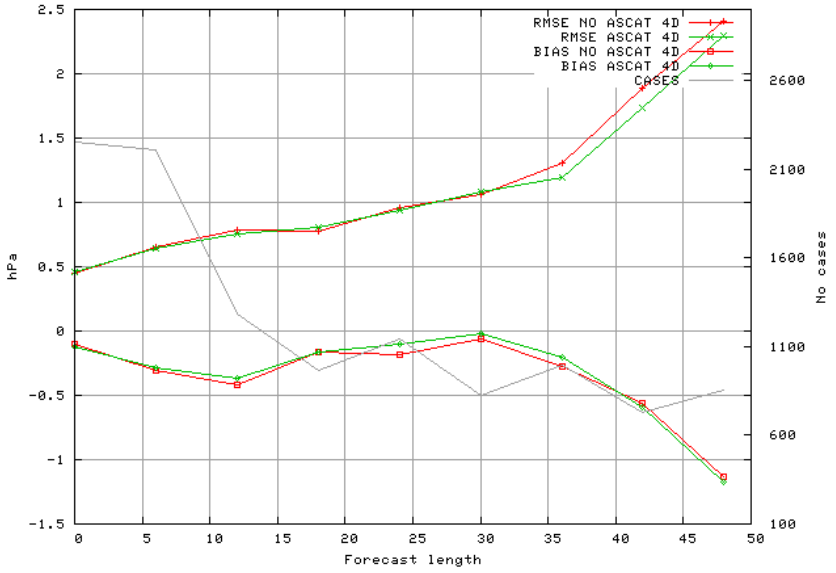
(f) 02.01. 00UTC

Figure 27: Synoptic evolution of mean sea level pressure between 00UTC 28.12.2008 and 00UTC 02.01.2009 from ECMWF analysis. Contours are drawn with 5hPa increments.

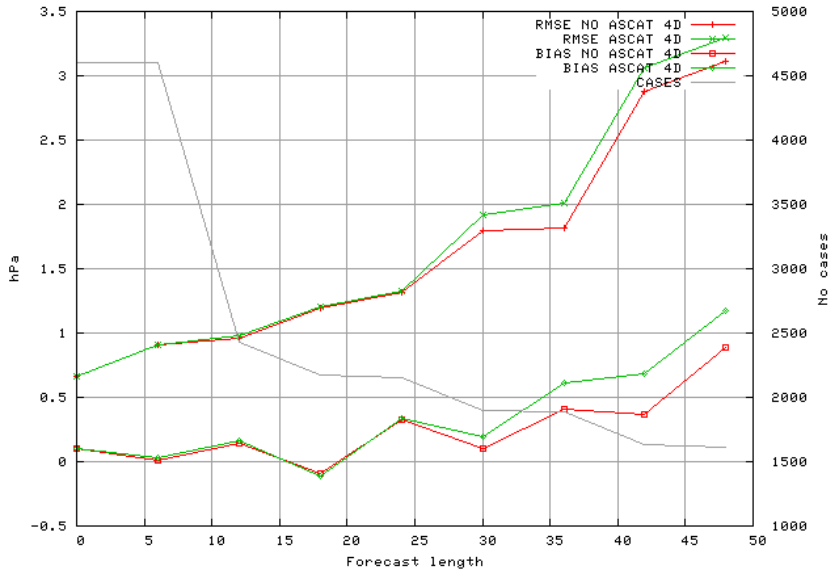
Negative bias impact from forecast hour 24 h onwards and a slightly negative rmse impact from forecast hour 36 h to 48 h is visible in the EWGLAM verification (Fig 28a). Ireland-England region verification (Fig 28b) indicates a positive-neutral bias impact and a positive rmse impact for forecast lengths 36 h, 42 h and 48 h. Both bias and rmse get a negative impact from ASCAT data on Scandinavian verification (Fig 28c) from forecast length 30 h onwards.



(a) EWGLAM



(b) Ireland and England



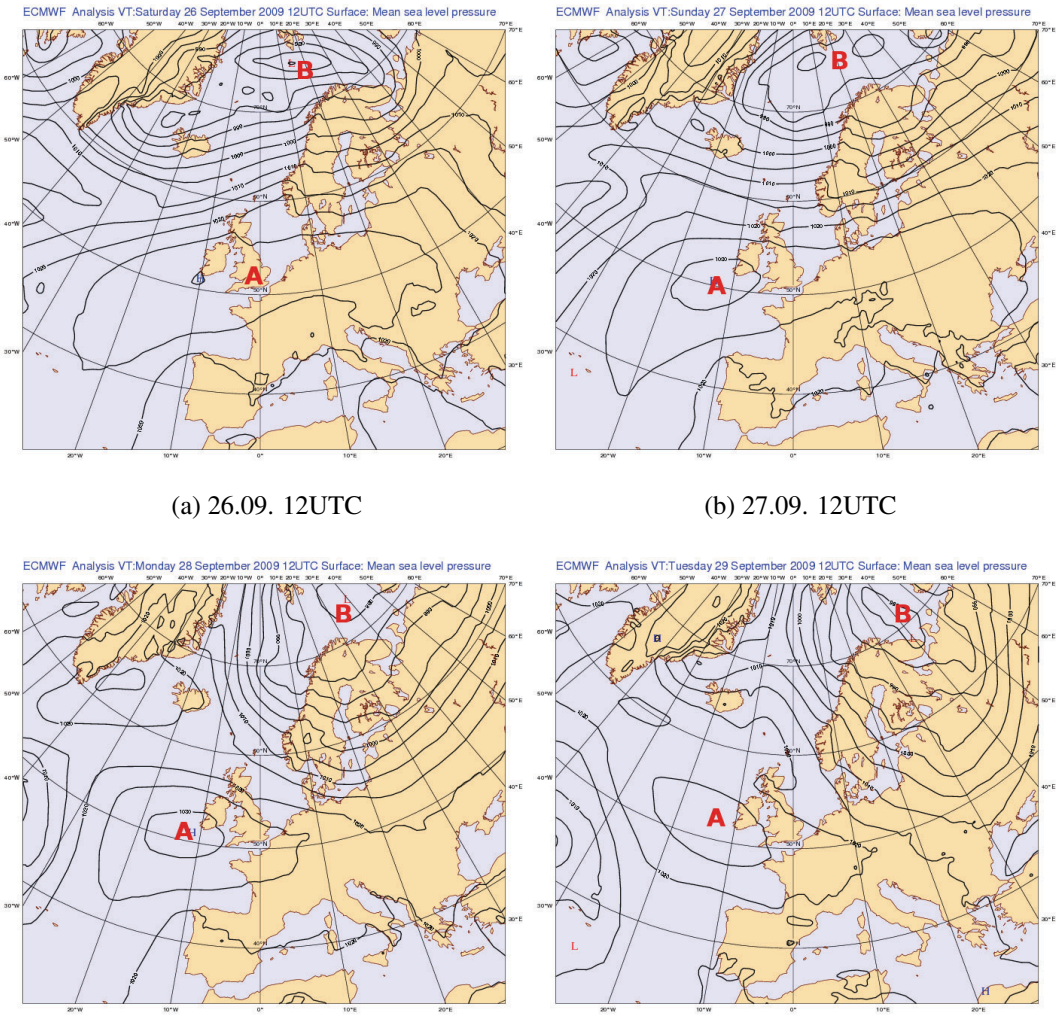
(c) Scandinavia

Figure 28: Verification results for mean sea level pressure from time period of 28.12.2008-1.1.2009. x-axis shows forecast length in hours and y-axis is deviation from observed pressure value in hPa. Red line with squares represent bias of control HIRLAM runs and green line with circles represents bias of ASCAT HIRLAM runs. Red line with +-signs represent rmse of the control run and green line with x-signs that of the ASCAT run.

5.3.2 Case 27.09.-29.09.2009

A vast high pressure system located around Britain and Central-Europe (A in Fig 29) and an eastward moving low pressure centre located north of Svalbard (B in the maps) determine the flow pattern during this period. As the high pressure system slowly advects north-westwards and the low pressure eastward, the flow to Scandinavia region changes from Atlantic into northern during 12UTC 28th. The polar air mass inflows to Scandinavia more vigorously as the low pressure moves further east.

