

UNIVERSITÀ DEGLI STUDI DI BOLOGNA

Facoltà di Scienze Matematiche, Fisiche e Naturali

Dipartimento di Scienze della Terra e Geologico–Ambientali

**DOTTORATO DI RICERCA IN MODELLISTICA FISICA PER LA
PROTEZIONE DELL'AMBIENTE**

XIX Ciclo

Tesi di dottorato

**ASSIMILATION OF METEOSAT SECOND GENERATION
(MSG) SATELLITE DATA IN A REGIONAL NUMERICAL
WEATHER PREDICTION MODEL USING A
ONE–DIMENSIONAL VARIATIONAL APPROACH**

Dottorando:
Marco Elementi

Relatore:
Prof. Stefano Tibaldi

Correlatori:
Dott. Francesca Di Giuseppe
Dott. Tiziana Paccagnella

Coordinatore del Dottorato:
Prof. Ezio Todini

Anno Accademico 2005–2006

A mia madre e mio padre che hanno reso possibile l'impossibile

Contents

0.1	The initial condition problem	13
0.2	Concept of analysis	14
0.3	Satellite data assimilation	15
0.4	Scientific questions to be addressed	17
1	1D-VAR FRAMEWORK	19
1.1	Basic concepts of data assimilation	19
1.2	Control space variables	20
1.3	Observations	20
1.4	The modelling of errors	21
1.5	Statistical interpolation with least-squares estimation	22
1.5.1	Notation	23
1.5.2	Hypotheses	23
1.5.3	BLUE theorem	24
1.6	The one-dimensional variational analysis	25
2	OBSERVATIONAL DATA AND MODEL DATA	27
2.1	Conventional data	27
2.2	Non-conventional data	28
2.3	Polar and geostationary satellites	29
2.4	The global system of operational meteorological satellites	31
2.5	The Meteosat series	32
2.6	Radiometric instrument specifications	33
2.6.1	SEVIRI	33
2.6.2	GERB	37

2.7	The Limited Area Model COSMO	38
2.7.1	Assimilation of observations in COSMO: the <i>nudging</i> technique . .	39
2.7.2	Advantages and disadvantages of nudging technique in comparison to other three-dimensional assimilation methods	41
3	1D-VAR SET-UP	44
3.1	The fast radiative transfer model RTTOV-7	47
3.2	CMA and CT products and the flag processing	50
3.3	The background error covariance matrix B and the observation error covari- ance matrix R	53
3.4	1D-VAR conceptual performance	55
3.5	The bias correction method	59
3.6	1D-VAR actual performance	60
3.6.1	Statistics of model departures from observations	61
3.6.2	Comparisons with independent observations	63
4	THREE-DIMENSIONAL CASE-STUDIES	68
4.1	A false alarm case study : 8th of July 2004	69
4.1.1	Assimilation	71
4.1.2	Forecast	80
4.2	A heavy precipitation case-study : 9th of April 2005	87
4.2.1	Assimilation	90
4.2.2	Forecast	97
4.3	A possible improvement of the B matrix and its use in the false alarm case-study	102
4.3.1	Calculation of the climatological B matrices	102
4.3.2	Calculation of the Ensemble Islands B matrices	104
4.3.3	Merging of the two B matrices	111
4.3.4	Assimilation	112
4.3.5	Forecast	114
5	CONCLUSIONS	117

List of Symbols

List of the most common symbols used in the thesis. Units are specified between square brackets.

- * A : analysis error covariance matrix
- * B : background error covariance matrix
- * bt_bcor : bias corrected brightness temperature [K]
- * bt_fg : first guess brightness temperature [K]
- * bt_obs : observed brightness temperature [K]
- * bt_ret : retrieved brightness temperature [K]
- * $d_{i,j}$: level to space optical depth from level j for channel i [-]
- * H : forward operator
- * \mathbf{H} : tangent linear of H
- * \mathbf{H}^T : adjoint of H
- * $L(v, \theta)$: upwelling radiance at the top of the atmosphere [W/(m² sr)]
- * J : cost function [-]
- * J_b : background cost function [-]
- * J_o : observational cost function [-]
- * K : gain matrix

- * $P_i(TB)$: modified Planck function for channel i at brightness temperature TB [W/(m² sr)]
- * $P(\nu, T)$: Planck function at frequency ν and temperature T [W/(m² sr)]
- * q : vertical profile of specific humidity [g/kg]
- * q_{2m} : 2 meters specific humidity [g/kg]
- * \mathbf{R} : observational error covariance matrix [K²]
- * SST : sea surface temperature [K]
- * T : vertical profile of temperature [K]
- * TB : brightness temperature [K]
- * TB_i : brightness temperature for channel i [K]
- * T_{2m} : 2 meters temperature [K]
- * T_s : surface temperature [K]
- * \vec{x}_a : analysis field
- * \vec{x}_b : background field
- * \vec{y}_o : observation vector [K]
- * ϵ : emissivity [-]
- * ϵ_s : surface emissivity [-]
- * τ : transmittance [-]
- * τ_s : surface transmittance [-]
- * $\tau_{i,j}$: level to space transmittance from level j for channel i [-]
- * θ : angle from zenith at the surface [rad]
- * ν : frequency [s⁻¹]

Acronyms list

- * 1D-VAR : One dimensional variational analysis
- * 3D-VAR : Three dimensional variational analysis
- * 4D-VAR : Four dimensional variational analysis
- * ARPA-SIM : Agenzia Regionale per la Prevenzione dell'Ambiente – Servizio Idro Meteorologico
- * CMa : Cloud Mask
- * COSMO : COnsortium for Small-scale MOdelling
- * COSMO-LAMI : COSMO Limited Area Model Italy
- * CT : Cloud Table
- * DWD : Deutscher WetterDienst
- * ECMWF : European Center for Medium range Weather Forecast
- * ERA-40 : Ecmwf Re-Analysis
- * EUMETSAT : European Organisation for METeorological SATellite
- * flg_pre : flag processing
- * GERB : Geostationary Earth Radiation Budget
- * GMS : Geostationary Meteorological Satellite
- * GOES : Geostationary Operational Environmental Satellite

- * GOMS : Geostationary Operational Meteorological Satellite
- * LAM : Limited Area Model
- * MSG : Meteosat Second Generation
- * MVIRI : Meteosat Visible and Infrared Resolution Imager
- * NCEP : National Centers for Environmental Prediction
- * NOAA : National Oceanic and Atmospheric Administration
- * NWP : Numerical Weather Prediction
- * OI : Optimum Interpolation
- * *RMS* : Root Mean Square
- * RTTOV : Radiative Transfer model for TOVs
- * SAFNWC : Satellite Application Facilities for NoWCasting
- * SEVIRI : Spinning Enhanced Visible and Infrared Resolution Imager
- * TOA : Top Of Atmosphere
- * TOVS : Television infrared orbiting satellite Operational Vertical Sounder

Abstract

The quality of temperature and humidity retrievals from the infrared SEVIRI sensors on the geostationary Meteosat Second Generation (MSG) satellites is assessed by means of a one dimensional variational algorithm. The study is performed with the aim of improving the spatial and temporal resolution of available observations to feed analysis systems designed for high resolution regional scale numerical weather prediction (NWP) models. The non-hydrostatic forecast model COSMO (COntortium for Small scale MOdelling) in the ARPA-SIM operational configuration is used to provide background fields. Only clear sky observations over sea are processed.

An optimised 1D-VAR set-up comprising of the two water vapour and the three window channels is selected. It maximises the reduction of errors in the model backgrounds while ensuring ease of operational implementation through accurate bias correction procedures and correct radiative transfer simulations.

The 1D-VAR retrieval quality is firstly quantified in relative terms employing statistics to estimate the reduction in the background model errors. Additionally the absolute retrieval accuracy is assessed comparing the analysis with independent radiosonde and satellite observations. The inclusion of satellite data brings a substantial reduction in the warm and dry biases present in the forecast model. Moreover it is shown that the retrieval profiles generated by the 1D-VAR are well correlated with the radiosonde measurements.

Subsequently the 1D-VAR technique is applied to two three-dimensional case-studies: a false alarm case-study occurred in Friuli-Venezia-Giulia on the 8th of July 2004 and a heavy precipitation case occurred in Emilia-Romagna region between 9th and 12th of April

2005. The impact of satellite data for these two events is evaluated in terms of increments in the integrated water vapour and saturation water vapour over the column, in the 2 meters temperature and specific humidity and in the surface temperature.

To improve the 1D-VAR technique a method to calculate flow-dependent model error covariance matrices is also assessed. The approach employs members from an ensemble forecast system generated by perturbing physical parameterisation schemes inside the model. The improved set-up applied to the case of 8th of July 2004 shows a substantial neutral impact.

INTRODUCTION

0.1 The initial condition problem

At the beginning of the meteorological science the Newtonian philosophy was dominant: assumed known the initial state and the physical laws describing the dynamical evolution of the atmosphere, its final state at any future instant could be forecasted. The deterministic approach is still used nowadays, but it is well known that the more carrying on in the forecast time the more there is probability that the outcoming fields are affected by errors. In fact the presence of non-linear terms in the dynamic equations is source of great differences for forecasts produced starting from very little different initial conditions, hence a little difference between the used initial condition and the “truth” determines an erroneous prediction in the atmospheric system evolution and the amplitude of these errors grow in time.

In mathematical terms the problem of doing a forecast is a problem at the initial conditions, which can at least numerically be solved when these three tasks are fulfilled:

- the definition of a network observing system collecting all available meteorological data;
- the definition of an analysis method to build the best approximation of the atmospheric initial state from all available data; this analysis will be the input of the meteorological model;
- the definition of a set of motion equations, i.e. a mathematical translation of the laws of conservation of energy, momentum, and mass together with the state equation of the gases.

The knowledge of the best initial atmospheric state would improve the forecast outcome, hence the increment of data from the network observing system and the production of more accurate analyses can be considered the new challenges of the future.

0.2 Concept of analysis

The analysis can be defined as the best accurate image of the true state of the atmosphere at a given time, produced by means of assimilation techniques, and it can be used as the initial state in a Numerical Weather Prediction (NWP) Model.

The basic information of the assimilation techniques consists in the collection of different kinds of observations. Under the hypothesis that the model is overdetermined by the observations the analysis problem reduces to an interpolation model, nevertheless this never happens in the reality since data are sparse and sometimes only indirectly related to the model variables. To avoid complications and to define a well-posed problem is necessary to introduce a background field, as a priori estimate of the model state. The collection of different kinds of observations together with the background field are taken as input to the assimilation techniques providing as output analysis increments that added to the background field result in the wanted analysis.

The different kinds of assimilation techniques can be divided in two main categories depending on the fact that they use statistical methods or not.

The most important technique where the statistical approach is missing is the Newtonian relaxation or nudging technique (Hoke and Anthes, 1976 [18]). This method involves adding a term to the prognostic model equations that effectively nudges the solution towards observations. The nudging term is time dependent and should be large enough that it has an effect on the solution, but small enough so that it does not dominate over other terms (Stauffer and Seaman, 1990 [41]). In practice the nudging term is often no larger than the magnitude of the least dominant term in the equation. This technique is widely used in Limited Area Model (LAM) where the application of other methods would result problematic.

The nudging technique can be applied only on the prognostic variables of LAMs and this is a very big limitation of this method. Hence the assimilation of remote sensing data, like radiances observed from satellite and reflectivity observed from radar, have to be performed in other ways.

In the statistical approach methods there is a mathematical representation of the uncertainty of the data, both of model data and of observational data. The error statistics are so modelled using probabilistic concepts and the analysis algorithm is defined on the basis that the analysis errors must be minimal in a root mean square sense. This is the core of the main variational assimilation techniques (1D-VAR, 3D-VAR and 4D-VAR) and of the OI (Optimum Interpolation) technique. Their capabilities are so general that they conceptually permit the assimilation of any kind of observation; that is why of their importance.

In particular the 1D-VAR approach is the simplest variational method (with respect to 3D-VAR, where the assimilation is done also in the other two horizontal directions, and to 4D-VAR, where beyond to the three spatial directions the time dimension is taken into account) and is used to assimilate data along a column over a grid point of the model. A very useful example is the satellite data retrieval problem, in which the 1D-VAR algorithm performs a local analysis of one atmospheric column (the model state) at the location of each satellite sounding.

0.3 Satellite data assimilation

The assimilation of satellite derived observations in NWP models makes it continuous progress with the advances in data assimilation techniques and the ever increasing quality and availability of remote observations (Daniels et al., 2004 [7]) . While some limited geostationary data are assimilated in the form of cloud track winds (Schmetz et al., 1993 [36]), up to now their main use in operational NWP models has been the update of radiances and the outcoming of derived products. The poor number of channels available, their broad spectral width and the high signal-to-noise ratio which imposes low spatial resolution has prevented wider

use. Nevertheless, with respect to radiance observations from polar-orbiting satellites, data from geostationary platforms offer higher temporal resolution which is crucial for regional NWP.

New generation geostationary satellites have improved capabilities. The MSG satellites, whose MSG-1 launched in August 2002 is the first prototype, provide measurements with double spatial resolution and temporal coverage compared with its predecessor Meteosat-7. Its 15 minutes repetition cycle and 3 km nominal spatial sampling permits the monitoring of localised phenomena and rapidly evolving weather systems making it a potentially good quality observational system not only for global atmospheric NWP models but also for high resolution LAM models.

Moreover the SEVIRI radiometer on board of MSG satellites measures radiation in 12 spectral channels, with respect to the only 3 channels of Meteosat-7, spanning the visible to the infrared with a much narrower band size. It is clearly desirable to exploit the wealth of radiance information from this new generation of geostationary satellites, particularly with the prospect of even higher spectral-resolution sounders such as the Geostationary Imaging Fourier Transform Spectrometer and the Advanced Baseline Sounder which are planned to be placed in geostationary orbit on GOES-R (Gurka and Smith, 2004 [16])

A possible way to evaluate the usefulness of any new type of observations in global or regional operational analysis system at a reasonable cost is to use a one dimensional variational approach (1D-VAR). To convert model data in forecasted satellite radiances fast radiative transfer models have been designed: they represent the crucial link between model prognostic variables and satellite data. In this case the assimilation is performed in two steps. The first step consists in the production of 1D-VAR retrieved profiles in which 1D-VAR seeks optimal, in a least square sense, model variables (e.g. temperature, humidity, etc.) which fit the observed variables within specified model and observational errors. The second step incorporates the 1D-VAR retrieved products as “pseudo-observations” into the assimilation system of the model itself.

One weakness of this methodology is to introduce correlations between these pseudo-observations and the model state since model variables are used as first guess in the minimisation performed during the 1D-VAR technique. However this technique has already been proven robust for operational use even when employing highly non-linear observation operators (Courtier et al., 1994 [6]) and the “1D-VAR + 4D-VAR” scheme implemented at ECMWF to assimilate rainy-radiances from two microwave sensors SSM/I and TRMM (Marécal and Mahfouf, 2002 [22]) is a demonstration of this fact. Moreover, since the final products of the 1D-VAR algorithm are basic thermo-dynamic/dynamic model variables, they can be introduced in less complex assimilation systems such as the nudging analysis, which do not use variational techniques and thus do not permit direct radiance assimilation.

The aim of this thesis is therefore to study the potential for a 1D-VAR retrieval using the fast radiative transfer model RTTOV-7. The retrieved variables are temperature and humidity profiles and surface variables derived from SEVIRI observations in clear sky conditions for their assimilation into the nudging scheme of the LAM COSMO-LAMI, which ARPA-SIM runs daily to provide weather forecasts for the Italian peninsula.

0.4 Scientific questions to be addressed

The work has the end to provide answers to these scientific questions:

- How to establish the best set of channels which maximise the “information content” brought by satellite data into the analysis and the analysis error reduction?
- How much is the impact of the 1D-VAR technique applied to specific selected case studies in terms of precipitation and surface variables?
- How the background error covariance matrix depends on the region and on the synoptic situation? How an improved construction of this matrix can be performed?

The structure of the thesis is as follows. In the chapter 1 the assimilation methods based on statistical approach are treated putting particular attention on the 1D-VAR technique. Chapter 2 gives a general view of the observational data already used inside the assimilation techniques in the main NWP models, showing the differences between conventional

and non-conventional data; subsequently firstly the new geostationary satellite data, which are used in this study, are presented and secondly information on LAM used in this thesis, COSMO, and its internal assimilation scheme, the nudging, are given. Chapter 3 explains the components of the 1D-VAR technique: the fast radiative transfer model used to simulate the brightness temperatures (section 3.1), the cloud detection method (section 3.2) and the background and observational error covariance matrices calculation (section 3.3); section 3.4 reports the 1D-VAR conceptual performance in terms of analysis error reduction, section 3.5 describes the bias correction method and finally section 3.6 describes the retrieval accuracy of the 1D-VAR technique comparing the analysis with independent radiosondes and satellite observations. In chapter 4 the use of the 1D-VAR method in the COSMO model is assessed for two three-dimensional case-studies, a false alarm case-study and a heavy precipitation case-study; in section 4.3 a possible improvement of the background error covariance matrix used in the 1D-VAR method is discussed. Finally in chapter 5 the conclusions obtained in this work are drawn.

Chapter 1

1D–VAR FRAMEWORK

1.1 Basic concepts of data assimilation

The basic objective information that can be used to produce the analysis is a collection of observed values provided by observations of the true state. If the model state is overdetermined by the observations, then the analysis reduces to an interpolation problem. Anyway in most cases the analysis problem is under-determined because data are sparse and some of them only indirectly related to the model variables. In order to make it a well-posed problem it is necessary to rely on some **background** information in the form of an a priori estimate of the model state. The background information can consist in a climatology or a trivial state; they can consist also in the output of a previous analysis, using some assumption of consistency in time of the model state, like stationarity (hypothesis of persistence) or the evolution predicted by a forecast model. In a well-behaved system one expects that this allows the information to be accumulated in time into the model state, and to propagate to all variables of the model: this is properly the basis concept of data assimilation. In other words the **data assimilation** is an analysis technique in which the observed information is accumulated into the model state by taking advantage of consistency constraints with laws of time evolution and physical properties.

The first step in the mathematical formalisation of the analysis problem is the definition of the work space. As in a forecast model, the collection of numbers needed to represent the atmospheric state of the model is collected as a column matrix called the **state vector** \vec{x} .

How the vector components relate to the real state depends on the choice of discretisation, which is mathematically equivalent to a choice of basis.

It is important to distinguish between reality itself, which is more complex than what can be represented as a state vector itself, and the best possible representation of reality as a state vector, which is usually denoted by \vec{x}_t , the **true state** at the time of the analysis. An other important value of the state vector is \vec{x}_b , the a priori **background** estimate of the true state before the analysis is carried out, valid at the same time. Finally the **analysis** is denoted by \vec{x}_a which is what is wanted.

1.2 Control space variables

In practice it is often convenient not to solve the analysis problem for all components of the model state, or it is necessary to reduce the resolution or domain of analysis because of insufficient computer power. In these cases the work space of the analysis is not the model space, but the space allowed for the corrections to the background, called **control variable space**. Then the analysis problem is to find a correction or **analysis increment** $\delta \vec{x}$ such that

$$\vec{x}_a = \vec{x}_b + \delta \vec{x} \quad (1.1)$$

is as close as possible to \vec{x}_t . So for the resolution of the analysis problem instead to look for \vec{x}_a one can look for $\vec{x}_a - \vec{x}_b$ in a suitable subspace.

1.3 Observations

For a given analysis a number of observed values is used; they are gathered into an **observation vector** \vec{y} . To use them in the analysis procedure it is necessary to be able to compare them with the state vector. It would be nice if each degree of freedom were observed directly, so \vec{y} could be regarded as a particular value of the state vector. In practice there are fewer observations than variables in the model and they are usually irregularly disposed, so that the only correct way to compare observations with the state vector is through the use of a

function from the model state space to the observation space called an observation operator that is denoted by H . This operator generates the values $H(\vec{x})$ that the observations would take if both they and the state vector were perfect, in the absence of any modelling error. In practice H is a collection of interpolation operators from the model grid to the observation points, and then conversion from model variables to the observed parameters. The key point in data analysis is the use of the discrepancies between observations and state vector; this is given by the vector of departures at the observation points:

$$\vec{y} - H(\vec{x}) \tag{1.2}$$

When calculated with the background \vec{x}_b these differences are called innovations, while with the analysis \vec{x}_a analysis residuals. Their study provides important information about the quality of the assimilation procedure.

1.4 The modelling of errors

To represent the fact that there is some uncertainty in the background, in the observations and in the analysis some model of the errors are assumed between these vectors and their true counterparts. The correct way to do this is to assume some **probability density function** (pdf) for each kind of error.

Given a background field \vec{x}_b just before doing an analysis, there is one and only one vector of errors that separates the background field from the true state:

$$\vec{\epsilon}_b = \vec{x}_b - \vec{x}_t \tag{1.3}$$

If it were possible to repeat each analysis experiment a large number of times, under exactly the same conditions, but with different realisations of errors generated by unknown causes, $\vec{\epsilon}_b$ would be different each time. Statistics such as averages, variances and histograms of frequencies of $\vec{\epsilon}_b$ can be calculated. In the limit of a very large number of realisations it is expected that the statistics converge to values which depend only on the physical processes responsible for the errors, not on any particular realisation of these errors. When another analysis will be done under the same conditions, it is not expected to know the error $\vec{\epsilon}_b$ but

at least its statistics. The best information about the distribution of $\vec{\epsilon}_b$ is given by the limit of the histogram when the classes are infinitely small, which is a scalar function of integral 1 called the probability density function of $\vec{\epsilon}_b$.

The errors in the background and in the observations are modelled as follows:

- **background errors:** $\vec{\epsilon}_b = \vec{x}_b - \vec{x}_t$, of average $\overline{\vec{\epsilon}_b}$ and covariances $\mathbf{B} = \overline{(\vec{\epsilon}_b - \overline{\vec{\epsilon}_b})(\vec{\epsilon}_b - \overline{\vec{\epsilon}_b})^T}$. They are the estimation errors of the background state, i.e. the difference between the background state value and the “true” value. They do not include discretisation errors.
- **observation errors:** $\vec{\epsilon}_o = \vec{y} - H(\vec{x})$ of average $\overline{\vec{\epsilon}_o}$ and covariances $\mathbf{R} = \overline{(\vec{\epsilon}_o - \overline{\vec{\epsilon}_o})(\vec{\epsilon}_o - \overline{\vec{\epsilon}_o})^T}$. They contain errors in the observation process (instrumental errors, due to the instrumental noise and to the limited scale resolution of the instrument itself) errors in the design of the operator H and representativeness errors i.e. discretisation errors which prevent \vec{x}_t from being a perfect image of the true state.
- **analysis errors:** $\vec{\epsilon}_a = \vec{x}_a - \vec{x}_t$, of average $\overline{\vec{\epsilon}_a}$. A measure $\|\overline{\vec{\epsilon}_a} - \overline{\vec{\epsilon}_a}\|$ of these errors is given by the trace of the analysis error covariance matrix \mathbf{A} ,

$$Tr(\mathbf{A}) = \|\overline{\vec{\epsilon}_a} - \overline{\vec{\epsilon}_a}\|^2 \quad (1.4)$$

They are the estimation errors of the analysis state which must be minimised. The averages of errors are called **biases** and they are the sign of a systematic problem in the assimilation system: a model drift, or a bias in the observations, or a systematic error in the way they are used.

1.5 Statistical interpolation with least–squares estimation

The fundamental equation for linear analysis in a general algebraic form is the least squares estimation, also called Best Linear Unbiased Estimator (BLUE). A brief explanation of this tool used nowadays in meteorological and oceanographic sciences is reported here.

1.5.1 Notation

The dimension of the model state is n and the dimension of the observation vector is p . Let denote

\vec{x}_t true model state of dimension n

\vec{x}_b background model state of dimension n

\vec{x}_a analysis model state of dimension n

\vec{y} vector of observations of dimension p

H observation operator from dimension n to p

B covariance matrix of the background errors $\vec{x}_b - \vec{x}_t$ of dimension $n \times n$

R covariance matrix of the observation errors $\vec{y} - H[\vec{x}_t]$ of dimension $p \times p$

A covariance matrix of the analysis errors $\vec{x}_a - \vec{x}_t$ of dimension $n \times n$

1.5.2 Hypotheses

The following hypotheses are assumed:

- **Linearized observation operator:** the variations of the observation operator in the vicinity of the background state are linear, i.e. for any \vec{x} close enough to \vec{x}_b , $H(\vec{x}) - H(\vec{x}_b) = \mathbf{H}(\vec{x} - \vec{x}_b)$, where \mathbf{H} is a linear operator;
- **Non-trivial errors:** B and R are positive definite matrices;
- **Unbiased errors:** the expectation of the background and observation errors is zero i.e. $\overline{\vec{x}_b - \vec{x}_t} = \overline{\vec{y} - H(\vec{x}_t)} = 0$
- **Linear analysis:** the aim is to find an analysis defined by corrections to the background which depends linearly on background observation departures.
- **Optimal analysis:** the analysis is considered that state closest to the “true” state in a root mean square sense, i.e. it is a minimum variance estimate.

1.5.3 BLUE theorem

Using the definitions and the hypotheses above explained the following theorem **least-squares analysis estimation** can be derived (Daley, 1991 [8]):

- a The *optimal least-squares estimator* or *BLUE analysis*, is defined by the following interpolation equations:

$$\begin{cases} \vec{x}_a = \vec{x}_b + \mathbf{K}(\vec{y} - H[\vec{x}_b]) \\ \mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \end{cases} \quad (1.5)$$

where the linear operator \mathbf{K} is called the **gain** or **weight matrix** of the analysis.

- b The analysis error covariance matrix is for any \mathbf{K} :

$$\mathbf{A} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}(\mathbf{I} - \mathbf{K}\mathbf{H})^T + \mathbf{K}\mathbf{R}\mathbf{K}^T \quad (1.6)$$

and if \mathbf{K} is the optimal least-squares gain, the expression becomes

$$\mathbf{A} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B} \quad (1.7)$$

- c The BLUE analysis is equivalently obtained as a solution to the **variational optimisation** problem:

$$\begin{cases} \vec{x}_a = \mathbf{Arg} \min J \\ J(\vec{x}) = (\vec{x} - \vec{x}_b)^T \mathbf{B}^{-1} (\vec{x} - \vec{x}_b) + (\vec{y} - H[\vec{x}])^T \mathbf{R}^{-1} (\vec{y} - H[\vec{x}]) \end{cases} \quad (1.8)$$

where J is called the *cost function* of the analysis, the first term on the left is called *background term* J_b and the second term on the left *observation term* J_o .

- d The analysis \vec{x}_a is *optimal*, i.e. it is closest in an r.m.s. sense to the true state \vec{x}_t
- e If the background and the observation error pdfs are Gaussian then \vec{x}_a is also the *maximum likelihood estimator* of \vec{x}_t .

1.6 The one-dimensional variational analysis

The advantage of the variational analysis methods (4D-VAR, 3D-VAR and 1D-VAR) is to avoid the computation of the gain matrix \mathbf{K} completely by looking for the analysis as an appropriate solution to the equivalent minimisation problem defined by the cost function J in the equation (1.8). The solution is sought iteratively by performing several evaluations of the cost function J and of its gradient in order to approach the minimum using a suitable descent algorithm; in particular for the 1D-VAR technique the gradient is :

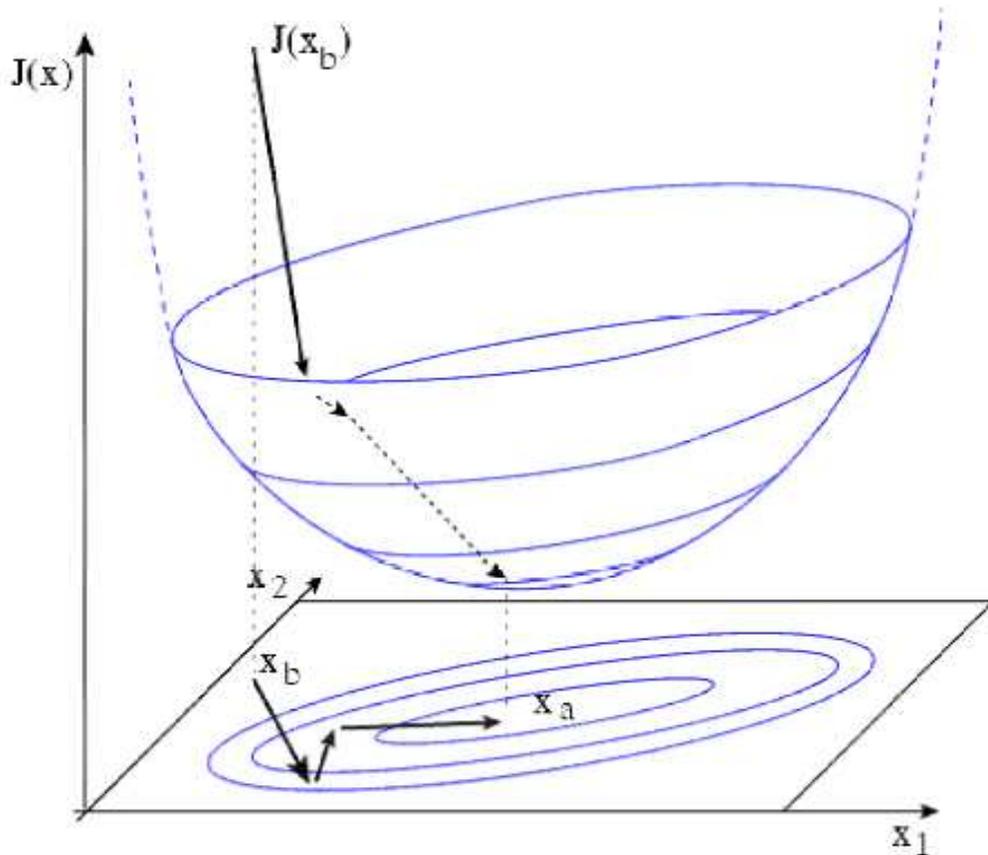


Figure 1.1: Schematic representation of the variational cost-function minimisation (here for simplicity in a two-variable model space) : the quadratic cost function has the shape of a paraboloid with the minimum at the optimal analysis \vec{x}_a . The minimisation works by performing several line-searches to move the control variables \vec{x} to areas where the cost-function is smaller, usually by looking at the local slope (the gradient) of the cost-function.

$$\nabla J = 2\mathbf{B}^{-1}(\vec{x} - \vec{x}_t) - 2\mathbf{H}^T\mathbf{R}^{-1}(\vec{y} - H[\vec{x}]) \quad (1.9)$$

The approximation lies in the fact that only a small number of iterations are performed. The minimisation can be stopped by limiting artificially the number of iterations, or by requiring that the norm of the gradient $\|\nabla J(\vec{x})\|$ decreases by a predefined amount during the minimisation, which is an intrinsic measure of how much the analysis is closer to the optimum than the initial point of the minimisation. The geometry of the minimisation is suggested in Fig. 1.1 (F. Bouttier and P. Courtier, 1999 [1]).

Usually the initial point of the minimisation, or **first guess**, is taken equal to the background \vec{x}_b . This is not compulsory, however, so it is important to distinguish clearly between the terms **background**, which is used in the definition of the cost function, and **first guess**, which is used to initiate the minimisation procedure. If the minimisation is satisfactory, the analysis will not depend significantly on the choice of first guess, but it will always be sensitive to the background.

Chapter 2

OBSERVATIONAL DATA AND MODEL DATA

Considering the types of available observations we can distinguish them in two main categories:

- conventional observations, i.e. in situ observations;
- non-conventional observations, i.e. remote sensing observations;

2.1 Conventional data

There are a lot of different types of conventional data used in meteorological assimilation systems. Here the most important are recalled:

- SYNOP data, i.e. conventional surface weather station reports, comprehending pressure, humidity, and 10 meter wind observations. They have usually a temporal resolution of 3 hours and a spatial resolution of some ten of kilometers;
- SHIP data, i.e. conventional weather reports from sounding ships, and DRIBU data, i.e. conventional weather reports from drifting buoys, comprehending pressure and wind observations;
- TEMP data, i.e. upper air data from radiosondes, comprehending temperature, wind and humidity data. The radiosondes are usually launched twice a day, at 00 UTC and at 12 UTC, and their horizontal resolution is some hundred of kilometers;

- PILOT data, i.e. wind measurements in the free atmosphere from station launching balloons;
- PROFILERS, measuring wind with remote sensing procedures, provide wind speed and wind direction observations at very high temporal resolution.
- AIREP data (manual aircraft reports), AMDAR (Aircraft Meteorological Data Relay) and ACARS (automatic aircraft reports) provide temperature and wind observations. The AMDAR and ACARS systems usually provide more information than AIREP; furthermore during landing and take-off the ACARS provide data in quantity, quality and location comparable to radiosondes.

2.2 Non-conventional data

Basically there are two types of non-conventional data : satellite data and radar data.

During the last twenty years there has been a significant increase in the quantity, quality and different types of satellite observations. Radiances observed by satellite instruments give indirect information on temperature and humidity in the atmosphere.

TIROS1 (Television InfraRed Observation Satellites), launched by United States the 1st April 1960, has been the first meteorological satellite used for meteorological applications. Even if this date can be considered the beginning of the use of satellite data in the meteorological science, only in the 70's and more in the 80's the meteorologists begun to realize the great potentiality of satellite data.

In fact satellite observations can cover areas which are inaccessible to conventional types of observations; furthermore satellite observations usually have a higher resolution both temporally and spatially than the conventional observations. Hence these more detailed observations, if opportunely used, can have a great positive impact on the definition of the initial condition, which is the mandatory step from which the forecast begins.

Radar data, with respect to satellite data, are used for meteorological applications from longer time. The radars, observing the interactions between electromagnetic radiations and the cloud microphysics, can give information about meteorological events for a circle area centered in the radar location for a radius ranging around to some hundred of kilometers. Their temporal resolution is comparable to satellite data, but their coverage is very much smaller.

In this thesis the main aim is to understand how much the use of satellite data can improve the forecasts in Numerical Weather Prediction (NWP) models, so in the next sections general information on the satellites is recalled.

2.3 Polar and geostationary satellites

Basically two types of meteorological satellites exist: geostationary satellites and polar satellites.

- A polar orbiting satellite circles the earth at a near polar inclination, meaning that it always passes almost exactly above the poles. The satellite passes the Equator and each latitude at the same local solar time each day, meaning that the satellite passes overhead at essentially the same solar time throughout all seasons of the year. The polar satellite orbit is much closer to the earth than a geostationary satellite orbit. Its altitude usually ranges from 700 to 800 km with orbital periods of 98 to 102 minutes, and thus can see a smaller part of the earth below than a geostationary satellite, but in a finer detail;
- The geostationary satellites circle the earth in a geosynchronous orbit which means they orbit the equatorial plane of the earth at a speed matching the earth's rotation. This allows to them to hover continuously over one position on the surface. The geosynchronous plane orbit is about 35800 km in altitude so they are high enough to take a full disk view of the earth.

The two types of weather satellite, polar and geostationary, can be seen as complementary. Each type has advantages and disadvantages, and an ideal observing system should combine

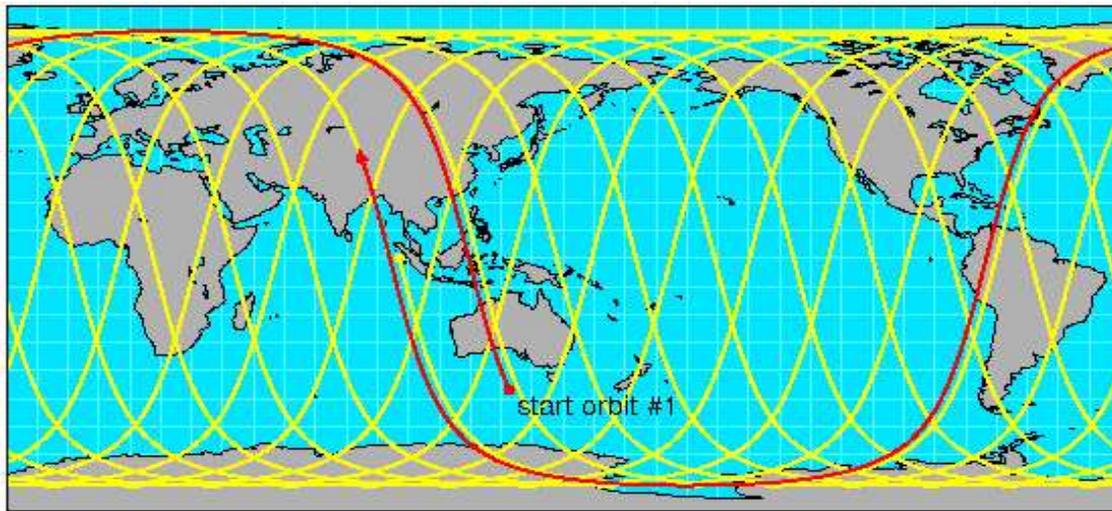


Figure 2.1: Illustration of the daily coverage of a polar satellite in sun-synchronous orbit.

both these elements.



Figure 2.2: Geographical area covered by geostationary satellites positioned at 0° longitude.

The polar satellites permit a global coverage of the earth in a given window of time, i.e. they register information from every point of the earth but not continuously in time. Each point on the earth's surface is observed at best every orbit for polar regions and at worst twice a day for equatorial regions. One of the positive advantages of the polar satellites is the good ground resolution because of their low orbit (see Fig. 2.1). The geostationary satellites give information of the crossed area of the earth, the disk, but in this case the observation of

each point in the disk is continuous in time with a capability of sampling which can reach few minutes. Furthermore only one ground station is necessary for satellite monitoring. The disadvantage of the geostationary satellites consists in the lacking of observations in the polar regions and in a lower spatial resolution, due to the height of the orbit, with respect to the polar satellites (see Fig. 2.2).

2.4 The global system of operational meteorological satellites

The global system of operational meteorological satellites was organized at the beginning of 90's and includes selected geostationary and polar orbit satellites:

- **GEOSTATIONARY SATELLITES**
 - Two NOAA's (National Oceanic and Atmospheric Administration's) geostationary satellites, called GOES (Geostationary Operational Environmental Satellites) are used together over America. One of these, GOES–East, is positioned at 75°W and monitors North and South America and most of the Atlantic Ocean, while the latter, GOES–West, is positioned at 135°W and monitors North America and the Pacific Ocean basin;
 - the Japanese geostationary satellite GMS (Geostationary Meteorological Satellite) positioned at 140°E ;
 - the Russian geostationary satellite GOMS (Geostationary Operational Meteorological Satellite)/Elecktro positioned at 76°E ;
 - the EUMETSAT (EUropean organisation for the exploitation of METeorological SATellites)'s geostationary satellite, METEOSAT, positioned at 0° that monitors Europe, the Mediterranean basin and Africa.

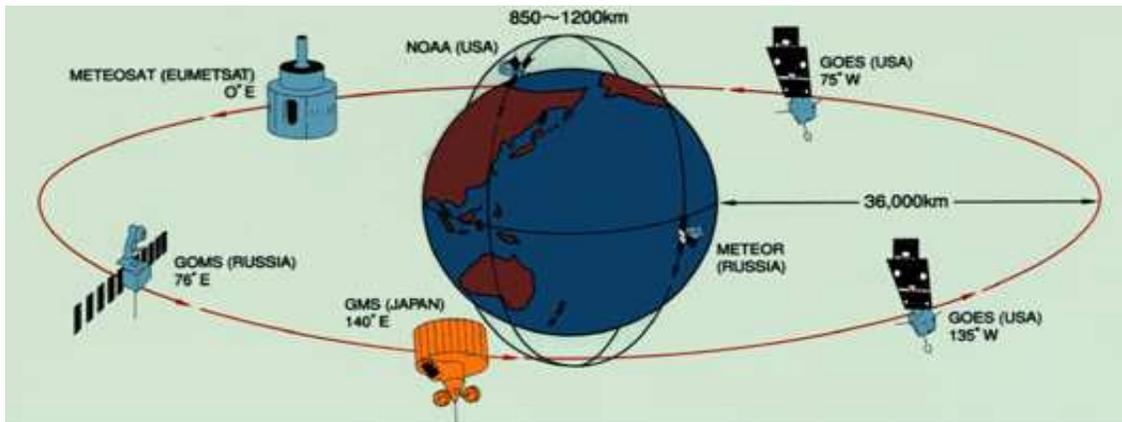


Figure 2.3: The global system of meteorological satellites.

Furthermore also Indian series of communication/meteorological geostationary satellites, INSAT, positioned at 74°E has operated for many years but unfortunately the data/products have not generally been made available to the global meteorological community.

- POLAR SATELLITES

- Two NOAA’s polar satellites, called simply NOAA, and one Russian polar satellite, called METEOR.

Fig. 2.3 shows the global operational satellites.

2.5 The Meteosat series

Meteosat series refers to whichever geostationary satellites has been or is currently operational over Europe, Mediterranean basin and Africa at an altitude of about 35800 km. EUMETSAT operates the Meteosat series of geostationary satellites which have served the meteorological community since 1977. The table 2.1 shows the Meteosat missions and their duration from Meteosat-1 to nowadays.

