ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

Scuola di Scienze
Corso di Laurea Magistrale in Fisica del Sistema Terra

Forecast of High-Impact Weather over Italy: performance of global and limited-area ensemble prediction systems

Relatore:
Prof.ssa Silvana Di Sabatino

Correlatori:
Dott. Andrea Montani
Dott.ssa Tiziana Paccagnella

Presentata da:
Giacomo Pincini

Sessione I
Anno Accademico 2017/2018

Sub lumine Matris

Sommario

Nella previsione del tempo l'approccio deterministico non permette di stabilire a priori se una previsione sará corretta o meno; d'altra parte, le previsioni probabilistiche forniscono un punto di vista piú completo, affidabile e accurato di che cosa potrebbe accadere nel futuro, dando idealmente informazioni sulla relativa frequenza dell'evento. Quindi apportano precisi vantaggi per coloro che devono prendere delle decisioni. Gli utenti delle previsioni possono sfruttare queste informazioni per esempio quando vogliono valutare le perdite associate a condizioni meteo avverse, rispetto ai costi delle azioni preventive. Lo scopo di questo lavoro é valutare il valore aggiunto di una migliore risoluzione orizzontale nella previsione probabilistica dei campi in quota e alla superficie. In particolare, sono state confrontate le performance di tre differenti sistemi di previsione di ensemble: ECMWF-ENS (51 membri, risoluzione orizzontale 18km), COSMO-LEPS (16 membri, risoluzione orizzontale 7km) e COSMO-2I-EPS (10 membri, risoluzione orizzontale 2.2km). Mentre i primi due sistemi di ensemble sono operativi, COSMO-2I-EPS é ancora in fase di sviluppo. Pertanto, la finestra di comparazione copre un periodo limitato: dal 20 al 27 giugno 2016. In questo lavoro, sono state analizzate sia le variabili in quota che al suolo. Le variabili in quota, considerate a tre differenti livelli di pressione, sono l'altezza di geopotenziale e la temperatura; per la verifica sono stati calcolati l'ensemble spread e la radice dell'errore quadratico medio, utilizzando i dati dei radiosondaggi italiani disponibili ogni 12/24 ore. Le variabili al suolo, temperatura a 2 metri e precipitazione cumulata su 6 ore, sono state verificate tramite la rete di stazioni non convenzionali fornite dal Dipartimento di Protezione Civile Nazionale. Per la temperatura a 2 metri sono stati calcolati l'ensemble spread e la radice dell'errore quadratico medio, mentre per la precipitazione sono stati considerati alcuni score probabilistici (Brier Skill Score, Ranked Probability Score, ROC-Area, Percentuale di Outliers ed altri). Sia per la verifica in quota che per quella alla superficie, gli score migliori sono stati ottenuti principalmente dai sistemi di ensemble del consorzio COSMO. Questi si caratterizzano per avere una risoluzione orizzontale piú alta e una popolazione di membri piú bassa. In particolare, il recentemente implementato COSMO-2I-EPS raggiunge spesso le performance piú soddisfacenti. A causa della limitata disponibilitá di dati, i risultati di questo studio pilota si basano su un periodo relativamente breve, pertanto sono necessarie ulteriori analisi. Ciononostante, nei sistemi di ensemble alla mesoscala, il valore aggiunto dell'alta risoluzione sembra giocare un ruolo cruciale nella previsione probabilistica dei campi atmosferici a tutti i livelli. In particolare, negli ensemble di COSMO, la descrizione piú dettagliata dei processi connessi all'orografia fornisce un valore aggiunto per la previsione di eventi meteorologici localizzati ad elevato impatto.

Abstract

The deterministic approach to weather prediction does not allow to establish a-priori whether a forecast would be skilful or unskillful; on the other hand, probabilistic forecasts provide a more complete, reliable and accurate view of what could happen in the future, ideally providing information on the relative frequency of the event. Therefore, they bring definite benefits for decision-makers. Forecast users can exploit such information for example when they want to weight the losses associated with adverse weather events against the costs of taking precautionary actions. The aim of this work is to assess the added value of the enhanced horizontal resolution in the probabilistic prediction of upper-level and surface fields. In particular, the performances of three different ensemble prediction systems were compared: ECMWF-ENS (51 members, 18 km horizontal resolution), COSMO-LEPS (16 members, 7 km horizontal resolution) and COSMO-2I-EPS (10 members, 2.2 km horizontal resolution). While the first 2 ensemble systems are operational, COSMO-2I-EPS is still in a development phase. Therefore, the intercomparison window covers a limited period, which ranges from 20 to 27 June 2016. In this work, both upper-level and surface variables are analyzed. As for upper-level, both temperature and the geopotential height at three different pressure levels are considered; the ensemble spread and the root mean square error are computed using the available Italian radiosounding data every 12/24 hours for verification. As for the surface, 2-metre temperature and precipitation cumulated over six hours are verified against the non-conventional station network provided by the National Civil Protection Department. The ensemble spread and the root mean square error of 2-metre temperature are computed, while a number of probabilistic scores (Brier Skill

Score, Ranked Probability Score, Roc-Area, Outliers Percentage and others) are considered for precipitation.

For both upper-level and surface verification, it turns out that the best scores are mainly obtained by the COSMO-based ensemble systems with higher horizontal resolution and lower ensemble size; in particular, the newly implemented COSMO-2I-EPS often achieves the most satisfactory performances.

Although the results of this pilot study are based over a relative short period due to limited data availability and further investigations is needed, the added value of high resolution in mesoscale ensemble systems seems to play a crucial role in the probabilistic prediction of atmospheric fields at all levels. In particular, the more detailed description of mesoscale and orographic-related processes in COSMO-ensembles provides an added value for the prediction of localised High-Impact Weather events.

Contents

List of Figures

List of Tables

List of Acronyms

1. *AISAM Associazione Italiana di Scienze dell'Atmosfera e del Clima*
2. *ARPA-SIMC Agenzia Regionale Prevenzione Ambientale - Servizio Idro Meteo Clima*
3. *BS Brier Score*
4. *BSS Brier Skill Score*
5. *CAPE Convective Available Potential Energy*
6. *CFL Courant-Friedrichs-Lewy*
7. *COSMO COnsortium for Small-scale MOdeling*
8. *DPCN Dipartimento di Protezione Civile Nazionale*
9. *DWD Deutscher Wetterdienst*
10. *ECMWF European Centre for Medium-Range Weather Forecast*
11. *EDA Ensemble of Data Assimilation*
12. *EMS European Meteorological Society*
13. *ENIAC Electronic Numerical Integrator And Computer*
14. *ENS ENSemble*
15. *EPS Ensemble Prediction System*
16. *ERA European Reanalysis*

17. *EUMETSAT EUropean METeorological SATellites*

18. *GPS Global Positioning System*

19. *GTS Global Telecommunications Systems*

20. *HE-VI Horizontally Explicit, Vertically Implicit*

21. *HIW High-Impact Weather*

22. *HNMS Hellenic National Meteorological Service*

23. *HRES High-RESolution*

24. *H-TESSEL Hydrology-Tiled ECMWF Scheme for Surface Exchange over Land*

25. *ICs Initial Conditions*

26. *IFS Integrated Forecasting System*

27. *LAI Leaf Area Index*

28. *LEPS Limited area Ensemble Prediction System*

29. *LM Lokal Modell*

30. *ME Mean Error*

31. *MPI Message Passing Interface*

32. *MPP Massively Parallel Processing*

33. *MSE Mean Square Error*

34. *MSLP Mean Sea Level Pressure*

35. *NCEP National Centre for Environmental Prediction*

36. *NWP Numerical Weather Prediction*

37. *OSTIA Operational Sea Surface Temperature and Sea Ice Analysis*

38. OUTL OUTLiers

39. PDF Probability Density Function

40. RASS Radio Acoustic Sounding System

41. RM Representative Member

42. RMSE Root Mean Square Error

43. ROC Relative Operating Characteristic Curve

44. RPS Ranked Probability Score

45. RPSS Ranked Probability Skill Score

46. $RPSS_D$ *debiased Ranked Probability Skill Score*

47. SPPT Stochastic Perturbations of Physical Tendencies

48. SPRD ensemble SPReaD

49. SST Sea Surface Temperature

50. SV Singular Vector

51. TVD Total Variation Diminishing

52. UTC Universal Time Coordinated

53. VAD Velocity Azimuth Display

Chapter 1

Introduction

The prediction of weather events related to strong winds, heavy rain and snowfall is still nowadays a serious challenge, especially when high spatio-temporal details are required. Despite Numerical Weather Prediction (NWP) modelling has made great progress in recent decades, thanks to the increases in model resolution, better understanding of atmospheric dynamical processes and advances in data assimilation techniques, the above-mentioned atmospheric events, usually referred to as "High-Impact Weather" (HIW), can have horizontal dimension too small to be explicitly resolved. HIWs provide the most dramatic examples of how the atmosphere affects people daily lives, since they may cause both human and economic costs. Therefore, there is a need of better ways to predict this type of phenomena, also accounting for their inherent degree of non-predictability.