EWGLAM area verification (Fig 30a) shows a negative impact for bias for nearly all forecast lengths and a declining rmse score from forecast length 18 h onwards. The bias for Ireland-England verification (Fig 30b) is worse for forecast lengths 6 h and 18 h and better for 36 h, 42 h and 48 h. The rmse impact is quasi-neutral for all forecast lengths. Positive bias impact can be seen for Scandinavia verification (Fig 30c) for forecast lengths greater than 30 h, while the rmse score is worse from forecast length 18 h onwards.



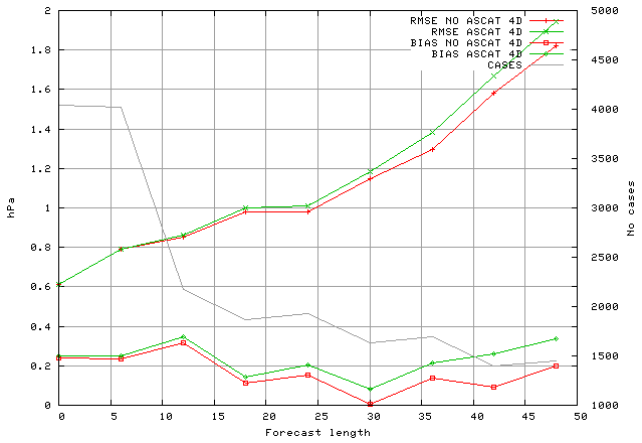
(a) 26.09. 12UTC

(b) 27.09. 12UTC

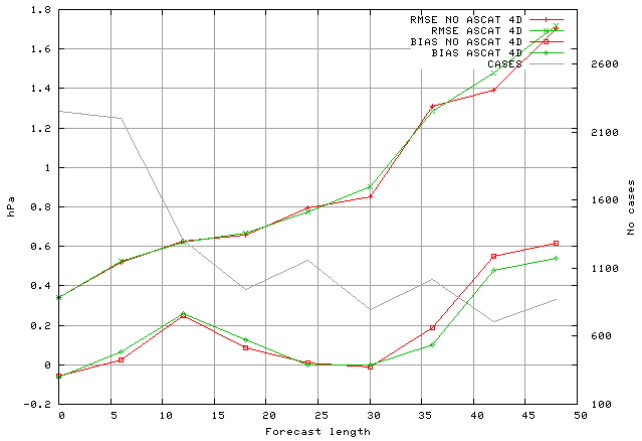
(c) 28.09. 12UTC

(d) 29.09. 12UTC

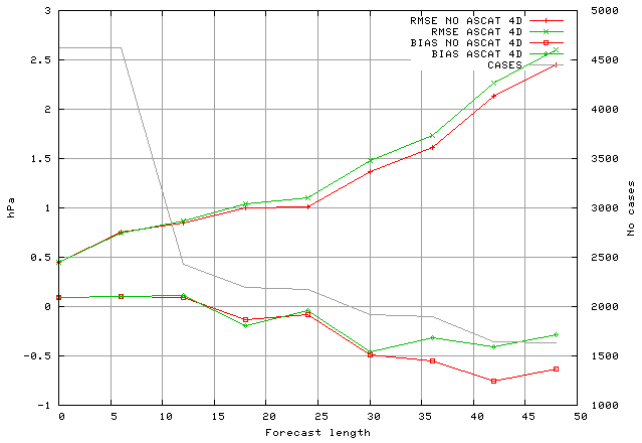
Figure 29: Synoptic evolution of mean sea level pressure between 12UTC 26.9. and 12UTC 29.9 from ECMWF analysis. Contours are drawn with 5hPa increments.



(a) EWGLAM



(b) Ireland and England



(c) Scandinavia

Figure 30: Verification results for mean sea level pressure from time period of 26.9.-30.9.2009. x-axis shows forecast length in hours and y-axis is deviation from observed pressure value in hPa. Red line with squares represent bias of control HIRLAM runs and green line with circles represents bias of ASCAT HIRLAM runs. Red line with +-signs represent rmse of the control run and green line with x-signs that of the ASCAT run.

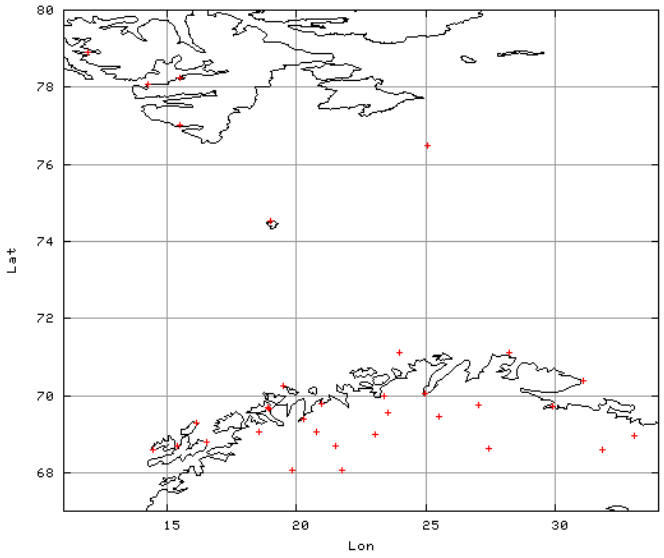
6 Discussion

6.1 Performance

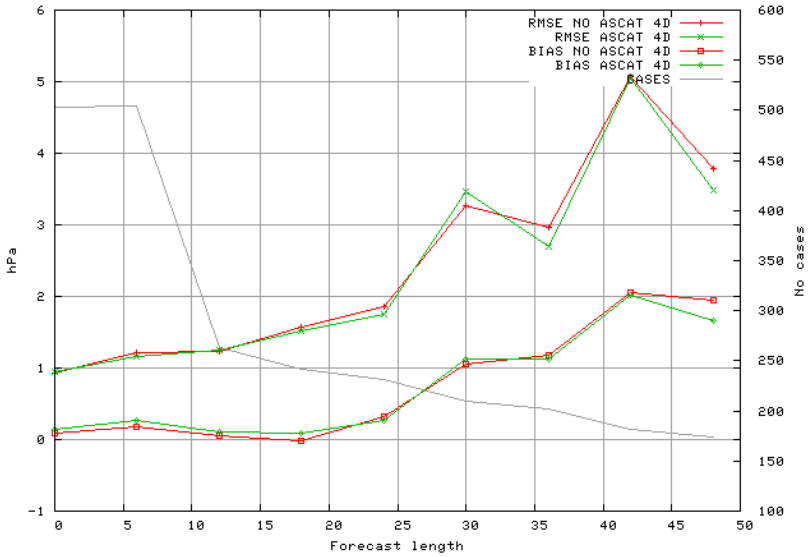
A main notice from the verification results is the difference between the Atlantic and northern flow situations. The EWGLAM area verifications suggest a neutral-positive impact from ASCAT assimilation for each of the three Atlantic flow cases for (nearly) all forecast lengths. On the other hand both northern flow case verifications for EWGLAM area show a negative impact for most of the forecast lengths. The difference is easy to understand as a larger area of Europe is affected when the winds are blowing from the Atlantic Ocean, whereas in northern flow cases only Scandinavia is directly affected. However the verification results of the two northern flow cases from Scandinavia area do not show similar improvements in the verification scores as the EWGLAM area verifications did for the Atlantic flows. Moreover a negative effect on rmse is noticeable in the verification graphs of the northern flow cases. The bad verification results of the northern flow cases are thus likely originated from problems in modelling situations where the flow direction is from the polar areas. The more dominant role of the less predictable polar masses over the northern areas, and of the more predictable weather in Ireland-England area, can also be seen in the verification graphs as the anti correlation between Ireland-England and Scandinavia scores: when the Ireland-England bias and rmse scores show positive impact the Scandinavia scores show a negative impact, and as the Scandinavia scores shows positive impact the Ireland-England scores show a negative impact. It should be noted that none of the cases represents a situation where the entire flow regime would be purely Atlantic or purely northern.