Meteosat satellite series		
Satellite	Launch date	Mission duration (years)
<i>Meteosat-1</i>	23/11/1977	8
<i>Meteosat-2</i>	19/06/1981	10
<i>Meteosat-3</i>	15/06/1988	7
<i>Meteosat-4</i>	06/03/1989	7
<i>Meteosat-5</i>	02/03/1991	ongoing
<i>Meteosat-6</i>	20/11/1993	ongoing
<i>Meteosat-7</i>	02/09/1997	ongoing
<i>MSG-1</i> → <i>Meteosat-8</i>	28/08/2002	ongoing
<i>MSG-2</i> — <i>Meteosat-9</i>	22/12/2005	ongoing

Table 2.1: Meteosat satellites missions since 1977, year of the launch of Meteosat-1, until today. The symbol → means that MSG-1 has been renamed Meteosat-8 when it became operational on February 2004. MSG-2 is now in the second part of the commissioning period, that is why of the symbol —, and will be renamed Meteosat-9 after the commissioning period itself. Furthermore the launches of MSG-3 and MSG-4 are planned for 2008 and 2012 respectively.

2.6 Radiometric instrument specifications

MVIRI (Meteosat Visible and Infrared Resolution Imager), a high resolution radiometer with three spectral bands, was the main payload of Meteosat-7 and its predecessors and it provided information on only three channels with a temporal resolution of 30 minutes. The new generation of Metosat geostationary, Meteosat Second Generation (MSG), satellites have two different instruments on board:

- the meteorological instrument, that is the main payload, SEVIRI (Spinning Enhanced Visible and Infrared Resolution Imager)
- the climate research instrument, GERB (Geostationary Earth Radiation Budget)

2.6.1 SEVIRI

SEVIRI is the optical imaging radiometer on board the MSG satellites. Its twelve different spectral channels provide twenty times more information than the precedent METEOSAT

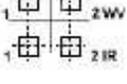
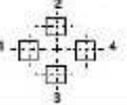
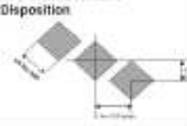
MOP / MSG PERFORMANCE EVOLUTION			
	1st Generation	2nd Generation	
Imaging Format			
Imaging Cycle	30 min	15 min	
Channels	Wavelength		
	Visible	0.5-0.9	HRV VIS 0.6 VIS 0.8 IR 1.6
	Water Vapour	WV 6.4	WV 6.2 WV 7.3
	IR Window	IR 11.5	IR 3.9 IR 8.7 IR 10.8 IR 12.0
	Air mass Analysis		IR 9.7 + WV IR 13.4
Sampling Distance	2.25 km (visible) 4.5 km (IR + WV)	1 km (HRV) 3 km (others)	
Pixel Size (MTP Square-shaped compared to MSG diamond-shaped)	2.25 km (visible)  5 km (IR + WV) 	 	
Number of Detectors	4	42	
Telescope Diameter	400 mm	500 mm	
Scan Principle	Scanning telescope	Scan mirror	

Figure 2.4: Comparison between MVIRI and SEVIRI performances

satellite (Meteosat-7), offering new capabilities in cloud imaging and tracking, fog detection, measurements of earth surfaces and cloud top temperatures, tracking ozone patterns, as well as many other improved performances. A new image is provided every 15 minutes instead of every 30 minutes (see Fig. 2.4). This, together with the enhanced imagery, results in an important increase in capabilities for monitoring weather patterns over Atlantic Ocean, Europe and Africa and for the prediction of severe storms and other potentially hazardous phenomena like hurricanes.

SEVIRI comprehends twelve imaging channels and for each of these it provides the observed radiance or the equivalent brightness temperature (bt_obs). There are two visible channels, respectively at 0.6 and 0.8 μm , which are essential for cloud detection, cloud tracking, scene identification and the monitoring of land surfaces and aerosol. The only near

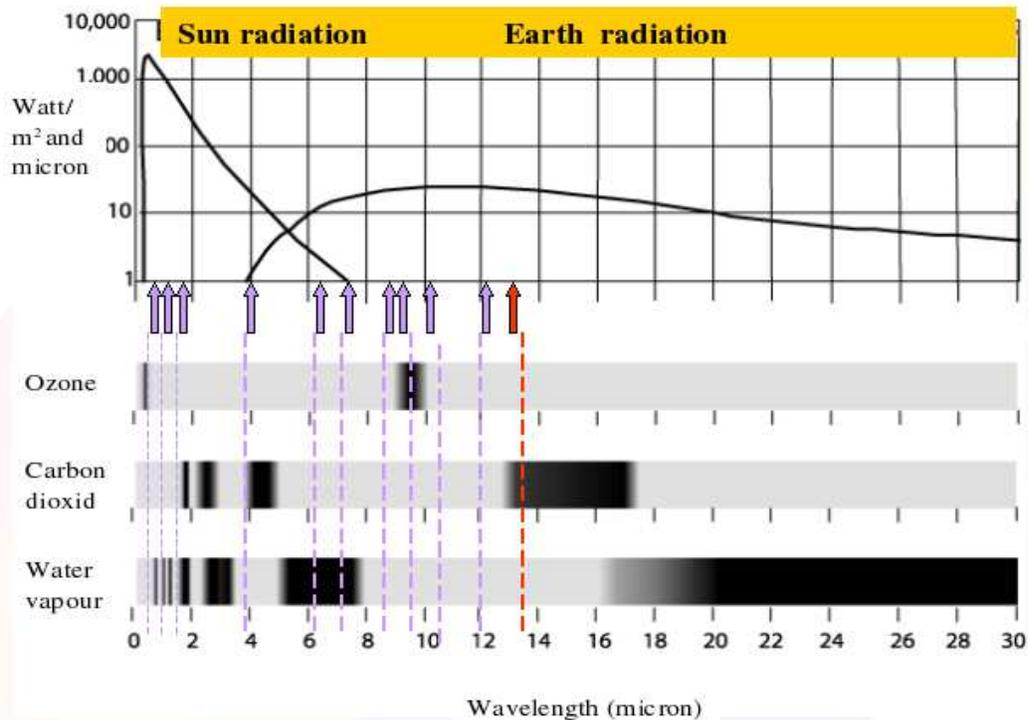


Figure 2.5: Absorption bands of water vapour, carbone dioxide and ozone in the spectrum covered by SEVIRI instrument

infrared channel is the one at $1.6 \mu\text{m}$, useful to discriminate between ice and water clouds and between snow and cloud. The water vapour in the atmosphere is monitored by two channels, at 6.2 and $7.3 \mu\text{m}$ respectively, which support height assignment for semi-transparent clouds.

Apart from the High Resolution Visible (HRV) channel, obtained integrating the information coming from the precedent two visible channels, the remaining channels fall in the infrared spectrum. Three of these infrared channels, at 8.7 , 10.8 and $12.0 \mu\text{m}$, are in the infrared window and give information on land and sea surface temperature, another one is at $3.9 \mu\text{m}$ and is primarily used for detection of low cloud and fog at night and furthermore is quite important for the detection of forest fires. Finally the other two infrared channels, at 9.7 and $13.4 \mu\text{m}$, are localised in the middle of the main O_3 absorption band and inside the main CO_2 absorption band respectively (see Fig. 2.5).

The operating principle of SEVIRI instrument is as follows: a series of mirrors within the radiometer direct the radiation originating from the earth surface, its atmosphere and

CHANNEL	SPECTRAL BAND IN μm		
	λ_{min}	λ_{cen}	λ_{max}
VIS 0.6	0.56	0.63	0.71
VIS 0.8	0.74	0.81	0.88
NIR 1.6	1.50	1.64	1.78
IR 3.9	3.48	3.90	4.36
WV 6.2	5.35	6.25	7.15
WV 7.3	6.85	7.35	7.85
IR 8.7	8.30	8.70	9.10
IR 9.7	9.38	9.66	9.94
IR 10.8	9.80	10.80	11.80
IR 12.0	11.00	12.00	13.00
IR 13.4	12.40	13.40	14.40

Figure 2.6: Spectral bands of SEVIRI channels

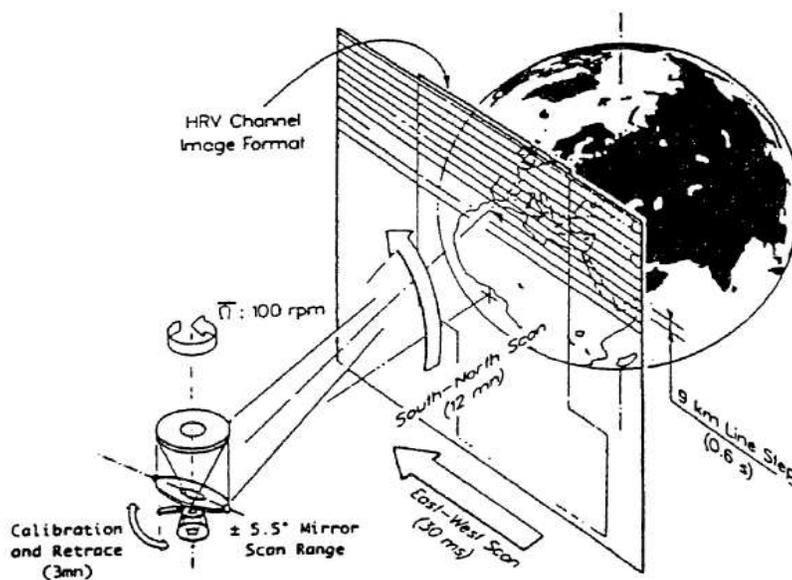


Figure 2.7: Image acquisition by the SEVIRI radiometer.

cloud systems, received via an opening in the side of the satellite, onto an array detectors. Readings are taken from the detectors approximately every 24 microseconds as the satellite spins, so that the spin is used to scan the earth in the East-West direction. After every scan line a mirror is stepped in the South-North direction in order to acquire subsequent scan lines.

One complete revolution of the satellite lasts 0.6 seconds, of which only about 30 milliseconds are available over the earth disk to acquire one scan. For each scan step several image lines are acquired (3 lines for nominal channels and 9 lines for the HRV channel). The remaining 570 milliseconds are used mainly for scan mirror stepping, data transmission and measurements directed at deep space, used for removal of noise from the data (see Fig. 2.7). The nominal repeat cycle for a complete scan of the full earth disk is 15 minutes, this includes measurement of on-board calibration sources and scan mirror retrace.

The nominal image size for all channels (Level 1.5 image) except for the High Resolution Visible (HRV) is 3712 by 3712 pixels (N-S by E-W), the sampling distance is defined to be exactly 3km by 3km at the sub-satellite point. For the HRV channel the image size is 11136 by 5568 pixels (N-S by E-W), the sampling distance defined to be exactly 1km by 1km at the sub-satellite point.

The scans of data are taken at constant angular steps, and this together with the natural curvature of the earth means that as the satellite scans away from the sub-satellite point the area covered by the pixels is greater than at the sub-satellite point. For example at mid latitudes the distance between two consecutive points for the nominal channels is about 5 km (more details can be found in the EUMETSAT web site www.eumetsat.de).

2.6.2 GERB

The GERB instrument is a broadband scanning radiometer used to derive the thermal radiation emitted by the earth. The Earth Radiation Budget is the balance between the radiation coming from the Sun and the reflected and scattered solar radiation, plus thermal infrared emissions, from earth to space. GERB measures the short wave and total radiation from the earth about every 6 minutes. Long-wave (thermal) radiation is also obtained, by subtraction of the two measurements.

At the core of the GERB instrument is a broadband, three-mirror telescope housed in the Instrument Optical Unit. The instrument's overall mass is 25 kg and its power consumption

is nominally less than 32 W (end of life) during operations.

2.7 The Limited Area Model COSMO

The regional forecast model used in this thesis is COSMO, a fully-compressible (non-hydrostatic) primitive equation model without any scale approximation, developed by the COSMO Consortium (COntortium for Small-scale MOdelling), which coordinates the cooperation of Germany, Italy, Switzerland, Greece and Poland (details can be found in the COSMO web site www.cosmo-model.org). Due to the unfiltered set of equations, the vertical momentum equation is not approximate, allowing a better description of non-hydrostatic phenomena such as moist convection, breezes circulations and some kinds of mountain-induced waves. Furthermore it includes an explicit Eulerian horizontal advection scheme and a full set of physical parameterisations for large scale condensation (Steppeler et al., 2003 [42]), convection (Tiedtke, 1989 [44]), radiation (Ritter and Geleyn, 1992 [30]) In this sense COSMO has been essentially designed for meso- β and meso- γ scales where non-hydrostatic effects begin to play an essential role in the evolution of atmospheric flows.

The Italian implementation of COSMO, called COSMO-LAMI (COSMO – Limited Area Model Italy) and managed by ARPA-SIM in the framework of an agreement among UGM (Ufficio Generale di Meteorologia), ARPA-SIM and ARPA-Piemonte, consists of two runs a day (at 00 and 12 UTC) for 72 hours with a spatial horizontal resolution of 7 km and 40 levels in the vertical between the surface and the top of the troposphere at around 30 hPa; the horizontal model domain is set to cover Italy, the major part of the Mediterranean sea and the Alpine region (see Fig. 2.8 for the operational domain specification). Hourly boundary conditions are provided by the DWD global model GME of the German Weather Service (Deutscher WetterDienst DWD) (Majewski et al, 2002 [20]) COSMO assimilation system is based on a nudging scheme (Schraff and Hess, 2003 [37]) in which Newtonian relaxation of prognostic variables is performed towards observed values with weighting functions which are function both of space and time coordinates.

All the NWP models use assimilation techniques to feed the available observations; anyway only the use of high resolution models can permit to assimilate high resolution satellite data, such as MSG radiances, with a remarkable improvement of the forecasts. Hence, detailed high-resolution analyses have to be able to produce frequently and quickly and this should be done thorough use of asynoptic and high-frequency observations such as aircraft and remote sensing data. Up to now the observations used in the COSMO assimilation technique, called nudging, were only conventional data:

- radiosonde observations including wind, temperature and humidity;
- surface level data from SYNOP, SHIP and BUOY stations providing pressure, wind and humidity observations;

2.7.1 Assimilation of observations in COSMO: the *nudging* technique

The nudging (or Newtonian relaxation) technique consists of relaxing the model's prognostic variables towards prescribed values within a predetermined time window. In the scheme used inside COSMO nudging is performed towards direct observations, which seems to be better than nudging towards three dimensional analyses (Schraff and Hess, 2003 [37]).

The basic equation of nudging consists of introducing a relaxation term into the prognostic variable $\psi(\vec{x}, t)$ and is given by:

$$\frac{\partial \psi(\vec{x}, t)}{\partial t} = F(\psi, \vec{x}, t) + G_\psi \sum_{k(obs)} W_k(\vec{x}, t) [\psi_{k_{obs}} - \psi(\vec{x}_k, t)] \quad (2.1)$$

where F denotes the model dynamics and physical parameterisations, ψ_k^{obs} the value of the k^{th} observation influencing the grid point \vec{x} at time t , \vec{x}_k the observation location, G_ψ a constant called nudging coefficient and W_k is an observation-dependent weight whose value is always between 0 and 1 (except for surface pressure). The difference in the square bracket between observed and model value is called observation increment, and the complete additional so-called nudging term determines the analysis increment which is defined as the change explicitly imposed on the model value by the nudging. Setting W_k equal to 1 and neglecting dynamics and physics contributions, the model value at the observation location

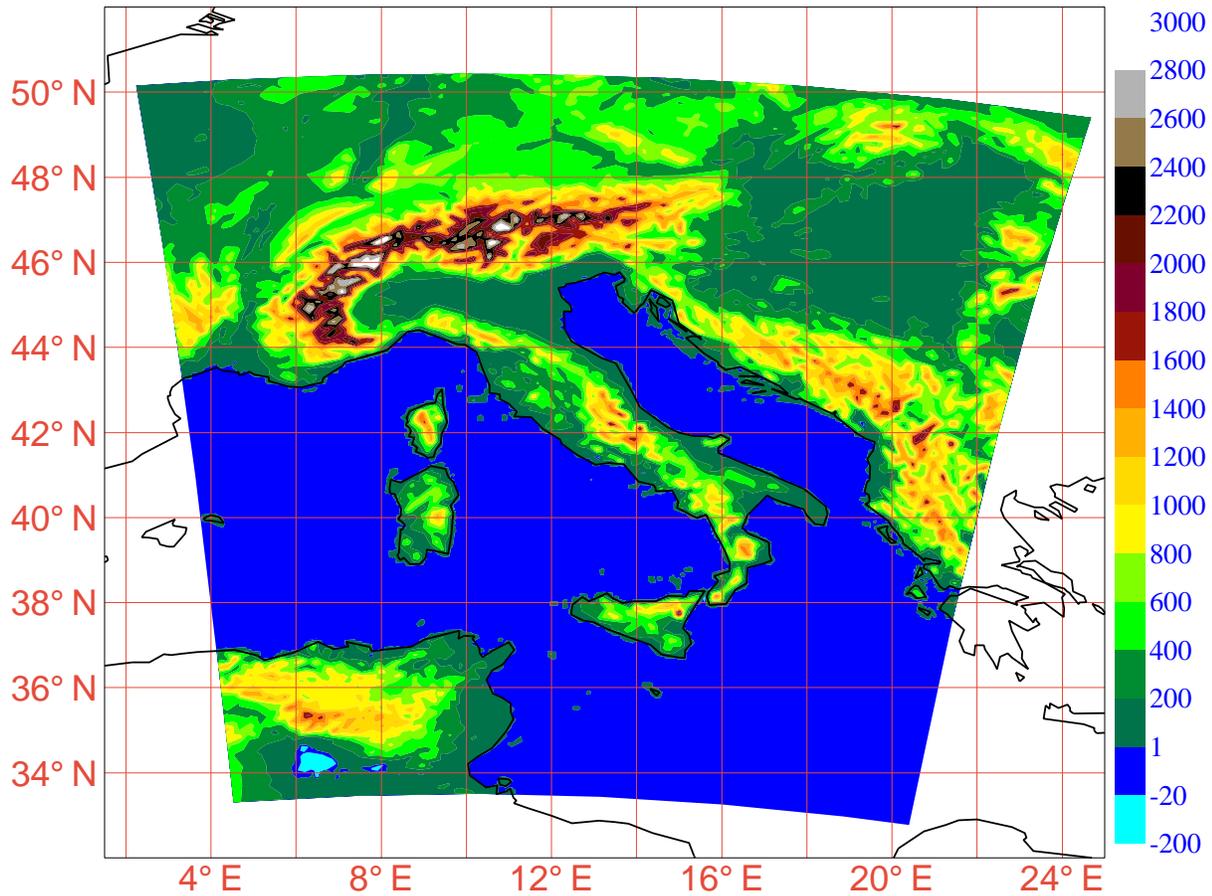


Figure 2.8: COSMO-LAMI operational domain and its correspondent orography in meters

relaxes exponentially towards the observed value with an e-folding decay rate of $1/G_\psi$, corresponding to about half an hour.

Supposing that the model fields have to be relaxed towards the observed values without significantly perturbing the dynamic balance of the model, the nudging term should remain smaller than the largest term of the dynamics (see Fig. 2.9) (Doms and Schattler, 1999 [9]).

The factors W_k take into account the weights w_k given to the different observations influencing the nudging term for a specific grid point. These weights w_k in their turn take into account the quality and representativeness of the observation ϵ_k , and the temporal distance w_t , the horizontal distances w_{xy} and the vertical distance w_z between the observation and the target grid point. In mathematical equation terms this is translated in this way:

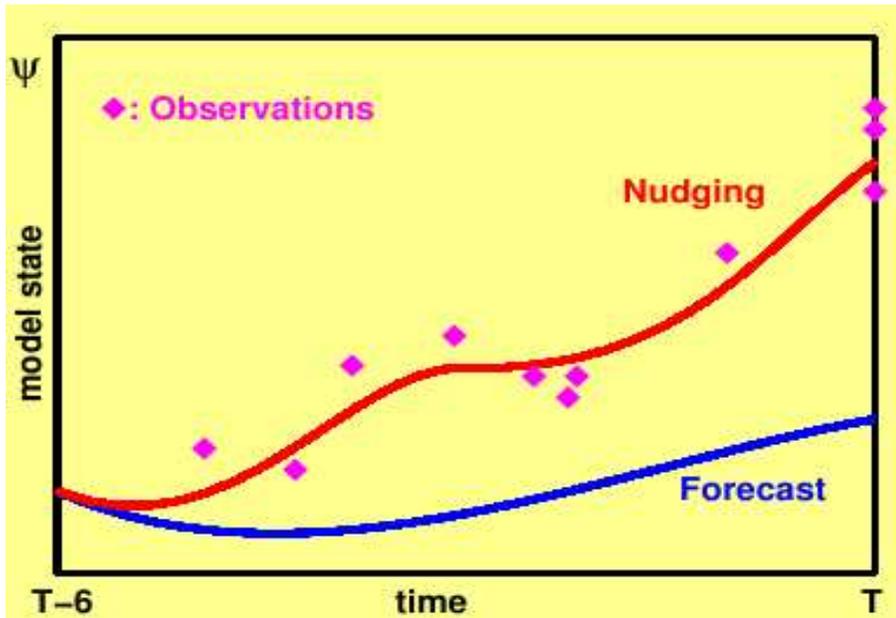


Figure 2.9: Conceptual illustration of the concept of nudging

$$W_k = \frac{w_k}{\sum_j w_j} w_k \quad (2.2)$$

$$w_k = w_t w_{xy} w_z \epsilon_k \quad (2.3)$$

2.7.2 Advantages and disadvantages of nudging technique in comparison to other three-dimensional assimilation methods

It is worth noting that the nudging approach only takes into account the data density at the target grid points and neglects the relative positions between the observation themselves in contrary to optimum interpolation (OI).

Furthermore in contrast to OI and multi-dimensional variational methods (3D-VAR, 4D-VAR) there is no mathematic formalism to determine a theoretically optimal solution to the analysis problem. Hence there are several free parameters and theoretical considerations can only provide rough estimations for their optimal specification. The appropriate values can only be determined by means of physical reasoning and tuning experiments.

Another problem in comparison to 3D-VAR and 4D-VAR is related to the fact that, similarly to OI, observation increments have to be expressed in model space rather than observation space. This means that for an observational information of any kind, observation increments have to be derived always in terms of the prognostic model variables in order to be used in the nudging. That limitation is not present in the variational schemes.

For example in order to assimilate radar reflectivity observations in a variational scheme, reflectivity values should be obtained from the model fields in order to compute observation increments of reflectivity which would be used for the analysis; instead in the nudging method increments of temperature, humidity, wind, etc. have to be deduced from the observed reflectivity. Nevertheless this operation comprises more degrees of freedom, requires more assumptions with less confidence and is therefore more prone to errors.

Nevertheless these disadvantages in the nudging technique with respect to OI, 3D-VAR and 4D-VAR, it allows to use the observations in the best way in COSMO. In fact OI and 3D-VAR do not allow to take into account the exact observation time of asynoptic data and they make it necessary to neglect most of the high-frequency data unless the analysis scheme is applied very frequently; even if OI or 3D-VAR were applied with a high temporal frequency this would increase the computational cost and could result in problem at asynoptic analysis times when data density may become very low and inhomogeneous. Moreover the geostrophic approximation, usually a key ingredient of such schemes, is of limited validity in the meso-scale.

Four-dimensional methods offer potential advantages since they include the model dynamics in the assimilation process directly. As variational methods allow to compare observations with the model state in observation space, they appear in principle best suited for the use of many types of remote sensing data. This would be right, but the 4D-VAR is too expensive for operational application in COSMO considering the small amount of time available to produce the analyses; in fact the calculations of tangent linear and of the adjoint of the COSMO model would require very big human resources, furthermore up to now these two products, the tangent linear and the adjoint of the model, are not available and this avoids

this approach at the basis.

These are the reasons why the nudging technique is operationally used as assimilation approach inside COSMO.

Chapter 3

1D–VAR SET–UP

In our approach to assimilate satellite data two steps are necessary: as first step the 1D–VAR algorithm is used to convert satellite data in vertical profiles of temperature and humidity and in other surface variables which together represent the control vector, while in the second step these profiles are included in the data to be nudged in the the COSMO model.

In this study the control vector \vec{x} contains vertical profiles of temperature and specific humidity, $2m$ temperature, $2m$ specific humidity and sea surface temperature (i.e. $\vec{x}_b = (T, q, q_{2m}, T_{2m}, SST)$) derived from the regional non-hydrostatic forecast model COSMO. H is the radiative transfer model RTTOV–7 (Radiative Transfer model for Television infrared orbiting satellite Operational Vertical Sounder) (Saunders et al., 2002 [33]) for which also the tangent linear and adjoint versions are available and which is explained in more detail in the section 3.1. The observation vector \vec{y} contains a selection of the eight available infrared SEVIRI channels.

The 1D–VAR minimisation uses an adapted version of a 1D–VAR scheme that was first developed at ECMWF (European Center for Medium range Weather Forecast) for the assimilation of rain affected microwave radiances (Marécal and Mahfouf, 2000 [21]) and employed for cloud analysis (Chevallier et al., 2004 [5]). It finds the solution \vec{x}_a through the minimisation of $J(\vec{x})$. Successful convergence is achieved employing a minimiser based on the routines M1QN3 (by Gilbert and Lemaréchal, 1989 [15]) which is designed to minimise functions depending on a very large number of variables. It implements a limited memory

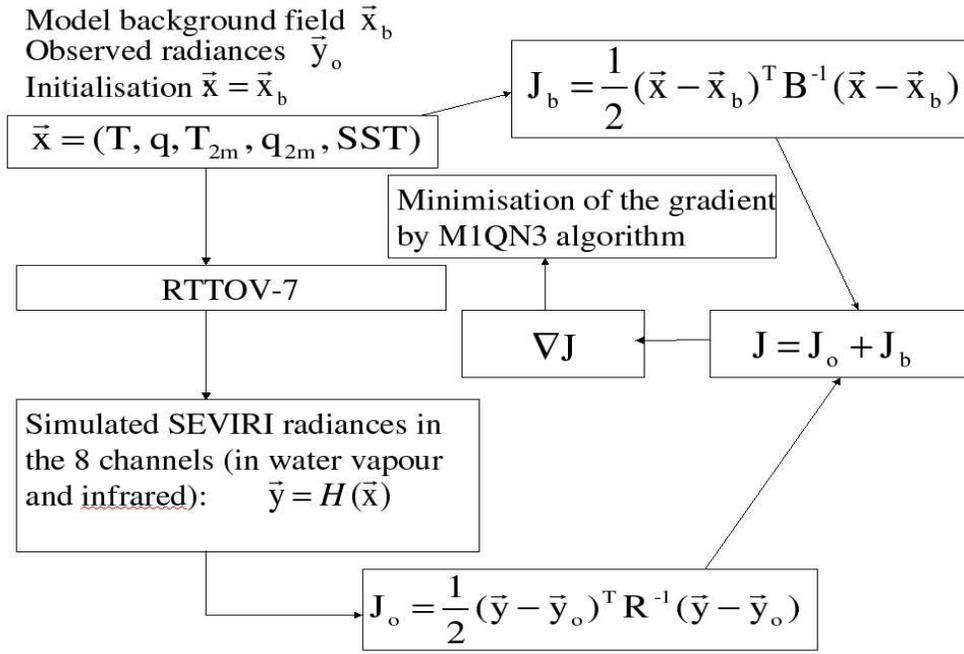


Figure 3.1: Sketch of the whole 1D–VAR scheme used in this thesis

quasi-Newtonian technique with a dynamically updated scalar or diagonal preconditioner. A simplified sketch of the whole 1D–VAR algorithm is given in the Fig. 3.1.

At a given model time–step and grid–point, the nudging procedure uses both past and future observations within $1 \Delta t$ with each observation weighted by its relative proximity in time and space. In contrast to other conventional observations, 1D–VAR retrievals are generated during the assimilation cycle and their values, even if required by the nudging scheme, are not known before the nominal observation time. To avoid doubling the analysis run or a time–consuming analysis backward integration a different approach has been followed to simultaneously produce and ingest 1D–VAR non–conventional observations into the nudging scheme.

The procedure implemented is schematised in Fig. 3.2. For each satellite observation two minimisations are performed using background profiles at $T_n = T_{obs}$ and $T_{n-1} = T_n - \alpha$ with $\alpha = 15min$. These two retrievals are then combined to provide the necessary continuity in time. More specifically, referring to Fig. 3.2, at T_{n-1} two 1D–VAR retrievals are created using background profiles $\bar{x}_b(T_{n-1})$; one with observation $\bar{y}(T_{n-1})$ and another with observation $\bar{y}(T_n)$. At each instant between T_{n-1} and T_n the two ‘pseudo–observations’ are

combined using a triangular function.

The spread over space of the resulting 1D-VAR increments uses Gaussian weighting functions. The choice of the time interval α and of the spatial weight is somehow arbitrary and will be tuned for the operational implementation. In this chapter only retrievals which are synchronous with the satellite observations are considered.

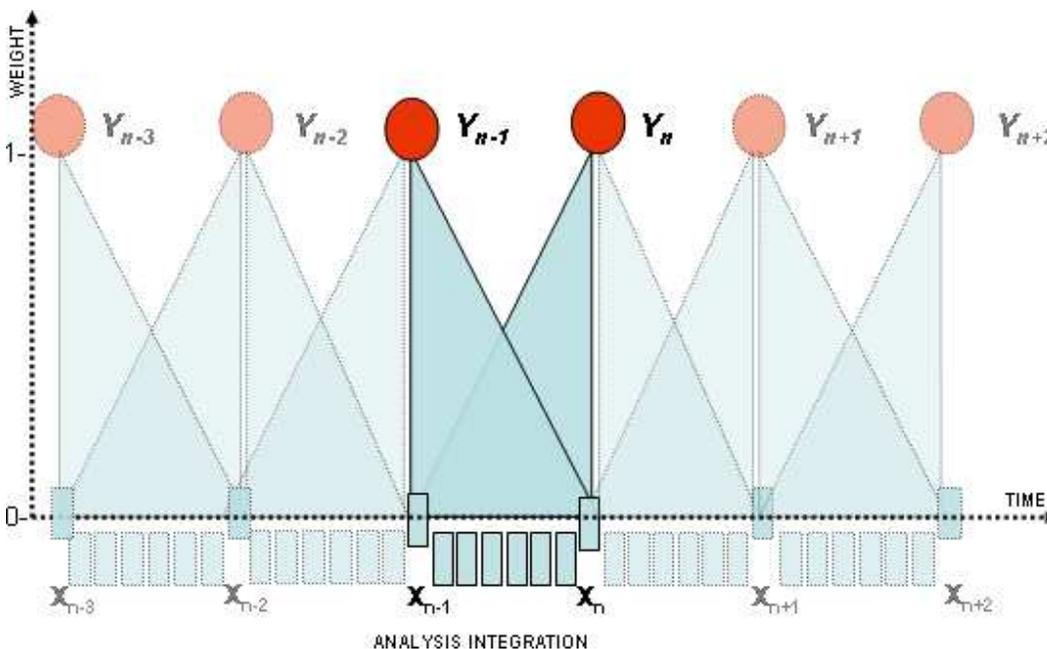


Figure 3.2: Sketch of the assimilation system designed for COSMO to include 1D-VAR retrieved profiles into the nudging scheme. Two successive minimisations are performed by using profiles generated at T_n and $T_{n-1} = T_n - 15min$ (see text for more details).

Before entering the 1D-VAR package the model temperature and humidity profiles are preprocessed by (i) adding stratospheric values up to 1hPa on top of COSMO standard levels using a climatological database extracted from the ERA-40 (Ecmwf Re-Analysis) reanalysis (Uppala et al., 2005 [46]), (ii) attributing a given profile to the observation location using a bilinear interpolation between the four closest grid points; (iii) performing quality checks to ensure the profiles are in the range of applicability of the radiation code RTTOV-7. The SST value is prescribed from a SST external analysis which is performed daily at 00 UTC. It uses Cressman-type weighting to blend NCEP (National Centers for Environmental Prediction) analysis (Reynolds et al., 2002 [29]) with buoy and ship observations. The $2m$ temperature and humidity are diagnostic variables. They are functions of the SST and the temperature

and humidity values of the first model level taking into account boundary layer stability.

3.1 The fast radiative transfer model RTTOV–7

To exploit satellite sounding data in a NWP model the use of a fast radiative transfer model is mandatory. In particular the radiative transfer model must satisfy these requirements:

- it must be fast enough so that it can be used in near real–time with meteorological data;
- it must be however highly accurate in its calculations;
- it must be able to perform the computations both in presence of clouds and in presence of clear sky.

The original basis for the fast radiative transfer model RTTOV–7 is based on Eyre and Woolf, 1988 [11]. This was successively modified by Eyre, 1991 [12], Rayer, 1995 [28], Rizzi and Matricardi ,1998 [31] and Saunders and Matricardi ,1999 [34] and [35].

The simulation of transmittances in RTTOV–7 is based on a regression scheme with a variety of predictors from the profile variables. The regression is performed in terms of departures from a reference profile, for mixed gases, water vapour or ozone. The formulation to predict the layer optical depth is:

$$d_{i,j} = d_{i,j-1} + \sum_{k=1}^K a_{i,j,k} X_{k,j} \quad (3.1)$$

where $d_{i,j}$ is the level to space optical depth from level j and channel i , K is the number of predictors, and $a_{i,j,k}$ are predefined coefficients. For the mixed gases there are now 10 predictors, for water vapour 15 and for ozone 11.

The optical depth so obtained are firstly corrected multiplying the optical depth by a factor γ_i (correction factor) for each channel i and then converted to transmittances:

$$d_{i,j}^c = \gamma_i d_{i,j} \quad (3.2)$$

$$\tau_{i,j} = \exp(-d_{i,j}^c) \quad (3.3)$$

The transmittances are calculated on 43 fixed pressure (in hPa) listed below:

.1, .29, .69, 1.42, 2.611, 4.407, 6.95, 10.37, 14.81, 20.4, 27.26, 35.51, 45.29, 56.73, 69.97, 85.18, 102.05, 122.04, 143.84, 167.95, 194.36, 222.94, 253.71, 286.6, 321.5, 358.28, 396.81, 436.95, 478.54, 521.46, 565.54, 610.6, 656.43, 702.73, 749.12, 795.09, 839.95, 882.8, 922.46, 957.44, 985.88, 1005.43, 1013.25

The radiances are theoretically calculated both in clear-sky and in cloudy-sky conditions. In clear-sky conditions the top of the atmosphere upwelling radiance, $L(v, \theta)$, at a frequency v and viewing angle θ from zenith at the surface, neglecting scattering effects, is written as:

$$L^{Clr}(v, \theta) = \tau_s(v, \theta)\epsilon_s(v, \theta)P(v, T_s) + \int_{\tau_s}^l P(v, T)d\tau + (1 - \epsilon_s(v, \theta))\tau_s^2(v, \theta) \int_{\tau_s}^l \frac{P(v, T)}{\tau^2} d\tau \quad (3.4)$$

where τ_s is the surface transmittance, ϵ_s is the surface emissivity and $P(v, T)$ represents the output of the Planck function at frequency v and temperature T .

The output of RTTOV-7 is used to simulate multilevel infrared and microwave cloudy radiance and the RTTOVCLD routines now supplied with RTTOV-7 provide this capability (see Fig. 3.3). Clouds are assumed to be grey bodies with their contribution to the radiances computed from their horizontal coverage n_i , and their emissivity ϵ_v^i in each vertical layer i of the user's model. ϵ_v^i is derived from the cloud liquid and/or ice water path l^i by the relationship:

$$\epsilon_v^i = 1 - \exp^{-l^i k_v^i} \quad (3.5)$$

where k_v^i is the extinction coefficient at frequency v . Its value varies according to the phase of the cloud water, the particle sizes and the temperature. This allows the radiances for semi-transparent cloud to be expressed as a linear combination of $L^{Clr}(v, \theta)$ and single layer black body clouds. The coefficients of the linear combination are functions of the n^i and ϵ_v^i and depend on the way the cloudy layers overlap. Cloud absorption is taken into account in

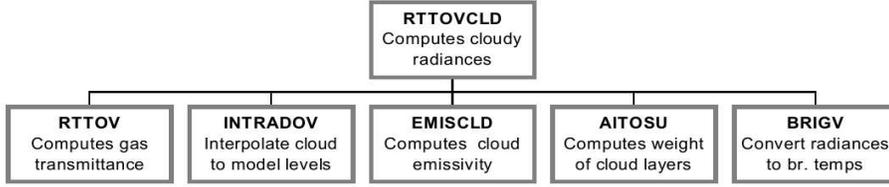


Figure 3.3: Simplified scheme of RTTOVCLD used to simulate cloudy radiances

the infrared spectrum following Ebert and Curry, 1992 [14] for ice water and Smith and Shi, 1992 [40] for liquid water (more details can be found in the RTTOV-7 user-guide present in Met-Office web site <http://www.metoffice.gov.uk/>).

The conversion of radiances in forecasted brightness temperatures is done using the concept of modified Planck function. If the Planck function does not vary too much over the channel response function, it can be evaluated at the central frequency of the channels. This can be accurate for narrow channels, but for the broad channels, like the SEVIRI channels, the concept of modified Planck function must be introduced and it takes account of the averaging of the true Planck function over the spectral response of channel i for the brightness temperature TB (Matricardi et al., 2001 [23]). The modified Planck function for channel i is given by:

$$P_i(TB) = \frac{c_{1,i}}{\exp\left[\frac{c_{2,i}}{a_i + b_i TB}\right] - 1} \quad (3.6)$$

where $c_{1,i}$, $c_{2,i}$, a_i and b_i are pre-computed coefficients for each channel; $c_{1,i}$ is equal to $c_1 v_i^3$ and $c_{2,i}$ is equal to $c_2 v_i$ where c_1 and c_2 are the normal Planck function coefficients, v_i is the central frequency of the channel, and finally a_i and b_i are the so called band correction coefficients (Weinreb et al., 1981 [48]).

The correspondent inversion of equation 3.6 converting radiances in forecasted brightness temperatures is:

$$TB_i = \left[\frac{c_{2,i}}{\ln\left(1 + \frac{c_{1,i}}{P_i}\right) - a_i} \right] / b_i \quad (3.7)$$

It is worth noting that up to now the RTTOV-7 scheme is applied for the 1D-VAR algorithm only in clear-sky conditions over the sea points. This limitation is due to the difficulty to accurately estimate the surface emissivity over the land, where the presence of asphalt, human buildings, roads and plants complicates this task. On the other hand the sea surface emissivity is well known for the different twelve SEVIRI channels. In section 3.2 the algorithm used to detect clear-sky points from SEVIRI data is described.

3.2 CMA and CT products and the flag processing

Since the retrieval of cloud contaminated pixels is very complicated, both for problems in the simulation of radiances in the radiative transfer model RTTOV-7 and for problems to detect which types of clouds are present in the pixels, the 1D-VAR algorithm at the moment is performed only for clear-sky points over sea.

Two special algorithms, the CMA (Cloud Mask) algorithm and the CT (Cloud Table) algorithm, are so used to detect if the pixel is cloud contaminated and to discriminate which type of cloud is present in the pixel respectively. They are developed within the SAFNWC (Satellite Application Facilities for NoWCasting) context and aim to support nowcasting applications, and additionally the remote-sensing of continental and oceanic surfaces. Both these algorithms are based on multispectral thresholds techniques applied to each pixel of the image.

The set of thresholds applied in the CT and CMA algorithms depends mainly on the illumination conditions whereas the values of the thresholds themselves may depend on the illumination, the viewing geometry, the geopotential location and NWP model data (more details can be found in SAFNWC web-site <http://nwcsaf.inm.es/>). While the output of the CMA product consists only in the detection if a pixel of a satellite image is contaminated by clouds, the output of the CT product consists in the following twenty categories:

0) : non processed due to the presence of corrupted data or to the absence of the data;

1) : cloud free land with no contamination by snow/ice covered surface;

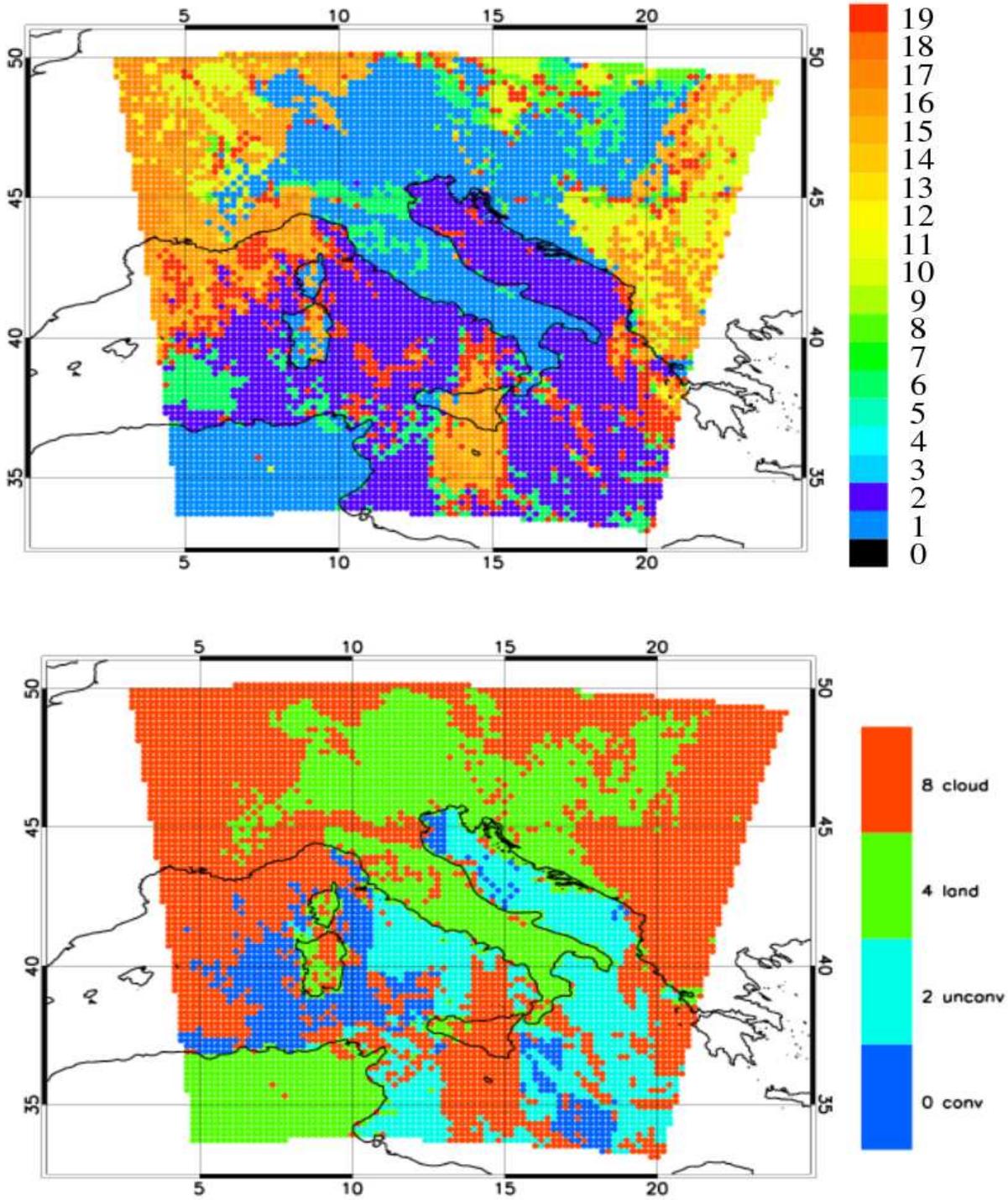


Figure 3.4: Example of SEVIRI CT (top panel) and the correspondent flg_prc (bottom panel) over the COSMO-LAMI domain at 04:15 UTC of 14th February 2007.