This paves the way to the introduction of a probabilistic approach via the ensemble forecasting, which was introduced at the beginning of the nineties, in order to provide a representation of model uncertainty, due to the imperfect knowledge of atmospheric initial conditions and the approximate model formulation. This approach has now become commonplace in operational weather forecasting by the major meteorological institutes. Instead of running just one forecast with an unknown error, an ensemble of slightly different forecasts are run, in order to integrate the deterministic forecast with an estimate of the "forecast of forecast skill" (Tennekes et al., 1986). Probabilistic forecasts provide a more complete, reliable and accurate view of what might happen in the future, ideally providing information on the relative frequency of an event occuring. Therefore, they bring definite benefits for decision-makers. Forecast users can exploit such information for cost-loss analysis. The estimation of uncertainty is even more crucial when local effects come into play and a high spatio-temporal detail is required as in the case of precipitation, where NWP limitations become more evident.

To understand the difficulty by medium–to–low horizontal resolution models in forecasting this type of phenomena, it is presented the case of 27^{th} June 2016. On this day, two storm lines have crossed at different times some Italian regions: those of the northeast during the first hours of the day, Umbria and Marches in the afternoon. In Fig. 1.1, it is possible to see both observed and forecast precipitation for the 27^{th} June 2016. The top left panel provides observed precipitation according to the rain gauges collected by the National Civil Protection Department (DPCN). On the other hand, the bottom left and right panels show 24-hour total precipitation as predicted by the model runs starting at 00 UTC of 26^{th} June, respectively by the member 1 of global ensemble system ECMWF ENS and of the COSMO-based limited-area ensemble prediction system COSMO-LEPS.

Looking at the figure, it will be evident that both forecasts show some critical issues: there are marked inaccuracies both in the spatial localization and in the intensity of precipitation. In particular, the problems relate to the excessive extension of precipitation over northern Italy, the absence or inaccurate location of heavier rainfall and the lack or different pattern of precipitation that hit the central Italy. Thus there is room for improvement. This improvement can be sought using an ensemble system with a higher horizontal resolution. The new national ensemble system satisfies this request. It is in a pre-operational phase, for this reason it is useful to study its performance against ensemble systems with a lower resolution.

Therefore the main purpose of this thesis is to assess the performance of the newly developed high-resolution ensemble prediction system for a number of HIW events similar to that reported in Fig. 1.1. It is planned to compare its performance against the above mentioned state-of-the-art ensemble prediction systems, both running on an operational basis. More precisely, the

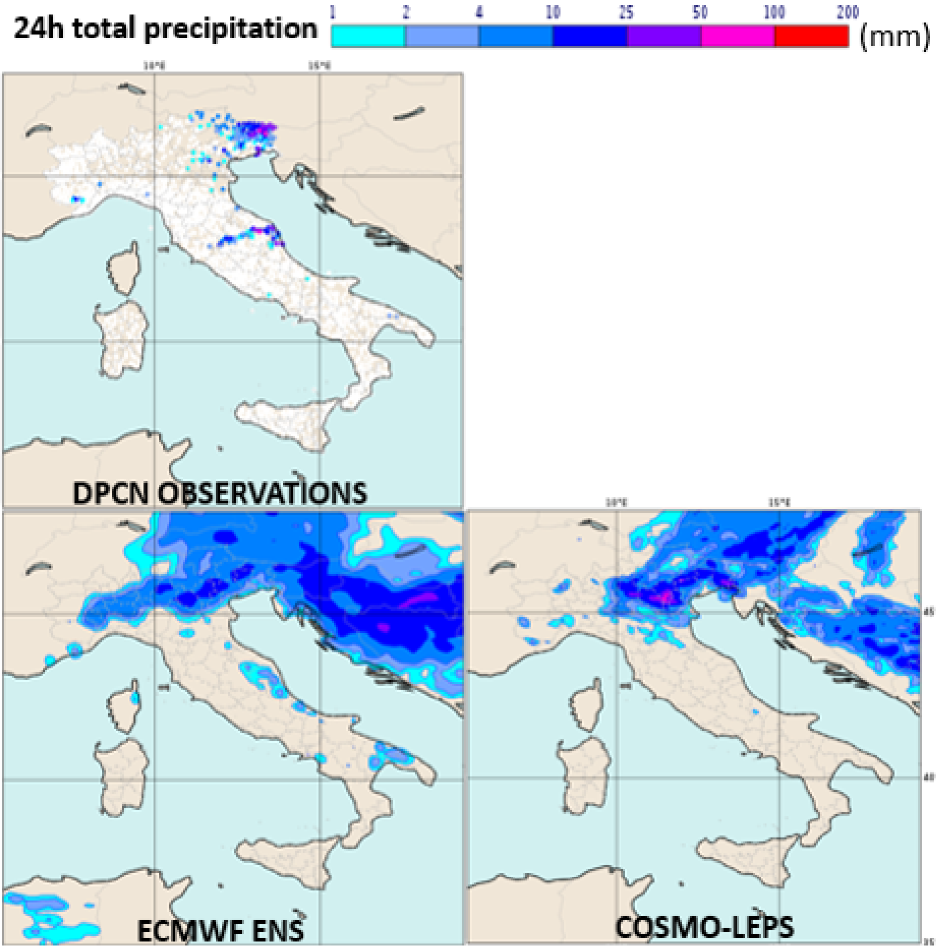


Figure 1.1: Maps of total precipitation cumulated over 24 hours for the 27th June 2016: on the top observations from rain gauges of DPCN, in the bottom left as was predicted by the run of 26th June 00 UTC from member 1 of ECMWF ENS, on the bottom right the same but for COSMO-LEPS.

main issues to be addressed in this work can be summarised as follows:

1. *How do the different ensemble systems behave in terms of prediction skill for both upper-level and surface variables?*
2. *What is the added value of high resolution? In which type of verification does it emerge more significantly?*
3. *Does the use of different verification methodologies provide an insight in the forecast skill of the ensemble systems based on COSMO model?*

In order to address these issues, different experiments will be performed in this study. After an introduction to weather prediction and ensemble systems in Chapter 2, a presentation of the different ensemble prediction systems follows in Chapter 3. Chapter 4 contains a synoptic description of the events and the methodology of evaluation of the ensemble systems, divided into deterministic and probabilistic scores. A detailed presentation of the main results is included in Chapter 5. These results are divided into four distinct sections: the upper-level variables (geopotential and temperature), the surface ones (2-metre temperature and 6-h total precipitation), the sensitivity of the scores to the verification methodology and a deterministic evaluation of the ensemble systems. Finally, conclusions are drawn in Chapter 6.

Chapter 2

Chaos and predictability

2.1 Numerical Weather Prediction

A dynamical system shows a chaotic behavior if most orbits exhibit sensitive dependence (Lorenz 1993). An orbit is characterized by sensitive dependence if most other orbis that pass close to it at some point do not remain close to it ad time advances. The atmosphere shows this behavior. The atmosphere is an intricate dynamical system with many degrees of freedom. The state of the atmosphere is described by the spatial distribution of wind, temperature and other weather variables (e.g. specific humidity and surface pressure). The mathematical differential equations describing the system time evolution include Newton's laws of motion used in the form "acceleration equals force divided by mass" and the laws of thermodynamics which describe the behavior of temperature and the other weather variables. Thus, generally speaking, there is a set of differential equations that describes the weather evolution, at least, in an approximate form.

Richardson(1922) can be considered the first one to have demonstrated that weather can be predicted numerically. In his work, he approximated the differential equations governing the atmospheric motions with a set of algebraic difference equations for the tendencies of various field variables at a finite number of grid points in space. By extrapolating the computed tendencies ahead in time, he could predict the field variables in the future. Unfortunately, his results were very poor, both because of deficient initial data, and because of the serious problems his approach implied.

After World War II the interest in numerical weather prediction revived, partly because of an expansion of the meteorological observation network, but also because of the development of digital computers. Charney (1947, 1948) developed a model applying an essential filtering approximation of Richardson's equations, based on the so-called geostrophic and hydrostatic equations. In 1950, an electronic computer (ENIAC) was installed at Princeton University and Charney, Fjørtoft and Von Neumann & Ritchmeyer (1950) made the first numerical prediction using the equivalent barotropic version of Charney's model. Charney's results led to the developments of more complex models of the atmospheric circulation, the so-called global circulation models.

With the introduction of powerful computers in meteorology, the meteorological community invested more time and efforts to develop more complex Numerical Weather Prediction (NWP) models of the atmosphere. Numerical Weather Prediction (NWP) is realised by integrating primitive-equation models. The equations are solved by replacing time-derivatives by finite differences and spatially either by finite difference schemes or spectral methods. The state of the atmosphere is described at a series of grid-points (Fig. 2.1) and vertical levels (Fig. 2.2) by a set of state variables such as temperature, velocity, humidity and pressure.