The choice of verification stations has a major impact on the verification scores. By focusing on a very confined area with only 31 observation stations (Fig 31a) the verification results for the polar low case 28.12.2008-1.1.2009 show much better bias and rmse scores for the ASCAT run (Fig 31b) when compared to the previously shown verification results of the longer time-period. Though these kind of small area examinations may give good information on longer period verification, their use with short time periods involves large uncertainties and the statistical significance is question-

able since rmse and bias are statistical quantities. Hence the use of larger number of observation stations ensures statistically more reliable results.



(a)



(b)

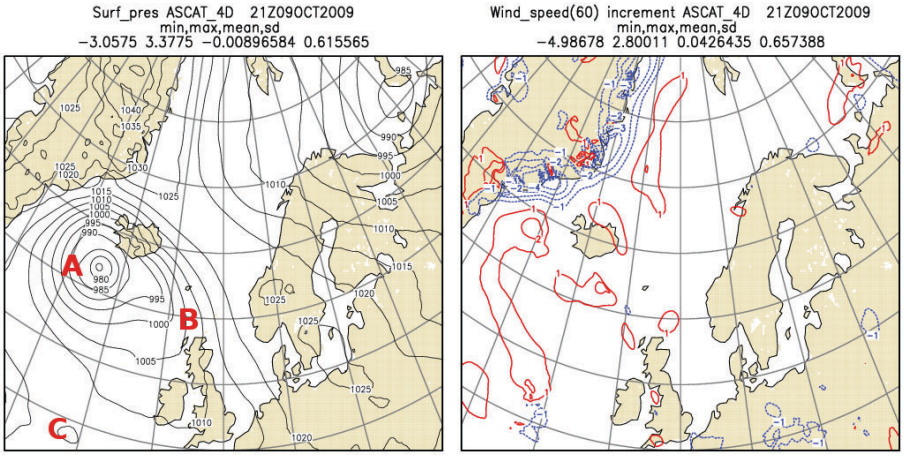
Figure 31: Verification results for mean sea level pressure for observation stations shown in a) from time period of 28.12.2008-1.1.2009. x-axis shows forecast length in hours and y-axis is deviation from observed pressure value in hPa. Red line with squares represent bias of control HIRLAM runs and green line with circles represents bias of ASCAT HIRLAM runs. Red line with +-signs represent rmse of the control run and green line with x-signs that of the ASCAT run.

6.2 Closer look on the case 08.10.-13.10.2009

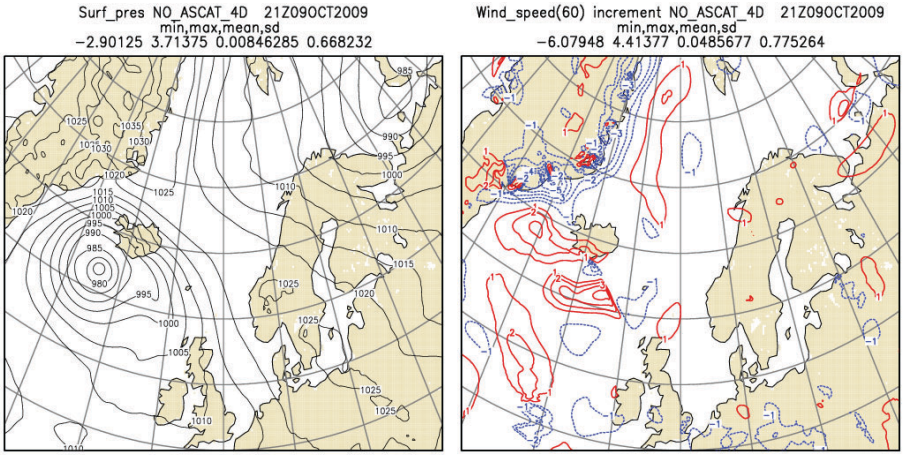
To get a better understanding how the ASCAT data effects analysis and forecast outcome a more detailed study from 10.10.2009 is presented. Figure 32 shows the mean sea level pressure (mslp) analysis and differences between background and analysis wind velocities for both HIRLAM runs at 00UTC 10.10. Figure 32e illustrates ASCAT data coverage from the corresponding observation window. As expected no major differences exist between ASCAT and control run mslp-values, but various smaller differences are evident (e.g. areas A, B and C). These anomalies also correspond well to the differences between the increment maps. The ASCAT increment map is clearly smoother than the control one. This is either due to ASCAT run background being closer to the analysis, i.e. ASCAT run forecast being more accurate, or could indicate ASCAT data influencing the analysis and keeping it from deviating too much from the background field. Or is most likely a combined effect of the two. This smoothing effect seems also to be present on increment maps of different runs (not shown).

The outcome of the two analyses is presented in Figure 33. The left-hand side figures (33a, 33c and 33e) represent ASCAT run forecasts of +24 h, +36 h and +48 h initiated from the 00UTC 10.10. analysis, and right-hand side figures (33b, 33d and 33f) that of the control run. The dashed grey line is later made analysis for the forecast hour, and thus represent the "real" state of the atmosphere. Therefore the closer the forecast and analysis contours are the better the forecast can be thought to be.

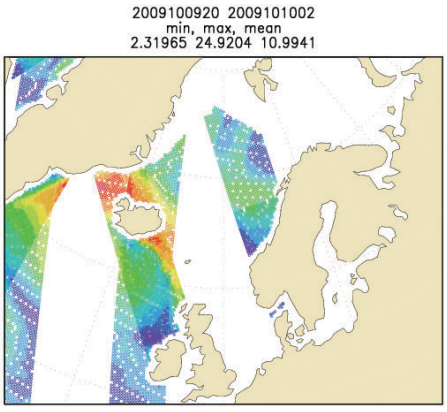
For forecast hour +24 h both of the runs agree fairly well with their respective analysis and no major differences between the two forecasts are noticeable. The differences become more visible in forecast hour +36 h: 1020 hPa isobar on top of Scandinavia (A in Figure 33c) is clearly more in accordance in north-south direction with analysis on the control run. Also the 1015 hPa isobar (B) is better forecasted by the control run. The ASCAT run seems to handle the low pressure centre (C) better than the control run but places the 1015 hPa isobar (D) too for to the west. When advancing to forecast hour +48 the problem over Scandinavia remains with the 1020 hPa and 1015 hPa isobars (A and B) more displaced in the ASCAT run than on the control. The ASCAT run also predicts the low pressure centre (C) too vast and deep, and places the isobars on top of Britain still too far to the west.



(a) ASCAT run, mean sea level pressure (b) ASCAT run, wind speed increments

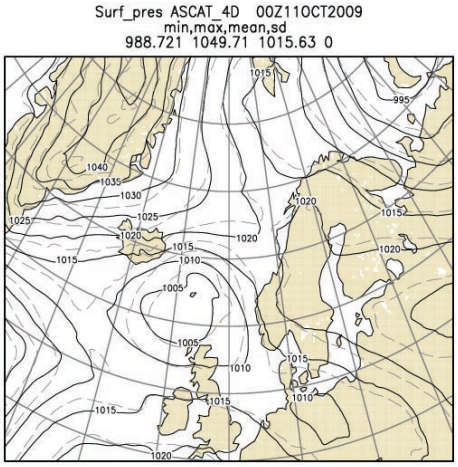


(c) Control run, mean sea level pressure (d) Control run, wind speed increments

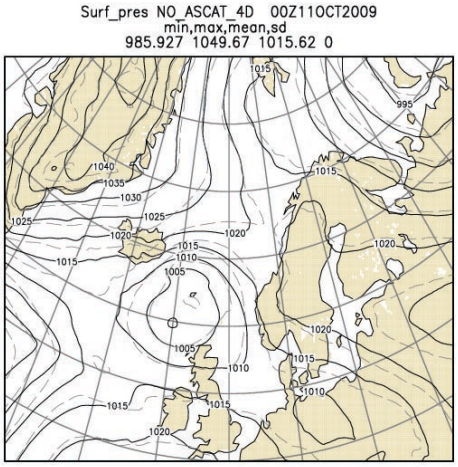


(e) Wind speed from ASCAT overpasses

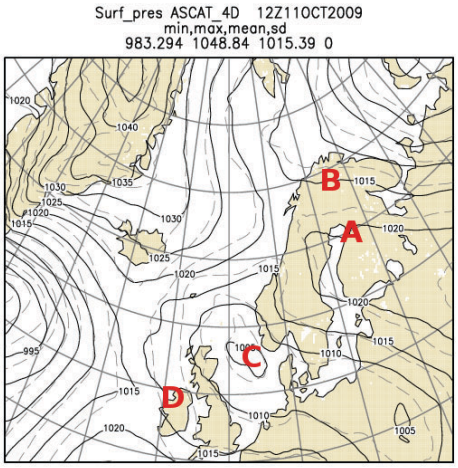
Figure 32: Comparison of ASCAT and control run outputs from 00UTC 10.10. and wind speed magnitude from ASCAT overpasses in the assimilation window 20UTC 9th to 03UTC 10th. Mean sea level pressures in 5 hPa contour interval. Wind speed increments refer to differences between background field (6 h forecast) and analysis (on lowest model level), with contours in 1 m/s.