2) : cloud free sea with no contamination by snow/ice covered surface;

3) : land contaminated by snow;

- 4) : sea contaminated by snow/ice;
- 5) : very low and cumuliform clouds;
- 6) : very low and stratiform clouds;
- 7) : low and cumuliform clouds;
- 8) : low and stratiform clouds;
- 9) : medium and cumuliform clouds;
- 10) : medium and stratiform clouds;
- 11) : high opaque and cumuliform clouds;
- 12) : high opaque and stratiform clouds;
- 13) : very high opaque and cumuliform clouds;
- 14) : very high opaque and stratiform clouds;
- 15) : high semitransparent thin clouds;
- 16) : high semitransparent meanly thick clouds;
- 17) : high semitransparent thick clouds;
- 18) : high semitransparent above low or medium clouds;
- 19) : fractional clouds (sub-pixel water clouds);

At this moment the 1D-VAR algorithm is applied only for clear-sky points over sea but anyway the convergence required by the minimisation of the cost function could be unsuccessful. Hence there are four possibilities for the flag processing value (`flg_prc`) (see Fig. 3.4):

- 0) it means that the convergence is reached;
- 2) it means that the convergence is not reached;
- 4) it means that the pixel is over land;

8) it means that the pixel is cloud-contaminated;

By definition the convergence is reached if the minimisation by the MIQN3 algorithm determines a reduction of an order of magnitude of the cost function.

3.3 The background error covariance matrix \mathbf{B} and the observation error covariance matrix \mathbf{R}

The background error covariance matrix \mathbf{B} define the relative weight given to each forecast variable in the definition of the analysed field and for this reason it is of vital importance to the 1D-VAR procedure. The \mathbf{B} matrix can be obtained in different ways:

- one of these ways is to compare the vertical profiles of temperature and water vapor observed by radio-soundings with respect to the ones forecasted by COSMO and extracted at the same observation/forecast time. The main disadvantage of this method result in the fact that the observed values do not completely cover the vertical range of COSMO profiles; in fact while the forecast profiles start close to the surface and reach about 30 hPa at the top of atmosphere (TOA), usually the observed values are missing over 300 hPa. This determines the impossibility to correctly estimate the \mathbf{B} matrix in the levels from 300 hPa to TOA. Furthermore another crucial disadvantage of the radio-soundings is their poor spatial resolution covering;
- the second way is to compare the vertical profiles forecasted by COSMO, for example the vertical profiles extracted from a COSMO forecast at +24h with the analysis at the same forecast time, but usually the analyses and the first forecast hours in the model are moderately affected by the spin-up problem;
- the third approach is to compare vertical profiles extracted from two different forecasts, for example at +36 hours extracted from the first forecast with the +12 hours extracted from the latter, but correspondent to the same forecast time.

The \mathbf{B} matrix is estimated in this thesis using the third approach, comparing different forecasts of temperature and humidity profiles, also called the Parrish-Derber method (Parrish and Derber, 1992 [25]). 12h and 36h forecasts are compared at the same nominal time,

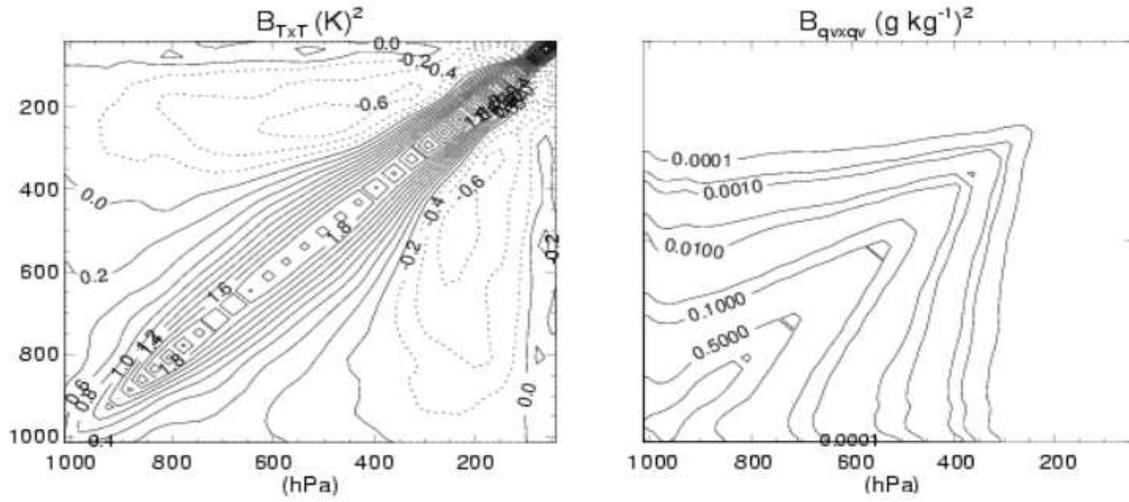


Figure 3.6: Background error covariance matrices for vertical profiles of COSMO-LAMI temperature and humidity background fields. The B matrix has been calculated using forecast comparisons at +12h and +36h averaged over four months worth of data.

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Figure 3.7: Structure of the observation error covariance matrix: assumed the uncorrelation of the errors between the different channels and that for every channel the observational and simulated error is of 1 K, R is a diagonal matrix with values 1 on the diagonal.

3.4 1D-VAR conceptual performance

Conceptually it is possible to calculate A and thus establish the expected performance of the retrieval scheme without explicitly performing simulated retrievals once it is assumed that a successful minimisation technique exists for finding a solution. For the case of a control vector \vec{x}_b , constituted of temperature and humidity variables and for clear-sky radiances it has already been proven that the linear approximation of the observation operator H is ver-

ified to guarantee convergence of the minimisation process (Eyre et al., 1993 [13]). The expected quality of the analysis can therefore be used as an a priori metric to define the best channel combination to use in an operation context. To this end it is also useful to introduce the concept of Data Resolution Matrix (DRM) and Model Resolution Matrix (MRM) which, if the assumption is made that $\vec{y}_{\text{true}} = H(\vec{x}_{\text{true}})$ i.e. that errors deriving from the observation operator are null, links the analysis to the true value either in physical (\vec{x}_{true}) or in observational (\vec{y}_{true}) space (Rodgers, 2000 [32]) :

$$\vec{x}_{\text{a}} - \vec{x}_{\text{b}} = \mathbf{KH}(\vec{x}_{\text{true}} - \vec{x}_{\text{b}}) = \mathbf{MRM}(\vec{x}_{\text{true}} - \vec{x}_{\text{b}}) \quad (3.8)$$

$$\vec{y}_{\text{a}} - \vec{y}_{\text{b}} = \mathbf{HK}(\vec{y}_{\text{true}} - \vec{y}_{\text{b}}) = \mathbf{DRM}(\vec{y}_{\text{true}} - \vec{y}_{\text{b}}) \quad (3.9)$$

Therefore, for those channels for which the radiative transfer simulation is accurate, the DRM quantifies the contribution of each wavelength to the global quality of the analysis. The diagonal elements of this matrix indicate how much weight a datum has in its own analysis and they can be regarded as a measure of the importance of that data. Another useful metric of the gain in information deriving from each channel is the “degrees of freedom for the signal” (DFS) [32] which can be calculated by normalising the Jacobian using $\mathbf{R}^{-1/2}, \mathbf{H}' = \mathbf{R}^{-1/2}\mathbf{H}$ and considering separately each vector $\vec{\mathbf{h}}$ which corresponds to the line of \mathbf{H}' linked to a given channel:

$$DFS = \frac{\vec{\mathbf{h}}^T \mathbf{B} \vec{\mathbf{h}}}{1 + \vec{\mathbf{h}}^T \mathbf{B} \vec{\mathbf{h}}} \quad (3.10)$$

Fig. 3.8 shows the diagonal elements of the DRM and the DFS calculated for a sample of COSMO–LAMI background profiles randomly extracted from the 20 days (the first twenty days of September 2006) dataset available. The DFS and DRM estimate the importance of a channel from a different point of view. The DRM which can also be calculated as $\mathbf{HAH}^T \mathbf{R}^{-1}$ uses \mathbf{A} to estimate which is the most useful channel in the analysis between *all* the ones used. Instead the DFS measures how much a channel in *isolation* is able to reduce the model error defined by \mathbf{B} in the observational space.

The two water vapour channels at 6.2 μm and 7.3 μm have the largest impact in the

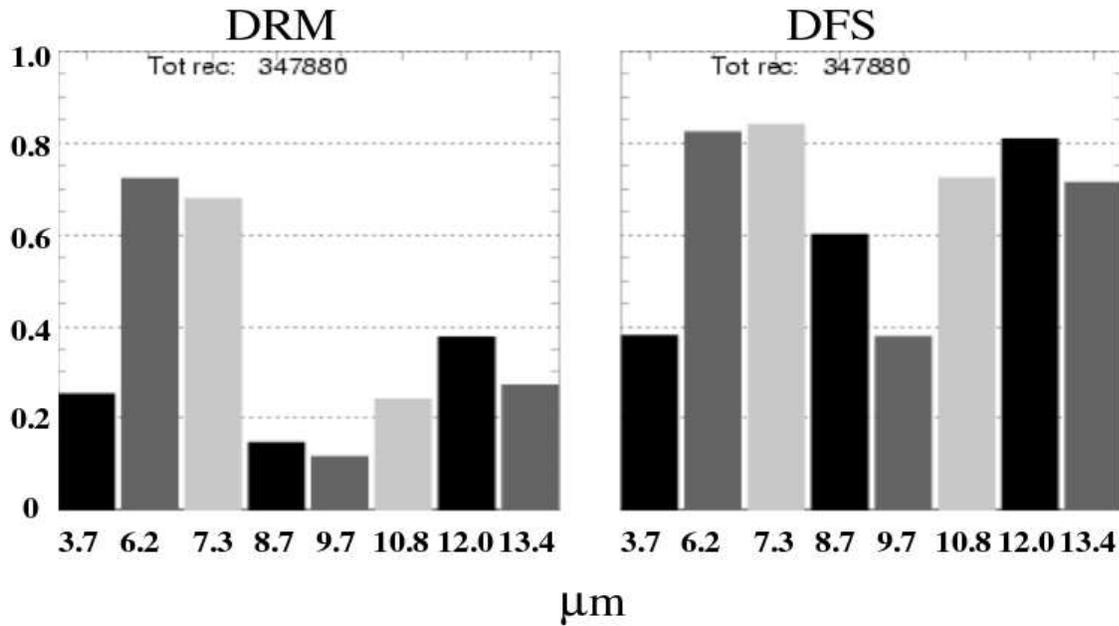


Figure 3.8: On the left panel averaged data resolution matrix (DRM) diagonal values for input background profiles randomly extracted from the 20 days dataset available, while on the right panel averaged degree of freedom for signals (DFS) calculated for each of the eight infrared SEVIRI channels and for the same dataset as for the DRM calculations.

accuracy of the analysis and in reducing the background errors and should be included in the 1D-VAR retrieval. Also the window channels at $10.8 \mu\text{m}$, $8.7 \mu\text{m}$ and $12.0 \mu\text{m}$ show non-negligible DFS values. Since the use of these additional wavelengths could be very efficient in reducing errors near the ground it is worth investigating their impact in changing the 1D-VAR retrievals.

Four set of wavelengths are selected by progressively adding to the two water vapour channels the window channels possessing the highest DRM and DFS values. These groups are then tested to define the best set to be used in the 1D-VAR retrieval scheme. Table 3.1 summarise the four groups. Set “ch8” in which all channels are included into the analysis is used as a benchmark to asses the quality of the other sets.

Fig. 3.9 and 3.10 report an estimation of the vertical resolution of the analysed profiles and the standard deviation of the analysis errors (i.e. square root of diagonal elements of matrix **A**) respectively for the channel combinations under examination. Following Purser et al., 1993 [26] the vertical resolution at each height is calculated as the ratio between the

Table 3.1: Sets of channels under test to define the best channel combination to be used in the 1D-VAR retrieval.

Set Id	<i>Channel Frequency (μm)</i>							
ch2	6.2	7.3						
ch3	6.2	7.3			10.8			
ch4	6.2	7.3			10.8	12.0		
ch5	6.2	7.3	8.7		10.8	12.0		
ch8	3.9	6.2	7.3	8.7	9.7	10.8	12.0	13.4

model vertical thickness and the corresponding diagonal element of the MRM. If the analysis is equal to the true atmospheric state, MRM reduces to the identity matrix and the model layer thickness equals the vertical resolution at each height. Otherwise the vertical resolution provides information on the vertical smoothing of the analysis compared to the original background fields.

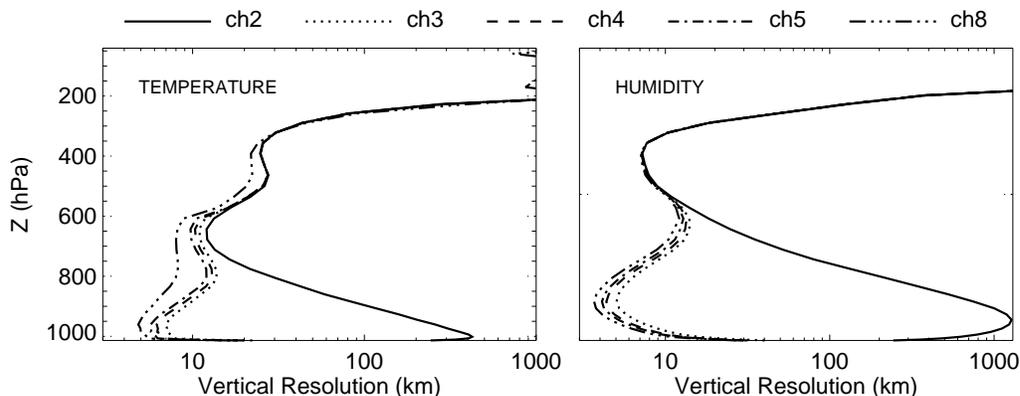


Figure 3.9: Vertical resolution of the analysed profiles calculated as ratio between the forecast model vertical thickness and the corresponding diagonal element of the Model Resolution Matrix (MRM).

The use of all 8 infrared channels leads to maximum temperature error reduction of 11% in the troposphere and a remarkable improvement in the humidity estimation. This is an upper theoretical limit of error reduction which can be achieved only if the radiative transfer simulation in each channel is correct. The inclusion of the window channel at $12.0 \mu m$ in set ch3 already produces a substantial decrease of the analysis errors both in temperature and humidity and an increase in the vertical resolution below 600 hPa. The additional use of channels $10.8 \mu m$ and $8.7 \mu m$ has the effect of increasing the accuracy of the humidity

analysis in the lower levels. For this added information it is therefore decided that set ch5 is selected as a standard for the 1D-VAR retrieval scheme.

The two gas monitoring channels at $9.7 \mu\text{m}$ and $13.4 \mu\text{m}$ do not hold a large information content (Fig. 3.8). Moreover, the radiances in these two channels are strongly influenced by the columnar total amount of ozone and carbon dioxide respectively. Since CO_2 and O_3 concentrations are neither diagnostically nor prognostically predicted by COSMO, large inaccuracies in the radiative transfer simulation are expected at these wavelengths. The assumptions of equations 3.9 and 3.10 could not be verified. The two channels are therefore discarded. Analogously the $3.9 \mu\text{m}$ channel which has been shown to detect a non-negligible portion of solar shortwave radiation during daytime is blacklisted (Schumann et al., 2002 [38]). It is worth noting that the overall small number of channels on SEVIRI and their quite broad band sizes make the expected 1D-VAR increment vertical resolution quite large. Sharp features, such as temperature inversions, won't be present in the analysis increments.

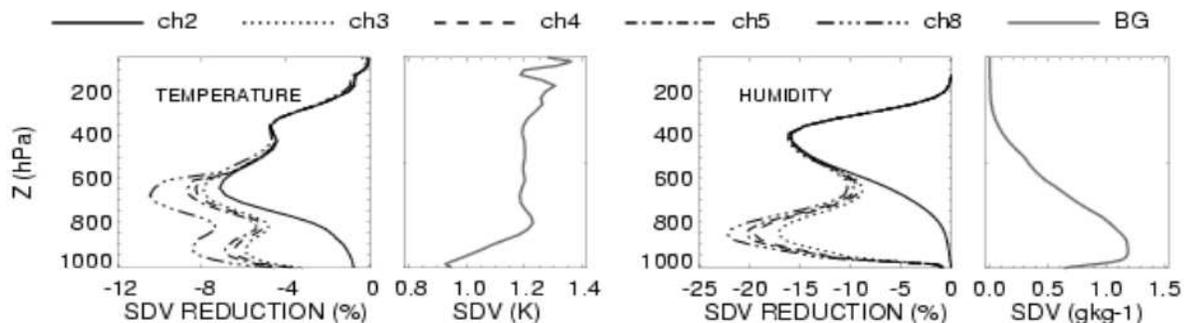


Figure 3.10: Expected error reduction in the standard deviation (SDV) of the background temperature and humidity profiles for the selected combinations of channels described in table 3.1. The analysis SDV is calculated as square root of the diagonal elements of \mathbf{A} .

3.5 The bias correction method

All the radiative transfer models involved to assimilate satellite sounding radiances (beyond to the NWP models themselves) are subject to errors. These errors, which arise mainly from errors in the spectroscopic data on which the radiative transfer models are based, can be comparable to or greater than the instrument noise, and so they must be taken into account.

The importance of the radiance bias problem has been recognised for many years and empirical correction schemes have been developed (Smith et al., 1984 [39], Chedin and Scott, 1984 [2] and Susskind et al., 1983 [43]). Some works have demonstrated that a successful bias correction scheme must take into account the spatially varying and air–mass dependent nature of the radiance biases (Kelly and Flobert, 1988 [19], McMillin et al., 1989 [24] and Uddstrom, 1991 [45]). In practice this imposes the application of an algorithm that reduces statistically the differences between the simulated brightness temperatures obtained by the first guess model fields (`bt_fg`) and the brightness temperature observed by the satellite radiometer (`bt_obs`). So before being used the satellite data (the brightness temperatures) are bias corrected (`bt_bcor`) using a scheme which is based on four air–mass dependent predictors (following the guidelines of Harris and Kelly, 2001 [17]):

- the vertical thickness between 900 and 300 hPa;
- the vertical thickness between 200 and 70 hPa;
- the 2 meters temperature;
- the water vapour content in the column of atmosphere.

In Fig. 3.11 the performance of the bias correction method applied to two weeks of available data is shown. It is evident for all the five channels a reduction of the bias.

Assuming that the model is unbiased, this guarantees that the 1D–VAR analysis works in an optimal way and that inaccuracy in the radiative transfer model and/or changes to the instrument characteristics over time do not spoil the information contained in the radiances when converted into temperature and humidity. In the following we will refer always to bias–corrected brightness temperatures when observed brightness temperatures are mentioned.

3.6 1D–VAR actual performance

The 1D–VAR retrieval performances have been assessed firstly in relative terms of error reduction in the background forecast fields, then in absolute terms comparing the analysis against radiosondes and *SST* independent observations.

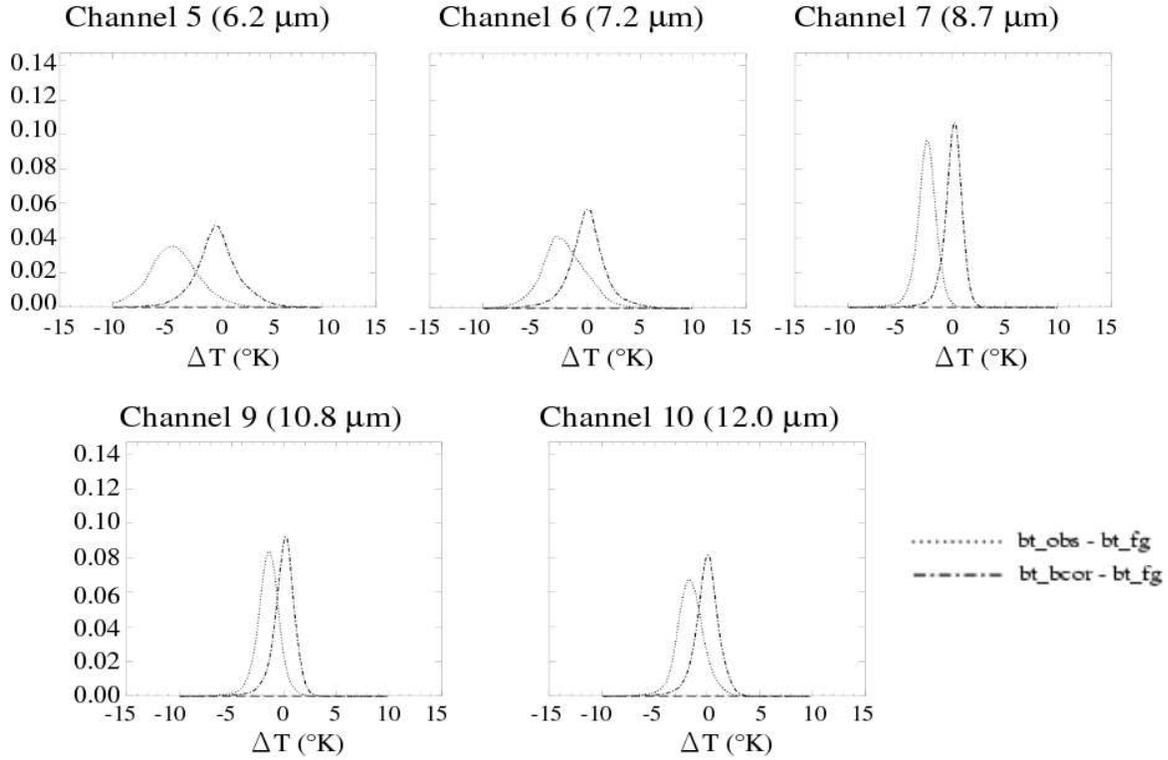


Figure 3.11: Statistics performance of the bias correction scheme applied to two weeks of available data for the 5 channels selected for the 1D-VAR.

3.6.1 Statistics of model departures from observations

The relative quality of the retrieval procedure is quantified by the reduction in the background departures (bt_bcor minus bt_fg) compared to the analysis departures (bt_bcor minus bt_ret). This expected reduction is a metric of the improved fit of the analysis TBs to the observations compared to the background. The two dimensional probability density functions (PDF) of the model TBs (calculated applying the operator H to \vec{x}_b) versus the observed ones for all channels indicate that the residual biases still present in the background after the bias correction procedure are reduced in the analysed TBs and the variances are greatly decreased by the 1D-VAR procedure (fig.3.12). The magnitude of the PDF's standard deviation can be regarded as a measure of the model variability. Large background TBs departures are present in the two water vapour channels, mostly due to inaccuracies in the forecast of high troposphere humidity. However the largest SDV reduction is obtained at 6.2 and 7.3 μm showing the significant information content provided by these two channels in the analysis.

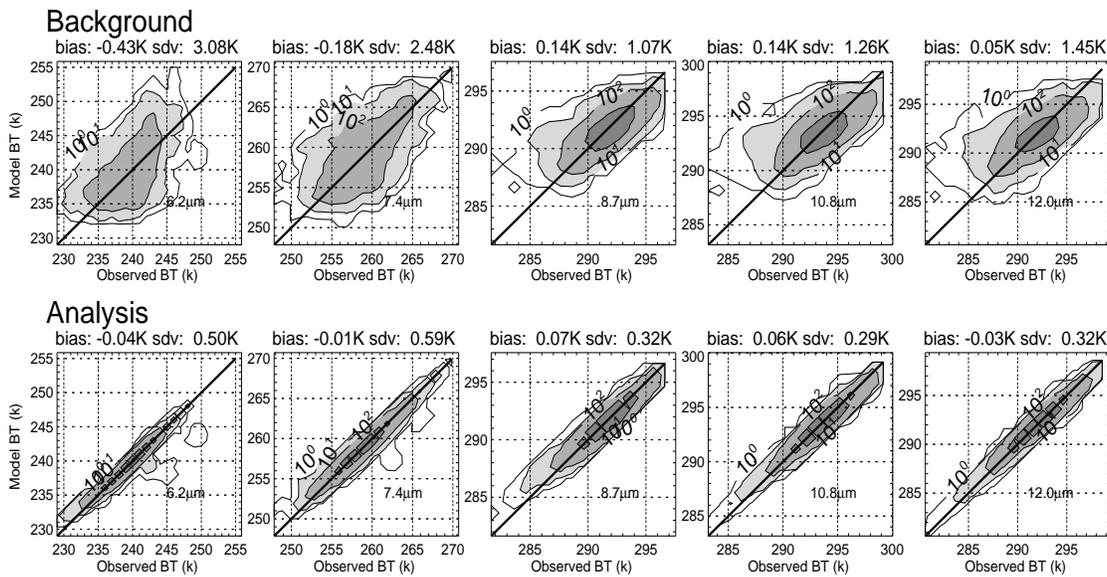


Figure 3.12: Two dimensional Probability Density Function (PDF) of background departures (upper panels) and analysis departures (lower panels) for an example data set taken on the 18 of September 2006 and for the 5 channels selected for the 1D-VAR.

Fig. 3.13 shows the mean 1D-VAR analysis increments in the temperature and humidity profiles. Higher values of standard deviation indicate a greater sensitivity of the model background to a particular satellite observation. As a confirmation of the expected impact on the temperature profiles (Fig. 3.10) most of the information content is spread over the whole troposphere with a maximum at around 600 hPa. Humidity profiles undergo most of the changes in the lower and upper troposphere at 800 hPa and 200 hPa. 1D-VAR has weak effect on the 2 meters temperature and humidity, while a non-negligible impact is recorded for the *SST* analysis increments with a SDV of 0.45 K (not shown).

It is worth noting that in variational assimilation system one should expect the whole domain averaged mean analysis increments to be null for each variable since observations are supposed to be unbiased. The fact that the temperature and humidity vertical mean increments are close to zero indicates that the bias correction procedure works well and that the final analysis will be in balance with the model minimising spin-up problems.

Nevertheless, the two dimensional mean increment distributions in Fig. 3.14 show how heat and water are redistributed by the 1D-VAR inside the domain. The increments have been averaged for the whole period under examination over boxes of 0.05 degrees and plotted

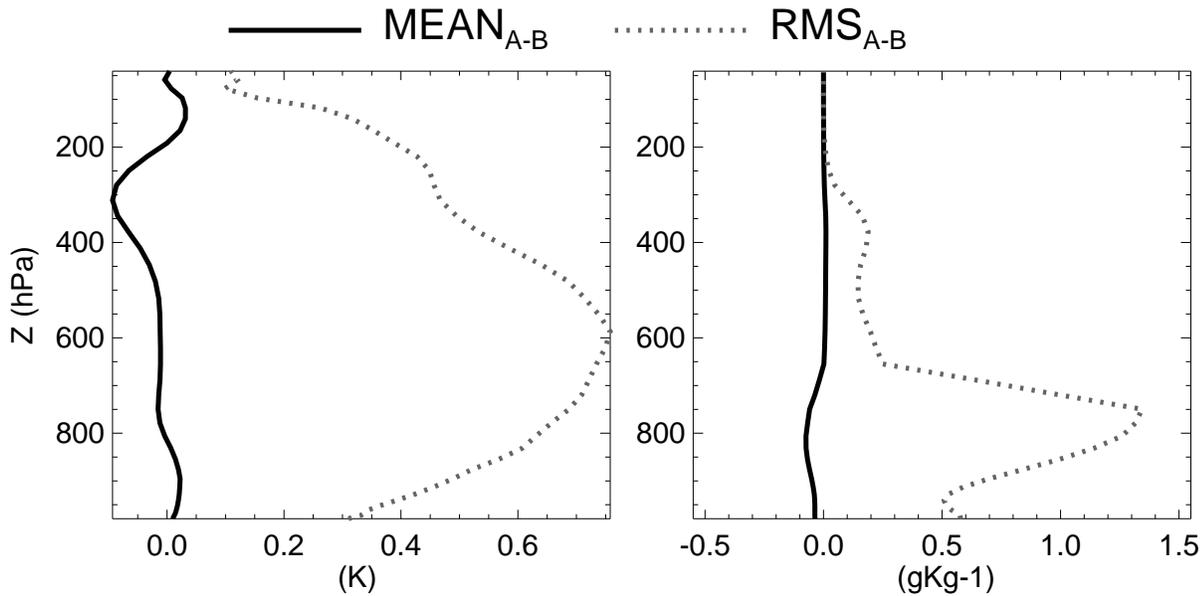


Figure 3.13: 1D-VAR analysis mean increments (analysis minus background) in the temperature (left panel) and specific humidity (right panel) profiles.

in terms of column integrated water vapour and saturation mixing ratio. Positive (Negative) differences between observed and background TB s are transformed in increases (decreases) of temperature profiles and decreases (increases) of water vapour. Since 1D-VAR increments lie in the column space of the kalman gain matrix (\mathbf{HB}), the relative contribution of temperature and humidity to the TB 's increments is defined by the minimisation process. q_v and q_{sat} are of the same order of magnitude showing that the most efficient convergence is obtained by changing both temperature and humidity in the column. It should be noted that in some circumstances the 1D-VAR increments can lead to supersaturated states. These may occur as a consequence of the presence of clouds that the cloud clearing algorithm has failed to detect, or a poor estimate of the SST with an error exceeding the estimated background error, for example. While these supersaturated states can be used to spot weaknesses in the cloud clearing algorithm, these profiles should however be discarded.

3.6.2 Comparisons with independent observations

As an independent validation of the 1D-VAR products, radiosonde measurements and independent SST datasets are employed. Radiosondes measurements are taken at three stations included in the ARPA-SIM operational domain of COSMO-LAMI close to the sea which

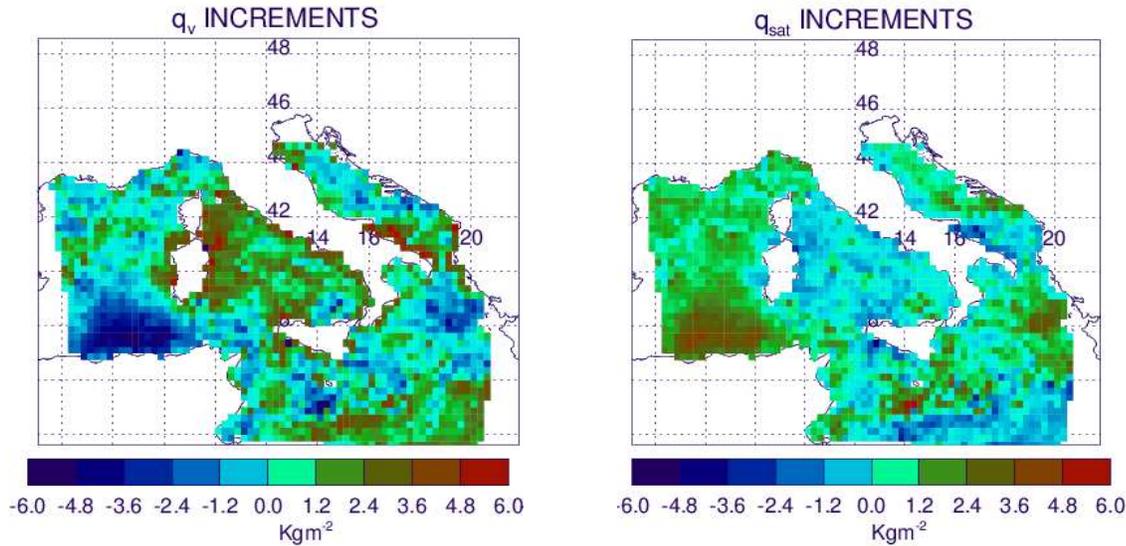


Figure 3.14: 1D-VAR analysis mean increments (analysis minus background) in the integrated column water vapour(left panel) and saturation water vapour (right panel).

Table 3.2: Radiosondes locations selected for validation of the 1D-VAR retrieved profiles.

ID	NAME	HEIGHT (m)	LAT (deg)	LON (deg)
16320	BRINDISI	15	40.6	17.9
16560	P CAGLIARI/ECOSMOAS	4	39.2	9.1
16429	TRAPANI/BIRGI	7	37.9	12.5

are blacklisted from the analysis so to represent a real independent source of validation. The radiosondes employed are standard VAISALA RS-92.

There is a significant reduction in mean-bias and root mean square (*RMS*) errors in the temperature profiles throughout the troposphere, both at 00 UTC (fig. 3.15) and 12 UTC (fig. 3.16). The improvement in the humidity estimation is less evident. Despite the questionable significance of around 100 soundings, it appears that the 1D-VAR introduces a moist bias in the humidity profiles especially during daytime (12 UTC) while still reducing the *RMS* error. However, it should be recalled that, while the RS-92 sounds have been shown to provide accurate humidity measurements at night, they suffer from a significant day-time bias of around 10 % in relative humidity due to sensor heating (Vömel et al., 2006, [47]). This could be the origin of the apparent day-time moist bias in the 1D-VAR analyses.

The effect of the 1D-VAR analysis to reduce the observed model bias can also be quan-

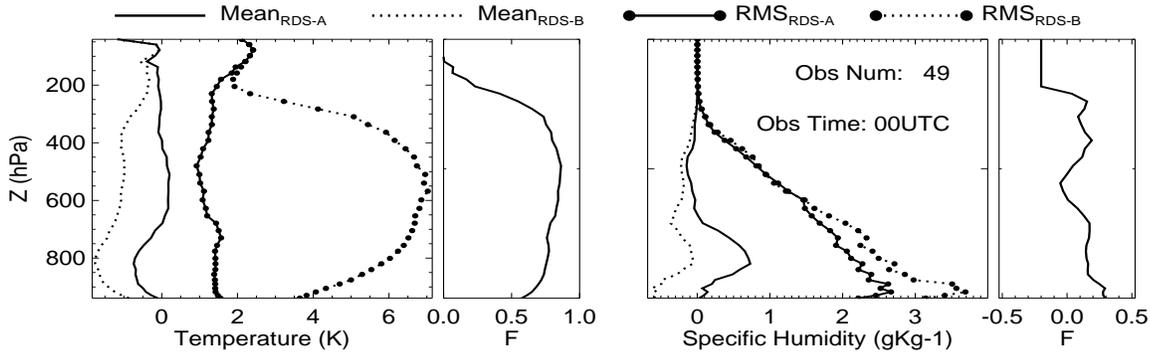


Figure 3.15: Model background and analysis biases (mean and RMS) when compared with radiosonde observations. The F factor is defined as $(RMS_{RDS-B} - RMS_{RDS-A})/RMS_{RDS-B}$ and quantifies the fraction of observed bias variability corrected by the analysis. See text for details.

tified in terms of the fraction of observed bias variability corrected by the analysis defined as:

$$F = \frac{RMS(X_{RDS} - X_B) - RMS(X_{RDS} - X_A)}{RMS(X_{RDS} - X_B)} \quad (3.11)$$

where $RMS(X_{RDS} - X_B)$ and $RMS(X_{RDS} - X_A)$ are the root mean square errors of the background and analysis departures, respectively. If $F = 1$ then the 1D-VAR analysis produces profiles which correlate perfectly with the radiosondes observations, if $F = 0$ then the effect of 1D-VAR is null, if it is negative then the 1D-VAR analysis procedure deteriorates the background.

This fraction F is overall positive with peaks of 0.8 for temperature and 0.4 for humidity. The negative values of F in the specific humidity profiles shows deficiencies in the 1D-VAR retrieval especially at high vertical levels but this may also be partially explained by sounding inaccuracies which are known to be non-negligible at these levels. For levels above 600 hPa the analysed 1D-VAR produces a substantial improvement to the humidity profiles.

As an independent source of validation of the analysed SST two different datasets for the Mediterranean and the Adriatic (ADR) seas are used: the optimally interpolated (OI) microwave (MW) SST products produced by Remote Sensing Systems (hereafter called MED) and the ADRICOSM produced by the ISAC-CNR institute in Rome, IT (hereafter called

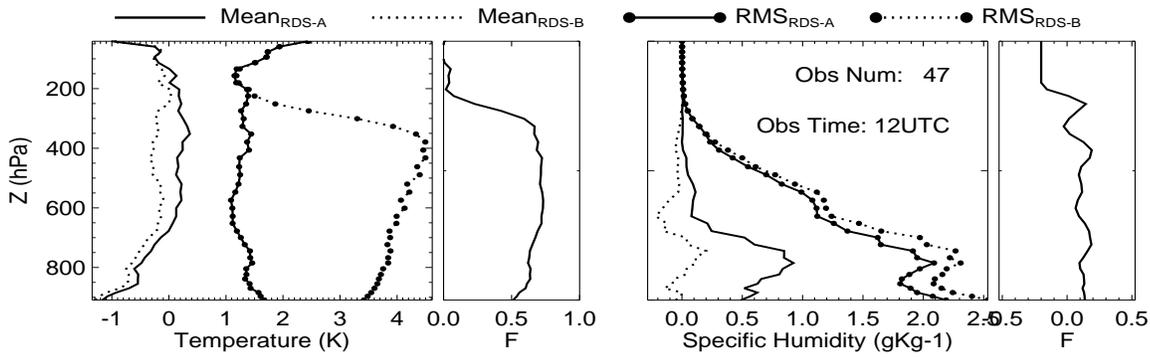


Figure 3.16: As for Fig. 3.16 but for 12 UTC observations

ADR). To guarantee their independence from the background analysed *SST* the two datasets employ microwave sensors and independent algorithms compared to the prescribed *SST* given to COSMO as a boundary condition which uses satellite infrared channels (Reynolds et al., 2002 [29]).

The MED dataset is available daily and MW *SST*s are 'normalised' to a daily minimum *SST*, defined to occur at approximately 8 AM, local time. It uses two microwave sensors (TMI and AMSR-E) sensors on board of TRM and AQUA (Chelton et al., 2000 [3], Donlon et al., 2002 [10]) and possesses a final resolution of a quarter degree ($25km$). Although these MW products are at a lower spatial resolution than standard infrared *SST*s, they have a much larger spatial and temporal coverage since retrieval is also possible for cloud contaminated spots. Other satellite observations deriving from microwave sensors used in the 1D-VAR can therefore improve the *SST* analysis. This is the main reason why *SST* has been included in the control vector.

Fig. 3.17 and 3.18 show two dimensional PDFs of the model *SST* versus the measured MED and ADR datasets, respectively. The background *SST*s are derived from an analysis which is performed at 00 UTC and is employed during all COSMO-LAMI integrations. The non-zero bias and *RMS* error between the model background *SST*s and the new datasets are not only due to differences in the data used but also to the fact that they effectively refer to two different analysis times. COSMO *SST*s do not evolve over time. The 1D-VAR analysis is able to reduce the bias only in the ADR dataset, while the *RMS* is reduced both in ADR

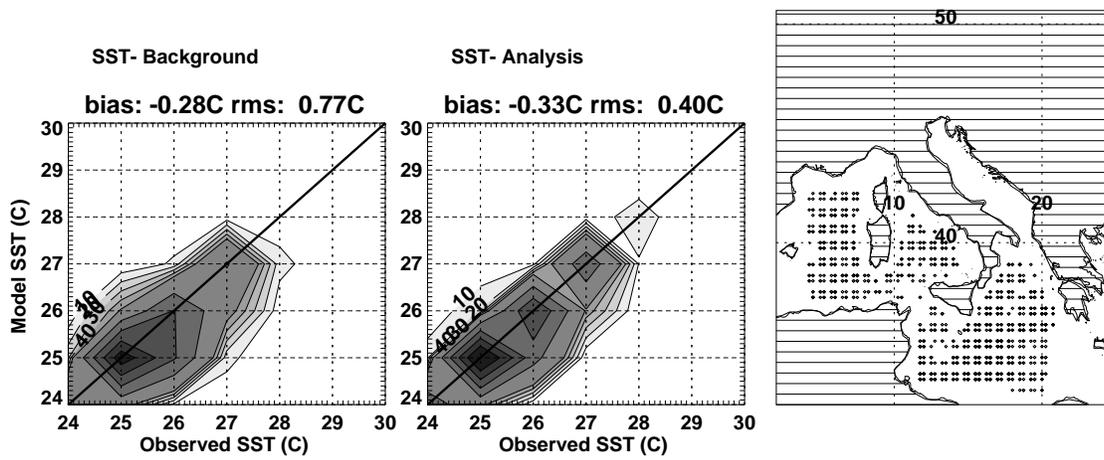


Figure 3.17: Two-dimensional probability density function of model (background and analysis) versus observed *SST* for the MED dataset. The right-endmost map shows the data localisation. The data cover the whole period of 20 days starting from the 1st September 2006.

and in MED dataset. Since in the ADR dataset the daily differences of the sea temperature can be more relevant, this result means that the infrared product is improved.