Meteorological observations made all over the world (Fig. 2.3) are used to compute the best estimate of the system initial conditions. Some of these observations, such as the ones from weather ballons or radiosondes, are taken at specific times at fixed locations (Fig. 2.4). Other data, such as the ones from aircrafts, ships or satellite, are not fixed in space. Thus the observations used for the analysis of the atmosphere can be divided roughly into conventional, in-situ observations and non-conventional, remote-sensing observations. The conventional observations consist of direct observations from surface weather stations, ships, buoys, radiosonde stations and aircraft, both at synoptic and, increasingly, at asynoptic hours. All surface and mean sea-level-pressure observations are used, with the exception of cloud cover, 2 m temperature and wind speed (over land). 2 m temperature and dew point

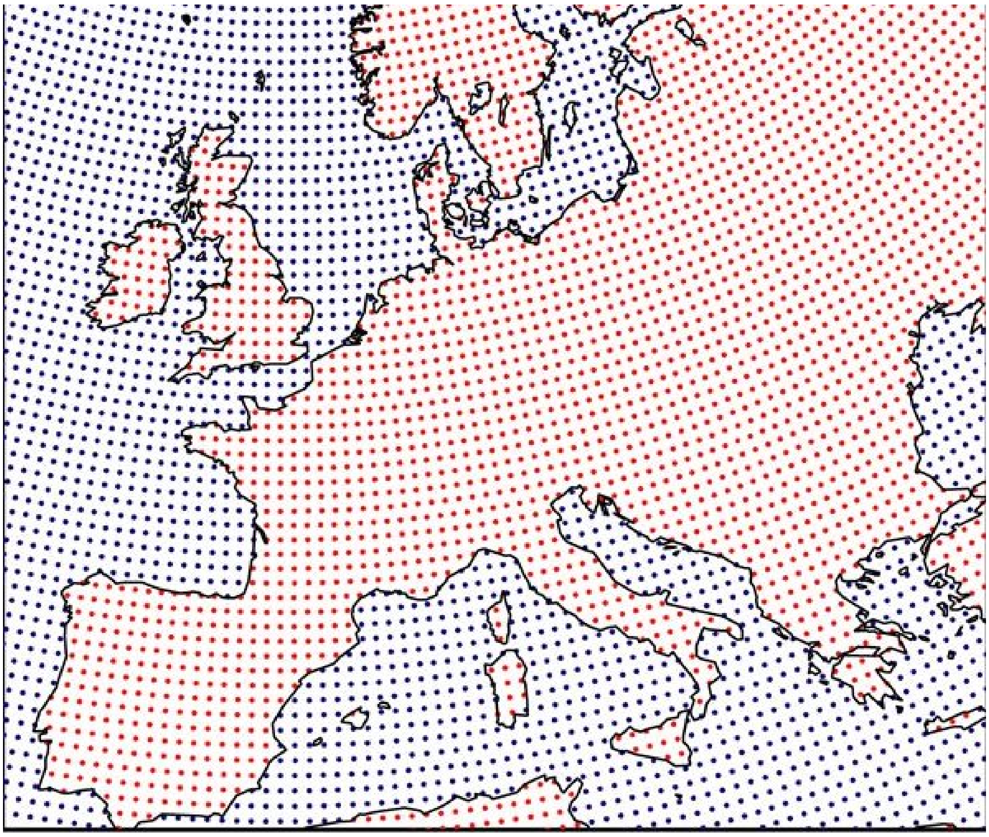


Figure 2.1: Grid points over Europe of ECMWF model (souce: ECMWF).

observations are used in the analysis of soil moisture. Observed winds are used from ships and buoys but not from land stations, not even from islands or coastal stations. The non-conventional observations are achieved in two different ways: passive technologies sense natural radiation emitted by the earth and atmosphere or solar radiation reflected by the earth and atmosphere; active technologies transmit radiation and then sense how much is reflected or scattered back. In this way surface-wind vector information is, for example, derived from the influence of the ocean capillary waves on the back-scattered radar signal of scatterometer instruments (Hersbach and Janssen 2007). Generally speaking, there is a great variability in the density of the observation network. Data over oceanic regions, in particular, are characterised by very coarse resolution.

Observations cannot be used directly to start model integration, but must

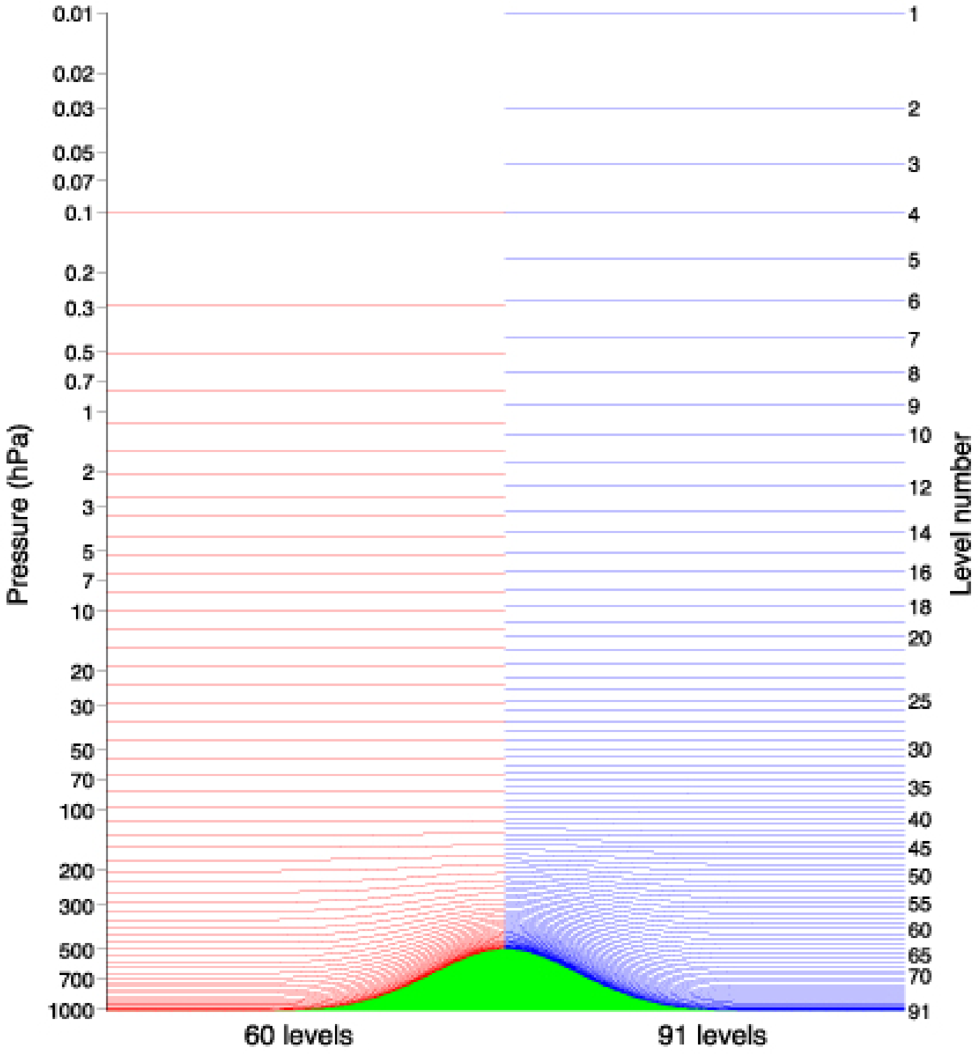


Figure 2.2: Vertical levels of the ECMWF model in previous versions (source: ECMWF).

be modified in a dynamical consistent way to obtain a suitable data set. This process is usually referred to as data assimilation.

In the ECMWF model, for example, dynamical quantities as pressure and velocity gradients are evaluated in spectral space, while computations involving processes such as radiation, moisture conversion, turbolence are calculated in grid-point space. This combination preserves the local nature of physical processes and retains the superior accuracy of the spectral method

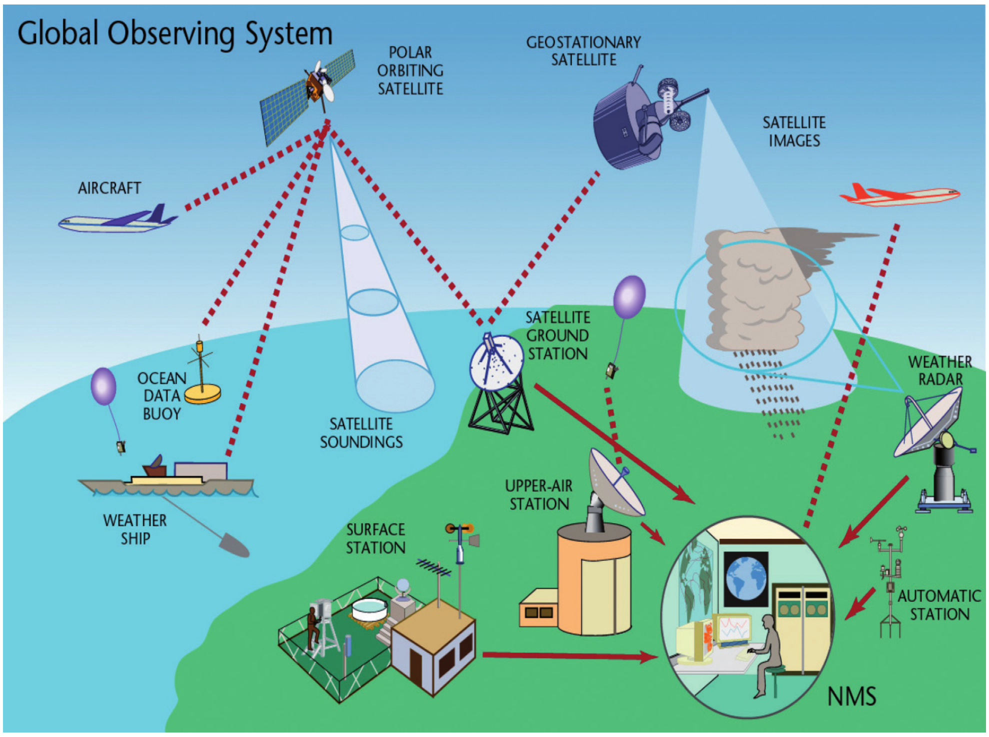


Figure 2.3: Type of observations used to estimate the atmosphere initial conditions in a typical day (source: ECMWF).