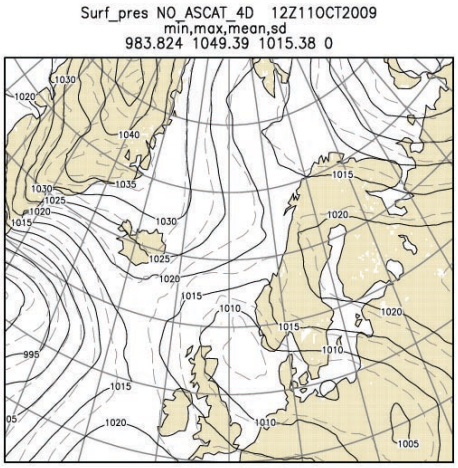
(a) ASCAT run, 10.10. 00UTC+24h



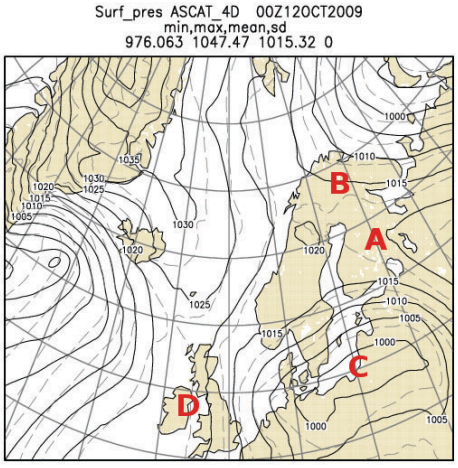
(b) Control run, 10.10. 00UTC+24h



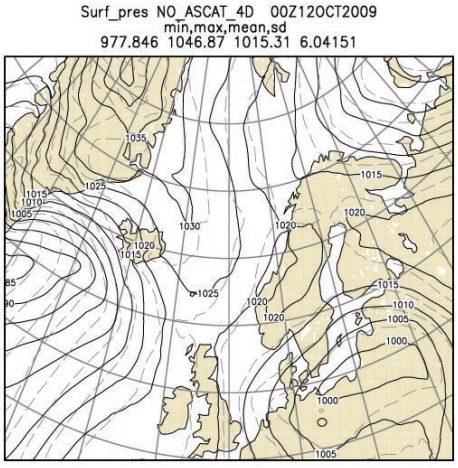
(c) ASCAT run, 10.10. 00UTC+36h



(d) Control run, 10.10. 00UTC+36h

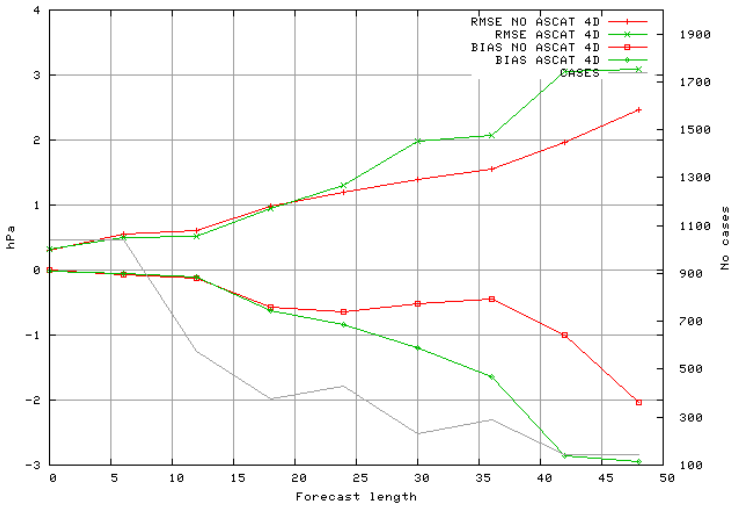


(e) ASCAT run, 10.10. 00UTC+48h

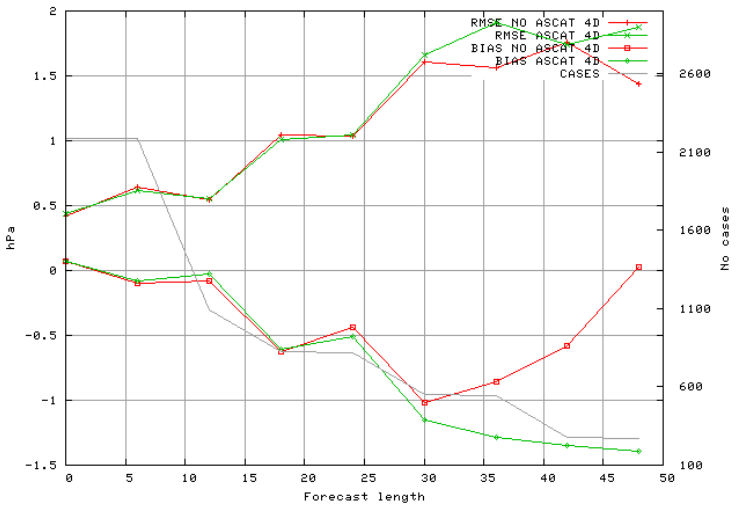


(f) Control run, 10.10. 00UTC+48h

Figure 33: Mean sea level pressure from ASCAT and control HIRLAM forecasts for 10.10. 00UTC +24 h, +36 h and +48 h (solid black line). Analysis from time of corresponding forecast time drawn as dashed grey contours. Contour interval for both curves 5 hPa.



(a) Ireland and England



(b) Scandinavia

Figure 34: Verification results for mean sea level pressure from time period of 10.10.-11.10.2009. x-axis shows forecast length in hours and y-axis is deviation from observed pressure value in hPa. Red line with squares represent bias of control HIRLAM runs and green line with circles represents bias of ASCAT HIRLAM runs. Red line with +-signs represent rmse of the control run and green line with x-signs that of the ASCAT run.

Figures 34a to 34b present verification results from shortened time period of 10.10 to 11.10. for Scandinavia and Ireland-England areas. Only one forecast for lengths +42 h and +48 h (since the verification time line starts at 00UTC 10.10. and ends at 00UTC 12.10.) and a reduced number of longer forecasts is included in the verification result. Though this is by no means a good or proper utilization of the verification tool, the origin of the bad verification scores for long forecasts in the longer period

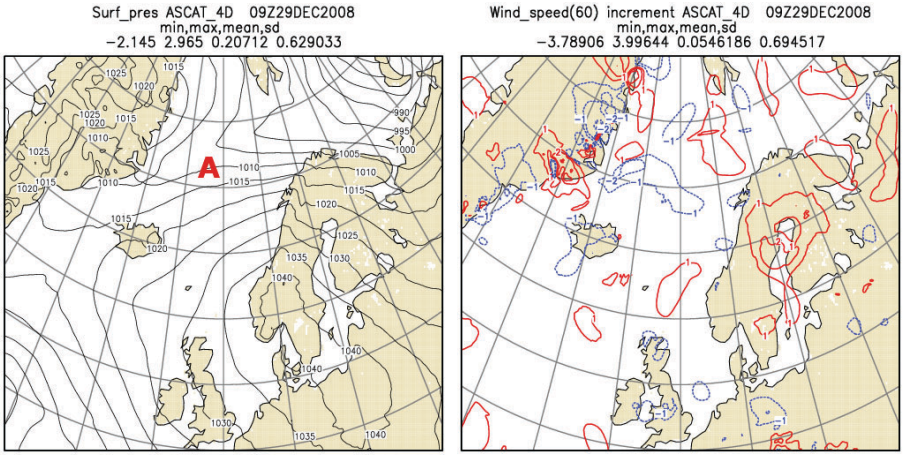
verifications shown earlier is observable. The outcome for the Scandinavia verification (Fig 34b) is clearly heavily influenced by the model displacement of isobars A and B, and the result of Ireland-England verification (Fig 34a) by the east-west displacement of the isobars (C). These scores from the two areas then not only affect their individual longer period verification scores, but also contribute in making the ASCAT run EWGLAM verification result worse for +48 h forecast length in the longer period verification.