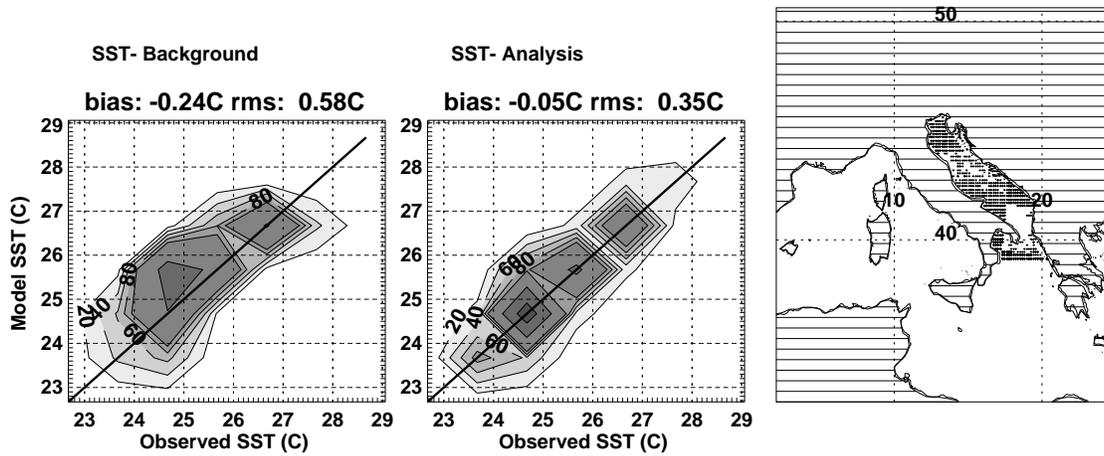


Figure 3.18: Same as for Fig. 3.17 but for the ADR dataset.

Chapter 4

THREE-DIMENSIONAL CASE-STUDIES

This chapter discusses the application of the 1D-VAR assimilation of satellite radiances for two selected case studies: a false alarm case occurred in Friuli-Venezia-Giulia region in July 2004 and a heavy precipitation case occurred in the Emilia-Romagna region in April 2005. For each of them a brief description of the synoptic situation is given and COSMO analysis and forecast fields produced with and without the contribution of the 1D-VAR retrieved profiles are compared.

The analysis of these differences have been divided in two phases. The fields are firstly analysed during the twelve hours assimilation cycle and then during the thirtysix hours model forecast. The initial conditions for the assimilation period are provided by the GME analysis. The hourly boundary conditions are provided by the GME forecasts both for the assimilation and for the forecast. It is worth noting that in both the case-studies the horizontal integration domain corresponds to the COSMO-LAMI operational one, but a different number of vertical levels is used: 35 levels for the Friuli-Venezia-Giulia case and 40 levels for the latter one. The use of only 35 vertical levels in the false alarm case-study is imposed for doing a coupling with ensemble members previously run with only 35 vertical levels (more details of this coupling is done in section 4.3).

Since the main aim of these case-studies is to evaluate the impact of the use of MSG-1

satellite data in the assimilation cycle, two different assimilations have been performed:

- the first one where only the conventional observations are assimilated;
- the second one where also MSG–1 satellite data are assimilated using the 1D–VAR approach as discussed in chapter 3.

As next step two different forecasts have been performed using the two different analyses one at a time. Hence both the assimilation, as well as the forecasts, are in the same temporal window and the only difference between each other is the presence/lacking of MSG–1 satellite data in input. To better evaluate the impact of MSG satellite data the differences of some output variables are plotted both for the assimilation and for the forecast period; the considered variables are:

- the integrated water vapour over the column;
- the integrated saturation water vapour over the column;
- the 2 meters temperature T_{2m} ;
- the 2 meters specific humidity q_{2m} ;
- the surface temperature T_s ;
- the precipitation field.

The inclusion in the considered variables of the integrated saturation water vapour over the column allows to translate humidity information into temperature increments and to perform in this way a direct comparison between these two variables. Furthermore the forecast wind fields at the end of the assimilation periods are plotted.

4.1 A false alarm case study : 8th of July 2004

The meteorological situation for this event can be explained as follows: a frontal system, linked to a large low depression localised on the Western Europe, in the night between the 7th and the 8th of July 2004 was moving towards to lower latitudes. The meteorological system reached firstly the North–Western regions of Italy and subsequently the other Northern

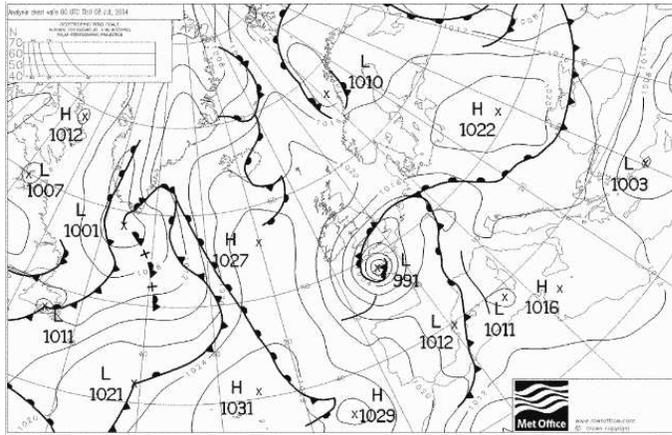


Figure 4.1: Synoptic situation at 00 UTC of the 8th July 2004

regions.

This synoptic situation occurs very frequently and usually determines major effects in terms of precipitation over the Western Alps and, after the front passage, over Friuli–Venezia–Giulia, Veneto and Trentino regions. Looking at several NWP model variables and in particular to the forecasted precipitation, a risk scenario was announced.

For this case study the temporal window of the two assimilations starts at 00 UTC of the 8th of July 2004 and lasts 12 hours. The forecasts are performed for the next 36 hours starting at 12 UTC of the 8th of July (see Fig. 4.2).

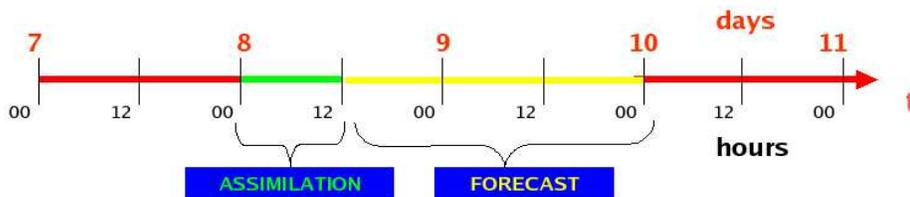


Figure 4.2: Scheme of the temporal windows for assimilations, in green colour, and forecasts, in yellow, for the false alarm case–study

To distinguish between the two assimilations and the two forecasts, with and without the use of the MSG–1 satellite data, the experiments are labelled as follows:

- ASS_MSG1_20040708 and FOR_MSG1_20040708 for the assimilation and the forecast where MSG–1 satellite data are assimilated using the 1D–VAR approach in COSMO;

for all the case–studies selected in this thesis with a temporal frequency of 1 hour.

The CT product (see Fig. 4.3), with the same colour palette defined in section 3.2, allows the view of the front passage over the north of Italy during the first hours of the 8th of July 2004.

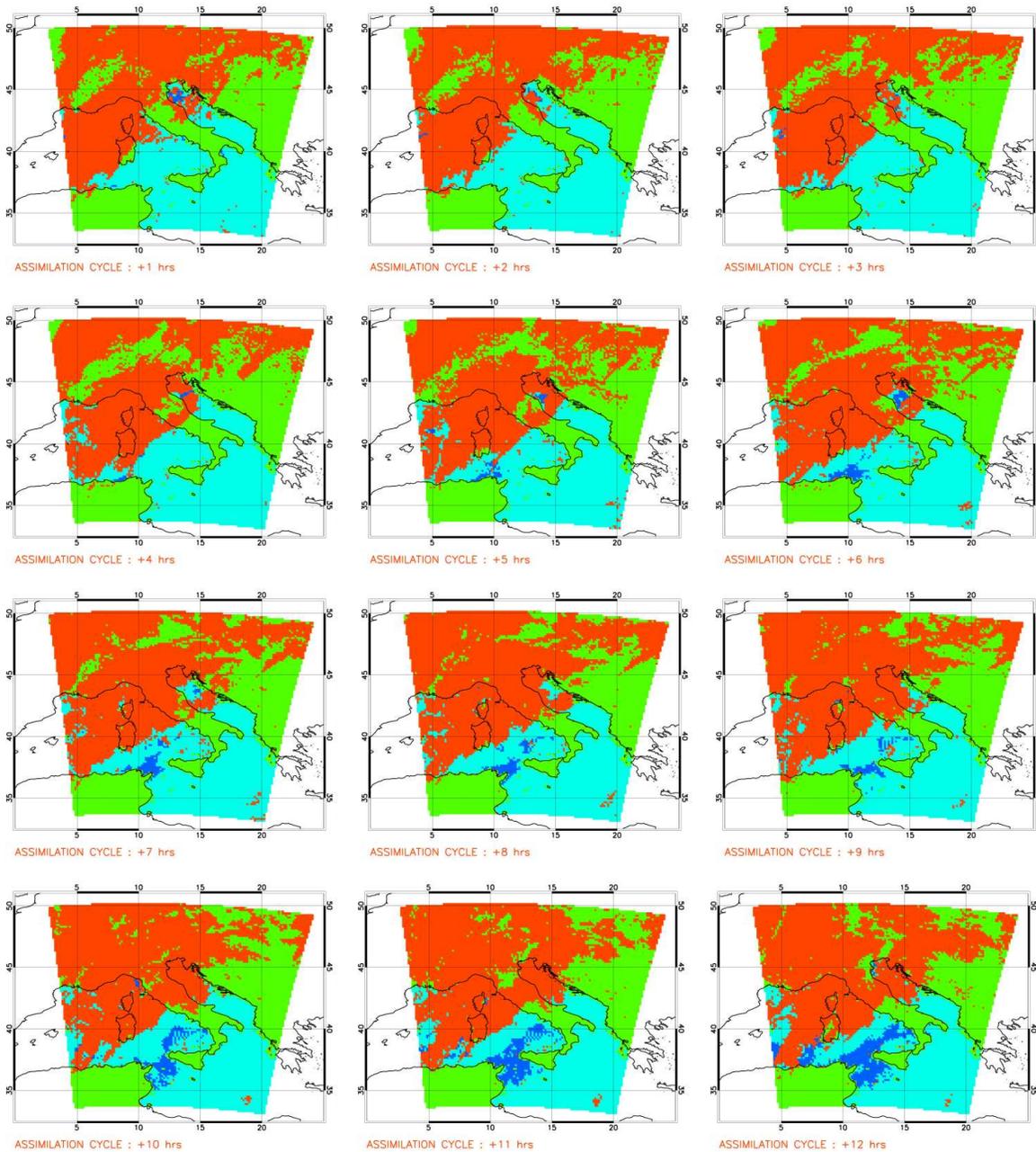


Figure 4.4: flg_prc during the assimilation temporal window with a temporal frequency of 1 hour.

The flg_prc , with the same colour palette defined in section 3.2, obtained for the assimilation temporal window (see Fig. 4.4) shows that the data assimilated during the first hours are mostly localised over the north of the Adriatic sea, while at the middle and at the end of the period the available data points are concentrated on the Tyrrhenian sea and on the Western Mediterranean sea close to the African coasts.

This data localisation has an impact on the initial position of the differences for the considered variables. In fact, except for T_s , the model variable increments are localised on the north of the Adriatic sea during the first hours of assimilation.

During the assimilation integration these increments are spread in space and time by the nudging scheme.

In particular the integrated water vapour increments (see Fig. 4.5) for the first assimilation hours remain localised where observed radiances reach convergency in the 1D-VAR and are thus assimilated. This produces significant positive increments of about $0.5\text{--}1.0\text{ kg/m}^2$ on the north of the Adriatic sea and in the Veneto region.

The magnitude of these differences reaches the maximum at the end of the assimilation period itself when the differences are appreciable almost over the integration domain; the most positive increments are present on the Liguria region and on the Ligurian sea, and in the Mediterranean sea north of Tunisia, while remarkable negative increments are present over the sea between Corsica and Sardinia islands and west of the Tuscany region.

For the integrated saturation water vapour over the column (see Fig. 4.6) the differences are significant at the beginning of the assimilation period only over the Veneto region, while in the middle of the period the largest differences are localised over the sea between Sicily, Sardinia and the African coasts. The spreading of these structures and their amplitude results smaller than for the integrated water vapour one.

The T_{2m} and q_{2m} differences over the assimilation cycle period (shown in Figg. 4.7 and

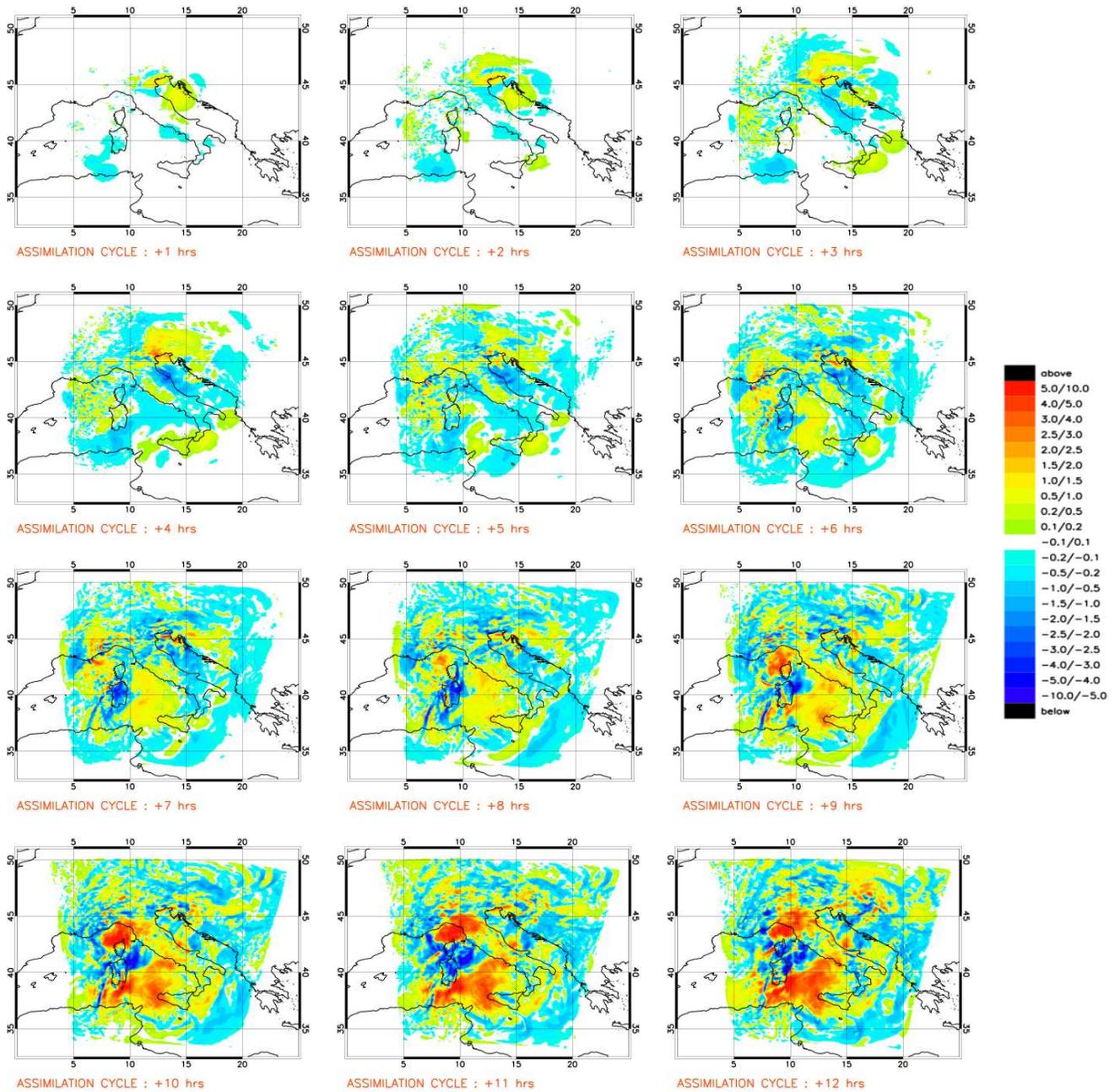


Figure 4.5: Differences in kg/m^2 between ASS_MSG1_20040708 and ASS_NUDG_20040708 of the integrated water vapour in the column during the assimilation temporal window with a temporal frequency of 1 hour.

4.8 respectively) show a progressive spreading of model differences in time. For the T_{2m} (see Fig. 4.7) differences between the two runs at the end of the assimilation period are generally negative over the Tyrrhenian sea while they are positive in the north of the Adriatic sea, but both these differences have generally an amplitude of 0.5 K. The most positive differences are present in the south-east of Sardinia, west of Corsica and over the Western Alps; the

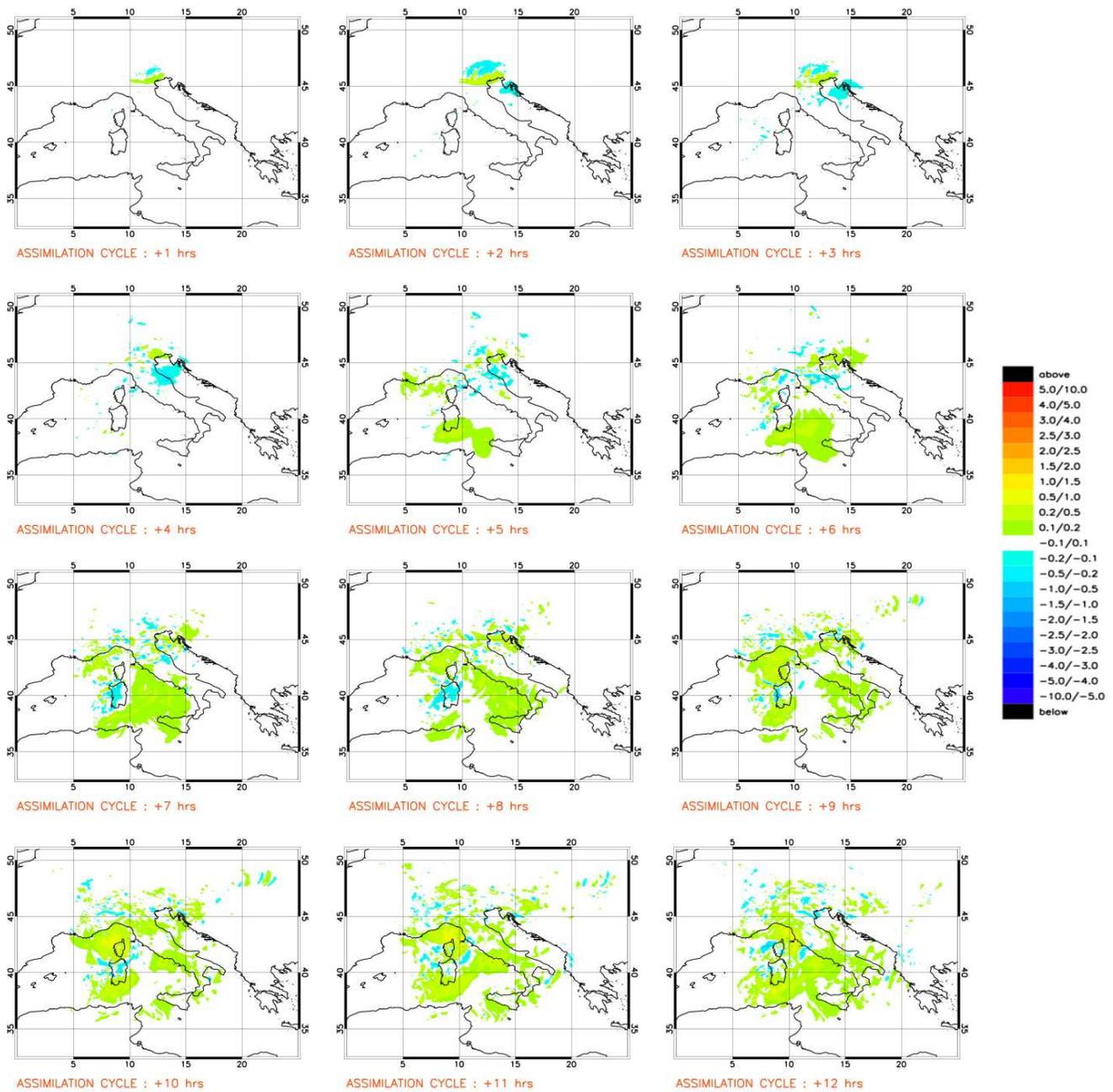


Figure 4.6: Differences in kg/m^2 between ASS_MSG1_20040708 and ASS_NUDG_20040708 of the integrated saturation water vapour in the column during the assimilation temporal window with a temporal frequency of 1 hour.

most negative values are instead present in the south–west of Sardinia, in Liguria, Tuscany, in the Central Alps and in Friuli–Venezia–Giulia region.

For the q_{2m} (see Fig. 4.8) the differences at the end of the assimilation period have the most positive values on the Tyrrhenian sea, south–west and east of Sardinia, while the most

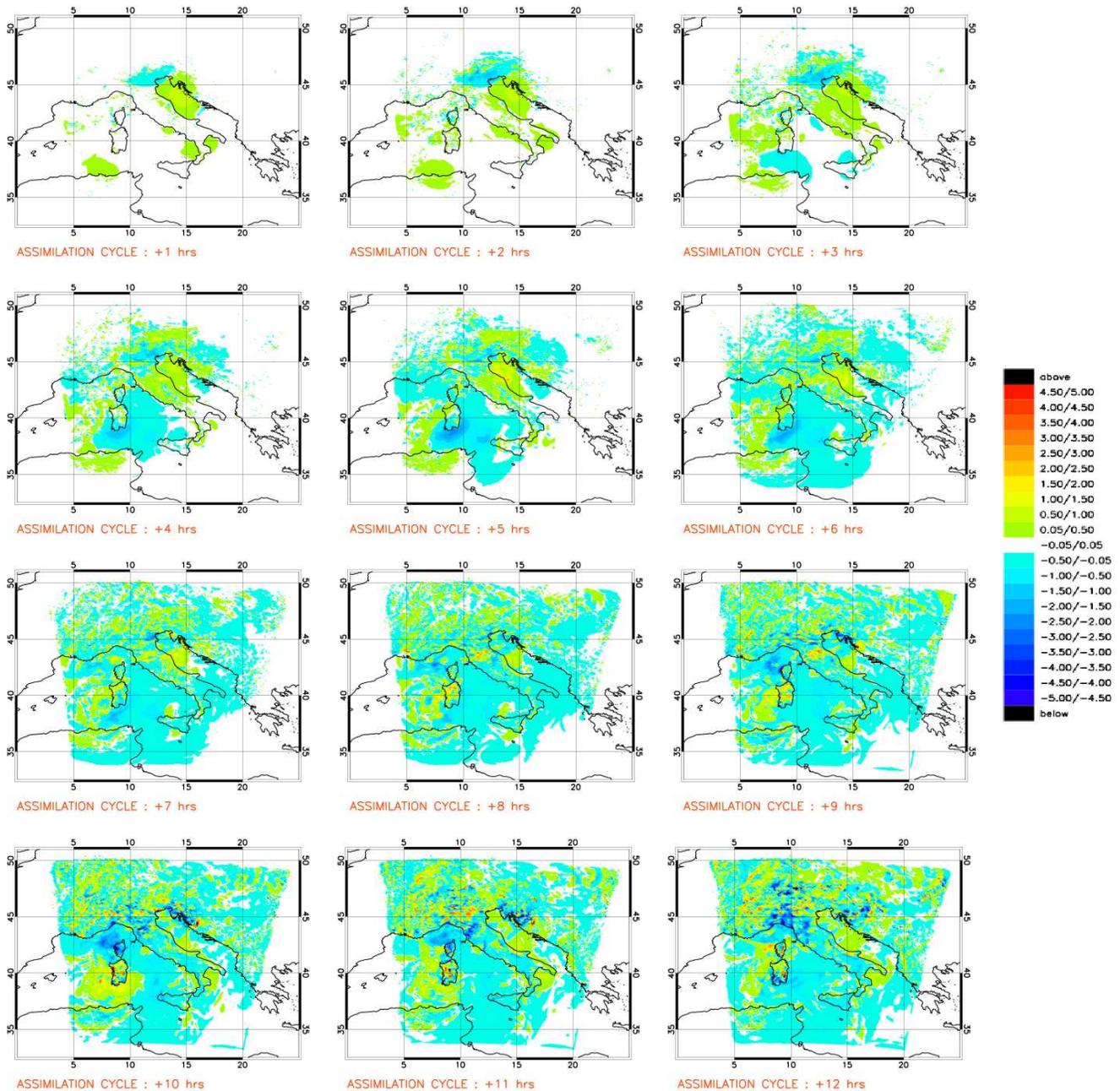


Figure 4.7: Differences in K between ASS_MSG1_20040708 and ASS_NUDG_20040708 of the forecasted T_{2m} during the assimilation temporal window with a temporal frequency of 1 hour.

negative values are in Sardinia, Corsica and in Friuli–Venezia–Giulia.

In the forecast model the assumption is made that the SST vary slowly during the forecast integration while this assumption cannot be considered valid for the temperature over land. Hence the SST values are kept constant to the analysed SST provided by the GME boundary conditions while T_s is diagnosed by a land–scheme. Fig. 4.9 shows that T_s differ-

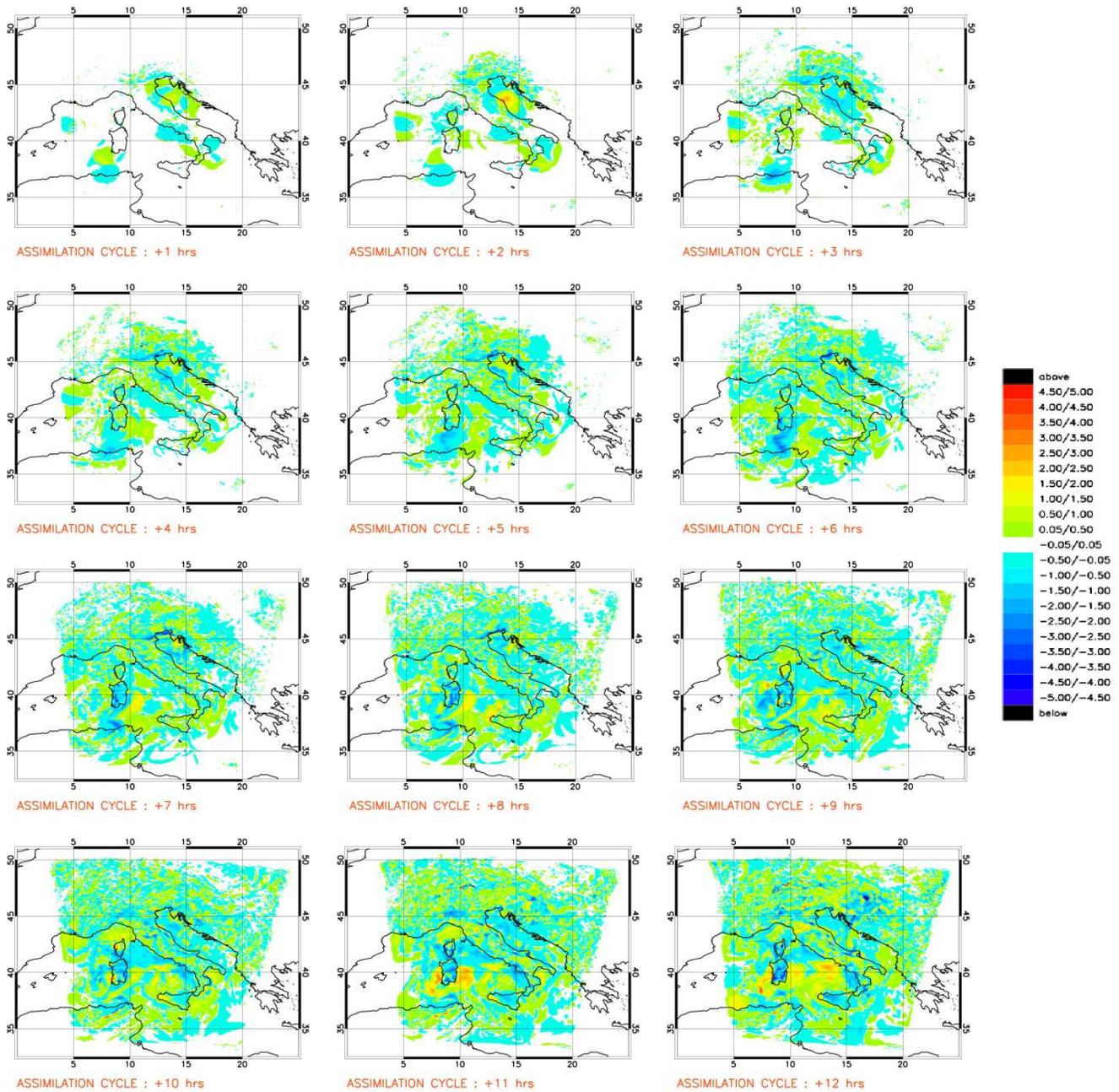


Figure 4.8: Differences in g/kg between ASS_MSG1_20040708 and ASS_NUDG_20040708 of the forecasted q_{2m} during the assimilation temporal window with a temporal frequency of 1 hour.

ences over the assimilation period result very similar to T_{2m} differences over land.

It is worth noting that the assimilation of clear-sky radiances over the north of Adriatic sea at the beginning of the assimilation is not able to remove the cloud cover present in the assimilation run ASS_NUDG_20040708 (not shown) and this will be subject of further developments of the COSMO model system in the next future.

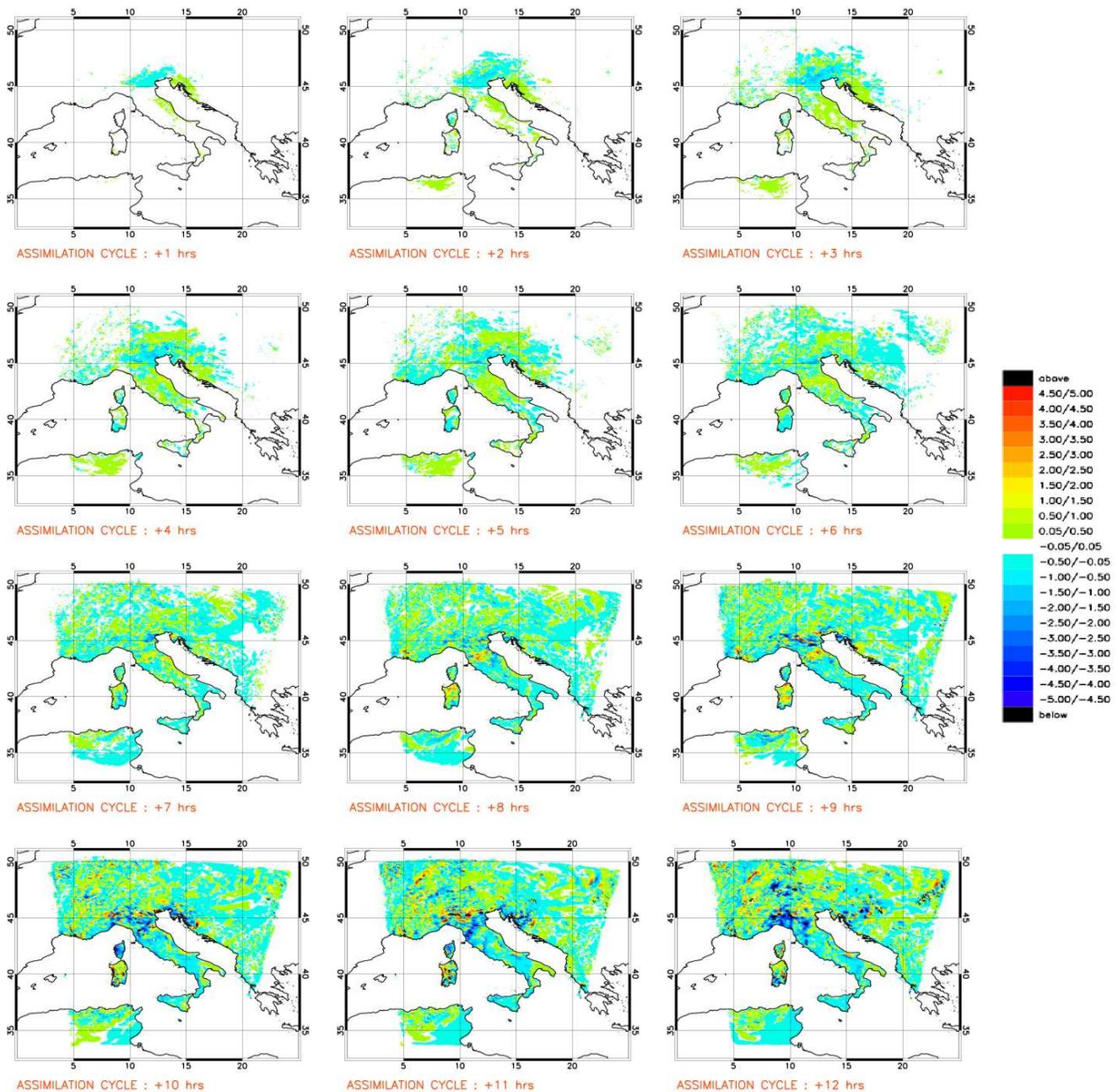


Figure 4.9: Differences in K between ASS_MSG1_20040708 and ASS_NUDG_20040708 of the forecasted T_s during the assimilation temporal window with a temporal frequency of 1 hour.

Since this case study concerns a false alarm case occurred in Friuli–Venezia–Giulia the forecast precipitations cumulated over the 12 hours of the two assimilations have been plotted only for the north of Italy (see Fig. 4.10). The overestimation of precipitation in Liguria region forecast by the run ASS_NUDG_20040708 is also present in the run ASS_MSG1_20040708. The main remarkable differences between the two forecast precipi-

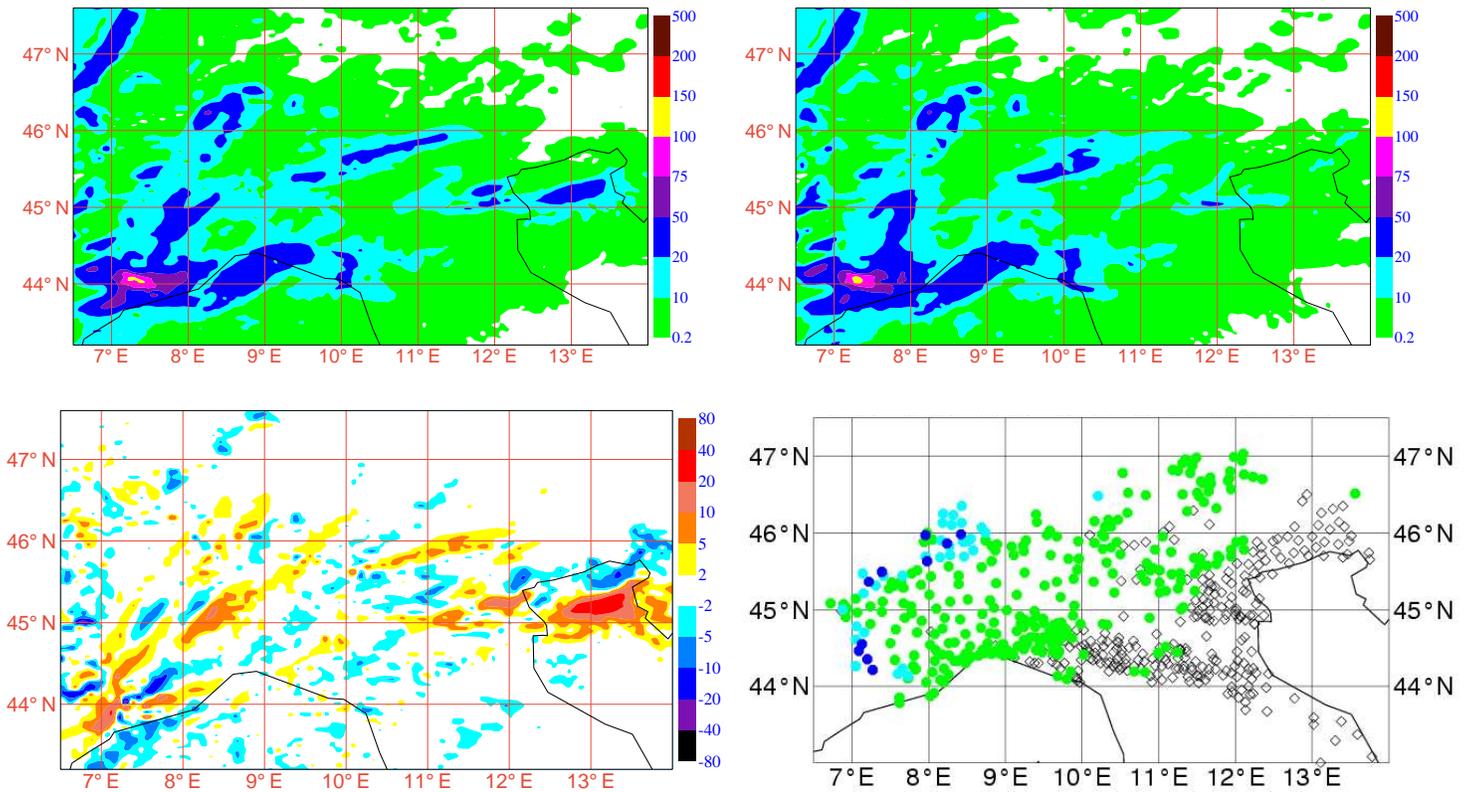


Figure 4.10: Twelve hours cumulated precipitation in mm from 00 UTC of the 8th of July 2004 forecasted by the runs ASS_MSG1_20040708 (top left panel) and ASS_NUDG_20040708 (top right panel), the differences (ASS_MSG1_20040708 minus ASS_NUDG_20040708) of the two forecast precipitation fields (bottom left panel) and the correspondent observations (bottom right panel).

tations are localised over the north of Adriatic sea where the run ASS_MSG1_20040708 increments the precipitation field. The reflectivity measured by the radar localised at Gattatico (one of the two radars of the Hydro–Meteorological Service of Emilia–Romagna region, ARPA-SIM) between 10:12 UTC and 11:42 UTC of the 8th of July 2004 with a temporal frequency of half an hour shows clearly the presence of hydrometeors in the north of the Adriatic sea and close to the Veneto coasts (see Fig. 4.11), just in the areas where the precipitation field forecasted by ASS_MSG1_20040708 is larger than the one forecasted by ASS_NUDG_20040708; this can be considered a qualitative confirmation that the assimilation of MSG–1 data has a positive impact in the forecasted precipitation and in the production of the analysis itself.