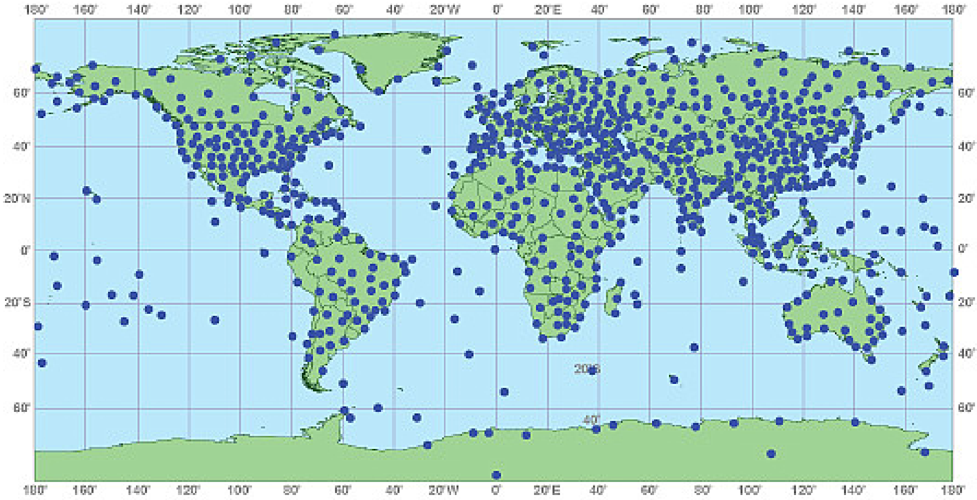


Figure 2.4: Map of radiosonde locations (source: ECMWF).

for dynamical computation.

The physical processes associated with radiative transfer, turbolent mixing, moist processes are active at smaller scales than the horizontal grid size. The approximation of unresolved processes in terms of model-resolved variables is referred to as parameterisation (Fig. 2.5). The parameterisation of physical processes is probably one of the most difficult and controversial area of weather modelling (Holton 1992).

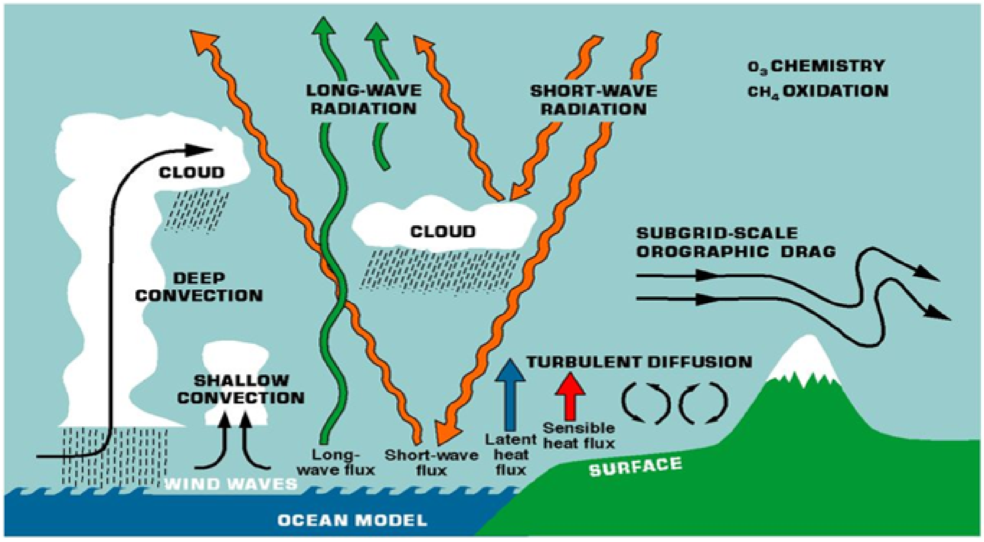


Figure 2.5: Schematic diagram of the different physical processes represented in the ECMWF model (source: ECMWF).

Nowadays, one of the most complex models used routinely for operational weather prediction is the one implemented at the European Centre for Medium-Range Weather Forecasts (ECMWF). The starting point, in mathematical terms known as the initial conditions, of any numerical integrations is given by very complex assimilation procedures that estimate the state of the atmosphere by considering all available observations. The fact that a limited number of observations are available (limited compared to the degrees of freedom of the system) and that part of the globe is characterized by a very poor coverage introduces uncertainties in the initial conditions. The initial conditions of a numerical weather prediction model can be estimated only within a certain accuracy. During a forecast some of these initial errors can amplify and result in significant forecast errors. Morover, the representation of the dynamics and physics of the atmosphere by numerical algorithms introduces further uncertainties associated for instance with truncation errors, with uncertainty of parameters describing sub-grib scale processes such as cumulus convection in a global model. We will refer to these two kind of errors as initial condition errors and model errors, respectively. For the prediction of the real atmosphere, these two kinds of errors are not really separable because the estimation of the initial conditions involves a forecast model and thus initial condition errors are affected by model errors. A requirement for skifull predictions is for numerical models to be able to accurately simulate the dominant atmospheric phenomena. Computer resources contribute to limit the complexity and the resolution of numerical models and assimilation, as long as, in order to be useful, numerical predictions need to be produced within a resonable time limit.

These two sources of forecast errors generate weather forecast deterioration with forecast time.

Initial conditions will always be known approximately, since each item of data is characterized by an error that depends on the instrumental accuracy. In other words, small uncertainties related to the characteristics of the atmospheric observing system will always characterize the initial conditions. As a consequence, even if the system equations were well known, two initial states only slightly differing would depart one from the other very rapidly as time progresses (Lorenz 1965). Observational errors, usually in the smaller scales, amplify and through nonlinear interactions spread to longer scales, eventually affecting the skill of these later ones (Somerville 1979).

The error growth of the 10-day forecast of the ECMWF model was analyzed in great detail by Simmons et al. (1995). It was concluded that 15 years of research had improved substantially the accuracy over the first half of the forecast range (say up to forecast day 5), but that there had been little error reduction in the late forecast range. While this applied on average, it has also been pointed out that there had been improvements in the skill of the good forecast. In other words, good forecast had higher skill now, than before. The problem was that it was difficult to assess a-priori whether a forecast would be skillful or unskillful using only a deterministic approach to weather prediction.

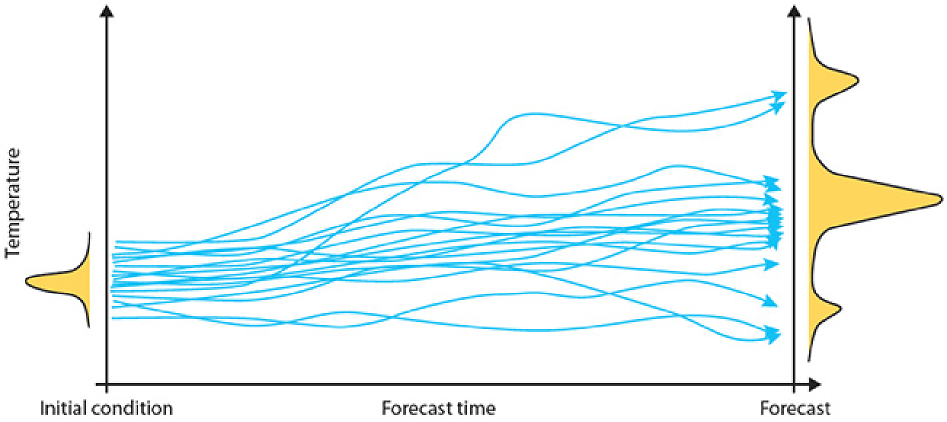


Figure 2.6: The predictability problem may be explained in terms of the time evolution of an appropriate probability density function (PDF). Ensemble prediction based on finite number of deterministic integration seems to be a feasible method to predict the PDF beyond the range of linear growth (source: ECMWF).

Generally speaking, a complete description of the weather prediction problem can be stated in terms of the time evolution of an appropriate probability density function (PDF) in the atmosphere's phase space (Fig. 2.6). Although this problem can be formulated exactly through the continuity equation for probability, ensemble prediction based on a finite number of deterministic integrations appears to be the only feasible method to predict the PDF beyond the range of linear error growth. Ensemble prediction provided a way to overcome one of the problems highlighted by Simmons et al. (1995), since it can be used to estimate the forecast skill of a deterministic forecast, or, in other words, to forecast the forecast skill.

Since December 1992, both the US National Centre for Environmental Predictions (NCEP) and ECMWF have complemented their deterministic high-resolution prediction with medium-range ensemble prediction (Tracton & Kalnay 1993, Palmer et al. 1993). These developments followed the theoretical and experimental work of, among others, Epstein (1969), Gleeson (1970), Fleming (1971a-b) and Leith (1974).

Both centres followed the same strategy of providing an ensemble of forecasts computed with the same model, one started with unperturbed initial conditions referred to as the "control" forecast and the others with initial conditions defined adding small perturbations to the control initial condition (Fig. 2.7). Generally speaking, the two ensemble systems differ in the ensemble size, specifically in the fact that at NCEP a combination of lagged forecasts is used, and in the definition of the perturbed initial. The reader is referred to Toth & Kalnay (1993) for the description of the 'breeding' method applied at NCEP and to Buizza & Palmer (1995) for a thorough discussion of the singular vector approach followed at ECMWF.