Naturally as the verification method is statistic in its nature its usage has limits. For getting a sensible result with the method either a long time period or very good observation station coverage should be used. Since neither is present in Figure 34 the verification results are rather indicative in their nature. This also applies to the previously shown verifications, although they are more statistically sound.

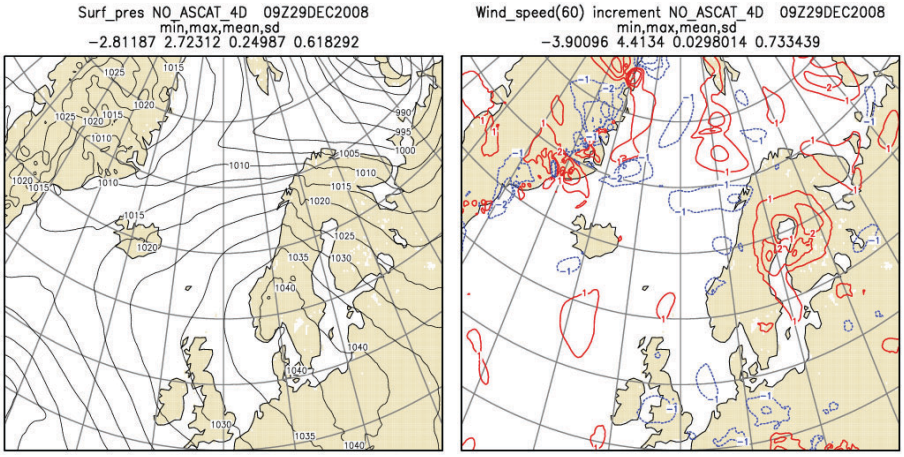
6.3 Closer look on the polar low

Figure 35 presents again the effect of ASCAT data on the analysis and the differences between the ASCAT and control run analyses. Adequate amount of ASCAT data was available on the observation window and the increment maps show again a smoother background versus analysis difference with the ASCAT run. No major differences exist between ASCAT and control run mslp maps, but some smaller differences are visible in e.g. area between Norway and Greenland (A in Fig 35a). Though small, the effects of the analysis differences can be seen clearly in forecasts for +24 h, +30 h and +36 h shown in Fig 36. The ASCAT run has clearly an edge over the control run in forecasting the polar low evolution: it forecasts the deepening of the polar low (A in Fig 36a) in an early stage and simulates the initial movement quite well. The ASCAT run also seems to perform better overall especially in the north.

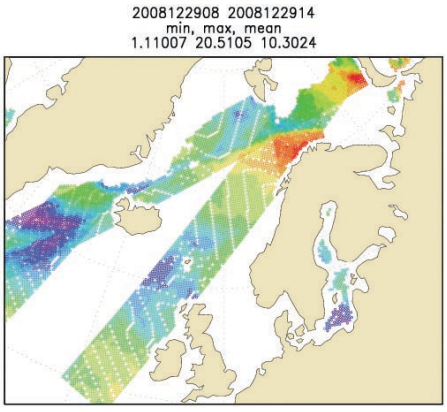
The cause for the poor verification results shown earlier is also evident: the model dynamics/physics speeds up the strong low pressure anomaly in the east-west direction, causing the polar low to move too fast eastward. This effect then dominates the long forecast lengths and outshines the improvements in forecasting the short development phase of the polar low.



(a) ASCAT run, mean sea level pressure (b) ASCAT run, wind speed increments

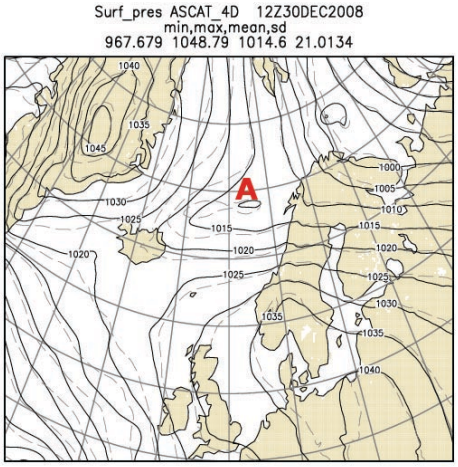


(c) Control run, mean sea level pressure (d) Control run, wind speed increments

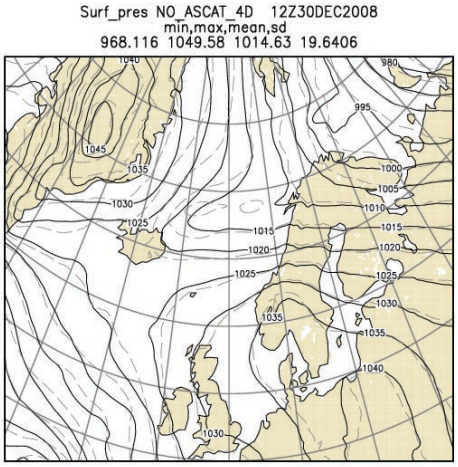


(e) Wind speed from ASCAT overpasses

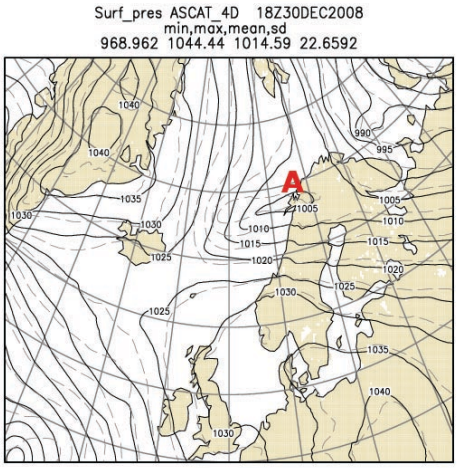
Figure 35: Comparison of ASCAT and control run outputs from 12UTC 29.12.2008 and wind speed magnitude from ASCAT overpasses in the assimilation window 08UTC to 14UTC 29th. Mean sea level pressures in 5 hPa contour interval. Wind speed increments refer to differences between background field (6 h forecast) and analysis (on lowest model level), with contours in 1 m/s.