Finally the ten meters wind fields forecasted by the two assimilations at 12 UTC of the 8th of July have similar structure with some changes in intensity and direction more evident

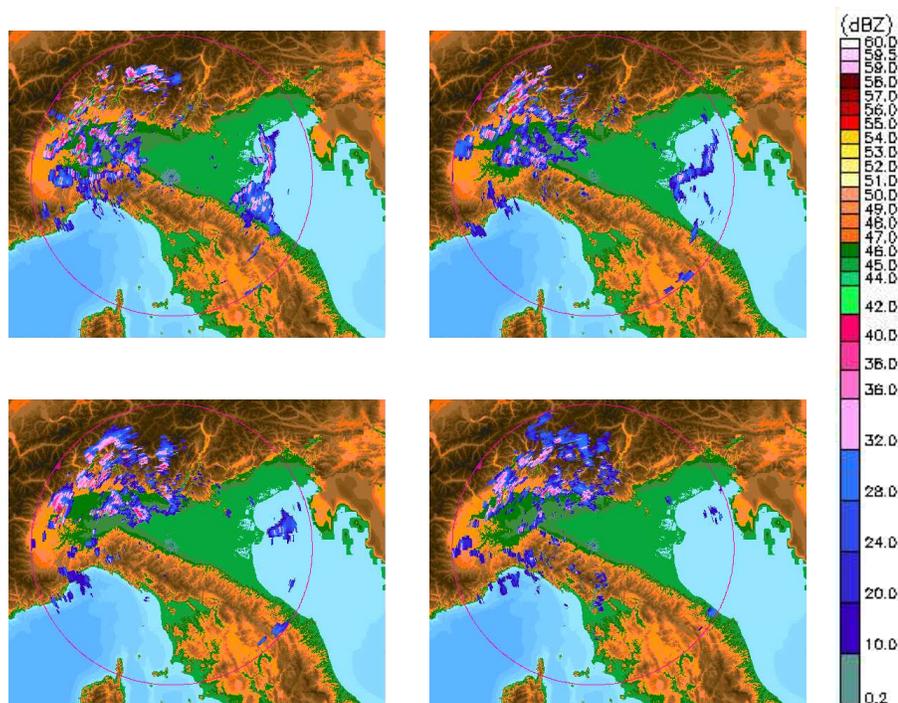


Figure 4.11: Radar reflectivity in dBZ at the medium range observed by the radar localised at Gattatico at 10:12 UTC (top left panel), 10:42 UTC (top right panel), 11:12 UTC (bottom left panel) and 11:42 UTC (bottom right panel) of the 8th of July 2004 respectively

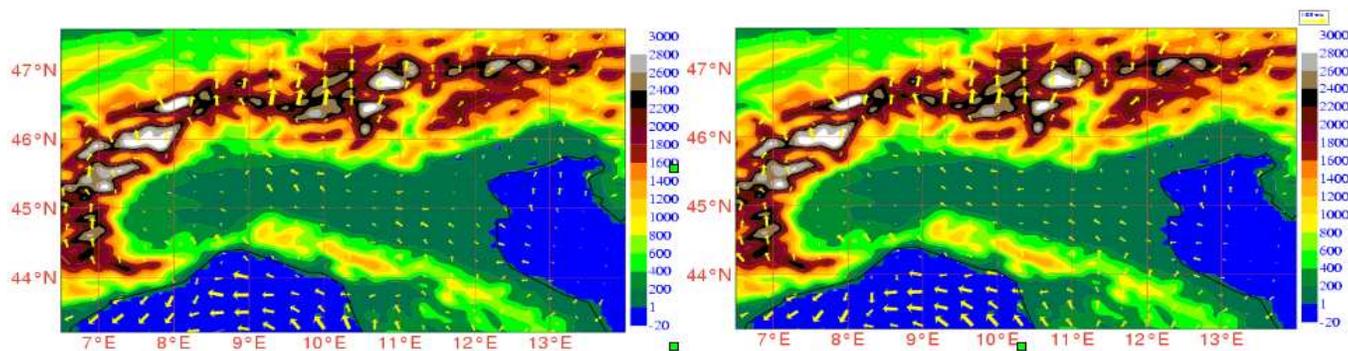


Figure 4.12: 10 meters wind field in m/s forecasted at the end of the assimilation window for ASS_MSG1_20040708 (left panel) and ASS_NUDG_20040708 (right panel) plotted together with the orography field (in meters).

in the Ligurian sea. (see Fig. 4.12)

4.1.2 Forecast

The amplitude of the differences of the integrated water vapour with respect to the saturation integrated water vapour ones remains greater also over the forecast period. This means that this greater impact for the integrated water vapour persists also during the forecast, where

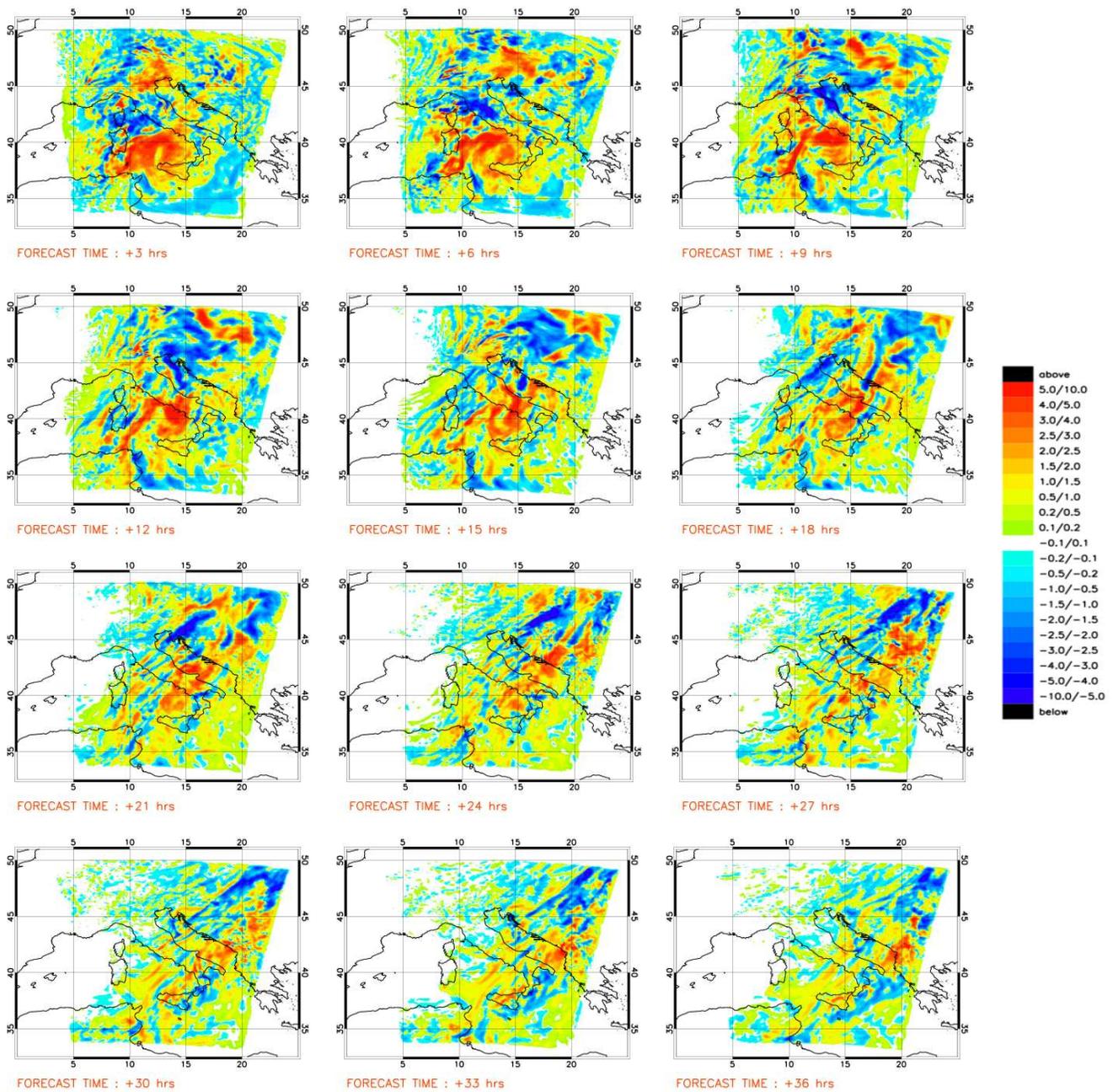


Figure 4.13: Differences in kg/m^2 between FOR_MSG1_20040708 and FOR_NUDG_20040708 of the integrated water vapour in the column during the forecast temporal window with a temporal frequency of 3 hours.

the data assimilation is not performed (see Fig. 4.13 and 4.14).

In particular the largest amplitude of the integrated water vapour differences is initially localised over the Tyrrhenian sea with positive differences in the south and negative differences in the north respectively. Carrying on in forecast time the differences are shifted

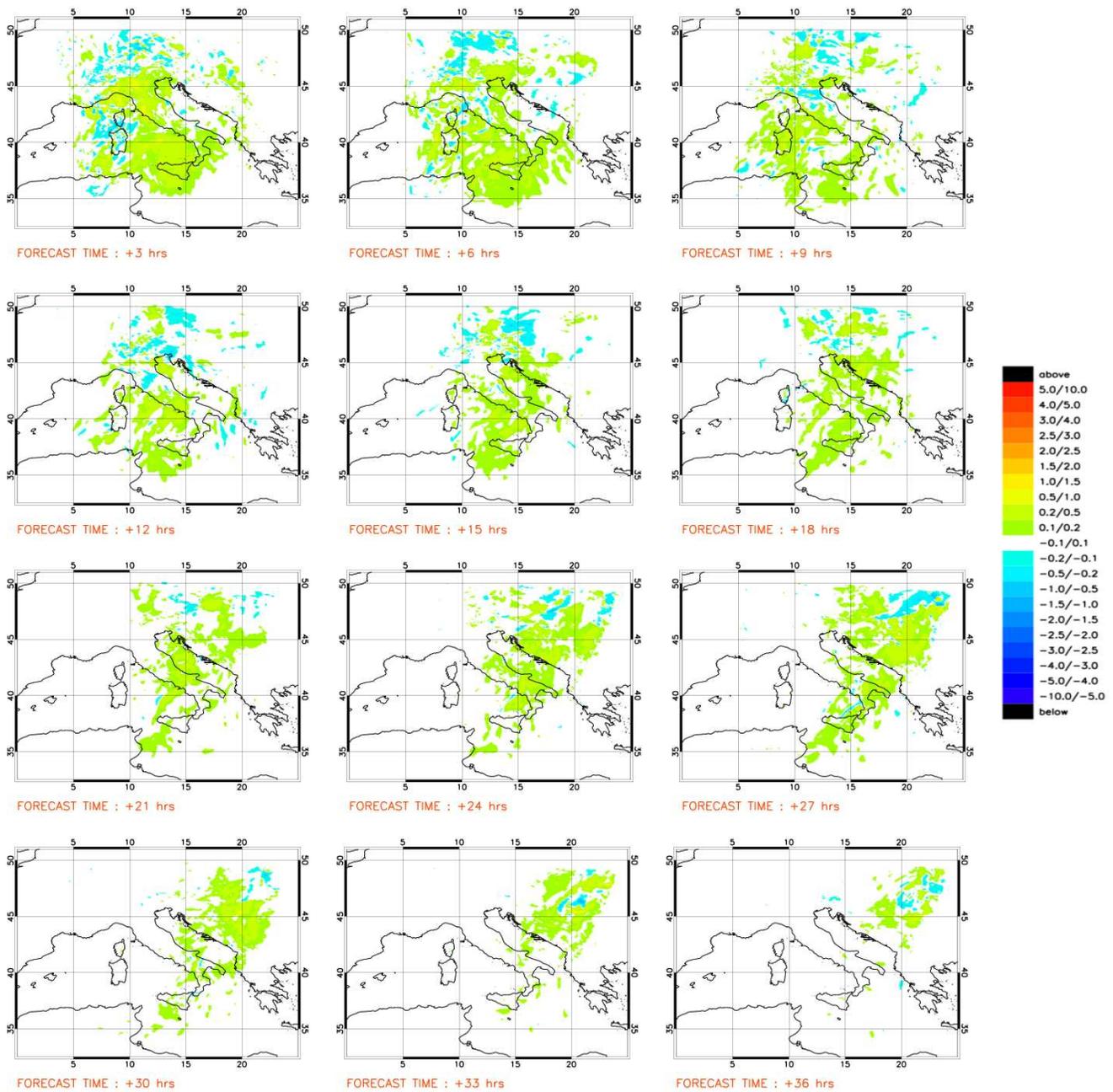


Figure 4.14: Differences in kg/m^2 between FOR_MSG1_20040708 and FOR_NUDG_20040708 of the integrated saturation water vapour in the column during the forecast temporal window with a temporal frequency of 3 hours.

towards the east north–east direction with a correspondent reduction of the differences in the north–west of COSMO–LAMI domain.

This effect is due to the fact that the boundary conditions have more impact on the North–Western part because the flow at high elevation comes from that direction. The reduction is

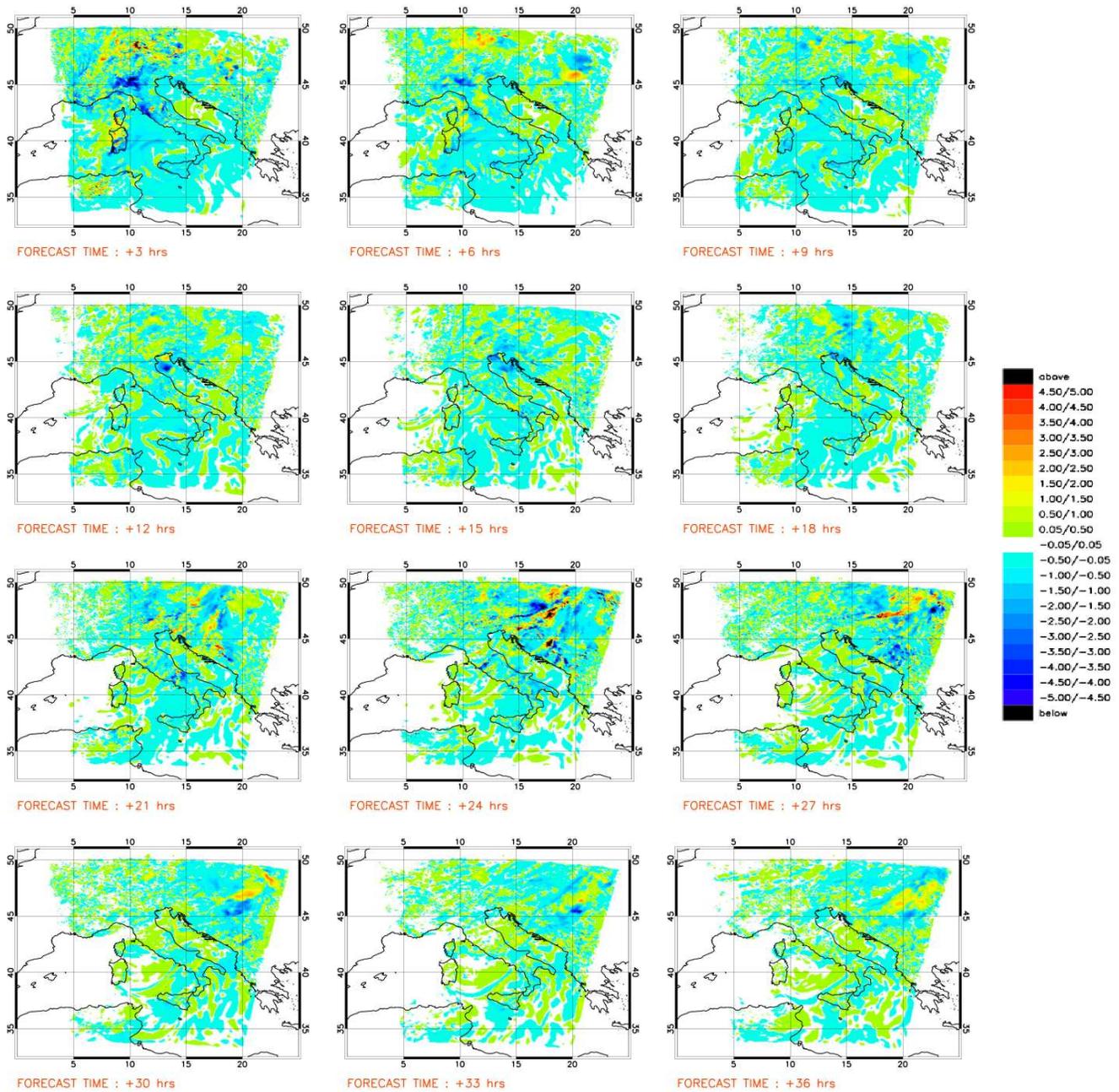


Figure 4.15: Differences in K between FOR_MSG1_20040708 and FOR_NUDG_20040708 of the T_{2m} during the forecast temporal window with a temporal frequency of 3 hours.

even more evident for the integrated saturation water vapour (see Fig. 4.14); the differences initially localised in the center part of COSMO-LAMI domain at the end of the forecast remain localised only in the north-east part of the domain itself.

The reduction of the differences is less evident for the other variables (T_{2m} , q_{2m} and T_s) (see Fig. 4.15, 4.16 and 4.17 respectively).

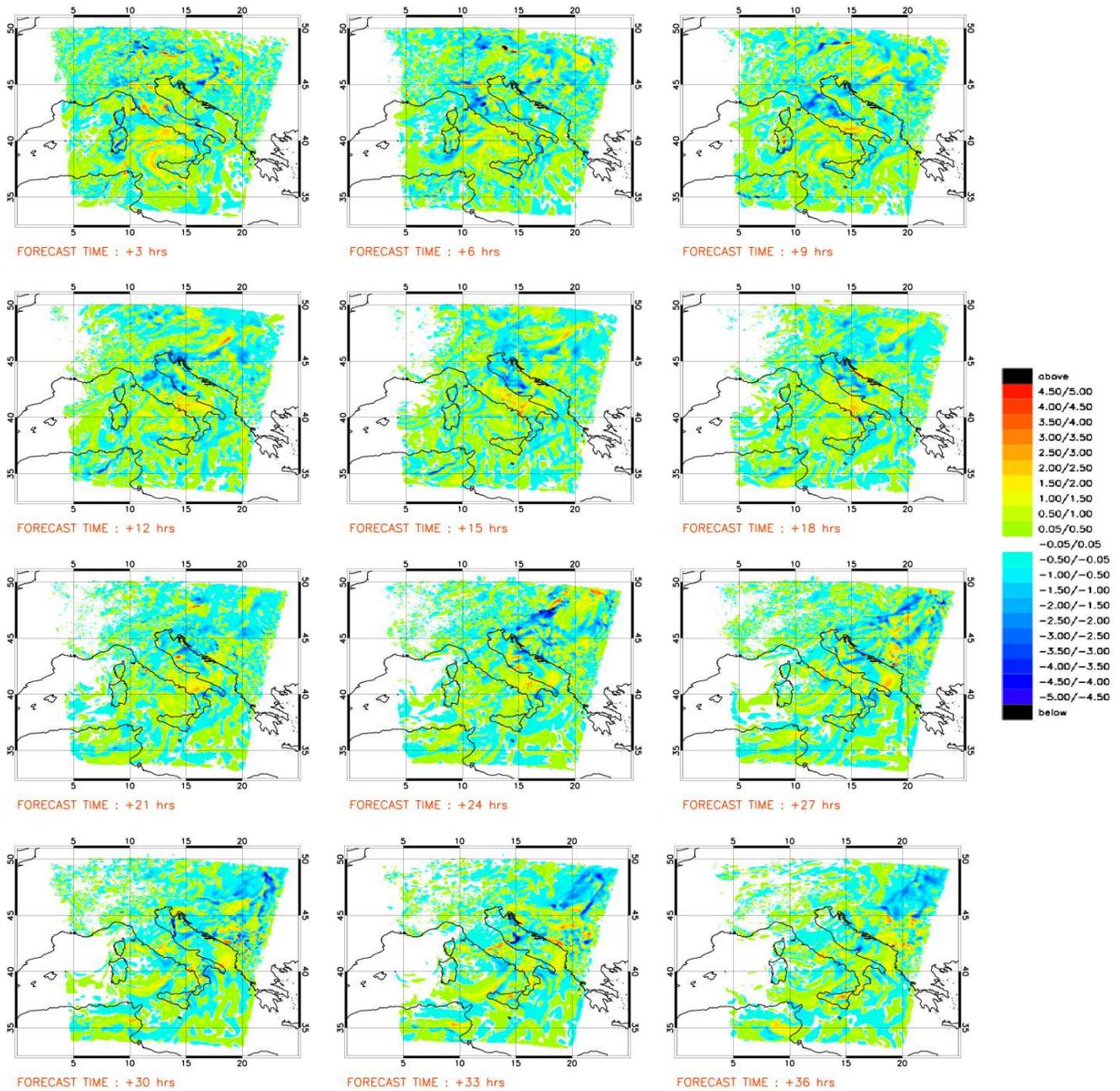


Figure 4.16: Differences in g/kg between FOR_MSG1_20040708 and FOR_NUDG_20040708 of the q_{2m} during the forecast temporal window with a temporal frequency of 3 hours.

For the 2 meters variables the difference reduction is greater over sea west of Corsica and Sardinia islands than over land. It is worth noting that the variables T_{2m} and q_{2m} are diagnosed in COSMO model from the temperature and the specific humidity at the ground and the lowest model level taking into account the atmospheric stability; this combined to the constant SST explains why the reduction of the 2 meters variables differences is larger

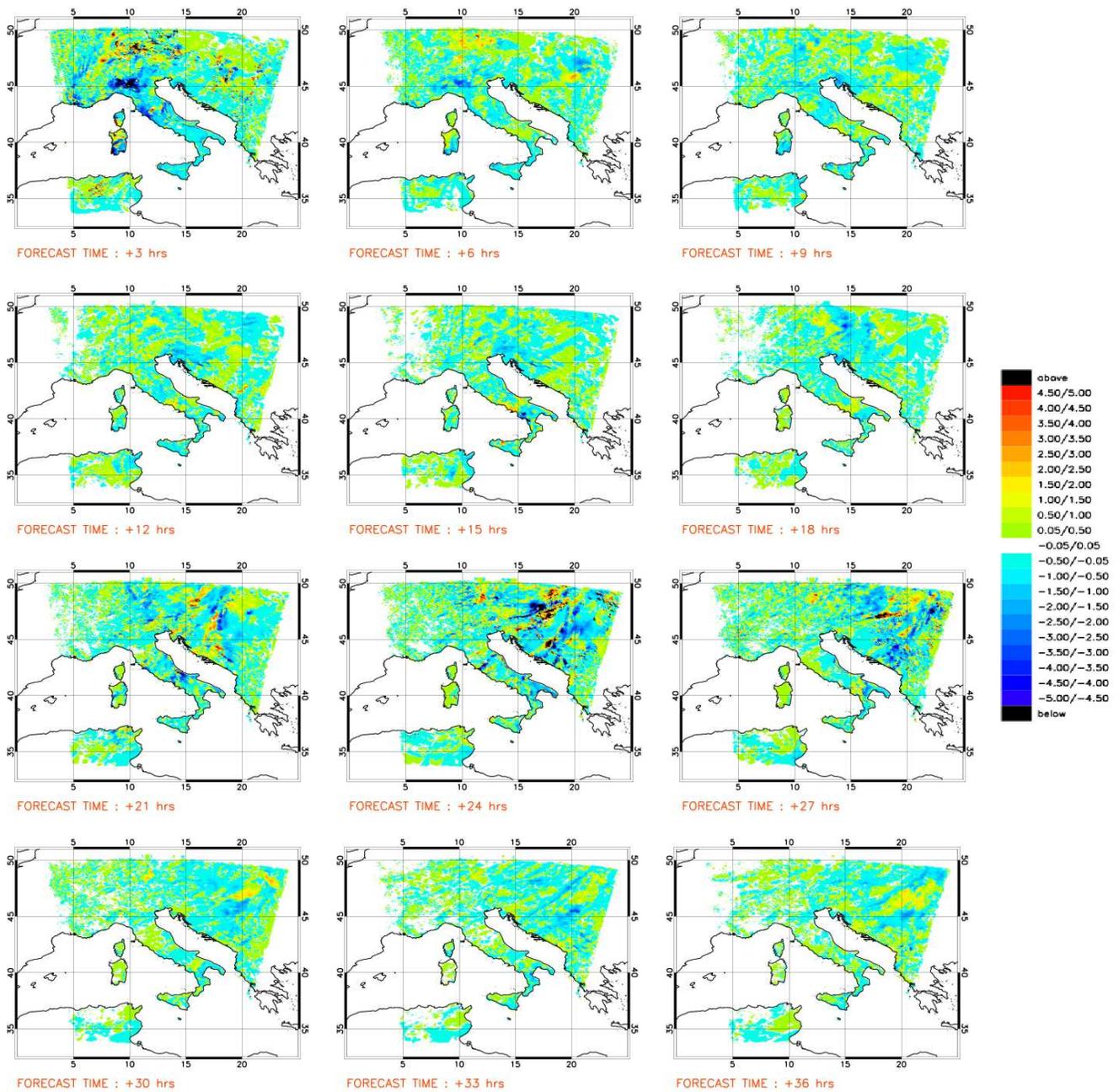


Figure 4.17: Differences in K between FOR_MSG1_20040708 and FOR_NUDG_20040708 of the T_s during the forecast temporal window with a temporal frequency of 3 hours.

over sea than over land.

The precipitation fields obtained by the runs FOR_MSG1_20040708 and FOR_NUDG_20040708 have very similar structures. The overestimation of the precipitation over the Central Alps is present in both the forecasts while the overestimations in Friuli-Venezia-Giulia, just the region where the false alarm was announced, present in FOR_NUDG_20040708 is remark-

ably reduced in FOR_MSG1_20040708. The difference of the forecast fields of precipitation shows the presence of a little phase displacement of the structures (see Fig. 4.18).

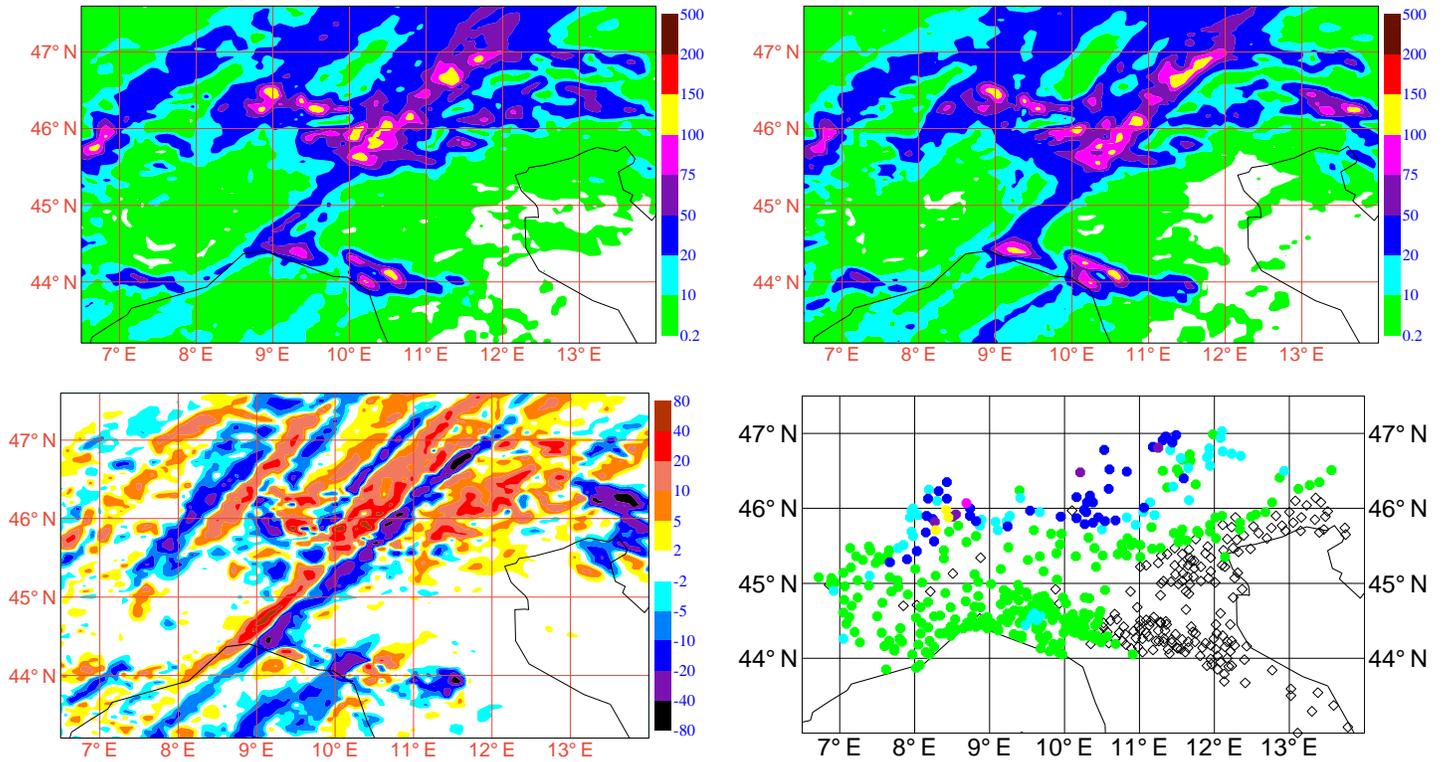


Figure 4.18: Twentyfour hours cumulated precipitation in mm from 12 UTC of the 8th of July 2004 forecasted by the runs FOR_MSG1_20040708 (top left panel) and FOR_NUDG_20040708 (top right panel), the differences (FOR_MSG1_20040708 minus FOR_NUDG_20040708) of the two forecast precipitation fields (bottom left panel) and the correspondent observations (bottom right panel).

4.2 A heavy precipitation case–study : 9th of April 2005

The meteorological situation for this event determined heavy precipitations over the whole Emilia–Romagna region during the two days of 10th and 11th of April 2005 due to the presence of a cyclonic depression over the Central–Western Mediterranean basin. This synoptic situation (see Fig. 4.19) was triggered by the arrival of a polar air mass moved from its original position in Greenland towards to lower latitude by a jet stream present on the north of the Atlantic.

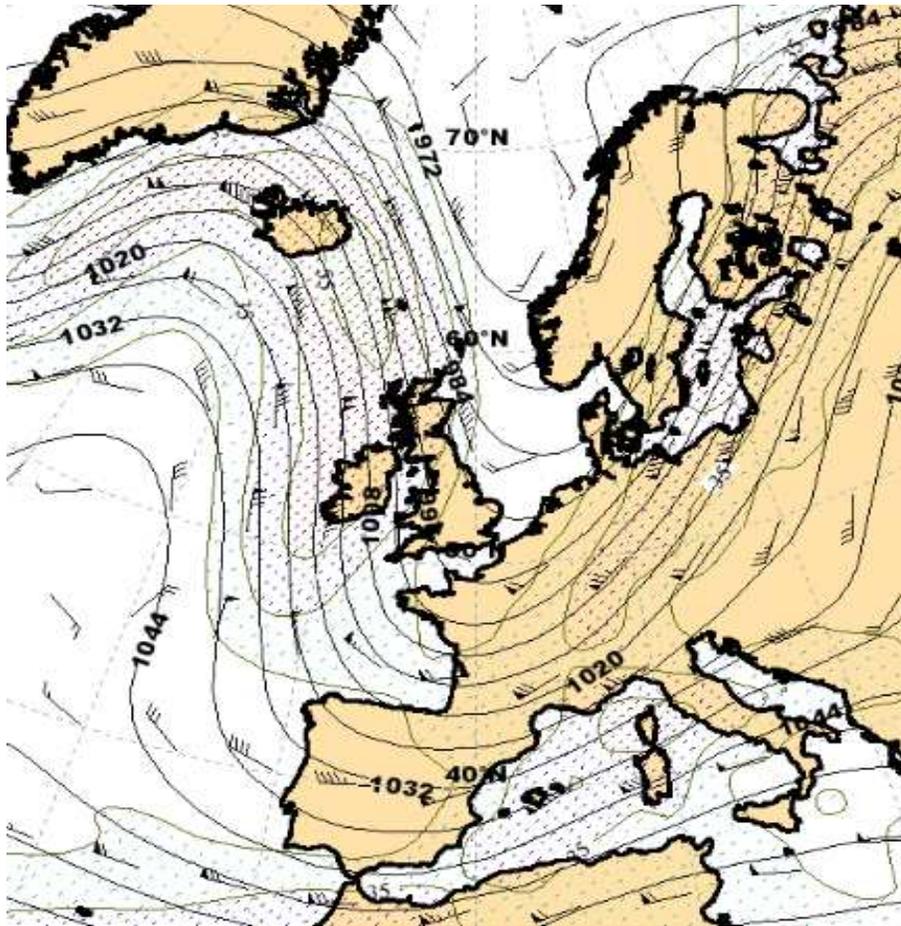


Figure 4.19: European Center for Medium range Weather Forecast (ECMWF) analysis at 250 hPa at 12 UTC of 8th of April 2005.

The arrival of the cold front happened in the night between the 8th and the 9th of April and produced the most significant effects in terms of precipitation in the following days. The cold air reached the Algerian coasts and triggered a great warm air flow from the African land towards the central part of the Mediterranean sea. This determined moist and intense

winds from south–east in the south–east part of Italy which impacted strongly with the cold air present a slightly more to the west. Furthermore the extensive cloud–system, driven from the meridional winds at high altitudes, was even more intensified in Emilia–Romagna region for the presence of north–east winds at the surface which impacted with the Apennine mountains. The baric minimum on the day of 11th of April reached the position between Sardinia and Sicily islands where it stationed for the next 36 hours constraining the cloud vortex to circle the center of low pressure (see Fig. 4.20).

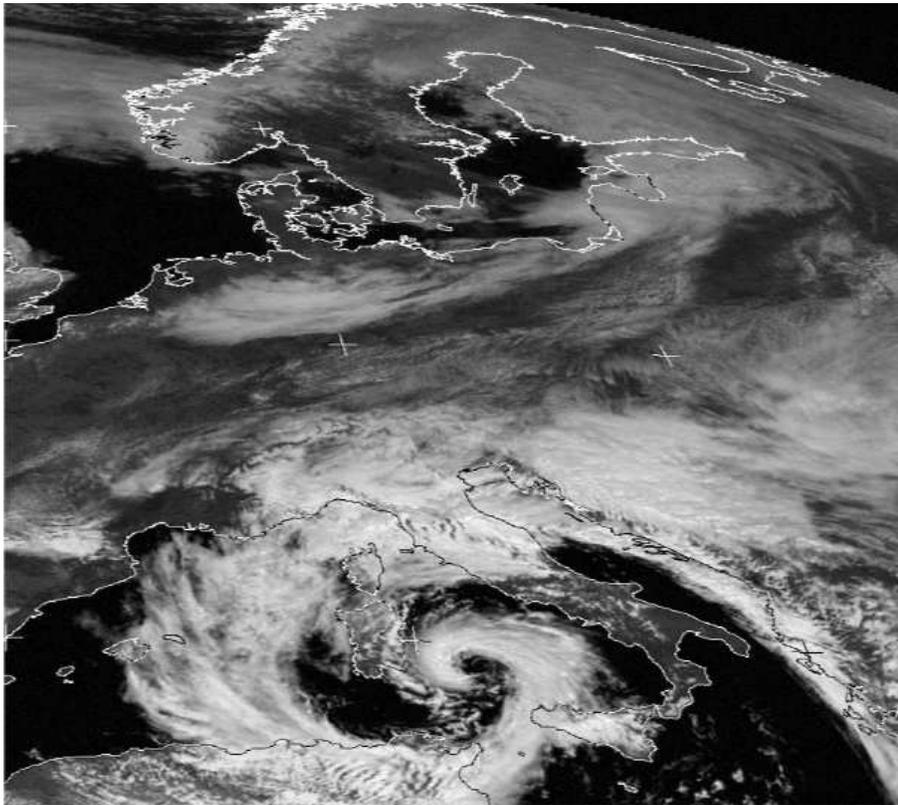


Figure 4.20: Image in the visible channel of Meteosat–7 at 12 UTC of the 11th of April 2005.

For this case–study the temporal window of the two assimilations start at 12 UTC of the 9th of April 2005 and lasts 12 hours. The forecasts are performed for the next 36 hours starting at 00 UTC of the 9th of April (see Fig. 4.21). Similarly for the false alarm case previously discussed, to distinguish between the two assimilations and the two forecasts, with and without the use of the MSG–1 satellite data, they are hereafter called respectively:

- ASS_MSG1_20050409 and FOR_MSG1_20050409 for the assimilation and the forecast where MSG–1 satellite data are assimilated using the 1D–VAR approach in COSMO;

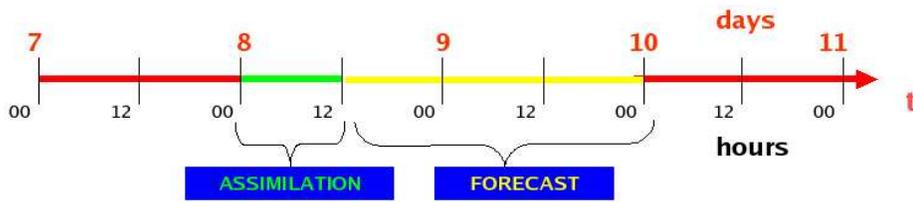


Figure 4.21: Scheme of the temporal windows for assimilations, in green colour, and forecasts, in yellow, for the heavy precipitation case–study

- ASS_NUDG_20050409 and FOR_NUDG_20050409 for the assimilation and the forecast where MSG–1 satellite data are not assimilated, so only the conventional data are assimilated by the nudging technique.

4.2.1 Assimilation

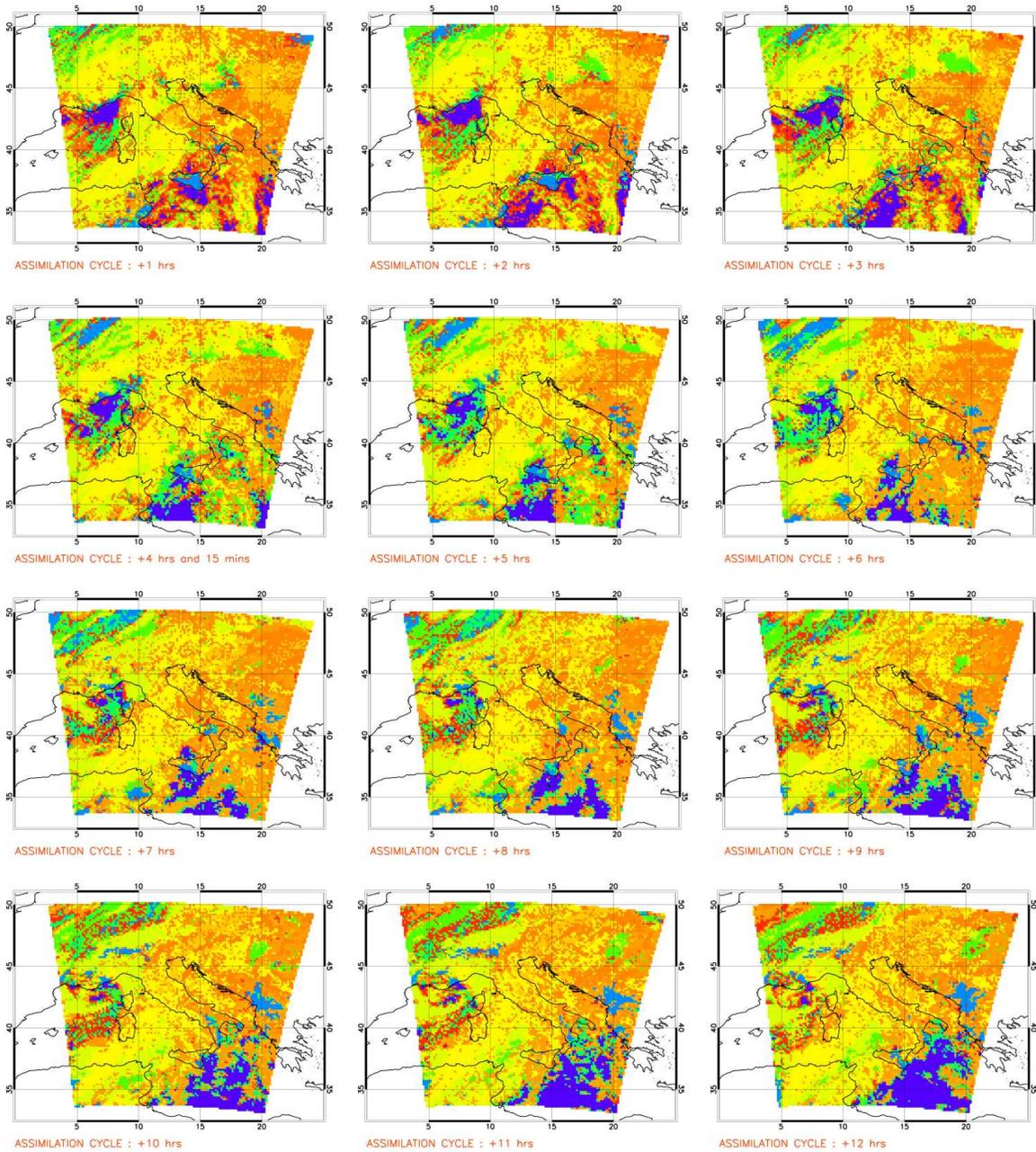


Figure 4.22: CT product for the temporal window of the assimilation with a temporal frequency of 1 hour.

The CT product (see Fig. 4.22), with the same colour palette defined in section 3.2, shows that the cloud cover in this case study is very much extended and this allows the assimilation of a small number of observed radiances as shown in the Fig. 4.23 (the colour palette used is the one defined in section 3.2). Most of the assimilated satellite observations are located in the Ligurian sea and south of Sicily island.