If forecast starting from perturbed analysis agrees more or less with the forecast from the non-perturbed analysis (the ensemble control forecast), then the atmosphere can be considered to be in a predictable state and any unknown analysis errors would not have a significant impact. In such a case, it would be possible to issue a categorical forecast with great certainty. On the other hand, if the perturbed forecasts (the ENSemble (ENS)) deviates significantly from the control forecast and from each other, it can be concluded that the atmosphere is in a rather unpredictable state. In this case, it would not be possible to issue a categorical forecast with great certainty. However, the way in which the perturbed forecast differs from each other may provide valuable indications of which weather patterns are likely to develop or, often equally importantly, not develop.

The ENS provides the ensemble mean (EM) forecast (or the ensemble median) where the less predictable atmospheric scales tend to be averaged out. In a well-costructed ensemble systems, the accuracy of the EM can be estimated a priori by the spread of the ensemble: the larger the spread, the larger the expected EM error, on average (Buizza, 2001). More importantly, the ENS provides information from which the probability of alternative developments is calculated, in particular those related to high-impact weather.

The characteristics of a good ensemble are:

- The ensemble forecasts should display no mean errors (bias), otherwise the probabilities will be biased as well;
- The forecasts should have the ability to span the full climatological range, otherwise the probabilities will either over-or under-forecast the

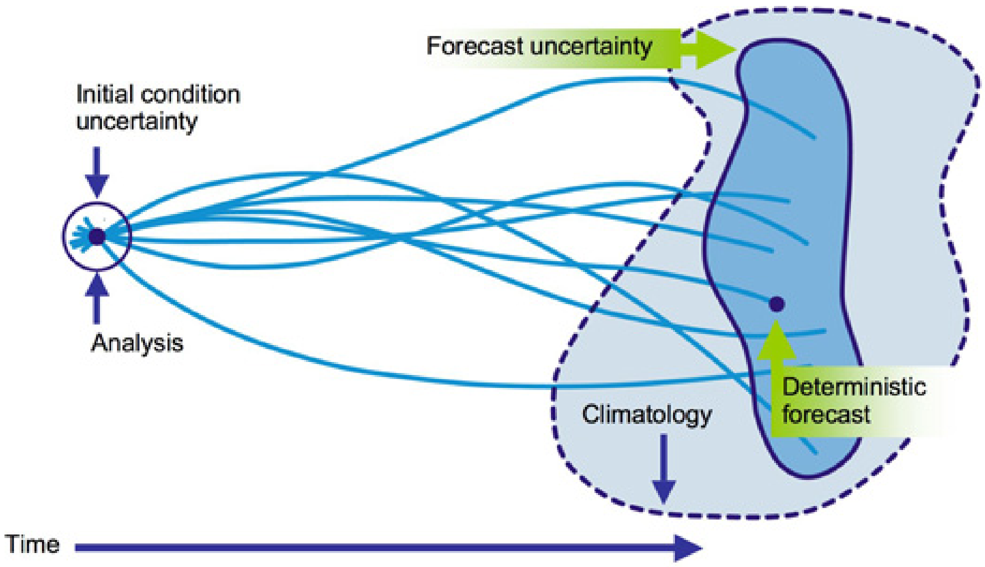


Figure 2.7: Schematic of a probabilistic weather forecast using initial condition uncertainties. The blu lines show the trajectories of the individual forecasts that diverge from each other owing to uncertainties in the initial conditions and in the representation of sub-grib scale processes in the model. The main goal is to explore all the possible future states of the atmosphere. The dashed, lighter blue envelope represents the range of possible states that the real atmosphere could encompass and the solid, dark blue envelope represents the range of states sampled by the model predictions. A good forecast is the one which analysis lies inside the ensemble spread (source: ECMWF).

risks of anomalous or extreme weather events.

Therefore, numerical weather prediction is, by its very nature, a discipline that has to deal with uncertainties. Over the past 15 years, ensemble forecasting became established in numerical weather prediction centres as a response to the limitations imposed by the inherent uncertainties in the prediction process. The ultimate goal of ensemble forecasting is to predict qualitatively the probability density of the state of the atmosphere at a future time. This is a nontrivial task because the actual uncertainty depends on the flow itself and thus varies from day to day.

2.2 The Lorenz system

Chaos Theory is a mathematical theory that can be used to explain complex systems such as weather. Although many complex systems appear to behave in a random manner, chaos theory shows that, in reality, there is an underlying order that is difficult to see. Many complex systems can be better understood through the lens of Chaos Theory. Henri Poincaré laid the groundwork for Chaos Theory. He was the first to point out that many deterministic systems display a sensitive dependence on initial conditions. Later, in the 1900s, Edward Lorenz (1963, 1965) studied Chaos Theory in the context of weather systems. When making weather predictions, he noticed that his calculations were significantly impacted by the extent to which he rounded his numbers. The end result of the calculation was significantly different when he used a number rounded to three digits as compared to a number rounded to six digits. His observations on Chaos Theory in weather systems led to his famous talk, which he entitled, "Predictability: does the flap of a butterfly's wings in Brazil set off a tornado in Texas?". Lorenz chaos model equations are:

$$\begin{cases} \dot{X} = -\sigma X + \sigma Y \\ \dot{Y} = -XY + rX - Y \\ \dot{Z} = XY - bZ \end{cases}$$

where σ is called the Prandtl number and r is called the Rayleigh number. All σ, r, $b > 0$, but usually $\sigma = 10$, b=8/3 and r is varied. The parameters σ, r, b are kept constant within an integration, but they can be changed to create a family of solutions of dynamical system defined by the differential equations. The particular parameter values chosen by Lorenz (1963) were: $\sigma = 10$, b=8/3 and r=28 -which result in chaotic solutions (sensitivity dependence on the initial conditions). Results from 3-dimensional Lorenz system illustrate the dispersion of finite time integrations from an ensemble of initial conditions (Fig. 2.8). The different initial points can be considered as estimates of the "true" state of the system (which can be thought of as any point inside the ellipsoid) and the time evolution of each of them as possible forecasts. Subject to the initial "true" state of the system, points close together at initial time diverge in time at different rates. Thus, depending on the point chosen to describe the system time evolution, different forecasts are obtained.

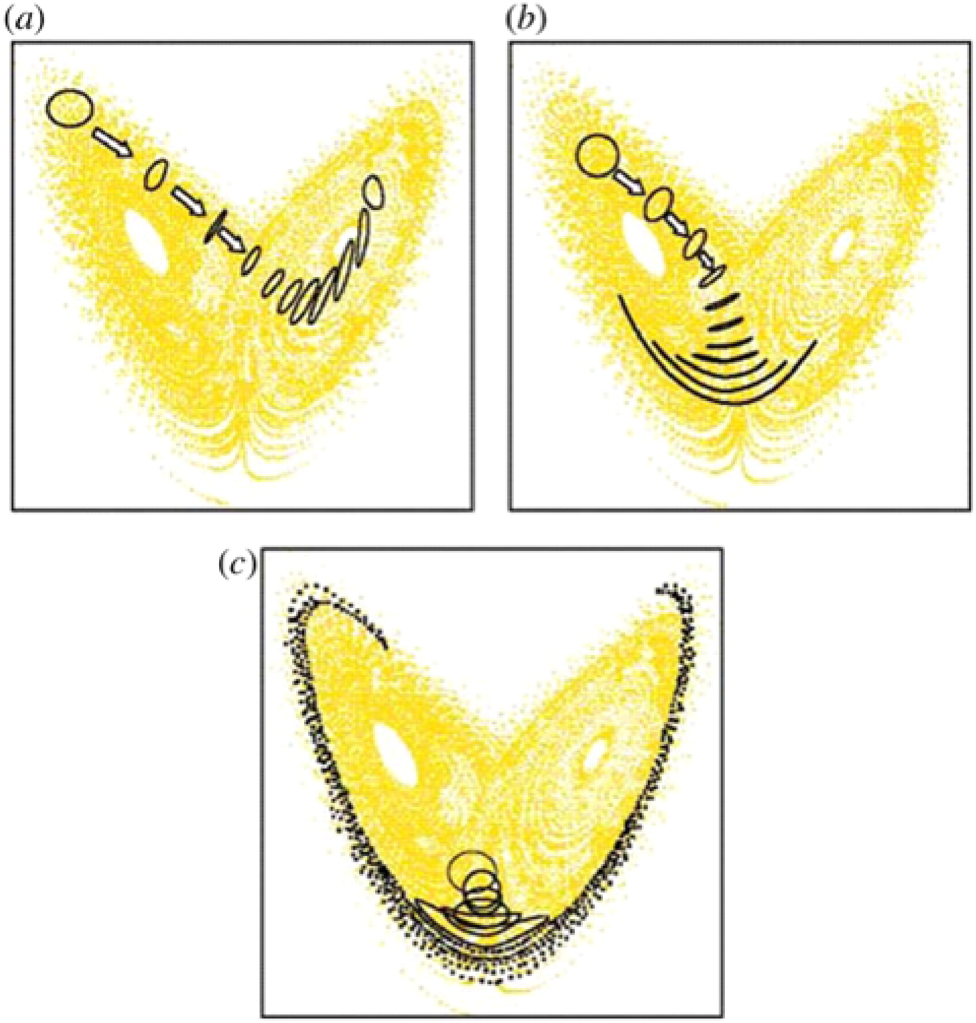


Figure 2.8: Lorenz attractor with superimposed finite-time ensemble integration (source: ECMWF).