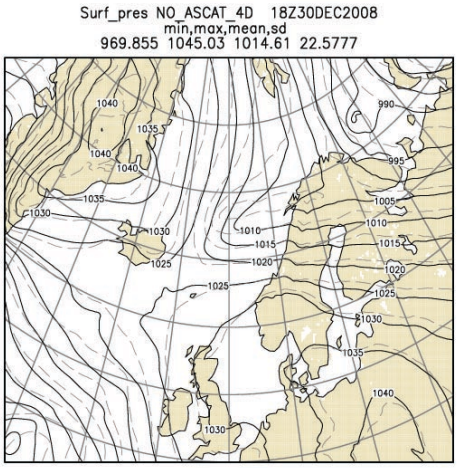
(a) ASCAT run, 29.12. 12UTC+24h



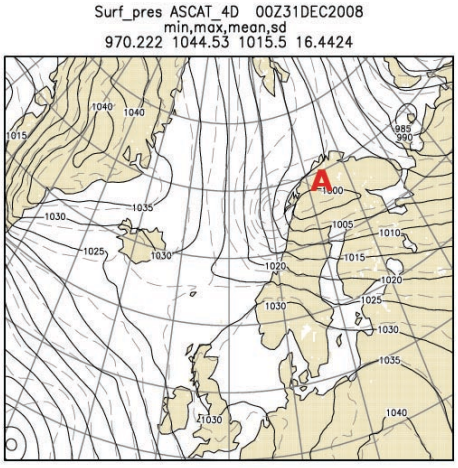
(b) Control run, 29.12. 12UTC+24h



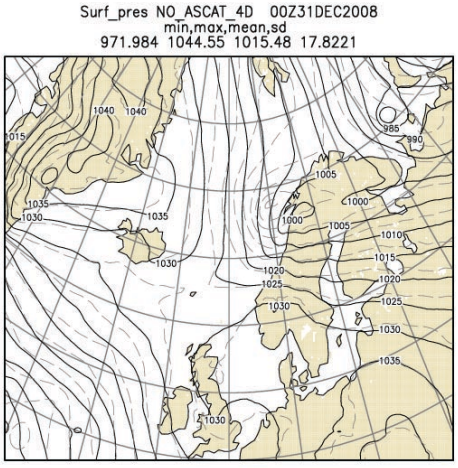
(c) ASCAT run, 29.12. 12UTC+30h



(d) Control run, 29.12. 12UTC+30h



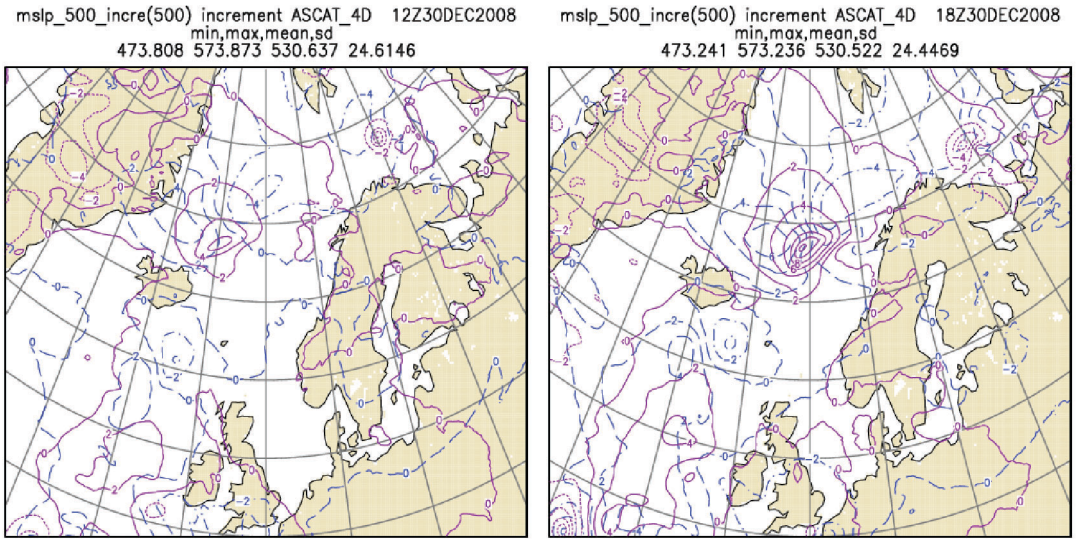
(e) ASCAT run, 29.12. 12UTC+36h



(f) Control run, 29.12. 12UTC+36h

Figure 36: Mean sea level pressure from ASCAT and control HIRLAM forecasts for 29.12.2008 12UTC +24h, +30h and +36h (solid black line). Analysis from time of corresponding forecast time drawn as dashed grey contours. Contour interval for both curves 5hPa.

Polar low is a three-dimensional phenomenon, but its early development is in a rough approximation defined by the 1000 hPa and 500 hPa topography (SatManu webpage). The differences between forecast and analysis geopotential fields from 1000 hPa and 500 hPa pressure levels are shown in Figure 37. The figure illustrates that the 500 hPa level anomaly was in the same wrong phase as the surface anomaly throughout the development phase of the polar low. Hence the forecast error in the upper level was not responsible for the too fast eastward advection of the polar low, and moreover the error was not caused by flaws in the implementation involving upward information spreading of ASCAT data. Thus the cause is rather in the model dynamics/physics.



(a) 10.10. 12UTC + 24h

(b) 10.10. 12UTC + 30h

Figure 37: Geopotential increments in 1000 hPa (purple solid and dotted contours) and 500 hPa (blue dashed and dash-dotted contours) pressure levels. Inrement contour spacing 2 m^2/s^2.

6.4 Error sources

Though no very long period verification was made it is easy to deduce that intensive storms dominate the RMSE score measured: the performance during a single storm can dominate an impact averaged over many weeks (Tveter, Frank T.). Scatterometer observations rely on the model first guess, which can be 180 degrees wrong if the storm centre is displaced. C-band scatterometer is not insensitive to rain, and is

affected directly by heavy precipitation from backscattering off the the rain-drops and probably also indirectly since the rain disturbs the ocean surface. Strong quality control is necessary for such data, but this removes observations from important areas of storms, making it more difficult to identify any overall impact (Tveter, Frank T.).

Another important error source for scatterometer measurements is the representativeness error: if the wind over a WVC area (25x25 km) is variable, it is hard to obtain a reliable wind for this WVC (see e.g. Portabella & Stoffelen (2006)). Near the centre of cyclonic phenomena, e.g. low pressure centres, the quality control sometimes flags some WVCs as unreliable, further reducing the information content from these important areas. Lack of validation sources especially for high wind speeds (> 25 m/s), arises an issue with calibration of the scatterometer winds, and may induce uncertainties in the wind speeds in this regime (Verhoef, Anton).

Those WVCs that are not flagged, are normally of good quality. Extensive validations with buoy data over a nearly two year period (see KNMI scatterometer webpage) and ongoing validation work (e.g. Verspeek *et al.* (2008)) provide means for guaranteeing this. Despite the errors which are inevitable, the scatterometer is the best data source for winds over sea and generally speaking the errors are quite small (Verhoef, Anton).

While doing this study I noticed that the implementation did not take into consideration that the input winds are calculated for *neutral* stratification, and hence no correction for real winds, corresponding to the *real* boundary layer stratification, are made in the model. The problem was discussed, but currently the HIRLAM vertical grid levels in the boundary layer are most likely not numerous enough (lowest model level at $\sim$30 m and second lowest at $\sim$90 m (FMI HIRLAM webpage)) for precise stability determination of the lowest part of the boundary layer. Thus implementing correction to the neutral winds with the HIRLAM boundary layer stability might in the worst case even correct the winds further from real case. Since the typical set-ups for polar lows comprise advection of cold air from over cold sea ice to warm sea surface (Savijärvi, 2006), the stability changes involved in the polar low initial evolution are substantial. Hence treating the ASCAT winds retrieved from these areas as neutral winds produce errors to the used ASCAT wind field.

The ASCAT quality control system in HIRLAM might also need tuning. The gross

error was by default small, thus suspicious observations are used. Though probably not a big problem, slightly increasing the gross error parameter might give more optimal use of the ASCAT observations (Tveter, Frank T.).

As already mentioned there arises a problem with the verification method used: a statistical tool is not designed for single cases, i.e. short time period and low observation count situations. This means that poor spatial coverage leaves room for coincidence in short time periods. Let us consider a situation where the ASCAT run forecast would be closer to the real atmospheric state than the control run forecast. The station coverage does not necessarily "see" this and wrong ASCAT run forecast values existing only in observation site area would lead to a poor verification score. Furthermore the observation stations are located on land areas and no oceanic observations are included. Since the improvements with ASCAT runs tend to occur on sea area from where no verification is available the scores are very biased. It is easy to justify the observation station placement as the weather occurring on top of land areas is of more importance to society. But since ocean transportation is dependent on oceanic weather forecasts and thus involves a big economic impact, sea areas should be included in the verification scoring.