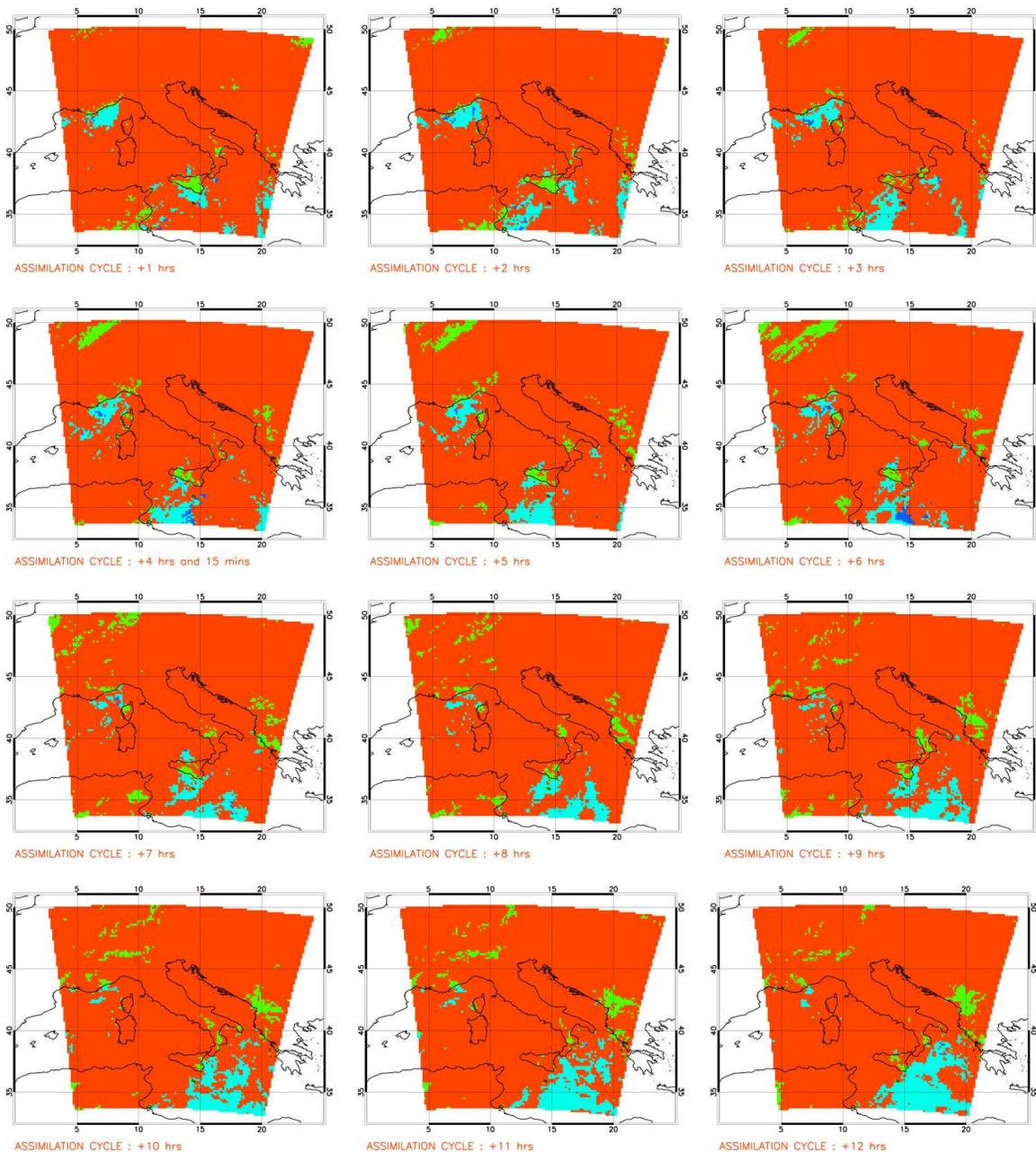


Figure 4.23: flg_prc for the temporal window of the assimilation with a temporal frequency of 1 hour.

For the integrated water vapour the initial differences, localised in the Ligurian sea and in the south of Italy, are negative; they are spread in space during the assimilation temporal window by the nudging. At the end of the assimilation cycle the most negative differences are present in the Ionian sea, while the most positive differences are present in the south of the Adriatic sea, south-west of Sardinia and in the Ligurian sea. The integrated saturation

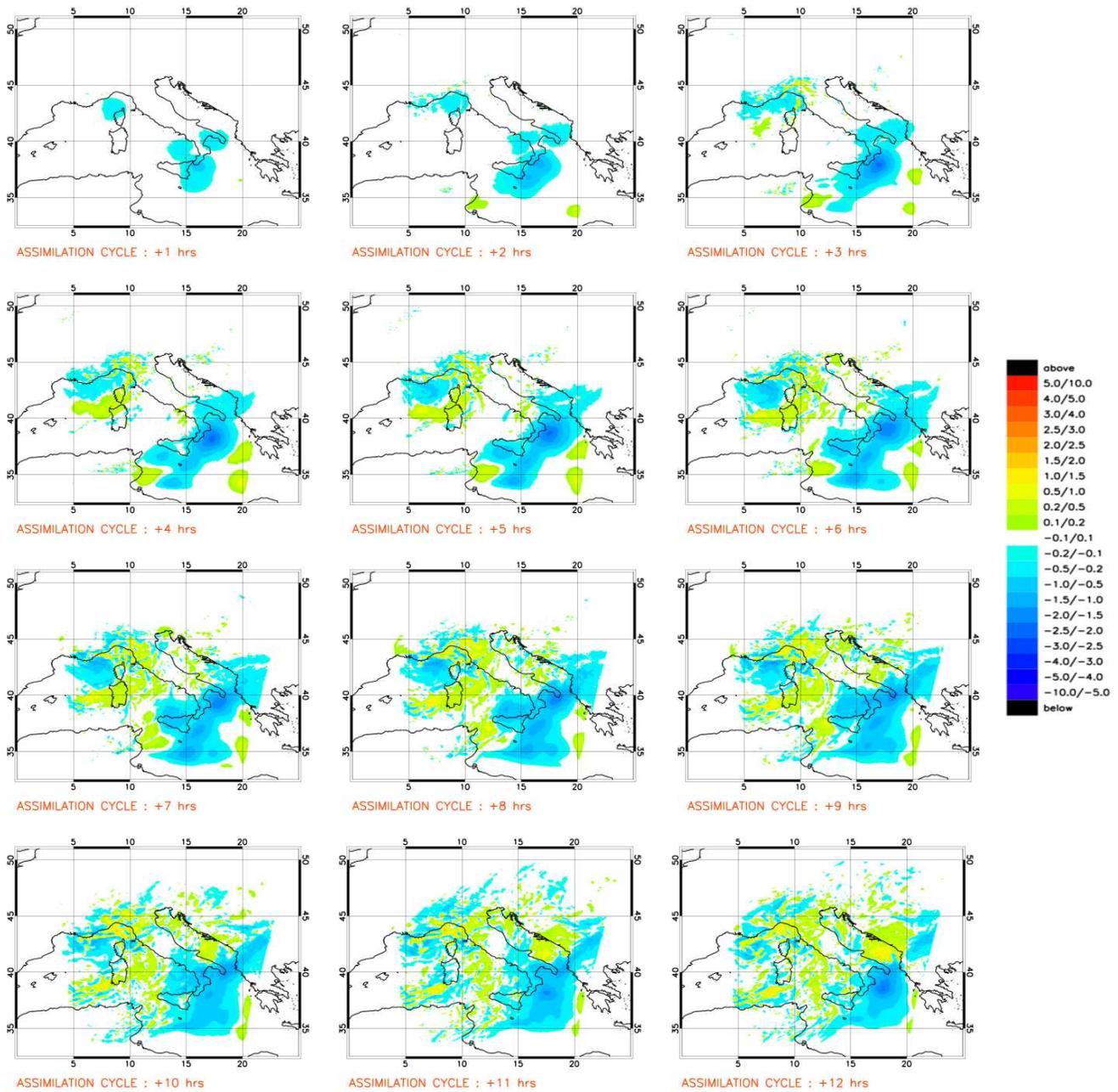


Figure 4.24: Differences in kg/m^2 between ASS_MSG1_20050409 and ASS_NUDG_20050409 of the integrated water vapour in the column during the assimilation temporal window with a temporal frequency of 1 hour.

water vapour differences (not shown) are almost negligible over the whole assimilation cycle.

For the 2 meters surface variables, T_{2m} and q_{2m} (see Fig. 4.25 and 4.26), initially the differences are concentrated in the south of Italy and west of Tunisian coasts and subsequently they are spread from this original position. For this case–study the amplitude of the

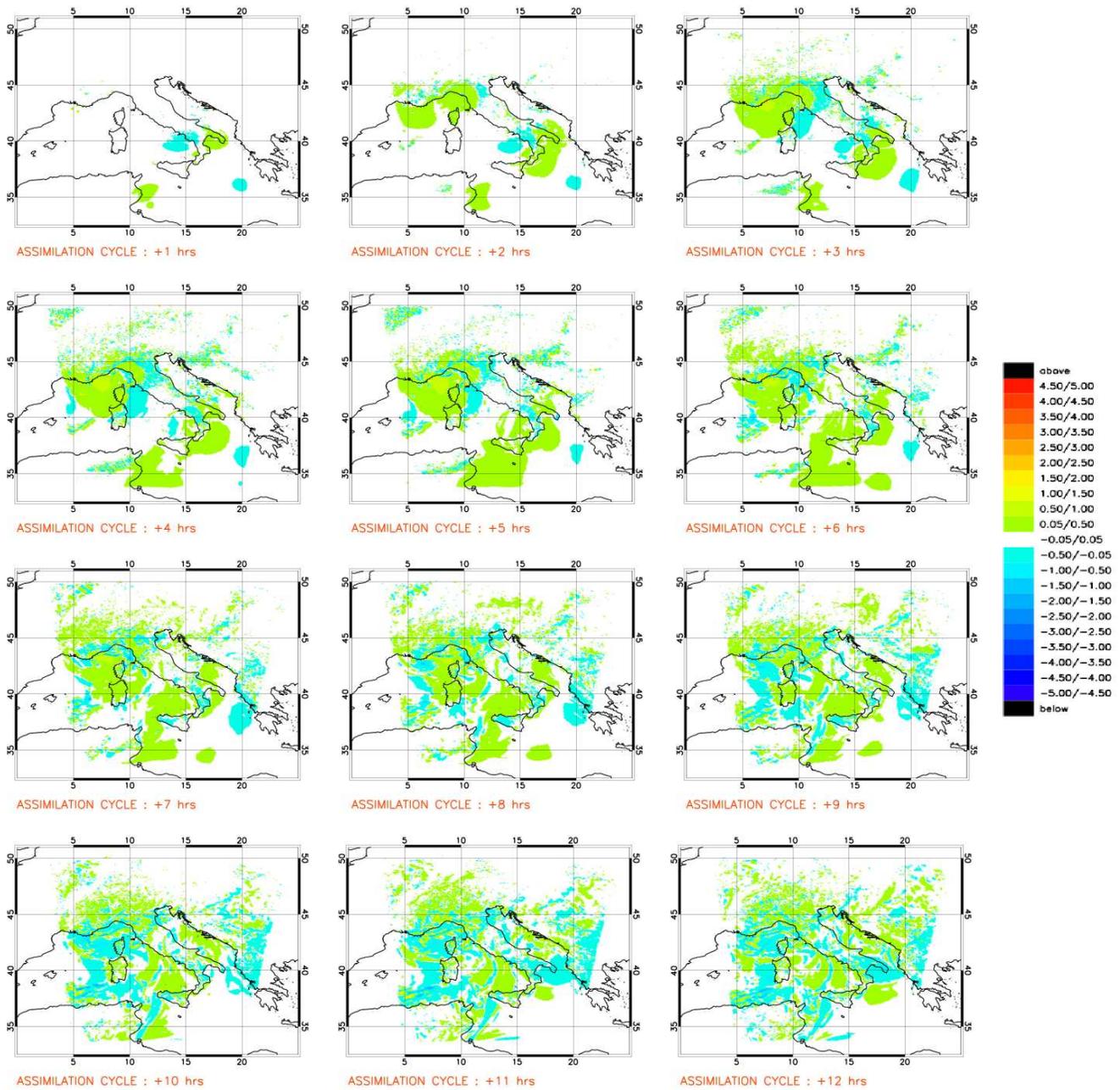


Figure 4.25: Differences in K between ASS_MSG1_20050409 and ASS_NUDG_20050409 of the forecasted T_{2m} during the assimilation temporal window with a temporal frequency of 1 hour.

differences at the end of the assimilation are rarely greater than 0.5 K for the T_{2m} and 0.5 g/kg for the q_{2m} respectively. It is worth noting the almost total absence of impact in the q_{2m} over the Alps.

Finally T_s differences (see Fig. 4.27) are very similar to the differences in the T_{2m} over land, and also for this variable the differences are rarely greater than 0.5 K.

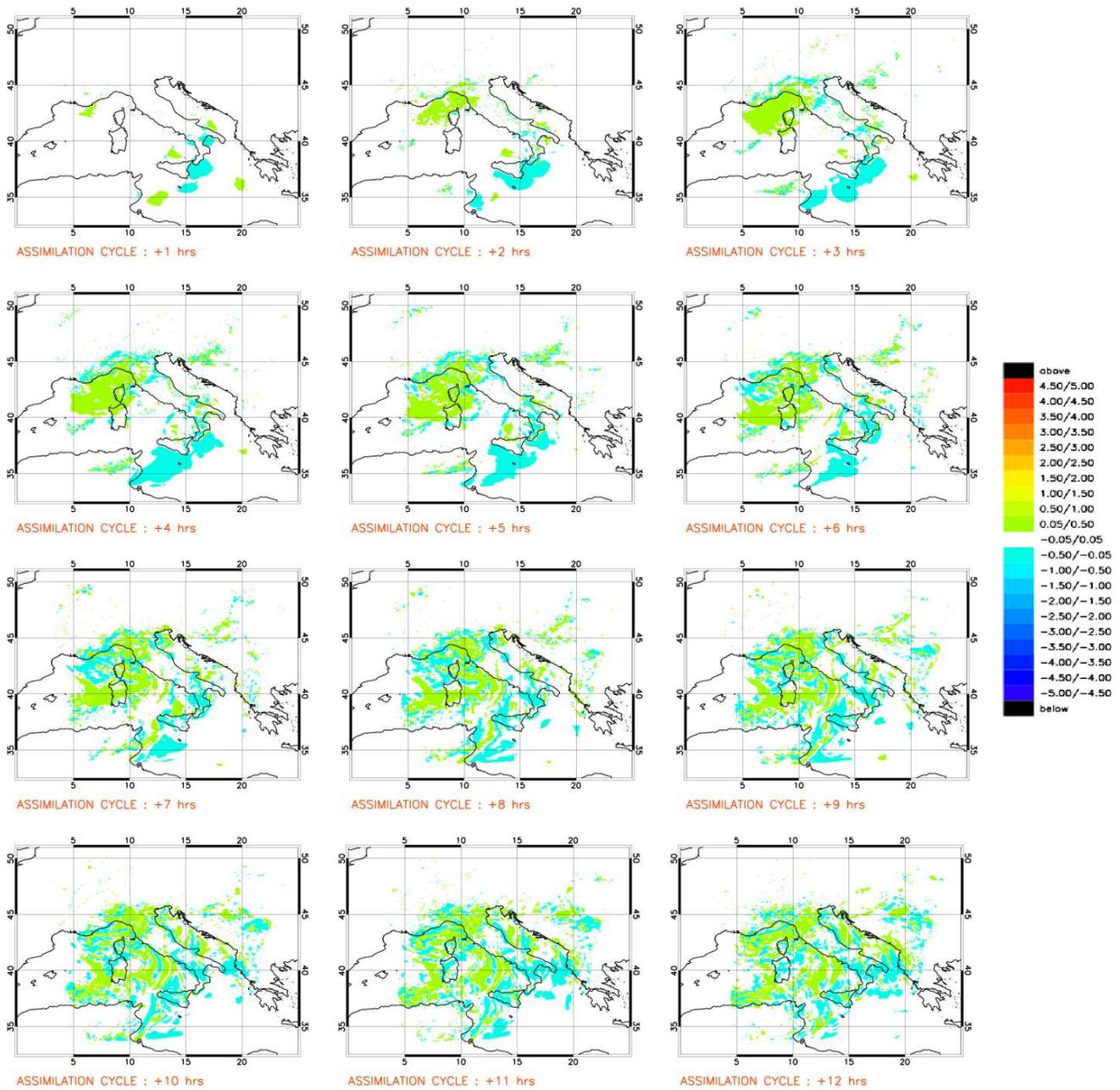


Figure 4.26: Differences in g/kg between ASS_MSG1_20050409 and ASS_NUDG_20050409 of the forecasted q_{2m} during the assimilation temporal window with a temporal frequency of 1 hour.

The forecast precipitation fields for the twelve hours of the two assimilations show similar structures with very small differences whose greatest amplitude ranges around 10 mm. The correspondent observations show that both the forecasts produce a slight overestimation of the precipitation in the Veneto region, while the other maxima of precipitation, present in Piemonte region and in the Po valley, are well localised (see Fig. 4.28).

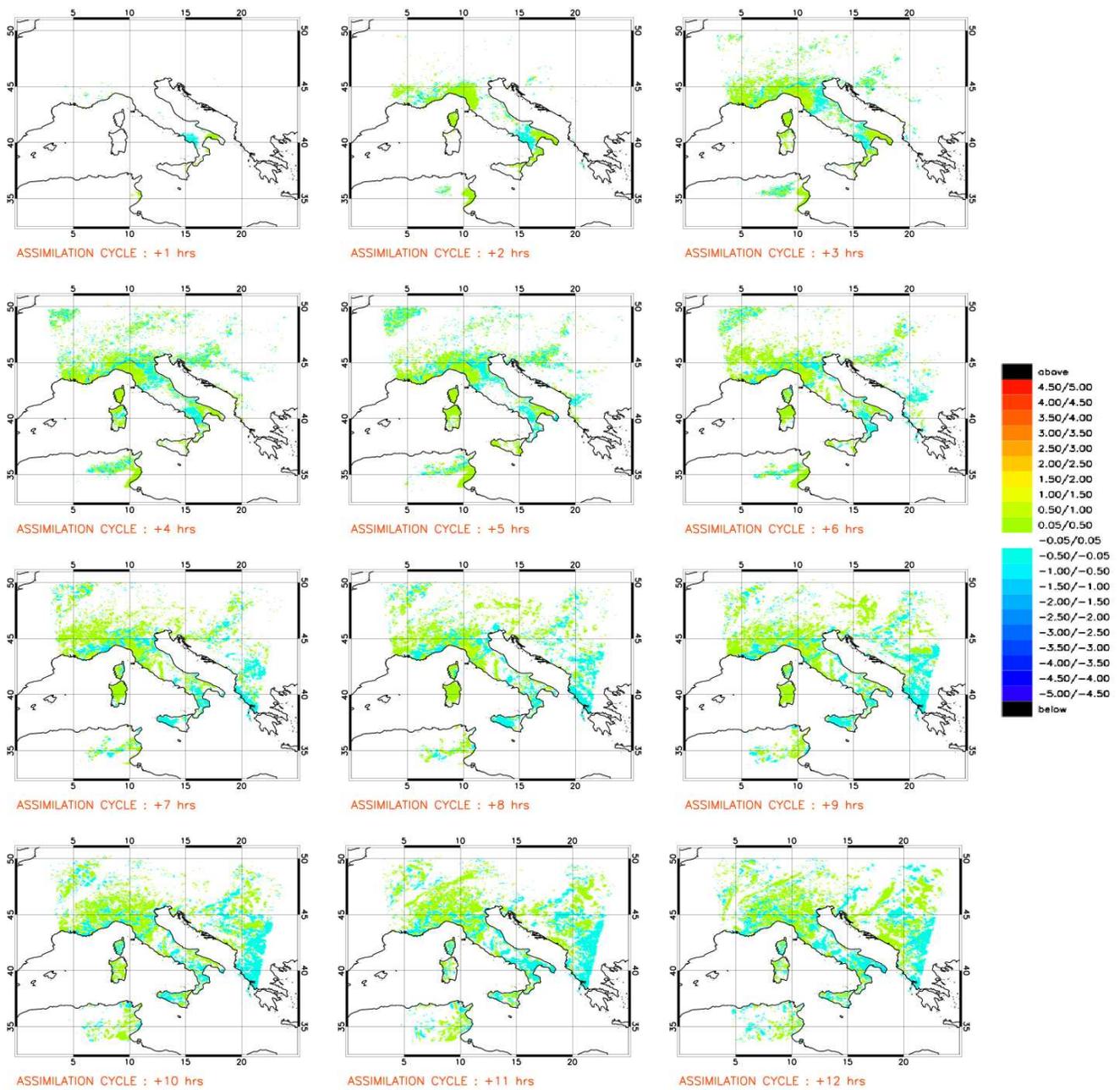


Figure 4.27: Differences in K between ASS_MSG1_20050409 and ASS_NUDG_20050409 of the forecasted T_s during the assimilation temporal window with a temporal frequency of 1 hour.

Finally the forecasted 10 meters wind fields forecasted by the assimilations at 00 UTC of 10th of April are identical and no such remarkable change can be reported (see Fig. 4.29).

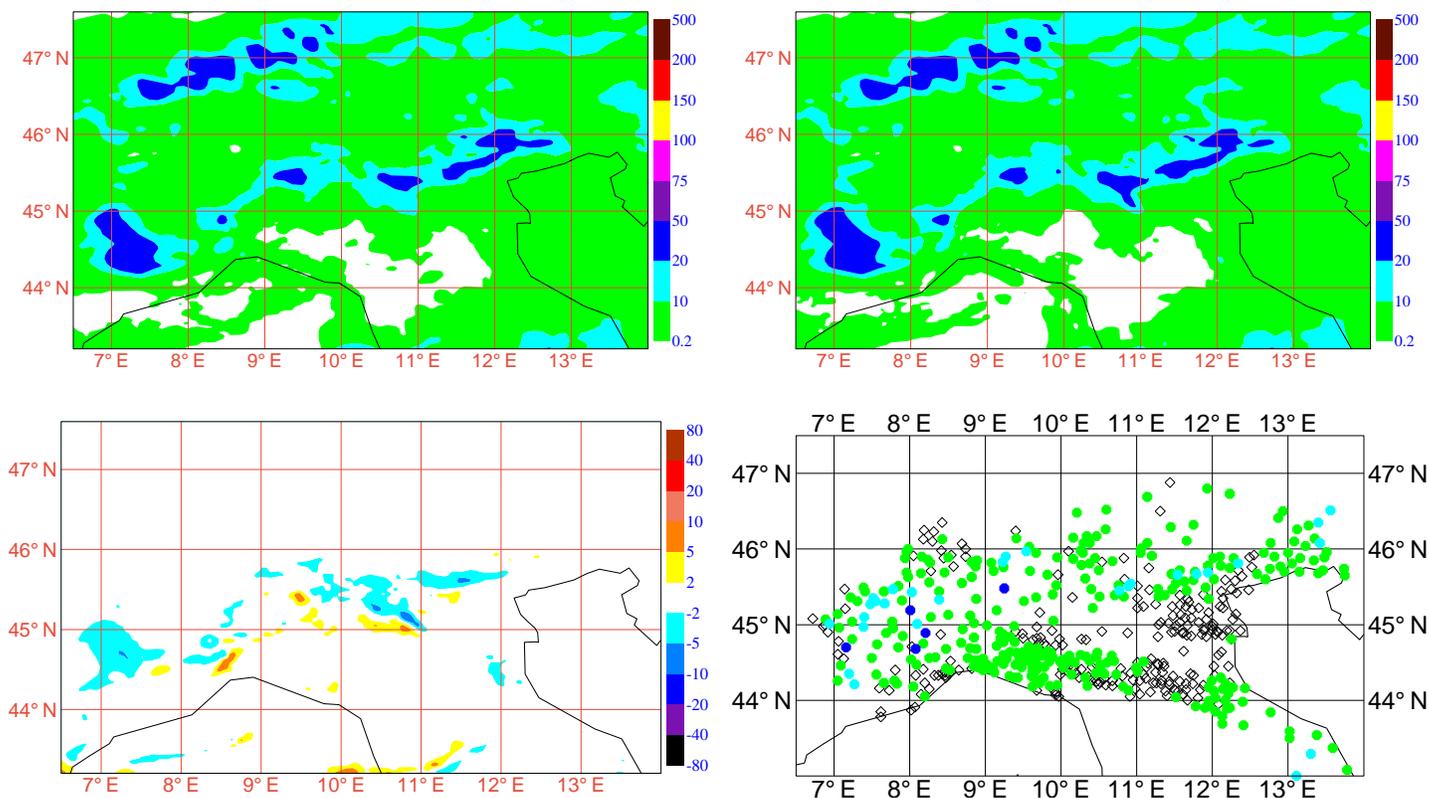


Figure 4.28: Twelve hours cumulated precipitation in mm from 12 UTC of the 9th of April 2005 forecasted by the runs ASS_MSG1_20050409 (top left panel) and ASS_NUDG_20050409 (top right panel), the differences (ASS_MSG1_20050409 minus ASS_NUDG_20050409) of the two forecast precipitation fields (bottom left panel) and the correspondent observations (bottom right panel).

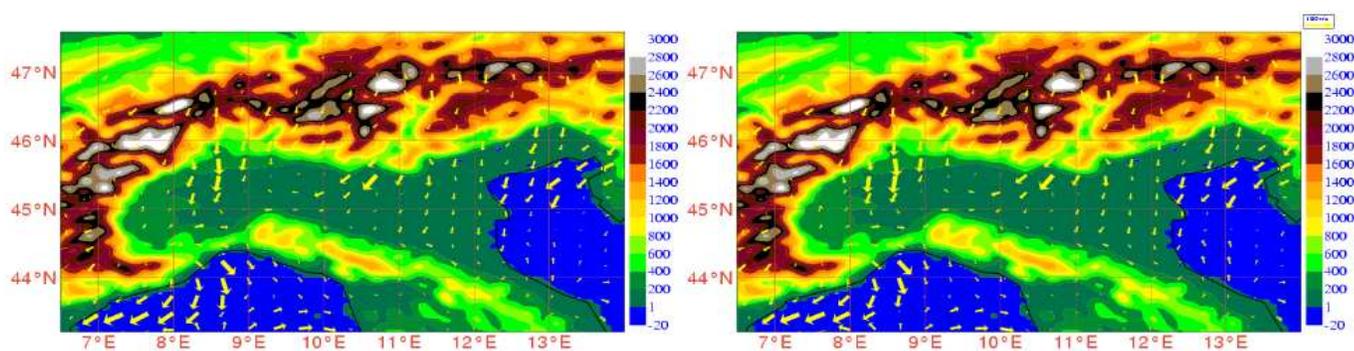


Figure 4.29: 10 meters wind field in m/s forecasted at the end of the assimilation window for ASS_MSG1_20050409 (left panel) and ASS_NUDG_20050409 (right panel) plotted together with the orography field (in meters).

4.2.2 Forecast

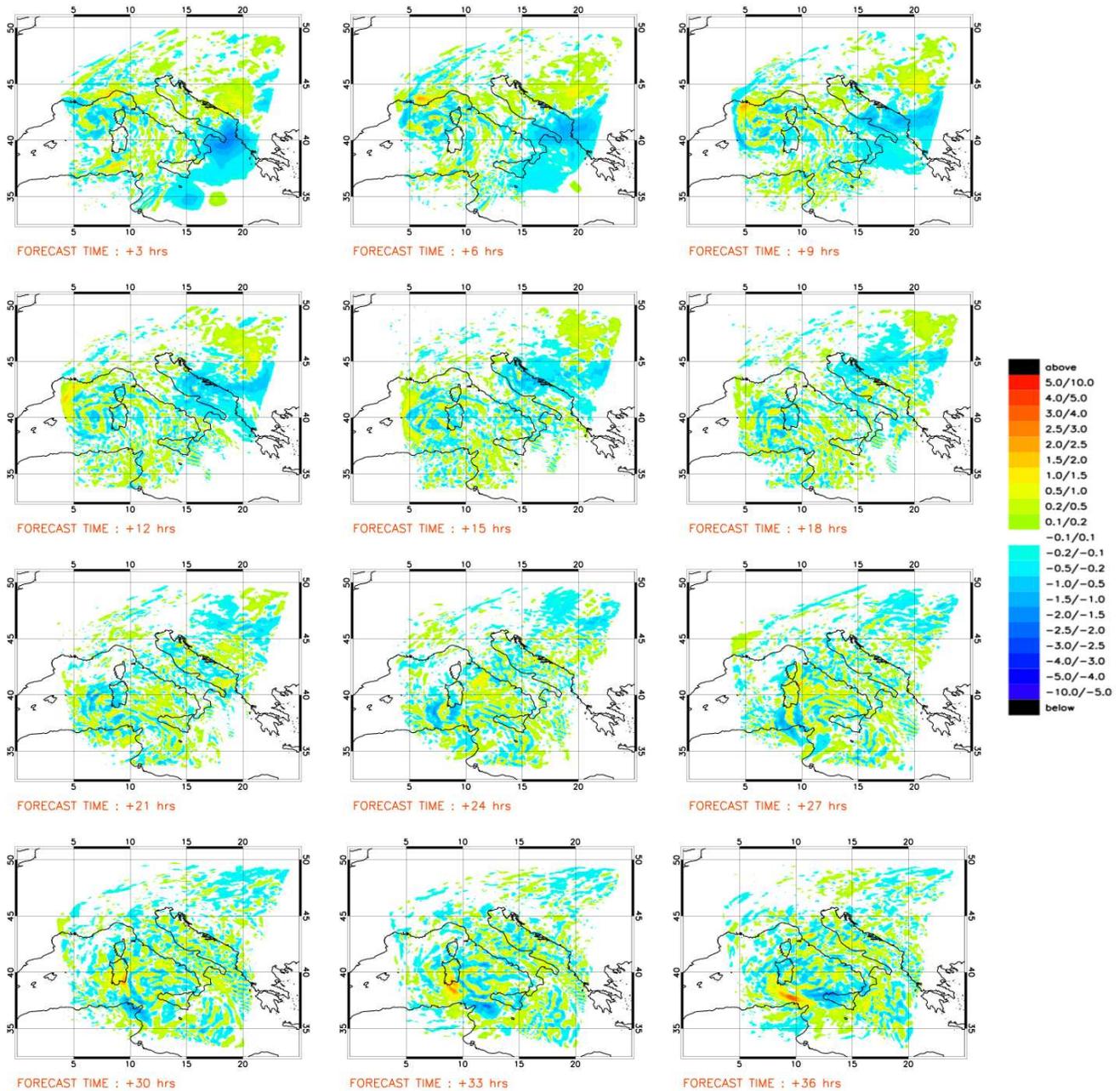


Figure 4.30: Differences in kg/m^2 between FOR_MSG1_20050409 and FOR_NUDG_20050409 of the integrated water vapour in the column during the forecast temporal window with a temporal frequency of 3 hours.

At the beginning of the forecast most of the negative integrated water vapour differences are localised in the Ionian sea and south of Sicily, while the most positive differences are present on the south-east of France. The differences seem to have a general cyclonic rotation in time and at the end of the forecast with the greatest differences in terms of amplitude

present close to the low pressure localised between Sicily and Sardinia islands (see Fig. 4.30). The integrated saturation water vapour, like during the assimilation cycle, have always values concentrated between -0.1 and 0.1 kg/m^2 (not shown).

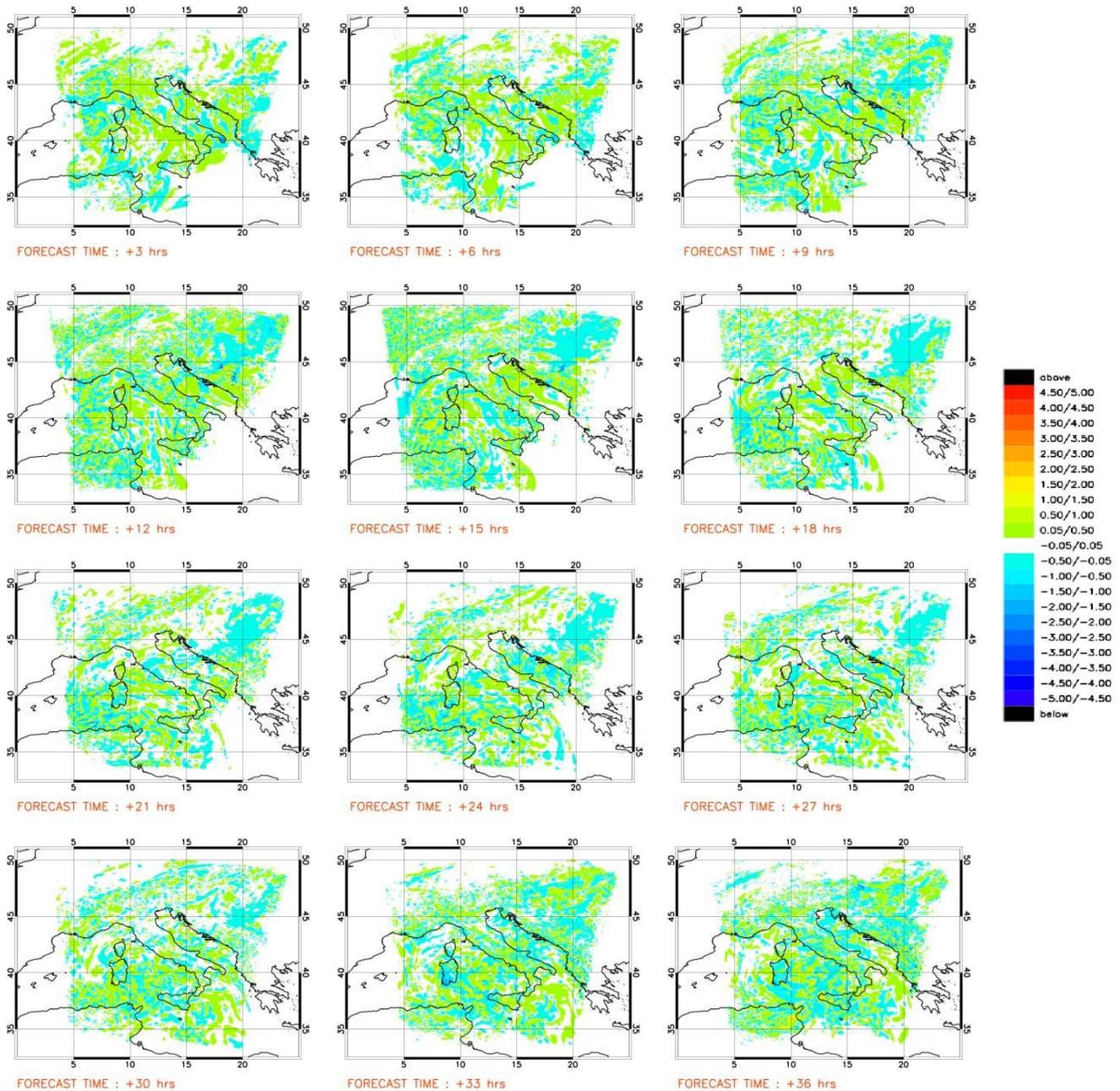


Figure 4.31: Differences in K between FOR_MSG1_20050409 and FOR_NUDG_20050409 of the T_{2m} during the forecast temporal window with a temporal frequency of 3 hours.

For the T_{2m} and the q_{2m} the amplitude of the differences is mostly of 0.5 K and 0.5 g/kg respectively, which means that the differences are not significant over the almost whole

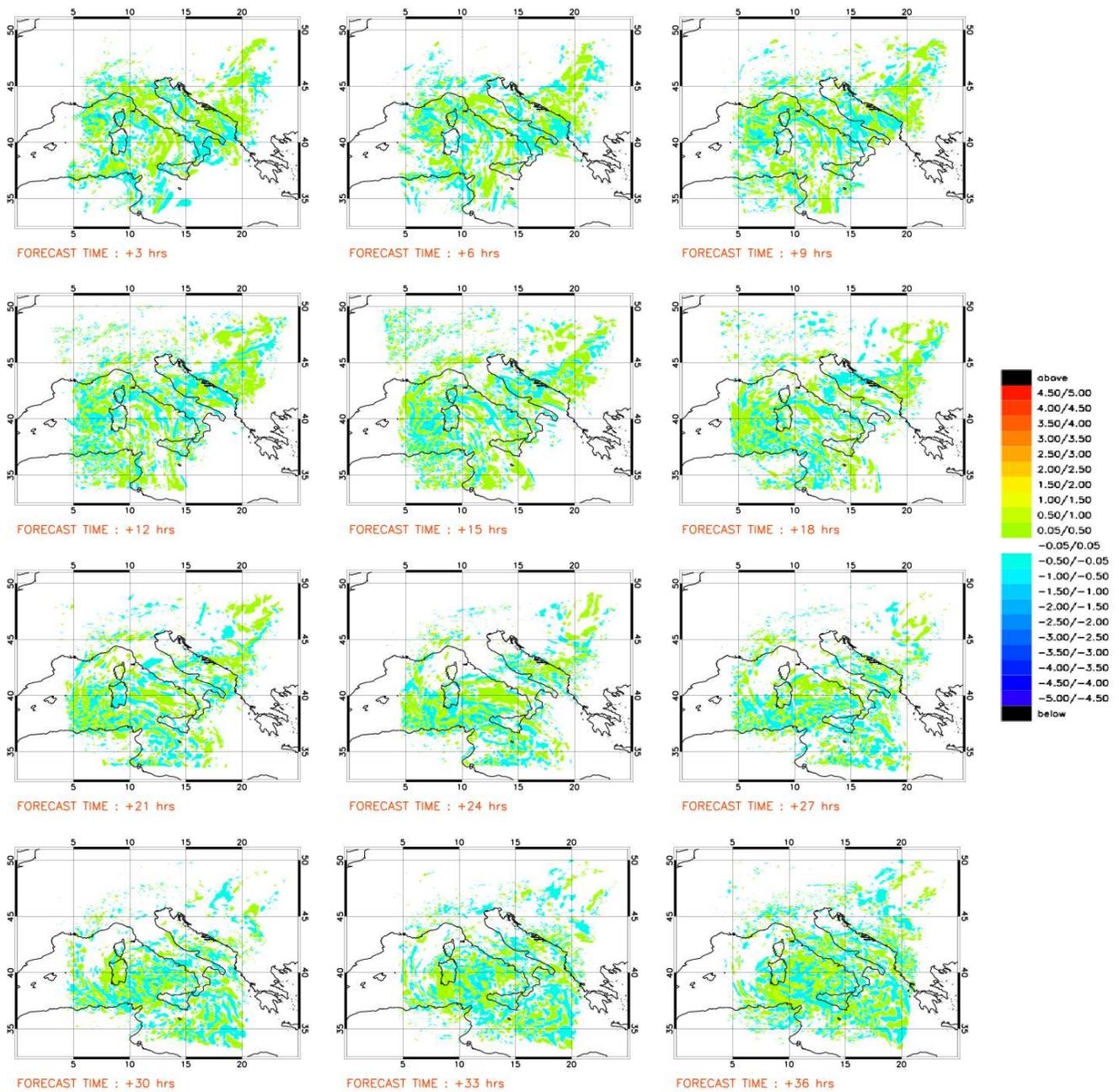


Figure 4.32: Differences in K between FOR_MSG1_20050409 and FOR_NUDG_20050409 of the q_{2m} during the forecast temporal window with a temporal frequency of 3 hours.

domain (see Fig. 4.31 and 4.32).

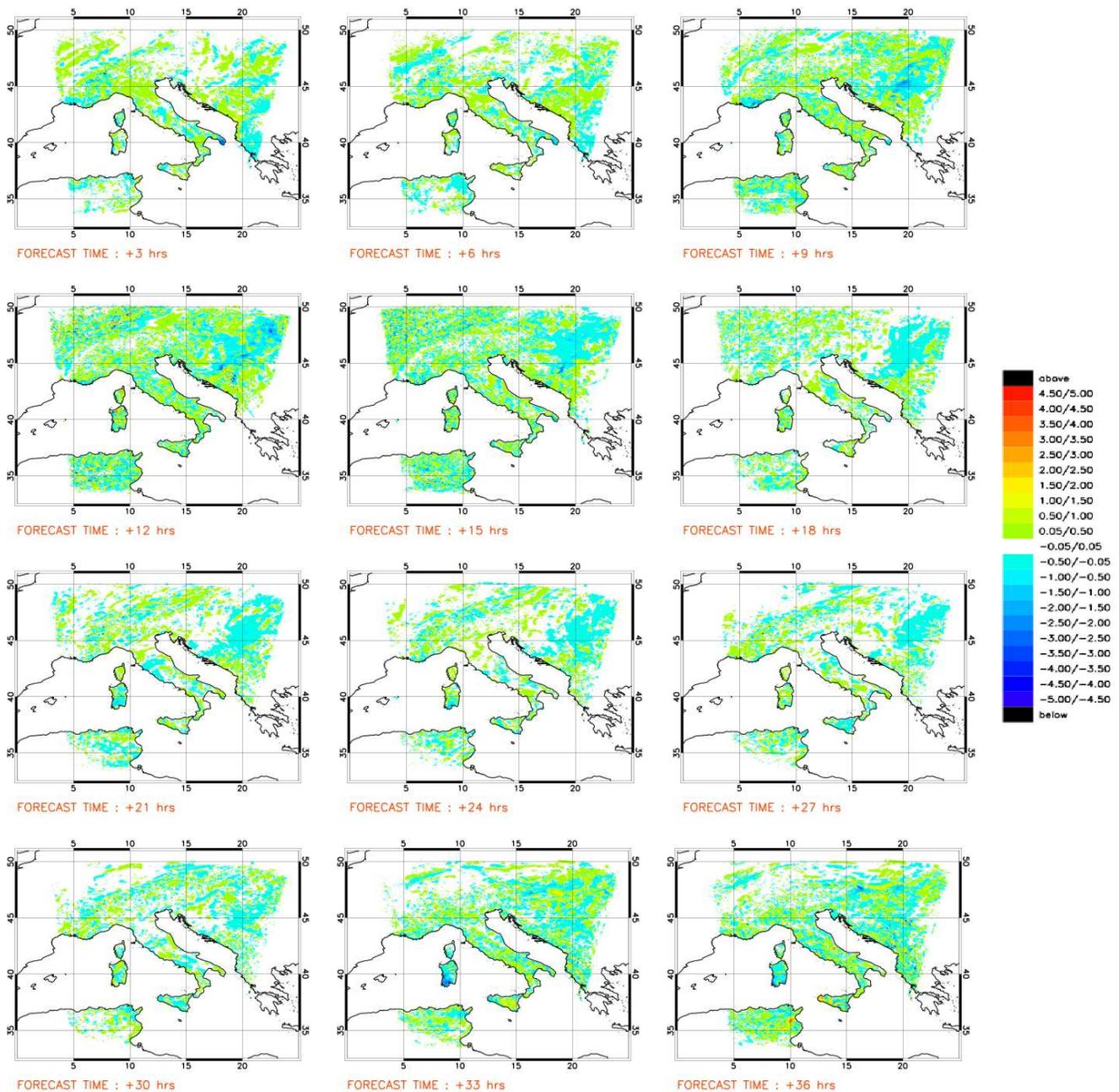


Figure 4.33: Differences in K between FOR_MSG1_20050409 and FOR_NUDG_20050409 of the T_s during the forecast temporal window with a temporal frequency of 3 hours.