The two wings of the Lorenz attractor can be considered as identifying two different weather regimes, for example one warm and wet and the other cold and sunny. Suppose that the main purpose of the forecast is to predict whether the system is going through a regime transition. When the system is in a predictable initial state (Fig. 2.8(a)), the rate of forecast divergence is small and all the points stay close together untill the final time. Whatever the point chosen to represent the initial state of the system, the forecast is characterised by a small error and a correct indication of a regime transition is given. The ensemble of points can be used to generate probabilistic forecast of regime transitions. In this case, since all points end in the other wing of the attractor, there is a 100% probability of regime transition. By contrast, when the system is in a less predictable state (Fig. 2.8(b)), the points stay close together only for a short time period and then start diverging. While it is still possible to predict with a good degree of accuracy the future forecast state of the system for a short time period, it is difficult to predict whether the system will go through a regime transition in the long forecast range. Fig. 2.8(c) shows an even worse scenario, with points diverging even after a short time period and ending in very distant part of the system attractor. In probabilistic terms, one could have only predicted that there is 50% chance of the system undergoing a regime transition. Morover, the ensemble of points indicates that there is a greater uncertainty in predicting the region of the system attractor where the system will be at final time in the third case (Fig. 2.8(c)).

The comparison of the points' divergence during the three cases indicates how ensemble prediction systems can be used to "forecast the forecast skill" (Tennekes et al., 1986). In the case of the Lorenz system, a small divergence is associated to a predictable case and confidence can be attrached to any of the single deterministic forecasts given by the single points. By contrast, a large diverge indicates low predictability.

Similar sensivity to the initial state is shown in weather prediction. Fig. 2.9 shows the forecasts for air temperature in London given by 33 different forecasts started from very similar initial conditions for two different dates, the 26^{th} of June of 1995 and the 26^{th} of June of 1994; in practice the image is about forecasts in the same place, one year apart. There is a clear difference degree of divergence during the two cases. All forecasts stay close together up to forecast day 10 for the first case (Fig. 2.9(a)), while they all diverge already at forecast day 3 in the second case (Fig. 2.9(b)). The level of spread among the different forecast can be used as a measure of the predictability of the two atmospheric states.

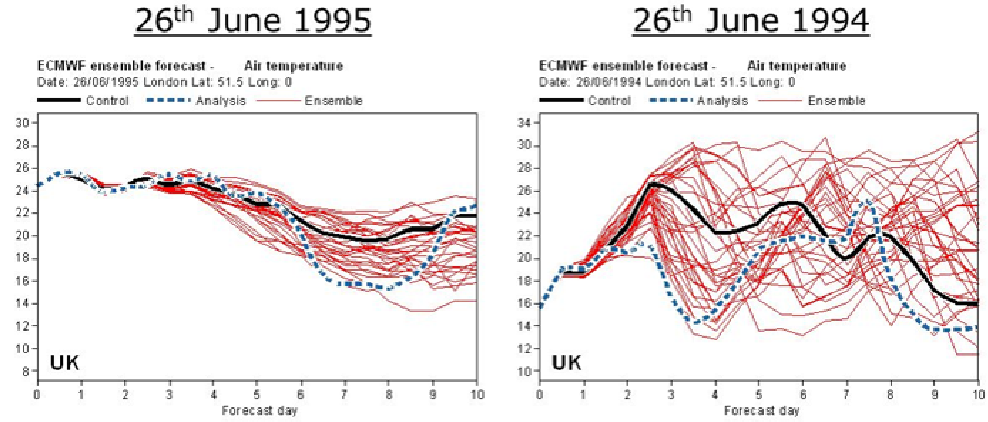


Figure 2.9: ECMWF forecasts for air temperature in London started from (a) 26 June 1995 and (b) 26 June 1994 (source: ECMWF).

2.3 Representation of model error

It has been already mentioned that uncertainties in the initial conditions and in the model are both sources of forecast error. Results from Harrison et al. (1999) indicate that the impact of model uncertainties on forecast error cannot be neglected. These results suggest that an ensemble system should try to describe not only the presence of uncertainties in the initial conditions, but also of model uncertainties. There is significant source of random error associated with the parameterized physical processes.

The laws of evolution which govern weather and climate, at least their physical aspects, are well known and are accurately represented by sets of partial differential equations. These equations nonlinearly couple circultations with different scales and are thus difficult to solve analytically. To solve the governing equations numerically, we project them onto some chosen orthonormal basis, thus determining a set of (up to 10^8) coupled ordinary differential equations. Within these equations, the nonlinear effect of unresolved scales of motion are traditionally represented by simplified deterministic formulae, known as parameterisations (Leutbecher and Palmer, 2007). These parameterisations represent the bulk effect of subgrid processes within a grid box and are justified in much the same way diffusive formulae are justified in statistical mechanics. Hence, for example, parameterisation of deep convection assumes the existence of an ensemble of deep convective plumes, in quasi-equilibrium with the larger scale environment. The associated parameterised convective tendency represents the bulk effect of these plumes in redistributing heat, momentum and water in the vertical column of a given grid box. Similarly, parameterisation of orographic gravity-wave drag assumes the existence of an ensemble of incoherent gravity waves which collectively are associated with a flux of momentum from the surface to some level of wave breaking.

Parameterisations are by their nature approximations. Hence the parameterised convective or orographic tendencies, which represent the mean effect of these processes over many realisations, are usually different from the tendencies associated with the actual convective or orographic subgrid flow. Since the latter is not known, the parameterisation process is necessarily a source of uncertainty in numerical forecasts and must therefore be represented explicitly in any ensemble forecast system. Without such a source of uncertainty either the ensemble will be underdispersive, or other sources of error, e.g. associated with observational uncertainty, will have to be inflated to prevent underdispersion. In this context, we note again that since the forecast model is used to assimilate observations in generating the initial conditions for a forecast, initial error includes a component due to model error. That is to say, when one speaks of forecast uncertainty as including initial error and model error, these two classes of errors are strictly interconnected.

There are several approaches to represent model errors: among them the most noticeable are the multi-model ensemble, the perturbed parameter ensemble and stochastic-dynamic parameterisation.

Chapter 3

Global and limited-area ensemble prediction systems

This chapter describes three ensemble systems with different characteristics that will be subjected to a careful verification work in this thesis. These are ECMWF ENS, COSMO-LEPS and COSMO-2I-EPS, whose performance will be evaluated in chapter 5, on the basis of a week of observed data.

3.1 The ECMWF global atmospheric model

The ECMWF Integrated Forecasting System (IFS) consists of several components: an atmospheric general circulation model, an ocean wave model, a land surface model, an ocean general circulation model and perturbation models for Data Assimilation (EDA) and forecast (ENS) ENSemble, producing forecasts from days to weeks and months ahead (ECMWF Forecast User Guide). The atmospheric general circulation model describes the dynamical evolution on the resolved scale and is augmented by the physical parameterisation, describing the mean effect of subgrid processes and the land-surface model. Coupled to this is an ocean wave model (Bechtold et al. 2008).

3.1.1 The IFS equations

The model formulation is based on a set of basic equations, of which some are *diagnostic*, describing the static relationship beetween pressure, density, temperature and height, and some are *prognostic*, describing the time evolution of the horizontal wind components, surface pressure, temperature and the water vapour contents of an air parcel. Additional equations describe changes in the hydrometeors (rain, snow, liquid water, cloud ice content etc). There are options for passive tracers such as ozone. The processes of radiation, gravity wave drag, vertical turbolence, convection, clouds and surface interaction are, due to their relatively small scales (unresolved by the model's resolution), described in a statistical way as parameterization processes (arranged in entirely vertical columns). Following Ritchie (1988, 1991), the first step in developing a semi-Lagrangian version of the ECMWF spectral model was to convert the existing Eulerian $\zeta - D$ (vorticity-divergence) model to a $U - V$ formulation, where U and V are the wind images defined by $U = u \cos\theta$, $V = v \cos\theta$ (u and v are components of the horizontal wind in spherical coordinates, and θ is latitude). We describe the Eulerian $U - V$ model. First we set out the continuous equations in (λ, θ, η) coordinates, where λ is longitude and η is the hybrid vertical coordinate introduced by Simmons and Burridge (1981). Therefore $\eta(p, p_s)$ is a monotonic function of the pressure p and also depends on the surface pressure p_s in such a way that

$$\eta(0, p_s) = 0 \quad \text{and} \quad \eta(p_s, p_s) = 1$$

The momentum equations are

$$\begin{aligned} &\frac{\partial U}{\partial t} + \frac{1}{a\cos^2\theta} U \frac{\partial U}{\partial \lambda} + V\cos\theta \frac{\partial U}{\partial \theta} + \dot{\eta}\frac{\partial U}{\partial \eta} \\ &-fV + \frac{1}{a}\frac{\partial \phi}{\partial \lambda} + R_{dry} T_V \frac{\partial}{\partial \lambda}(\ln p) = P_U + K_U \end{aligned} \tag{3.1}$$