A verification method comparing forecast with analysis of the corresponding forecast time would be extremely helpful to verify these kind of short time period impact studies, and would also provide an objective method for comparing contour maps. The verification could either be done with calculating increments between forecast and analysis of the same run (ASCAT forecast - ASCAT analysis, control forecast - control analysis) or comparing both forecasts against a neutral analysis field (e.g. ECMWF operational system or ECMWF Re-analyses (ERA) Interim archive). As an example, Figure 38 illustrates verification scores for 10.10.2009 obtained by comparing forecasts against their individual analysis over the whole RCR area. When compared to the previously shown verification scores for the same short period (Fig 34) and to the longer period verification showed it is clearly seen that this new method shows the oceanic effect of ASCAT data and gives the ASCAT runs larger positive impact. Though these kind of methods already exist, their use is not promoted since the comparison of model against model can be biased from the real atmospheric state. Nevertheless this approach could e.g. use analysis made with even more weight given

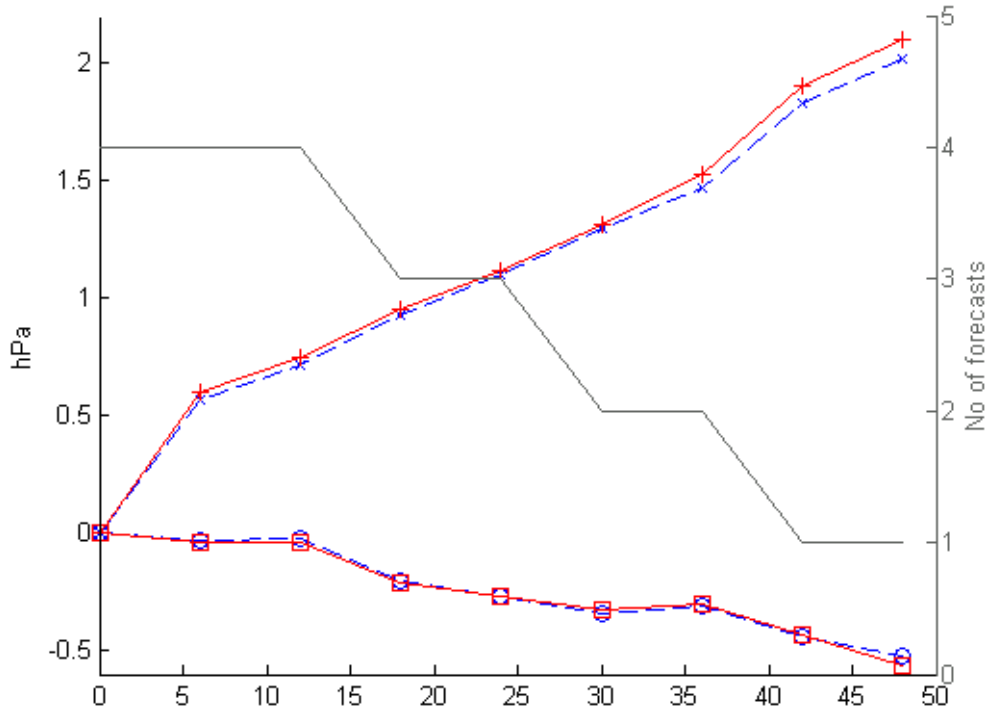


Figure 38: Whole RCR area verification result for mean sea level pressure from time period of 10.10.-11.10.2009. x-axis shows forecast length in hours and y-axis is deviation from observed pressure value in hPa. Red line with squares represent bias of control HIRLAM runs and blue dashed line with circles represents bias of ASCAT HIRLAM runs. Red line with +-signs represent rmse of the control run and blue dashed line with x-signs that of the ASCAT run.

to the observations. Or further verify the used analysis against the observations.

It remains unclear what kind of an impact using host model upper air analysis for redoing background (LSMIX) would have with the ASCAT assimilation. Since the ECMWF analysis used for doing LSMIX already contains ASCAT data the effect of the ASCAT implementation might lessen. It could also induce a too big impact of the ASCAT data since the data would be assimilated “twice” (once in the ECMWF analysis and second time in the HIRLAM analysis).

7 Conclusions

The previous impact studies of including scatterometer winds in analysis systems have also been conducted from the viewpoint of case studies, since the impact of previous generation scatterometers was close to neutral in longer period verifications. The focus of the case studies, conducted by ECMWF, has been on tropical cyclones (Isaksen & Stoffelen (2000); Isaksen & Janssen (2000); Leidner *et al.* (2002)) but a single case of a polar low has also been studied by Isaksen & Janssen (2000). The tropical cyclone studies showed a major positive impact from assimilating scatterometer data. Also the polar low study, conducted with ERS-scatterometer data, displayed an improvement to the analysis of the polar low for areas with higher wind speed, but did not present any forecast comparison or further evolution of the polar low. Tveter (2006) conducted an impact study over four-month period of HIRLAM runs, for which assimilation of QuickScat observations gave an overall small positive impact in mslp verification scores for all forecast lengths. The advantages of long time period verification comes from the statistical smoothing it offers, as single erroneous forecasts do not dominate the verification scores. However, conducting NWP model runs over a long period is extremely time and data storage consuming. Thus focusing on single case studies and studying the impact through them is practical.

Implementation of ASCAT into the ECMWF NWP system revealed an improved accuracy of the first-guess and analysis surface winds (Hersbach & Janssen, 2007). The positive verification results from the "normal" flow pattern with prevailing westerly winds shown in this study would also suggest that long term impact from assimilating ASCAT data into HIRLAM NWP is positive. The implementation also showed the usefulness of ASCAT assimilation by providing an analysis field from which the polar low evolution was predicted more accurately.

ASCAT data provides a good basis for better simulation of the flows from northern areas. ASCAT data coverage will be enhanced in near future (Verhoef, Anton), as an extension to the ASCAT GMF will provide data much closer to the coast (20-25 km off the coast) compared to the current 50-75 km (Portabella *et al.* , 2008). Future satellite missions like the Polar Communication and Weather mission (PCW) would broaden the basis provided by ASCAT even more. PWC would offer a continuous

24/7 weather monitoring in the entire Arctic region and provide wind profiles from the atmosphere (see PWC webpage for further details). With these two information channels predicting phenomena like polar lows more accurately would presumably become feasible.

As suggested by Vihma (2009) increasing the HIRLAM model vertical resolution to improve better simulation of the boundary layer would provide also vital information for converting the neutral winds now assimilated into real winds. This would likely improve the ASCAT assimilation effect even further. While the boundary layer resolution stays in its present form, adding a bias-correction scheme to correct the ASCAT winds for systematic errors could help improving the ASCAT effect. This could be done by using a rough stability parameter provided by HIRLAM as a predictor (Tveter, Frank T.).

It would be extremely interesting to compare the verification results from the same cases with forecast runs conducted with the ECMWF NWP system, and see if the two different ASCAT assimilations have similar impacts. Also a more thorough study of the polar low(s), focusing on model behaviour with *all* important parameters for the polar low evolution, should be conducted. Conducting a long period verification with slightly increased probability of gross error would be useful for determining whether the parameter has been too small or not. Improving the verification tool from the viewpoint I suggested and showed would provide a less spatially biased validation method for NWP forecasts. Finally, it would be insightful to compare the *predictability* of the atmosphere for all the cases, and see whether the assimilation impact had any correlation with it.

Based on this study, ASCAT assimilation shows a lot of potential. Before implementing the ASCAT assimilation into operational HIRLAM longer time period studies (two months or more) from different times of the year should be conducted.

Acknowledgements

Special thanks to Frank Tveter from met.no who provided the assimilation implementation and gave much needed guidance into the assimilation scheme, without him this study would not have materialized. Thanks to Anton Verhoef from KNMI who discussed the scatterometer uncertainties with me and to Sami Niemelä for his help with HIRLAM and expertise on related things. My gratitudes also to all co-workers helping with related and unrelated stuff.

Last but certainly not least, warm thanks to Mom and Dad for supporting me in whatever I do and always being there for me!

Bibliography

Bader, J.M., Forbes, G.S., Grant, J.R., Lilley, R.B.E., & Waters, A.J. (eds). 1995. *Images in weather forecasting - A practical guide for intepreting satellite and radar imagery*. Cambridge University Press.

Daley, Roger. 1991. *Atmospheric Data Analysis*. Cambridge Univesity Press.

Donnelly, William J., Carswell, James R., McIntosh, Robert E., Chang, Paul S.and Wilkerson, John, Marks, Frank, & Black, Peter G. 1999. Revised ocean backscatter models at C and Ku band under high-wind conditions. *Journal of Geophysical Research*, **104**(C5), 11,485–11,497.

Eerola, Kalle. 2010. Personal communication.

Figa-Saldaña, J., Wilson, J.J.W., Attema, E., Gelsthorpe, R., Drinkwater, M.R., & Stoffelen, A. 2002. The advanced scatterometer (ASCAT) on the meteorological operational (MetOp) platform: A follow on for European wind scatterometers. *Can. J. Remote Sensing*, **28**(3), 404–412.