The T_s differences are very similar to the 2 meters temperature differences over land with differences which almost always range from -0.5 to 0.5 K (see Fig. 4.33).

It is worth noting that for this case study, on the contrary of what happened for the false alarm case previously discussed, the variables differences, and in particular the integrated water vapour differences, do not reduce very much in the forecast. This could be due to

the presence of a stationary vortex of low pressure between Sicily and Sardinia islands; in fact this vortex reduces the impact of the boundary conditions hourly provided by the GME model and determines a more persistence in the differences themselves.

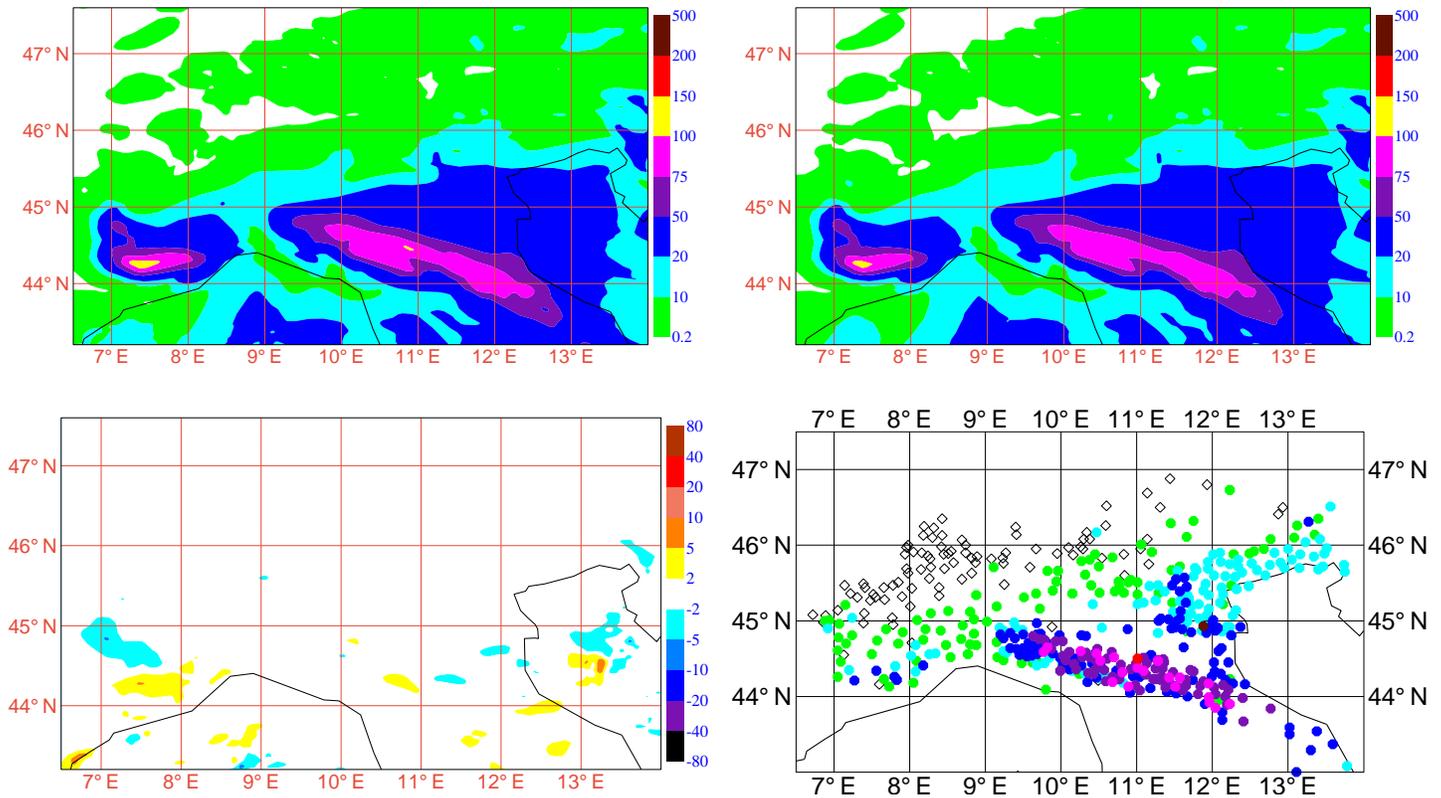


Figure 4.34: Twentyfour hours cumulated precipitation in mm from 00 UTC of the 10th of April 2005 forecasted by the runs FOR_MSG1_20050409 (top left panel) and FOR_NUDG_20050409 (top right panel), the differences (FOR_MSG1_20050409 minus FOR_NUDG_20050409) of the two forecast precipitation fields (bottom left panel) and the correspondent observations (bottom right panel).

The precipitation field forecasted by the two runs FOR_MSG1_20050409 and FOR_NUDG_20050409 are very similar; it is worth noting that even if the amount of forecasted precipitation is very high, especially in the Apennine areas of Emilia–Romagna region, the differences in the north of Italy are very small and ranging between -10 and 10 mm. Comparing the observations with the forecasted fields it is clear the good localisation of the peaks of the precipitation over the Emilia–Romagna Apennines but both the forecasts overestimate the precipitation in Liguria region (see Fig. 4.34).

4.3 A possible improvement of the B matrix and its use in the false alarm case–study

Up to now following the Parrish–Derber methodology the 12 hours forecast is assumed to be the “truth” and it is used to compute the background error covariance matrix \mathbf{B} . Nevertheless this statistical method does not take into account the synoptic situation of the moment which represents the main forcing of the meteorological system. Hence the best possible approach in the definition of \mathbf{B} is the condensation inside \mathbf{B} itself of both climatological and real–time information. Hence the calculation of “climatological” \mathbf{B} matrices has to be coupled with the calculation of synoptic situation dependent \mathbf{B} matrices. The calculation of this improved \mathbf{B} matrix consists in the following three steps:

- calculation of the climatological \mathbf{B} matrices;
- calculation of the “Ensemble Islands” \mathbf{B} matrices;
- merging of the climatological and ensemble islands \mathbf{B} matrices.

This method has been applied to the false alarm case–study of 8th of July 2004 and the several steps are explained in the following three subsections (4.3.1, 4.3.2 and 4.3.3).

4.3.1 Calculation of the climatological B matrices

The climatological \mathbf{B} matrix, originally calculated only over the whole domain, is now also calculated over twelve subdomains containing the same number of grid points (see Fig. 4.35) using the Parrish–Derber method. For simplicity the climatological \mathbf{B} matrices for a given variable will be indicated by the symbol $\mathbf{B}_{clim}^{variable}$, hence the temperature climatological \mathbf{B} matrices will be indicated by the symbol \mathbf{B}_{clim}^T , while the water vapour climatological \mathbf{B} matrices by \mathbf{B}_{clim}^q .

Figg. 4.36 and 4.37 show the \mathbf{B}_{clim}^T s and the \mathbf{B}_{clim}^q s respectively for the twelve subdomains.

All these twelve \mathbf{B}_{clim}^T s show the highest positive values on the diagonal in the last ten levels; in particular for the subdomain number 2 the variances at these levels reach the

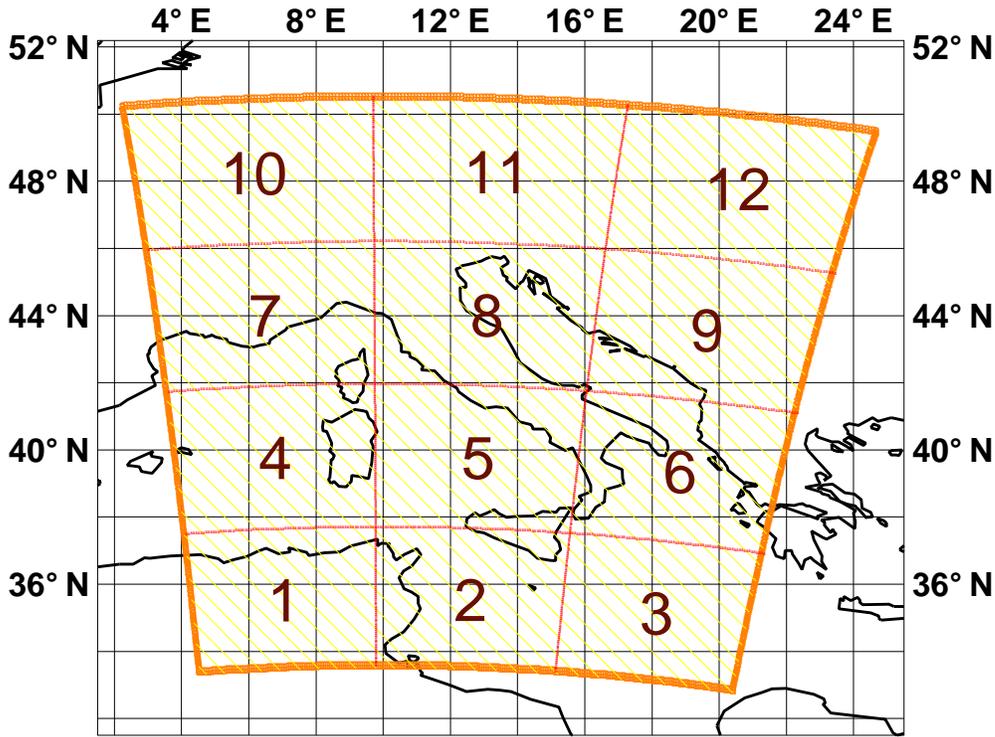


Figure 4.35: The twelve subdomains where the climatological \mathbf{B} matrices are calculated.

maximum value of 2.47 K^2 . Furthermore all the \mathbf{B}_{clim}^T s have negative values in the cross-correlations between the 5th and the 10th level and the levels at lower altitude, with negative values which are not lower than -0.65 K .

The \mathbf{B}_{clim}^q s calculated on the twelve subdomains show significant values greater than zero localised down the 15th level; the covariances for the higher levels have values lower in amplitude, since the presence of water vapour at these altitudes results scarce. The highest positive value of covariance is reached in the subdomain number 5 around the 30th level with a value of 2.30 (g/kg)^2 . Significant but very little negative values, non lower than -0.1 (g/kg)^2 , are present only in the subdomain number 3.

It is worth noting that for both the \mathbf{B}_{clim}^T s and \mathbf{B}_{clim}^q s on the twelve subdomains the positive covariances are greater in amplitude than the negative ones. Finally the \mathbf{B}_{clim}^T s and \mathbf{B}_{clim}^q s are calculated over the whole integration domain; their shapes are more regular than in the correspondent matrices calculated on the twelve subdomains (see Fig. 4.38).

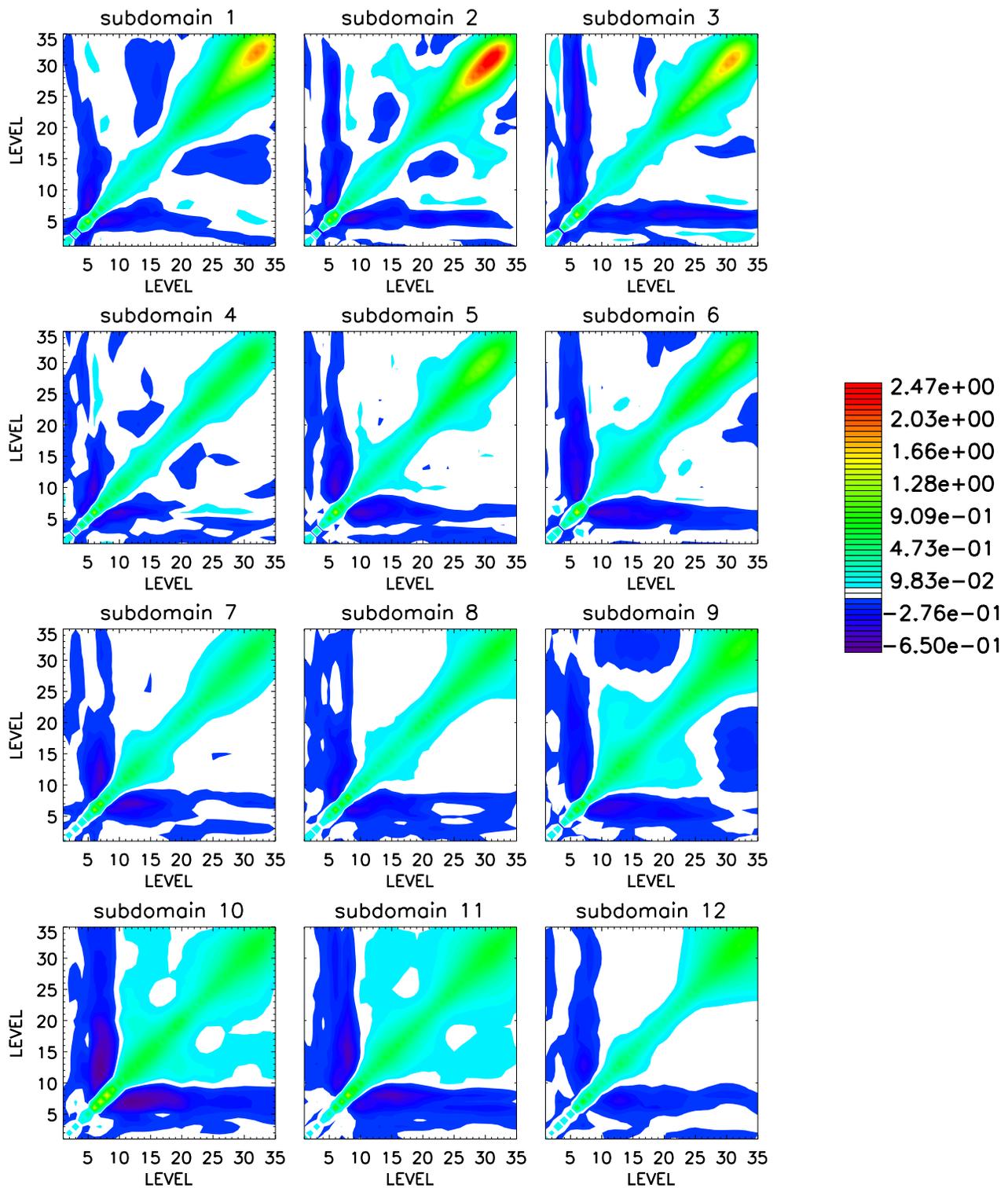


Figure 4.36: The B_{clim}^T s in K^2 on the twelve subdomains.

4.3.2 Calculation of the Ensemble Islands B matrices

The use of the ensemble members to compute Ensemble Islands B matrices is based on the assumption that the ensemble spread can give an important contribution to define the synoptic

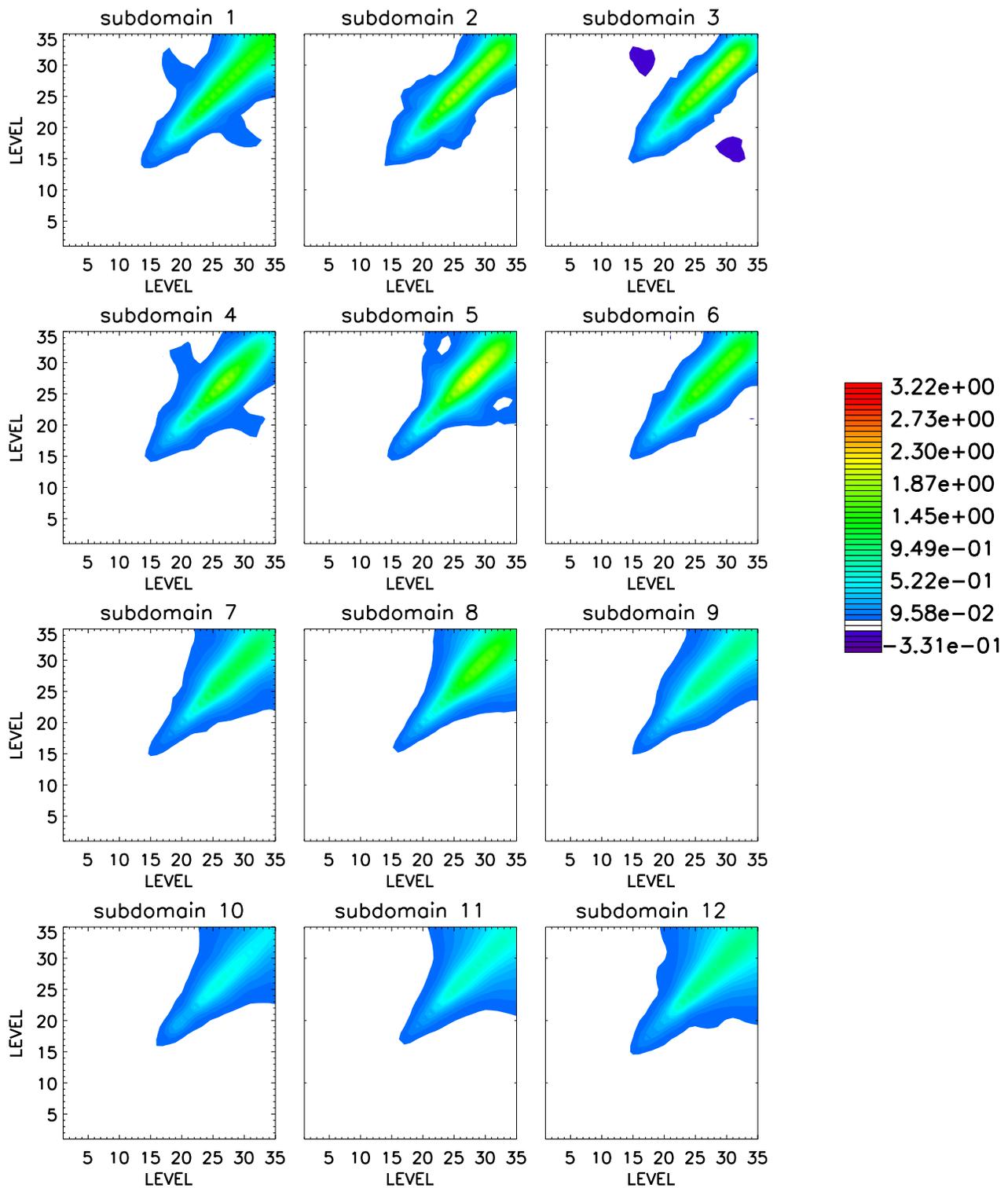


Figure 4.37: The B_{clim}^q s in $(g/kg)^2$ on the twelve subdomains.

dependent component of the background error covariances. For the false alarm case–study of the 8th of July 2004 five ensemble members were run. The initial conditions and the boundary conditions used to produce the ensemble members is provided by GME model of 00 UTC

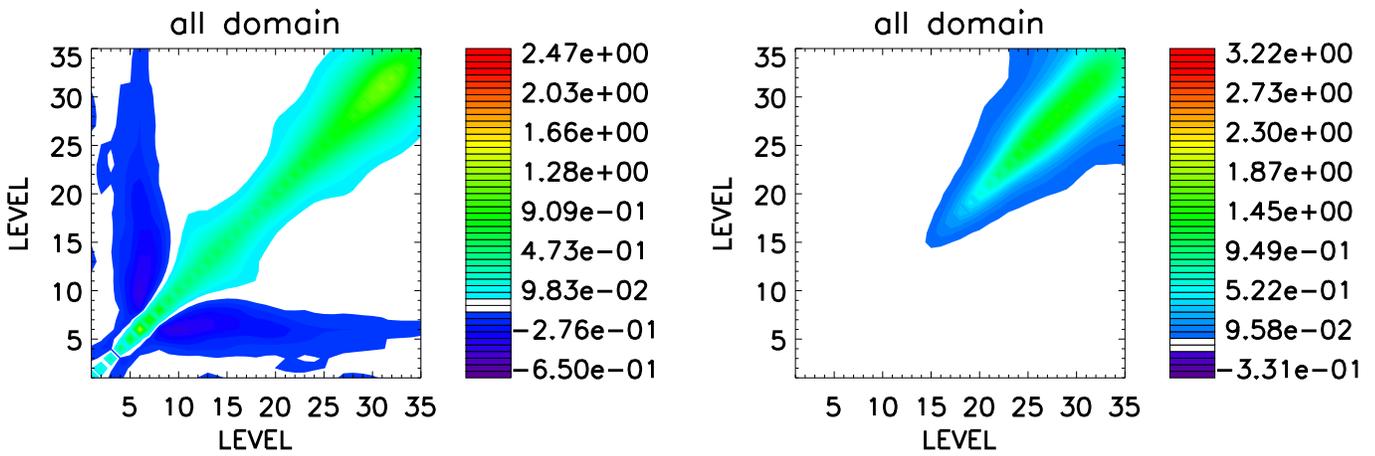


Figure 4.38: All domain B_{clim}^T in K^2 (left panel) and B_{clim}^q in $(g/kg)^2$ (right panel).

of the 8th of July 2004, while the perturbations are obtained varying some parameterisations inside the COSMO model. One of the five members is obtained running the operational COSMO model, two of the other four members are obtained replacing the Tiedtke convection scheme, which is used operationally in COSMO model, by the Kain-Fritsch convection scheme and removing any type of convection parameterisation scheme (explicit convection) respectively. The remaining two are instead obtained setting a parameter related to the turbulence scheme (`rlam_heat`), which has value 1 for the operational COSMO model, to the values 0.1 and 50 respectively (see Fig. 4.39).

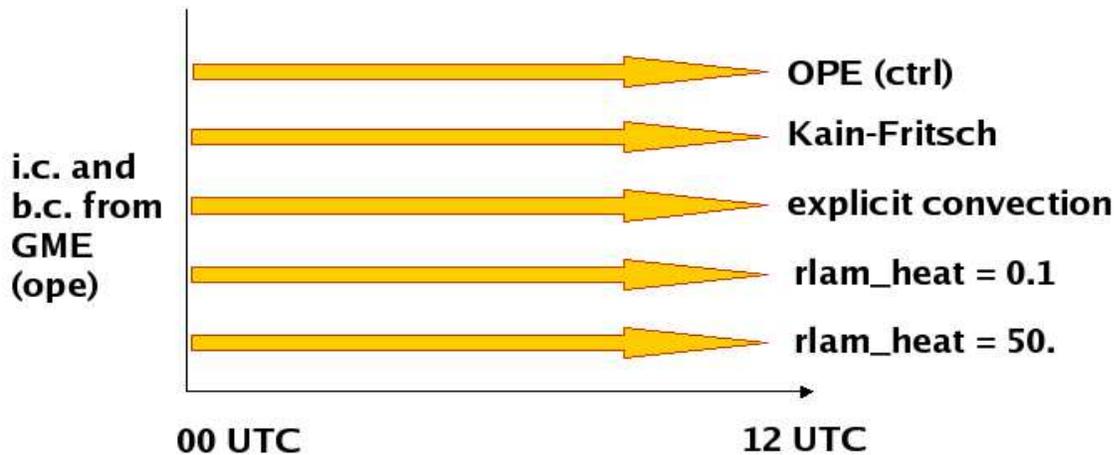


Figure 4.39: The different five ensemble members produced starting from the GME initial conditions and boundary conditions varying the COSMO parameters concerning the convection parameterisation scheme or the turbulence parameters (see the text for more details).

The ensemble runs start at 00 UTC of the 8th of July 2004 and the 12 hours forecast fields are used in the calculations. The mean of the ensemble members is considered our “truth”

and the departures from this “truth” is used to estimate the spread of the ensemble members for the variable temperature in each point and at each level of the integration domain. This three–dimensional field is then convolved with a normalised weighting function which has maximum value at 3 km at the altitudes where the use of satellite data has the major impact; the weighting function (see Fig. 4.40) has been chosen to have the following form:

$$W(z, z_0, A) = \frac{z}{z_0} \exp\left(\frac{-(z - z_0)^2}{A}\right) \quad (4.1)$$

where z_0 and A are parameters allowed to be changed (see Fig. 4.40).

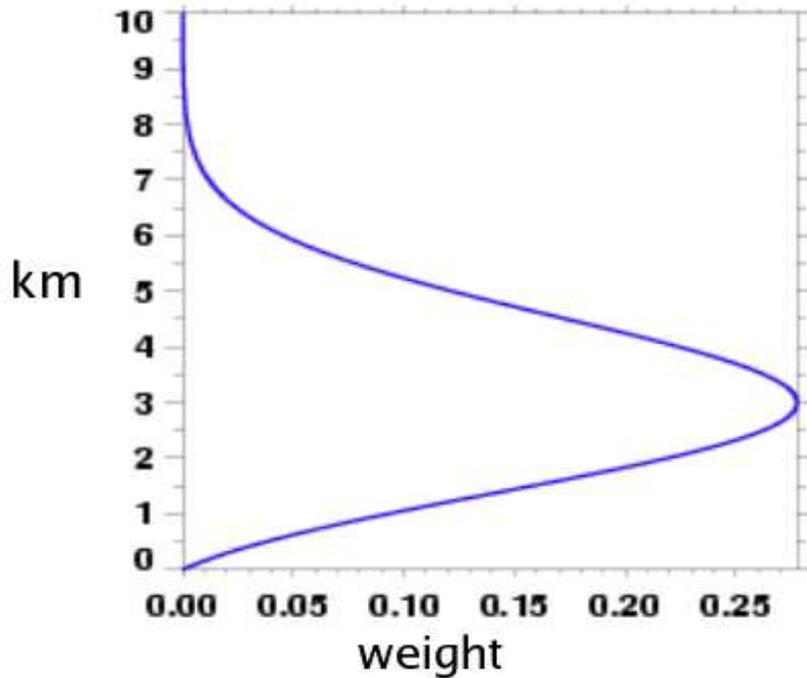


Figure 4.40: Weighting function of the temperature spread used to determine the ensemble islands ($z_0 = 2.0$, $A = 6.0$).

Finally a percentile technique is applied to this two–dimensional field to detect geographical areas, called “islands”, where the ensemble **B** matrices will be calculated. The percentile technique can be explained as follows: first of all the p_1 th percentile and the p_2 th percentile, with $0 < p_1 < p_2 < 100$, of the two dimensional field are calculated; then starting from the left corner of the domain and proceeding in west–east direction firstly and in south–north direction secondly the points where the two–dimensional field has value bigger than the p_2 th percentile are searched. If one points satisfies this condition, then this point is considered a

point of an island, and the size of this island is obtained proceeding from this point in each of the four cardinal directions until the two-dimensional field reaches a value lower than the p_{1th} percentile. The island so obtained is dismissed if the number of points inside itself is lower than a minimum number, otherwise the points constituting the island are not allowed to be inside other islands (see Fig. 4.41).

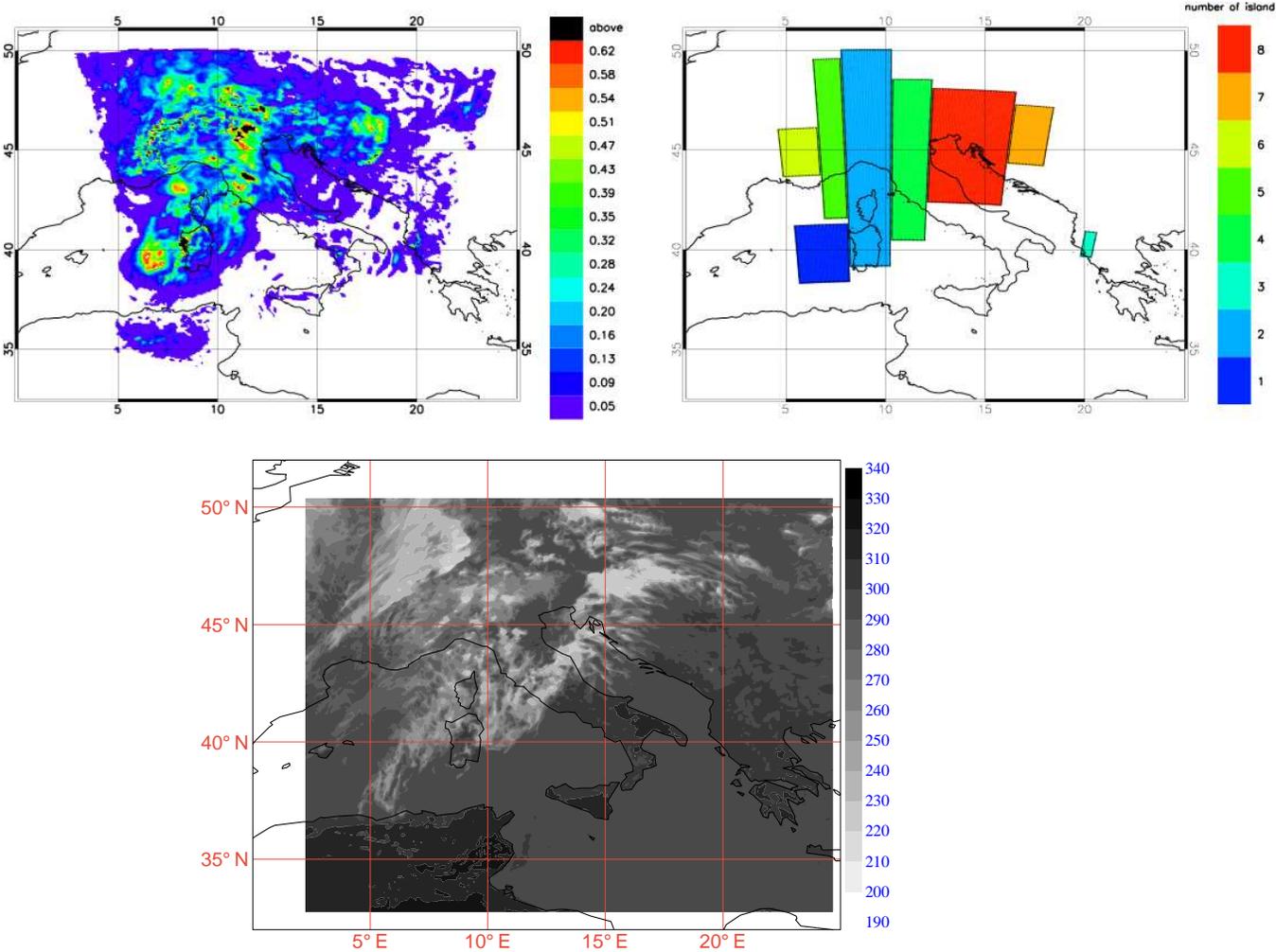


Figure 4.41: Ensemble temperature two-dimensional field obtained convolving the weighting function with the ensemble spread field of the temperature (top left panel) in K^2 , the islands obtained by the percentile technique (the minimum number of points constituting the islands is set to 150, p_1 is set to 70, and finally p_2 is set to 99) (top right panel), and the brightness temperature, in K, of the $10.8 \mu m$ infrared channel at 12 UTC of the 8th of July 2004 (bottom panel).

The islands are principally localised where the cloud systems are present. The ensemble **B** matrices will be calculated on these “islands”.

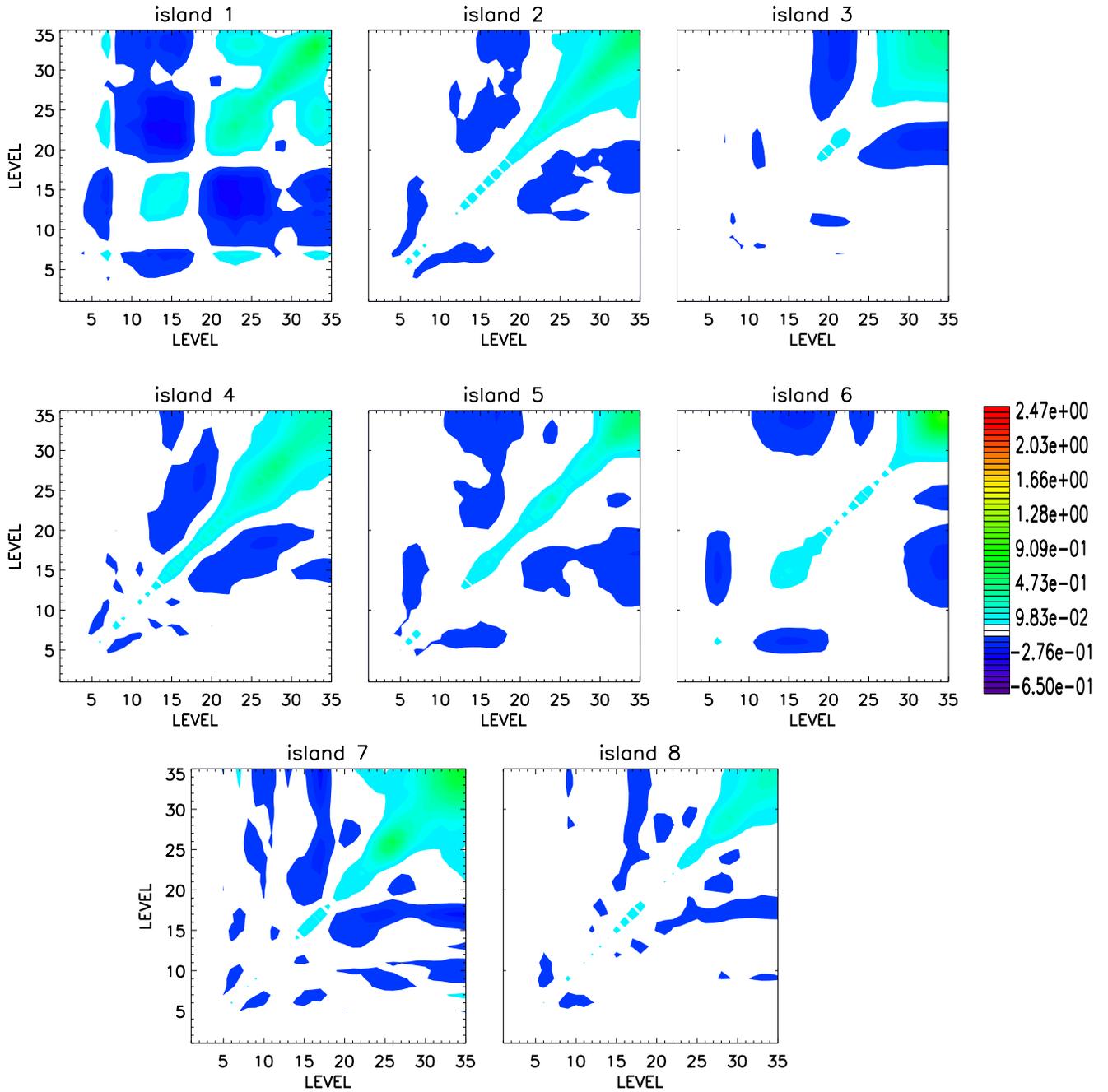


Figure 4.42: The B_{ens}^T s in K^2 on the eight islands.

For simplicity the ensemble B matrices for a given variable will be indicated by the symbol $B_{ens}^{\text{variable}}$, hence the temperature ensemble B matrices will be indicated by the symbol B_{ens}^T , while the water vapour ensemble B matrices by B_{ens}^q .

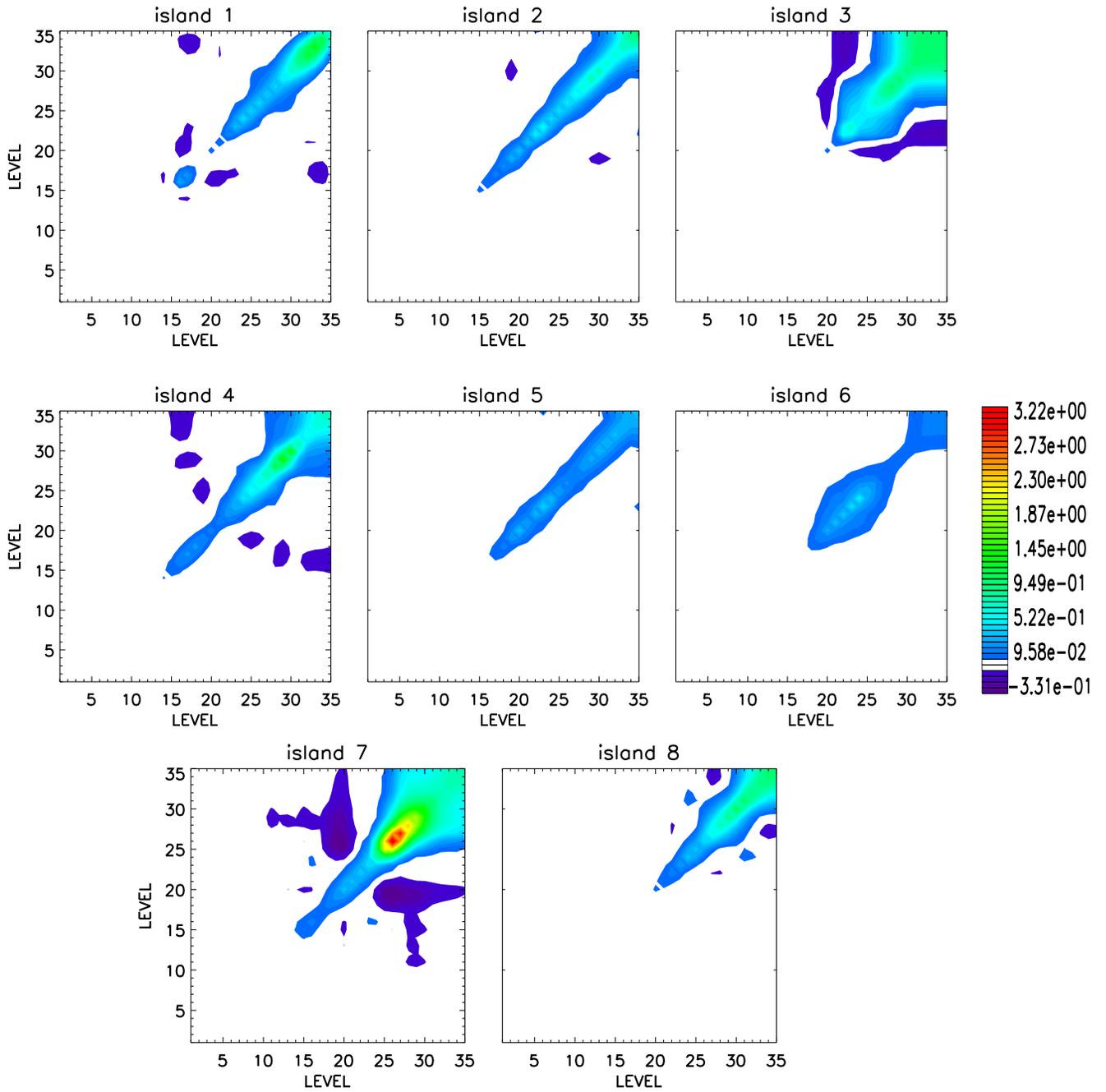


Figure 4.43: The B_{ens}^q s in $(g/kg)^2$ on the eight islands.

Both the B_{ens}^T s and B_{ens}^q s calculated on the ensemble islands result more irregular than the correspondent B_{clim}^T s and B_{clim}^q s matrices (see Figg. 4.42 and 4.43 respectively). The values of B_{ens}^T s calculated on the ensemble islands are always smaller than the values of B_{clim}^T s over the twelve subdomains, anyway the highest values are as before present on the diagonal in the levels close to the surface (see Fig. 4.42). The B_{ens}^q s calculated on the en-

semble islands have values whose order of magnitude is the same of the B_{clim}^q s, in particular in the island number 7 is reached the maximum value of 3.22 (g/kg)^2 in the diagonal between the 25th and the 30th level (see Fig. 4.43).