$$\begin{aligned} &\frac{\partial V}{\partial t} + \frac{1}{a\cos^2\theta} U \frac{\partial V}{\partial \lambda} + V\cos\theta \frac{\partial V}{\partial \theta} + \sin\theta(U^2 + V^2) + \dot{\eta}\frac{\partial V}{\partial \eta} \\ &+fU + \frac{\cos\theta}{a}\frac{\partial \phi}{\partial \theta} + R_{dry} T_V \frac{\partial}{\partial \theta}(\ln p) = P_V + K_V \end{aligned} \tag{3.2}$$

where a is the radius of the earth, $\dot{\eta}$ is the η-coordinate vertical velocity ($\dot{\eta} = \frac{\partial \eta}{\partial t}$), ϕ is geopotential, R_{dry} is the gas constant for dry air and T_V is the virtual temperature defined by

$$T_V = T1 + [(R_{vap}/R_{dry}) - 1]q$$

where T is temperature, q is specific humidity and R_{vap} is the gas constant for water vapour. P_U and P_V represent the contributions of the parameterized physical processes, while K_U and K_V are the horizontal diffusion terms. The thermodynamic equation is

$$\frac{\partial T}{\partial t} + \frac{1}{a\cos^2\theta} U \frac{\partial T}{\partial \lambda} + V\cos\theta \frac{\partial T}{\partial \theta} + \dot{\eta}\frac{\partial T}{\partial \eta} - \frac{kT_V\omega}{(1+(\delta-1)q)p} = P_T + K_T \quad (3.3)$$

where $k = \frac{R_{dry}}{c_{p_{dry}}}$ (with $c_{p_{dry}}$ the specific heat of dry air at constant pressure), ω is the pressure-coordinate vertical velocity ($\omega = \frac{\partial p}{\partial t}$) and $\delta = \frac{c_{p_{vap}}}{c_{p_{dry}}}$ (with $c_{p_{vap}}$ the specific heat of water vapour at constant pressure). The moisture equation is

$$\frac{\partial q}{\partial t} + \frac{1}{a\cos^2\theta} U \frac{\partial q}{\partial \lambda} + V\cos\theta \frac{\partial q}{\partial \theta} + \dot{\eta}\frac{\partial q}{\partial \eta} = P_q + K_q \quad (3.4)$$

In the momentum equations, there is no vertical velocity, which therefore is not a prognostic variable of the model, but is diagnostic. In (3.2) and (3.3), P_T and P_q represent the contributions of the parametrized physical processes, while K_T and K_q are the horizontal diffusion terms. The continuity equation is

$$\frac{\partial}{\partial t}(\frac{\partial p}{\partial \eta}) + \nabla \cdot (\mathbf{v}_H \frac{\partial p}{\partial \eta}) + \frac{\partial}{\partial \eta}(\dot{\eta}\frac{\partial p}{\partial \eta}) = 0 \quad (3.5)$$

where ∇ is the horizontal gradient operator in spherical coordinates and $\mathbf{V}_H = (u, v)$ is the horizontal wind. The geopotential ϕ, which appears in (3.1) and (3.2), is defined by the hydrostatic equation

$$\frac{\partial \phi}{\partial \eta} = -\frac{R_{dry}T_V}{p}\frac{\partial p}{\partial \eta} \quad (3.6)$$

while the vertical velocity ω in (3.3) is given by

$$\omega = -\int_0^\eta \nabla \cdot (\mathbf{v}_H \frac{\partial p}{\partial \eta})\, d\eta + \mathbf{v}_H \cdot \nabla p \quad (3.7)$$

This equation for verical velocity is diagnostic: ω is obtained from the divergence of the horizontal wind. In this ensemble system the vertical equation is approximated with the hydrostatic equation, therefore it is possible define this model as hydrostatic. Expression for the rate of change of surface pressure and for the vertical velocity $\dot{\eta}$, are obtained by integrating (3.5), using the boundary conditions $\dot{\eta}$ at $\eta = 0$ and at $\eta = 1$

$$\frac{\partial p_S}{\partial t} = -\int_0^1 \nabla \cdot (\mathbf{v}_H \frac{\partial p}{\partial \eta})\, d\eta \tag{3.8}$$

$$\dot{\eta}\frac{\partial p}{\partial \eta} = -\frac{\partial p}{\partial t} - \int_0^\eta \nabla \cdot (\mathbf{v}_H \frac{\partial p}{\partial \eta})\, d\eta \tag{3.9}$$

Since we use $\ln(p_s)$ rather than p_s as the surface pressure variable, it is convenient to rewrite (3.8) as

$$\frac{\partial}{\partial t}(\ln p_s) = -\frac{1}{p_s}\int_0^1 \nabla \cdot (\mathbf{v}_H \frac{\partial p}{\partial \eta})\, d\eta \tag{3.10}$$

3.1.2 The numerical formulation

The model equations are discretized in space and time and solved numerically by a semi-Lagragian advection scheme. This ensures stability and accuracy, using a time-step as large as possible to progress the computation of the forecast within an acceptable time. For the *horizontal representation* a dual representation of spectral components and grid points is used. All fields are described in grid point space. Due to the convergence of meridians, computational time can be saved by applying a "reduced Gaussian grid". This keeps the east-west separation between points almost constant by gradually decreasing the number of grid points towards the poles at every latitude in the extra-tropics. This also prevents the development of large gradient at high latitudes. For the computation of horizontal derivatives, a spectral representation, based on a series expansion of spherical harmonics, is used for a subset of the prognostic variables. The *vertical resolution* is the finest in geometric height in the planetary boundary layer and the coarsest close to the model top. The "σ-levels" follow the earth's surface in the lower-most troposphere, where the Earth's orography displays large variations. In the upper stratosphere and lower mesosphere there are surfaces of constant pressure with a smooth transition in between (ECMWF Forecast User Guide).

3.1.3 Topographical and climatological fields

The model orography is derived from a data set with a resolution of about 1 km which contains values of the mean elevation above the mean sea level, the fraction of land and the fractionale cover of different vegetation types. These detailed data are aggregated (”upscaled”) to the coarser model resolution. The resulting mean orography contains the values of the mean elevation above the mean sea level. In montainous areas it is supplemented by subgrib orographic fields, to enable the parametrization of the effects of gravity waves and to provide flow-dependent blocking of air flow. For example, cold air drainage in valleys makes the cold air effectively ”lift” the orography. The land-sea mask is a geographical field that contains the percentage of land and water between 0 (100% sea) and 1 (100% land) for every grid point. A grid point is defined as a land point if its value indicates that more than 50% of the area within the grid-box is covered by land. The albedo is determined by a combination of background monthly climate fields and forecast surface fields (e.g. from snow depth). Continental, maritime, urban and desert aerosols are provided as monthly means from data bases derived from transport models covering both the troposphere and the stratosphere. Soil temperatures and moisture in the ground are prognostic variables. There is a lack of observational data, so observed 2m temperature and relative humidity act as very efficient proxy data for the analysis. The snow coverage depth is analysed every six hours from snow-depth observations, satellite snow extent and a snow-depth background field. The snow temperature is also analysed from satellite estimates. They are forecast variables. Sea Surface Temperature (SST) and ice concentration are based on analyses received daily from the Met Office (OSTIA, 5km). It is updated during the model integration, according to the tendency obtained from climatology. The temperature at the ice surface is variable and calculated according to a simple energy balance/heat budget scheme, where the SST of the underlying ocean is assumed to be -1.7 ^{o}C. The sea-ice cover, which is kept constant in the 10-day forecast integration, is relaxed towards climatology between days 10 and 30, with a linear regression. Beyond day 30 the sea-ice concentration is based on climatological values only (from the ERA 1979-2001 data) (ECMWF Forecast User Guide).

3.1.4 The formulation of physical processes

Many physical processes occur at horizontal scales which are not resolved in the model. Their "bulk effect" is expressed in terms of resolved model variables by parametrisation schemes. This involves both statistical methods and simplified mathematical-physical models such as adjustment processes. For example, the air closest to the earth's surface exchanges heat with the surface through turbolent diffusion or convection, which adjusts unstable air towards neutral stability (Jung et al. 2010). The convection scheme does not predict individual convective clouds, only their overall physical effect on the surrounding atmosphere, in terms of latent heat release, precipitation and the associated transport of moisture and momentum. The scheme differentiates between deep, shallow and mid-level convection. Only one type of convection can occur at any given grid point at one time (see Fig. 3.1).