FMI. 2008. *Joint WMO technical progress report on the global data processing and forecasting system and numerical weather prediction research activities for 2008*. Tech. rept. Finland / FMI.

FMI HIRLAM webpage. URL: http://fminwp.fmi.fi/docs/.

Gelsthorpe, R.V., Schied, E., & Wilson, J.J.W. 2000. *ASCAT - Metop s Advanced Scatterometer*. Tech. rept. ESA bulletin 102.

Grasso, Lewis D. 2000. The differentiation between grid spacing and resolution and their application to numerical modeling. *Bulletin of the American Meteorological Society*, **81pp**(3), 579–580.

Haltiner, George J. 1971. *Numerical Weather Prediction*. John Wiley & Sons.

Haltiner, George J., & Williams, Roger Terry. 1980. *Numerical prediction and dynamic meteorology*. John Wiley & Sons.

Hersbach, Hans, & Janssen, Peter. 2007. Operational assimilation of surface wind data from the MetOp ASCAT scatterometer at ECMWF. *ECMWF Newsletter*, **113**, 6–8.

HIRLAM webpage. URL: http://hirlam.org.

Holopainen, Eero, Fortelius, Carl, Ruosteenoja, Kimmo, & Räisänen, Jouni. 2008. *Ilmakehän yleinen kiertoliike I*. University of Helsinki. Course literature.

Holton, James .R. 2004. *An Introduction to Dynamic Meteorology*. Elsevier Academic Press.

Isaksen, L., & Janssen, Peter A.E.M. 2000. *Impact of ERS Scatterometer Winds in ECMWF s Assimilation System*. Tech. rept. European Centre for Medium-Range Weather Forecasts.

Isaksen, L., & Stoffelen, A. 2000. *ERS-scatterometer wind data impact on ECMWF s tropical cyclone forecasts*. Tech. rept. European Centre for Medium-Range Weather Forecasts.

Järvinen, Heikki. 2003. *Numeerinen meteorologia II*. University of Helsinki. Course literature.

Jolliffe, Ian T., & Stephenson, David B. (eds). 2003. *Forecast Veri cation - A Practioner s Guide in Atmospheric Science*. John Wiley & Sons.

Kalnay, Eugenia (ed). 2003. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press.

Kidder, Stanley Q., & Vonder Haar, Thomas H. 1995. *Satellite Meteorology : An introduction*. Academic Press, Inc.

KNMI scatterometer webpage. URL: http://www.knmi.nl/scatterometer/.

Leidner, Mark S., Isaksen, Lars, & Hoffman, Ross N. 2002. Impact of NSCAT winds on Tropical Cyclones in the ECMWF 4DVAR Assimilation System. *Monthly Weather Review*, **131**, 3–26.

O&SI-SAF. 2010. *ASCAT Wind Product User Manual version 1.8*. Ocean and Sea Ice SAF. SAF/OSI/CDOP/KNMI/TEC/MA/126.

Phillips, N. A. 1959. A coordinate system having some special advantages for numerical forecasting. *J. Met. Soc.*, **14**, 184–185.

Portabella, M., & Stoffelen, A. 2004. A probabilistic approach for SeaWinds data assimilation. *Q. J. R. Meteorol. Soc.*, **130**, 1793–1814.

Portabella, M., Verhoef, A., & Stoffelen, A. 2008. *ASCAT coastal AWDP prototype*. Tech. rept. NWP SAF.

Portabella, Marcos. 2002. *Wind Field Retrival from Satellite Radar Systems*. Ph.D. thesis, University of Barcelona.

Portabella, Marcos, & Stoffelen, Ad. 2006. Scatterometer backscatter uncertainty due to wind variability. *IEEE. Trans. Geosci. Rem. Sens.*, **44**(11), 3356–3362.

PWC webpage. URL: http://www.asc-csa.gc.ca/eng/satellites/pcw/.

Saltikoff, Elena, Huuskonen, Asko, Hohti, Harri, Koistinen, Jarmo, & Järvinen, Heikki. *Quality Assurance in the FMI Doppler Weather Radar Network*. Boreal Env. Res 2009 (in print).

SatManu webpage. *Manual of synoptic satellite meteorology*. URL: http://www.zamg.ac.at/docu/Manual/SatManu/main.htm.

Savijärvi, Hannu. 2006. *Mesometeorologia*. University of Helsinki. Course literature.

Savijärvi, Hannu. 2007. *Numeerinen meteorologia I*. University of Helsinki. Course literature.

Savijärvi, Hannu, & Vihma, Timo. 2001. *Rajakerroksen fysiikka I*. University of Helsinki. Course literature.

Stephens, Graeme L. 1994. *Remote sensing of the Lower Atmosphere*. Oxford university press.

Stoffelen, A. 1998. *Scatterometry*. Ph.D. thesis, Universiteit Utrecht.

Storm archive. *FMI storm archive*.

Talagrand, O. 1997. Variational assimilation of observations, an introduction. *J. Met. Soc. Japan*, **Special issue 75**(1B), 191–209.

Tveter, Frank T. 2006. Assimilating ambigouous QuikScat scatterometer observations in HIRLAM 3-D-Var at the Norwegian Meteorological Institute. *Tellus*, **58A**, 59–68.

Tveter, Frank T. 2010. Personal communication.

Ulaby, Fawwaz T., Moore, Richard K., & Fung, Adrian K. (eds). 1981. *Microwave Remote Sensing: Active and Passive. Volume I. Microwave Remote Sensing Fundamentals and Radiometry*. Arctec House.

Ulaby, Fawwaz T., Moore, Richard K., & Fung, Adrian K. (eds). 1982. *Microwave Remote Sensing: Active and Passive. Volume II. Radar Remote Sensing and Surface Scattering and Emission Theory*. Arctech House.

Ulaby, Fawwaz T., Moore, Richard K., & Fung, Adrian K. (eds). 1986. *Microwave Remote Sensing: Active and Passive. Volume III. From Theory to Applications*. Arctech House.

Únden, Per, Rontu, L., Järvinen, H., Lynch, P., Calvo, J., Cats, G., Cuxart, J., Eerola, K., Fortelius, C., Garcia-Moya, J.A., Jones, C., Lenderlink, G., McDonald, A., McGrath, R., Navascues, B., Nielsen, N. W., Ø degaard, V., Rodriguez, E., Rummukainen, M., Rõ om, R., Sattler, K., Sass, B. H., Savijärvi, H., Schreur, B. W., Sigg, R., The, H., & Tijm, A. 2002. *HIRLAM-5 Scienti c Documentation*.

Verhoef, Anton. 2010. Personal communication.

Verspeek, Jeroen, Portabella, Marcos, Stoffelen, Ad, & Verhoef, Anton. 2008. *Calibration and Validation of ASCAT Winds*. Tech. rept. O&SI SAF.

Vihma, Timo. 2009. Modelling of the atmospheric boundary layer over sea ice. *HIRLAM Newsletter*, **48**, 112–119.

WMO. 2006. *Guide to Meteorological Instruments and Methods of Observation, No. 8 (7th edition)*. Secretariat WMO, Geneva, Switzerland.

Yang, Xiaohua. 2009. Analysis blending using spatial filter in grid-point model coupling. *Hirlam Newsletter*, **48**, 49–55.

A Abbreviations

3D-Var Three-dimensional variational assimilation

4D-Var Four-dimensional variational assimilation

ASCAT Advanced Scatterometer

EARS EUMETSAT Advanced Retransmission Service

ECMWF European Centre for Medium-Range Weather Forecasts

EPS EUMETSAT s Polar System

ERA ECMWF Re-analyses

ESA The European Space Agency

EUMETSAT European Organization for the Exploitation of Meteorological Satellites

GMP Geophysical Model Function

HIRLAM High Resolution Limited Area Model

LAM Limited Area Model

Metop Meteorological Operational

mslp Mean sea level pressure

NRCS Normalized Radar Cross Section

NWP Numerical Weather Prediction

O&SI-SAF Ocean & Sea Ice Satellite Application Facility

PCW Polar Communication and Weather

rmse Root mean squared error

SYNOP Surface Synoptic Observations

WVC Wind Vector Cell