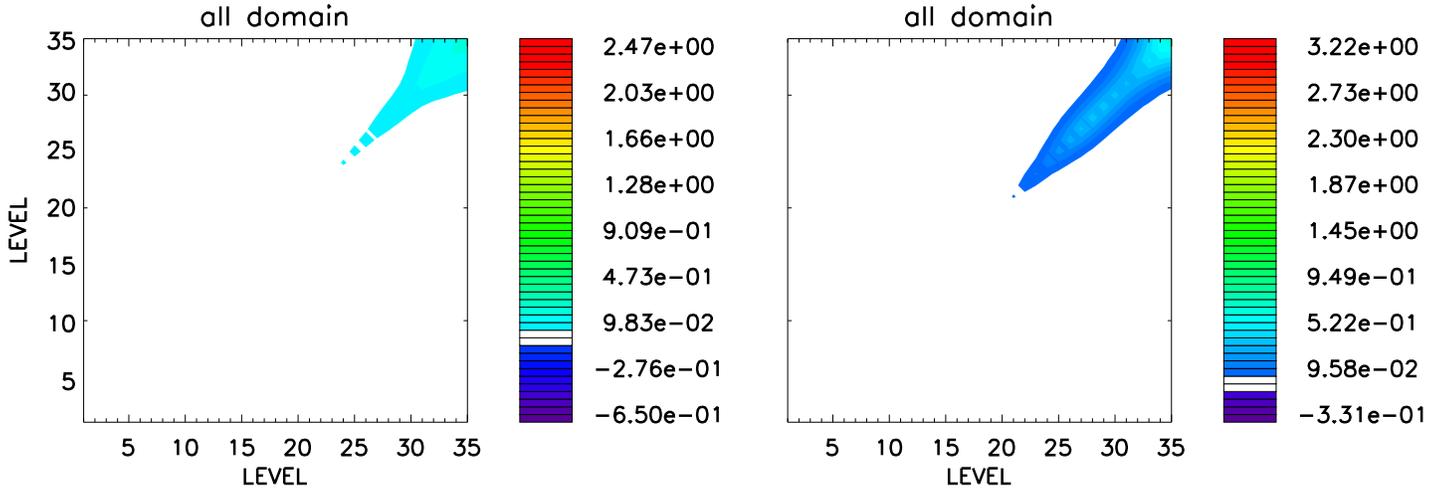


Figure 4.44: All domain B_{ens}^T in K^2 (left panel) and B_{ens}^q in $(g/kg)^2$ (right panel).

Over the whole domain the B_{clim}^T s and B_{clim}^q s are not only more homogeneous but also contain lower information than the correspondent B matrices calculated on the ensemble islands (see Fig. 4.44).

4.3.3 Merging of the two B matrices

Once the climatological B matrices and the ensemble B matrices are calculated, they are merged for each point of the integration domain. There are two possibilities:

- if the point falls in the subdomain number i and in the ensemble island number j , then the flow-dependent B matrix for this point is:

$$B_{\text{flow-dependent}} = 0.7B_{clim} + 0.3B_{ens} \quad (4.2)$$

- if the point falls in the subdomain number i but no ensemble island includes it, then the flow-dependent B matrix for this point is:

$$B_{\text{flow-dependent}} = B_{clim} \quad (4.3)$$

Using this “improved” B matrix assimilation and forecast are performed for the false alarm case–study with the same manner discussed at the beginning of this chapter. The assimilation and the forecast are called ASS_MSG1_20040708_ens and FOR_MSG1_20040708_ens respectively.

4.3.4 Assimilation

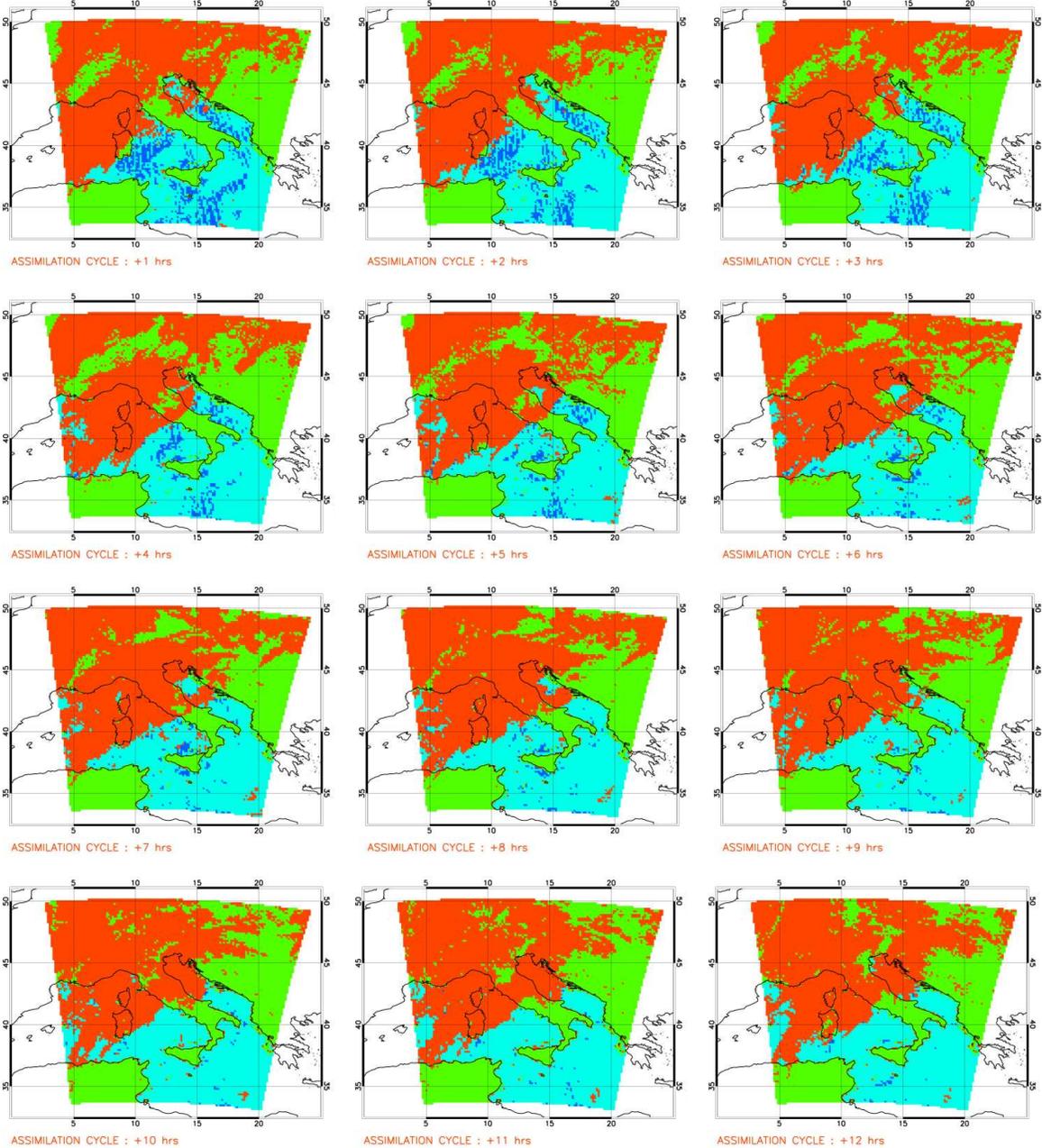


Figure 4.45: fg_prc during the assimilation temporal window with a temporal frequency of 1 hour.

Fig. 4.45 shows the flg_prc (the used colour palette is defined in section 3.2) for the assimilation period; most of the assimilated radiances are concentrated during the first hours of the assimilation and their number decreases carrying on in the assimilation time itself.

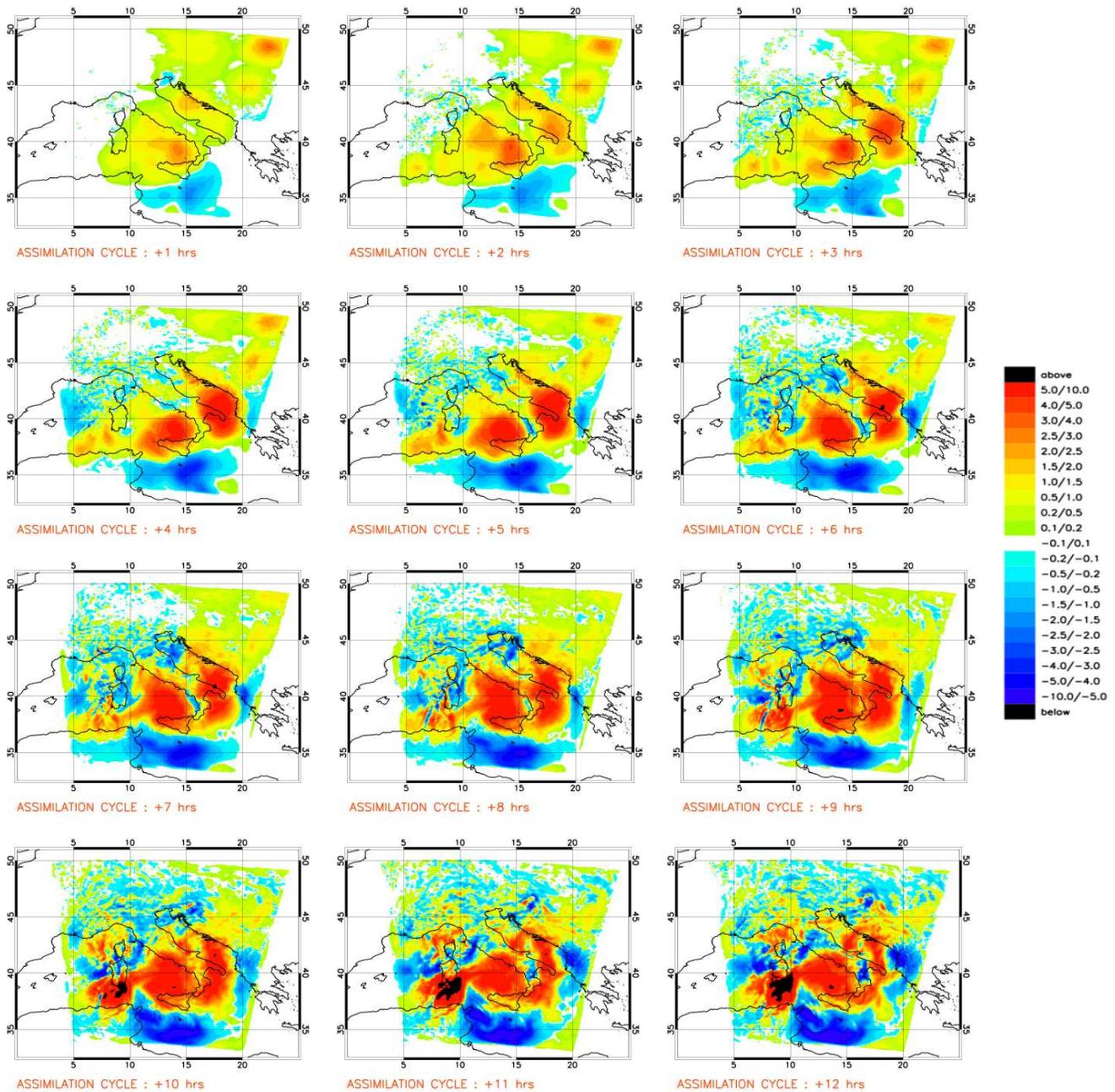


Figure 4.46: Differences in kg/m^2 between $ASS_MSG1_20040708_ens$ and $ASS_NUDG_20040708$ of the integrated water vapour in the column during the assimilation temporal window with a temporal frequency of 1 hour.

The integrated water vapour differences have significant values already in the first hours of the assimilation with positive values between 2.5 and 3.0 kg/m^2 north of Sicily island and

in the north–east part of the COSMO–LAMI domain. During the assimilation the differences are spread over the whole domain and at the end of the assimilation reach values greater than 10 kg/m^2 south of Sardinia island, with the most positive values in the south of the Tyrrhenian sea and the most negative ones south of Sicily island (see Fig. 4.46).

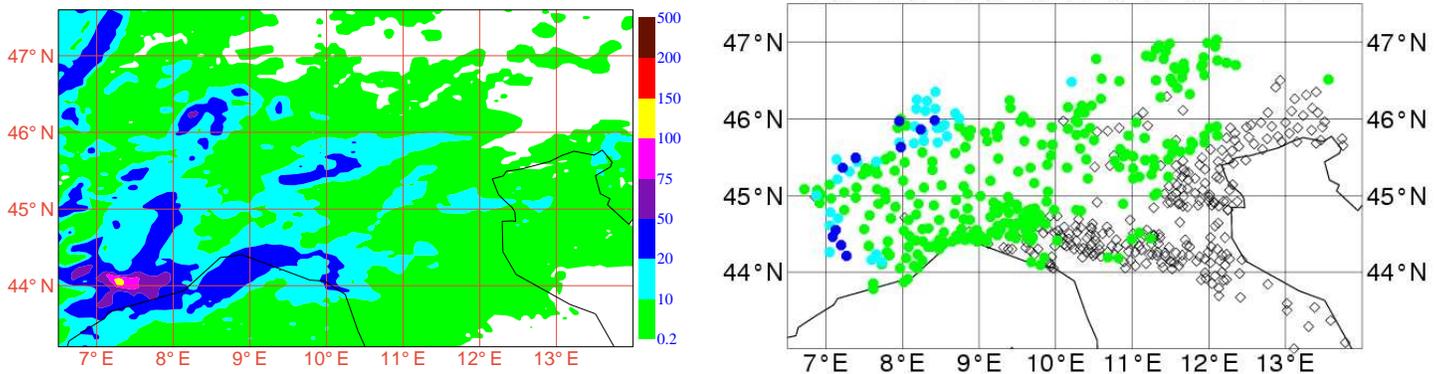


Figure 4.47: Twelve hours cumulated precipitation in mm from 00 UTC of the 8th of July 2004 forecasted by the run ASS_MSG1_20040708_ens (left panel) and the correspondent observations (right panel).

The forecasted precipitation presents overestimation of precipitation in Liguria region and in the north–west part of the Po Valley, while the maxima of precipitation in the Western Alps are well localised (see Fig. 4.47).

4.3.5 Forecast

At the beginning of the forecast the integrated water vapour differences have values greater than 10 kg/m^2 south of Sardinia island and negative values south of Sicily. These differences are reduced and shifted towards the east direction during the forecast. At the end of the forecast the most significant differences are present only on the Eastern part of the domain. This is due to the effect of the boundary conditions of the GME model which drive the COSMO model; in fact the synoptic situation was characterised at high altitude by intense north–west winds (see Fig. 4.48).

Finally the forecasted precipitation field shows an overestimation of precipitation over the Central Alps, in the Liguria region and also over Friuli–Venezia–Giulia where the false alarm case was announced (see Fig. 4.49).

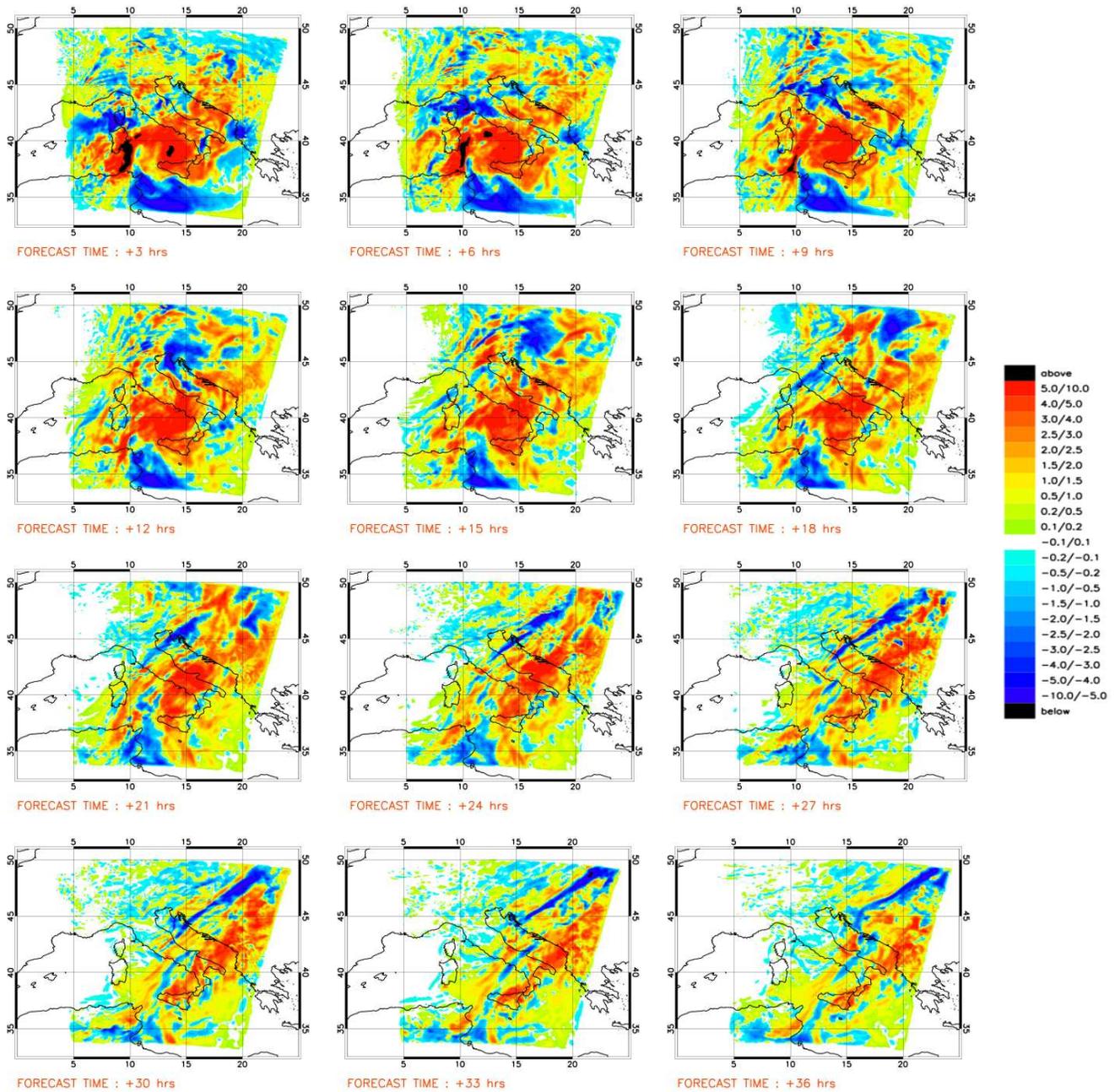


Figure 4.48: Differences in kg/m^2 between FOR_MSG1_20040708_ens and FOR_NUDG_20040708 of the integrated water vapour in the column during the forecast temporal window with a temporal frequency of 3 hours.

The non-remarkable improvement of the forecast in Friuli–Venezia–Giulia region using the flow-dependent **B** matrix could be due to the fact that in this case the assimilation of points in the north of the Adriatic sea at the beginning of the assimilation is not performed at the contrary of the run ASS_MSG1_20040708. This does not permit to produce the precipitation in the north of the Adriatic sea during the assimilation and results in an overestimation

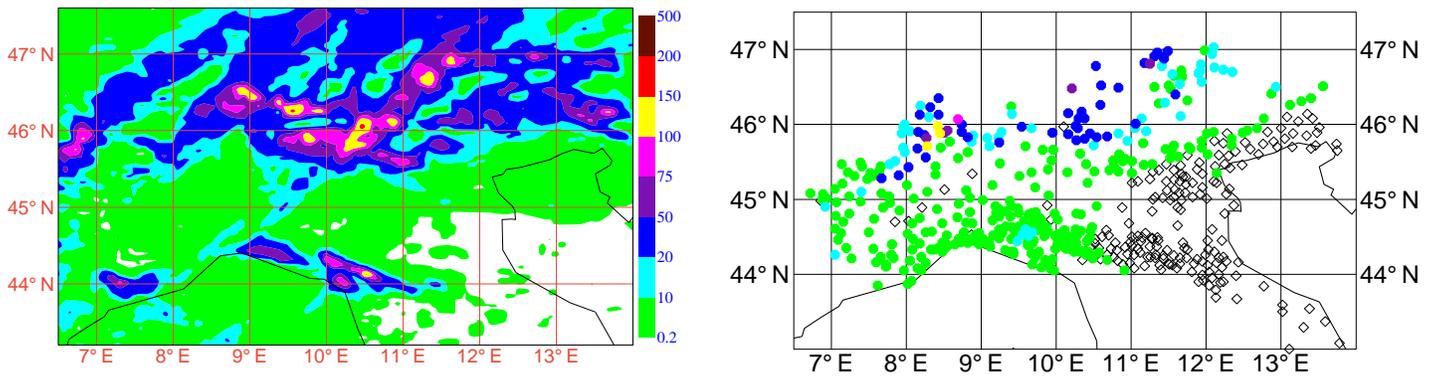


Figure 4.49: Twentyfour hours cumulated precipitation in mm from 12 UTC of the 8th of July 2004 forecasted by the run FOR_MSG1_20040708_ens (left panel) and the correspondent observations (right panel).

of precipitation in Friuli–Venezia–Giulia region in the forecast.

Chapter 5

CONCLUSIONS

The requirement to improve the prediction of severe weather events and localised heavy precipitation regimes driven by deep convection and complex orography has lead many national forecast centres to steadily increase the resolution of regional scale models, posing unprecedented problems to widely used variational data assimilation systems. On the one hand, analysis at 1 *km* resolution requires dense and frequent observations to capture spatially incoherent and quickly evolving structures typical of the meso- γ scales. On the other hand the applicability of variational techniques such as 4D-VAR imposes the linearity of the observation operator and the validity of balances imposed by the large scale flow (Rabier et al., 1998 [27]), which are violated at high resolutions. As the role of non-conventional observations such as radar or new high resolution satellite platforms becomes crucial for regional modelling at the kilometre scale, the suitability of 3D-VAR/4D-VAR analysis systems which allow direct radiance/reflectivity assimilation remains widely debated (Chevallier et al., 2003 [4] and Chevallier et al., 2004 [5]).

A possible way to overcome the linearity problem at high resolution while still retaining a variational approach is by performing an intermediate step in which part of the satellite information is transferred to prognostic variable increments which can then be used as “pseudo-observations” to be ingested into the assimilation system (Marécal and Mahfouf, 2000 [21]). The tractability of the assimilation problem is enhanced and detailed studies of model adjustments related to a certain observation are possible. Here this simplified approach is used and a 1D-VAR technique is employed to investigate the quality of temperature and humidity

retrievals from SEVIRI on board of the geostationary Meteosat Second Generation (MSG) satellite. Since observations from MSG platforms are available at 3 *km* resolution and with 15' repetition cycle they can substantially enrich assimilation systems based on rapidly updating cycles. The study has been performed for clear sky measurements over sea points and for a period of 20 days starting from 1st September 2006. Background data fields were provided by the regional non-hydrostatic model COSMO–LAMI during its assimilation cycle integration.

Between the eight available infrared channels of SEVIRI the two water vapour wavelengths at 6.2 and 7.3 μm have been shown to possess the largest 'information content', both in terms of absolute reduction of background errors (largest DFS) and relative performance compared to the other available wavelengths (largest DRM) (Rodgers, 2000 [32]). They produce a substantial reduction of background errors between 200 hPa and 600 hPa. For the reduction of both temperature and humidity errors below 600 hPa the inclusion of wavelengths at 8.7 μm , 10.8 μm and 12.0 μm is beneficial. The ozone and carbon dioxide monitoring channels are instead blacklisted. They were demonstrated to possess a poor information content (low values of DRM and DFS). Moreover inaccuracies in the radiative transfer calculations for these two wavelengths are expected as the concentration of both gases is not predicted by the model. Therefore, the theoretical evaluation of their impact into the analysis obtained using values of the DFS and MRM can be not reliable.

With this five channel configuration an expected maximum reduction of 11% in temperature and 20% in specific humidity standard deviations has been estimated. The vertical resolution of the expected analysis increments has been found quite coarse at roughly 10 *km* in the troposphere and worse above. Finally while the 1D–VAR algorithm is not able to reduce the 2*m* temperature and humidity errors it improved the background *SST*.

The quality of the 1D–VAR products has been assessed both on model departure statistics and comparing the final analysis to independent observations coming from collocated radiosondes which were not included in the analysis, and microwave based satellite *SST* products. From the background and analysis statistics the 1D–VAR methodology in the se-

lected configuration was shown to be a robust technique to extract temperature and humidity information from satellite radiances. The mean expected variable increments were close to zero showing the ease of performing air-mass dependent bias corrections.

The warm and dry biases revealed in COSMO–LAMI background fields by the comparison to the TEMP observations were reduced substantially in the 1D–VAR analysis. Moreover the 1D–VAR was able to correct for most of the forecast model bias variability producing profiles which were highly correlated with the radiosonde observations. In particular the analysed temperature profiles showed a maximum correlation of 80 % and the humidity around 40 %.

The assimilation of MSG satellite data is assessed for two three–dimensional case–studies: a false alarm case occurred in Friuli–Venezia–Giulia region on the 8th of July 2004 and a heavy precipitation case occurred in Emilia Romagna region between the 9th and the 12th of April 2005. Both these case–studies are analysed in two steps: the first one is the assimilation, with initial condition and boundary conditions provided hourly by the GME model, for twelve hours and the second one is the forecast, with boundary conditions provided hourly by the GME model, for the subsequent thirtysix hours. The impact of the 1D–VAR technique is evaluated in terms of variable differences: the integrated (over the column) water vapour and saturation water vapour, the T_{2m} , the q_{2m} and the T_s . The integrated saturation water vapour translates humidity information into temperature differences allowing a direct comparison between these two variables.

During the assimilation period the differences, initially localised in the points where MSG satellite data are assimilated, are spread; the extension and the speed of this spreading depends on the considered variable and on the number of MSG satellite data that are assimilated. For example during the assimilation period of the false alarm case–study the amount of MSG satellite data assimilated is larger than in the heavy precipitation case–study and this determines a more intense spreading of the differences in the first case for all the considered variables. It is worth noting that in both the case–studies during both the assimilation and the forecast the amplitude of the integrated water vapour differences is larger than the inte-

grated saturation water vapour in the column produced by the variation of temperature in the column using the 1D-VAR interfaced with the nudging technique.

During the forecast period the evolution of the differences for the two case-studies is different and affected by the different synoptic situations. In fact in the false alarm case-study, characterised by intense north-west winds at high altitudes, the differences of all the considered variables decrease in amplitude mainly in the north-west part of COSMO-LAMI domain. This behaviour is more evident for the integrated differences of water vapour and saturation water vapour where the impact of the winds at high altitude is bigger. The differences of the 2 meters fields, T_{2m} and q_{2m} , show a main reduction over the sea at west of Corsica and Sardinia islands. It is worth noting that the variables T_{2m} and q_{2m} are obtained in COSMO model combining the temperature and the specific humidity at the surface and at the level closest to the surface respectively; furthermore the SST is kept constant in COSMO during the forecast. Hence these two constraints, since the impact of the boundary conditions is bigger on the north-west part of domain, explain why the biggest reduction of the T_{2m} and q_{2m} is just over the sea.

The heavy precipitation case-study was instead characterised by a low pressure center which stationed between Sicily and Sardinia islands; this synoptic situation determines a weak impact of the boundary conditions in the forecast and in fact the amplitude of the differences in the forecast time remain the same for all the variables.

The impact of MSG satellite data is anyway more evident in the false alarm case-study where the assimilation of these new generation data allows the localisation of a precipitation cell in the north of the Adriatic sea during the assimilation and a subsequent reduction of the overestimation of precipitation in Friuli-Venezia-Giulia region during the forecast. The presence of an extended cloud system in the heavy precipitation case-study instead permits the assimilation of a very scarce number of MSG data avoiding a possible improvement of the forecast. This can be considered a proof of the limitation in the assimilation of MSG satellite data only in clear-sky points over sea.

An improvement in the definition of the **B** matrix is discussed using a flow-dependent approach. This approach tries to take into account the errors of the model depending also on the synoptic situation. On the one hand Parrish–Derber **B** matrices are calculated over twelve COSMO–LAMI subdomains, obtained dividing the domain in twelve parts containing the same number of grid points. These matrices are also called “climatological” **B** matrices. On the other hand ensemble members are used to select the areas where the model errors can vary remarkably with respect to the climatological values. The selection of these areas, also called “ensemble islands”, is performed convolving a predefined weighting function with the temperature ensemble spread field and applying to the two-dimensional field so obtained a percentile technique. Finally these two types of **B** matrices are appropriately merged together. The flow-dependent **B** method has been applied to the false alarm case–study but no remarkable improvement with respect to the run without the assimilation of MSG satellite data is obtained.

In summary, it has been shown that high spatial and temporal MSG observations can be exploited with a 1D–VAR approach resulting in an improved analysis of T and q profiles in clear sky over oceans. Furthermore these new generation data can be used in operational forecasts to detect possible scenarios of risk as it is demonstrated here for a false alarm case–study. The next step will be to extend the use of these MSG channels over land and investigate their impact on the subsequent forecasts using deterministic scores.

Bibliography

- [1] Bouttier, F. and P. Courtier, March 1999 : Data assimilation concepts and methods, Training course in Data Assimilation, ECMWF, Reading, United Kingdom.
- [2] Chadin, A. and N. A. Scott, 1984 : Improved Initialization Inversion Procedure (“3I”): Tech Proc 1st Int TOVS Study Conf; Igls, Austria; 29 August – 2 September 1983; Report of CIMSS, Univ of Wisconsin – Madison, Ed W P Menzel, pp. 14–79.
- [3] Chelton, D., F. Wentz, C. Gentemann, R. de Szoeko, and M. Schlax, 2000: Satellite Microwave SST Observations of Transequatorial Tropical Instability Waves, *grl*, **27**, pp. 1239–1242.
- [4] Chevallier, F., P. Lopez, A. M. Tompkins, M. Janisková, and E. Moreau, 2003: Can 4D-VAR systems assimilate cloud-affected satellite infrared radiances?, Technical Report NWPSAF-EC-TR-008, Satellite Application Facility for Numerical Weather Prediction, available at <http://www.ecmwf.int/publications/>.
- [5] Chevallier, F., P. Lopez, A. M. Tompkins, M. Janisková, and E. Moreau, 2004: The capability of 4D-Var systems to assimilate cloud-affected satellite infrared radiances, *Q. J. R. Meteorol. Soc.*, **130**, pp. 917–932.
- [6] Courtier, P., J.-N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach, *Q. J. R. Meteor.Soc.*, **120**, pp. 1367–1387.
- [7] Daniels, J., Velden C., Bresky, N. and Irving, A., 2004: Status of the NOAA/NESDIS Operational Satellite Wind Product System: Recent Im-

- provements, New Products, Product Quality, and Future Plans. Proceeding of the Seventh International Winds Workshop 14-17 June 2004, Helsinki.
- [8] Daley, R. : Atmospheric Data Analysis, 1991, Cambridge University Press, 457 pp.
- [9] Doms, G. and U. Schattler: The Nonhydrostatic Limited-Area Model LM (Lokal-Modell) of DWD: Part I, Scientific Documentation, available from <http://www.cosmo-model.org/>, 1999.
- [10] Donlon, C. J., P. Minnett, C. Gentemann, T. J. Nightingale, I. J. Barton, B. Ward, and J. Murray, 2002: Towards Improved Validation of Satellite Sea Surface Skin Temperature Measurements for Climate Research, *J. Climate*, **15**, pp. 353–369.
- [11] Eyre, J.R. and H. M. Woolf, 1988 : Transmittance of atmospheric gases in the microwave region: a fast model. *Applied Optics*, **27**, pp. 3244-3249
- [12] Eyre, J.R., 1991 : A fast radiative transfer model for satellite sounding systems. ECMWF Research Dept. Tech. Memo. **176** (available from the librarian at ECMWF).
- [13] Eyre, J. R., G. A. Kelly, A. P. McNally, E. Andersson, and A. Persson, 1993: Assimilation of tovs radiance information through one-dimensional variational analysis, *Q. J. R. Meteorol. Soc.*, **119**, pp. 1427–1463.
- [14] Ebert, E.E. and J. A. Curry, 1992 : A parametrization of ice optical properties for climate models, *J. Geophys. Res.*, **97**, pp. 3831–3836.
- [15] Gilbert, J. and C. Lemaréchal, 1989: Some numerical experiments with variable-storage quasi-Newton algorithms, *Mathematical Programming*, **45**(1), pp. 407–435.
- [16] Gurka, J. and T. Schmit, 2004: Baseline instruments planned for the GOES-R series, *Proceedings of SPIE*, **5549**, 96 pages.
- [17] Harris, B. and G. Kelly, 2001: A satellite radiance-bias correction scheme for data assimilation, *Quarterly Journal of the Royal Meteorological Society*, **127**(574), pp. 1453–1468.

- [18] Hoke, J. E., and R. A. Anthes, 1976: The initialization of numerical models by a dynamic initialization technique, *Mon. Wea. Rev.*, **104**, pp. 1551–1556.
- [19] Kelly, G. A., and J. F. Flobert, 1988: Radiance tuning, Tech Proc 4th Int TOVS Study Conf; Igls, Austria; 16–22 March 1988; Report of CIMSS, Univ of Wisconsin–Madison, Ed W P Menzel, pp. 99–117.
- [20] Majewski, D., D. Liermann, P. Prohl, B. Ritter, M. Buchhold, T. Hanisch, G. Paul, W. Wergen, and J. Baumgardner, 2002: The Operational Global Icosahedral–Hexagonal Gridpoint Model GME: Description and High-Resolution Tests, *Mon. Wea. Rev.*, **130**(2), pp. 319–338.
- [21] Marécal, V. and J. F. Mahfouf, 2000: Variational retrieval of temperature and humidity profiles from TRMM precipitation data, *Mon. Wea. Rev.*, **128**, pp. 3853–3866.
- [22] Marécal, V. and J. F. Mahfouf, 2002: Four-dimensional variational assimilation of total column water vapour in rainy areas, *Mon. Wea. Rev.*, **130**, pp. 43–58.
- [23] Matricardi, M., F. Chevallier and S. Tjemkes, 2001 : An improved general fast radiative transfer model for the assimilation of radiance observations. ECMWF Research Dept. Tech. Memo. **345** (available from the librarian at ECMWF).
- [24] McMillin L. M., L. J. Crone, and D. S. Crosby, 1989: Adjusting satellite radiances by regression with an orthogonal transformation to a prior estimate, *Journal of Applied Meteorology*, **28**, pp. 969–975.
- [25] Parrish, D. and J. Derber, 1992: The National Meteorological Center’s Spectral Statistical-Interpolation Analysis System, *Mon. Wea. Rev.*, **120**(8), pp. 1747–1763.
- [26] Purser, R. and H. Huang, 1993: Estimating effective data density in a satellite retrieval or an objective analysis, *Journal of Applied Meteorology*, **32**(6), pp. 1092–1107.
- [27] Rabier, F., A. McNally, E. Andersson, P. Courtier, P. Uden, J. Eyre, A. Hollingsworth, and F. Bouttier, 1998: The ECMWF implementation of

- the three dimensional variational assimilation (3D-Var). Part II: Structure functions, *Q. J. R. Meteorol. Soc.*, **124**, 1809–1829.
- [28] Rayer, P.J. 1995 : Fast transmittance model for satellite sounding, *Applied Optics*, **34**, pp. 7387–7394.
- [29] Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, 2002: An improved in situ and satellite SST analysis, *J. Climate*, **15**, pp. 1609–1625.
- [30] Ritter, B. and J. Geleyn, 1992: A comprehensive radiation scheme for numerical weather prediction models with potential applications in climate simulations, *Monthly Weather Review*, **120**(2), pp. 303–325.
- [31] Rizzi, R. and M. Matricardi, 1998: The use of TOVS clear radiances for numerical weather prediction using an updated forward model, *Quart. J. Roy. Meteor. Soc.*, **124**, pp. 1293–1312
- [32] Rodgers, C. D., 2000: *Inverse methods for atmospheric sounding*, World Scientific.
- [33] Saunders, R. W., P. Brunel, F. Chevallier, G. Deblonde, S. J. English, M. Matricardi, and P. Rayer, 2002: RTTOV-7 science and validation report, Technical report, UK Met Office, Met Office Forecasting and Research Technical Report No. 387.
- [34] Saunders, R.W., M. Matricardi and P. Brunel, 1999: A fast radiative transfer model for assimilation of satellite radiance observations - RTTOV-5. ECMWF Research Dept. Tech. Memo. 282 (available from the librarian at ECMWF).
- [35] Saunders, R.W., M. Matricardi and P. Brunel, 1999 : An Improved Fast Radiative Transfer Model for Assimilation of Satellite Radiance Observations, *Quart. J. Roy. Meteor. Soc.*, **125**, pp. 1407–1425.
- [36] Schmetz, J., K. Holmlund, J. Hoffman, B. Strauss, B. mason, V. Gaertner, A. Koch, and L. Van De Berg, 1993: Operational cloud-motion winds from Meteosat infrared images, *J. Applied. Met.*, **32**, pp. 1206–1225.

- [37] Schraff, C. and R. Hess, 2003: A description of the nonhydrostatic regional model LM. Part III: Data Assimilation, Technical Report 4, Deutscher Wetterdienst, available at <http://cosmo-model.cscs.ch/>.
- [38] Schumann, W., H. Stark, K. McMullan, D. Aminou, and H. Luhmann, 2002: The MSG System, Technical report, European Space Agency.
- [39] Smith, R. L., H. M. Woolf, C. M. Hayden, A. J. Schreiner, and J. F. Le Marshall, 1984: The physical retrieval TOVS export package. Tech Proc 1st Int TOVS Study Conf; Igls, Austria; 29 August – 2 September 1983; Report of CIMSS, Univ of Wisconsin–Madison, Ed W P Menzel, pp. 227–278.
- [40] Smith, E.A. and L. Shi, 1992 : Surface forcing of the infrared cooling profile over the Tibetan plateau. Part I: Influence of relative longwave radiative heating at high altitude, *J. Atmos. Sci.*, **49**, pp. 805–822.
- [41] Stauffer, D. R., and N. L. Seaman, 1990: Use of four-dimensional data assimilation in a limited-area mesoscale model. Part I: Experiments with synoptic-scale data, *Mon. Wea. Rev.*, **118**, pp. 1250–1277.
- [42] Steppeler, J., G. Doms, U. Schättler, H.-W. Bitzer, A. Gassmann, U. Damrath, and G. Gregoric., 2003: Meso-gamma scale forecasts using the non-hydrostatic model LM, *Meteorol. Atmos. Phys.*, **82**, pp. 75–96.
- [43] Susskind, J. , J. Rosenfield, and D. Reuter, 1983: An accurate radiative transfer model for use in the direct physical inversion of HIRS–2 and MSU temperature sounding data, *J. Geophys. Res.*, **88**, pp. 8550–8568.
- [44] Tiedtke, M., 1989: A comprehensive mass flux scheme for cumulus parameterization in large-scale models., *Mon. Wea. Rev.*, **117**, pp. 1779–1800.
- [45] Uddstrom, M. J., 1991: Forward model errors., Tech Proc 6th Int TOVS Study Conf; Arlie, Virginia; 1–6 May 1991; Report of Univ of Wisconsin–Madison, Ed W P Menzel, pp. 501–516.
- [46] Uppala, S., P. Kållberg, A. Simmons, U. Andrae, V. da Costa Bechtold, M. Fiorino, J. Gibson, J. Haseler, A. Hernandez, G. Kelly, et al., 2005: The ERA-40 re-analysis, *Quarterly Journal of the Royal Meteorological Society*, **131**(612), 2961–3012.

- [47] Vömel, H., H. Selkirk, L. Miloshevich, J. Valverde, J. Valdés, E. Kyrö, R. Kivi, W. Stolz, G. Peng, and J. A. Diaz, 2006: Radiation dry bias of the Vaisala RS92 humidity sensor, *J. Atmos. Oceanic Technol.*.
- [48] Weinreb, M. P., H. E. Fleming, L. M. McMillin, and A. C. Neuendorffer, 1981: Transmittances for the TIROS Operational Vertical Sounder, *NOAA Tech. Report* , NESS 85.

Ringrazio in primo luogo i miei familiari, ed in particolare i miei genitori il cui supporto morale è stato la risorsa che mi ha permesso di superare i momenti più difficili.

Un grazie a tutti i miei amici, in particolare un grazie speciale lo dedico a Luca Delli Passeri e Alessandra Cetro che sono stati fondamentali per arrivare fino a qui.

Ringrazio calorosamente i miei colleghi di lavoro dell'ufficio, in particolare Davide Cesari, Anna Fornasiero, Chiara Marsigli, Andrea Montani, Antonella Morgillo e Paolo Patruno, e dell'intero Servizio Idro-Meteorologico.

Un grazie grande grande a Tiziana Paccagnella per avermi affiancato con la sua grande competenza scientifica e con la sua grande personalità sia nella mia tesi di laurea che nel ciclo di dottorato. A great thanks also to Dr. Francesca Di Giuseppe whose work is at the basis of this thesis and can be defined a great step in the improvement of the next generation of limited area numerical weather prediction models for the whole scientific community.

Ich danke Dr. Christoph Gebhardt, Dr. Reinhold Hess und Dr. Werner Gräsle für die Zusammenarbeit während meines kurzen Arbeitsaufenthaltes beim Deutscher Wetterdienst.

SEVIRI data are provided by EUMETSAT and licensed by the Ufficio Generale per la Meteorologia, Rome ,IT. Microwave OI SST data are produced by Remote Sensing Systems and sponsored by National Oceanographic Partnership Program (NOPP), the NASA Earth Science Physical Oceanography Program, and the NASA REASoN DISCOVER Project. Data from the ADRICOMS database are courtesy of ISAC-CNR, Gruppo di Oceanografia da Satellite (Roma, IT). I am grateful to Jacopo Chiggiato for arranging this provision. Reinhold Hess, Christoph Schraff and Blazej Krzeminski contributed in the implementation of the 1DVAR system into Lokal Modell.

The INRIA (Institut National de Recherche en Informatique et Automatique) provided the M1QN3 minimisation code. Thanks to Frederic Chevallier for providing the core 1DVAR algorithm.