As for clouds, both convective and non-convective clouds are handled by explicit equations for cloud water, ice and cloud cover. Liquid and frozen precipitation are strongly coupled to other parameterized processes, in particular the convective scheme and the radiation. The scheme also takes into account important clouds processes, such as the clouds that form in the lowest model level. The radiation spectrum is divided into a long-wave part (thermal) and a short-wave part (solar radiation). Since it has to take the cloud-radiation interaction into account in considerable detail, it makes use of a cloud-overlap algorithm which calculates the relative placement of clouds across levels. For the sake of computational efficiency, the radiation scheme is called less frequently than the model time step on a reduced grid. Nevertheless, it accounts for a considerable fraction of the total computational time. For the precipitation and hydrological cycles both convective and stratiform precipitation are included in the ECMWF model (ECMWF Forecast User Guide). Evaporation of the precipitation, before it reaches the ground, is assumed not to take place within the cloud, only in the cloud-free, non-

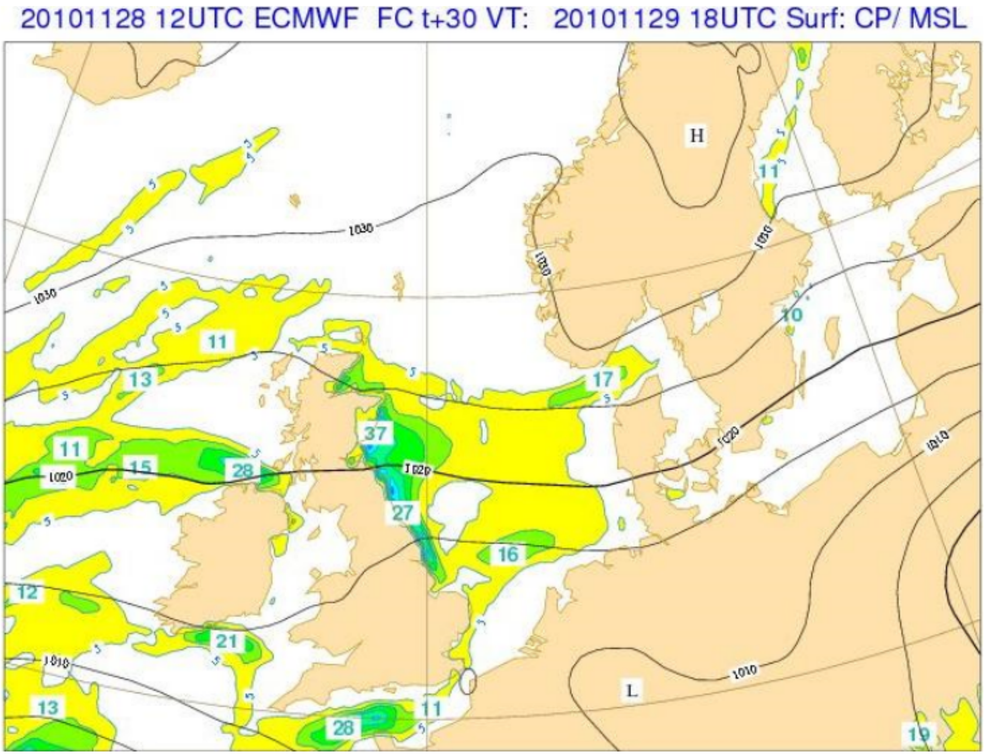

20101128 12UTC ECMWF FC t+30 VT: 20101129 18UTC Surf: CP/ MSL

Figure 3.1: The ECMWF total convective rainfall forecast from 28 November 2010 12 UTC +30h. The convection scheme has difficulty in advecting wintery showers inland over Scotland and northern England from the relatively warm North Sea. The convection scheme is diagnostic and works on a model column, so cannot produce large amounts of precipitation over the relatively dry and cold (stable) wintery land areas. In nature these showers succeed in penetrating inland through a convectively induced upper-level warm anomaly leading to large-scale lifting and saturation (source: ECMWF).

saturated air beside or below the model clouds. The melting of falling snow occurs in a thin layer of a few hungred metres below the freezing level. It is assumed that snow can melt in each layer, whenever the temperature exceeds 0^oC. The cloud-overlap algorithm is also important for the "life history" of falling precipitation: from level-with-cloud to level-with-clear-sky and vice versa. The near-surface wind forecast displays severe weaknesses in some mountain areas, due to the difficulty in parameterizing the interaction between the air flow and the highly varying sub-grid orography (see Fig. 3.2). As with many other sub-grid-scale physical processes that need to be treated in simplified ways, this problem will ultimately be reduced when the air-surface interaction can be described explicitly, thanks to a higher and appropriate resolution. The system also produces wind-gust forecasts as part of post-processing (Balsamo et al. 2011).

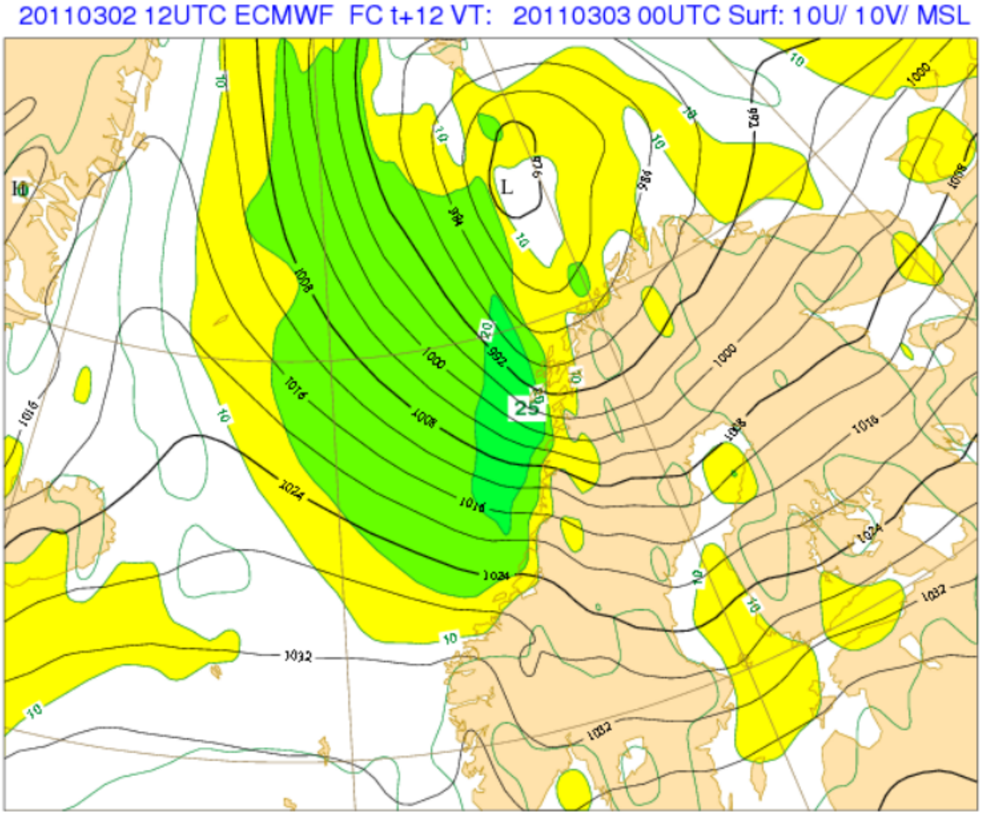

20110302 12UTC ECMWF FC t+12 VT: 20110303 00UTC Surf: 10U/ 10V/ MSL

Figure 3.2: MSLP and 10 m wind forecast from 2 March 2011 12 UTC +12h. The 10 m winds are unrealistically weak over the rugged Norwegian mountains. Value of 10 m/s might be realistic in sheltered valleys, but not on exposed mountain ranges (source: ECMWF).

3.1.5 Overview of the ECMWF Ensemble Prediction System

The ECMWF ENS Prediction System (hereafter ENS) contains 51 members, an unperturbed control forecast and 25 pairs of twin forecast with positive and negative initial perturbations. This yields a total of 50 global perturbations for 50 alternative analyses and forecasts. Therefore, consecutive members have pair-wise anti-symmetric perturbations. The anti symmetry may, depending on the synoptic situation and the distribution of the perturbations, disappear after one day or so, but they can occasionally be noticed 3-4 days into the perturbed forecasts. The horizontal resolution of the ENS is about 18 km, but it increases to 32 km after the 15^{th} day of the forecast range; the vertical resolution is 137 levels, which divide the atmosphere into many layers up to the isobaric hight of 0.01 hPa. The ENS configuration can be considered as an attempt to simulate random model errors due to parametrized physical processes. It is based on the notion that random errors due to parameterized physical processes are coherent between the different parameterization modules and have a certain coherence on the space and time scales represented by the model. The scheme assumes that the larger the parameterized tendencies, the larger the random error component. In the ENS, each ensemble member e_j can be seen as the time integration

$$e_j(t) = \int_{t=0}^{t} [A(e_j, t) + P'_j(e_j, t)]\, dt$$

of the perturbed model equations

$$\partial e_j/\partial t = A(e_j, t) + P'_j(e_j, t)$$

starting from the perturbed initial conditions

$$e_j(t = 0) = e_0(t = 0) + \delta e_j(t = 0)$$

where A and P' identify the contribution to the full equation tendency of the non-parameterized and parameterized physical processes. For each grid point $x = (\lambda, \phi, \sigma)$ (identified by its latitude, longitude and vertical hybrid coordinate), the perturbed parameterized tendency (of each state vector component) is defined as

$$P'_j(e_j, t) = [1+ < r_j(\lambda, \phi, \sigma) >_{D,T}]P(e_j, t)$$

where P is the unperturbed diabatic tendency and $< \cdots >_{D,T}$ indicates that the same random number r_j has been used for all grid points inside a

DxD degree box and over T time steps. The notion of space-time coherence assumes that the organized systems have some intrinsic space and time-scales that may span more than one model time step and more than one model grid point. Making the stochastic uncertainty proportional to the tendency is based on the concept that the stronger organization (away from the notion of a quasi-equilibrium ensemble of sub-grid processes) is likely to be, the stronger is the parameterized contribution. A certain space-time correlation is introduced in order to have tendency perturbations with the same spatial and time scales as observed organization. The performance of the EPS has improved steadily since it became operational in the mid 1990s, as shown in Fig.3.3.

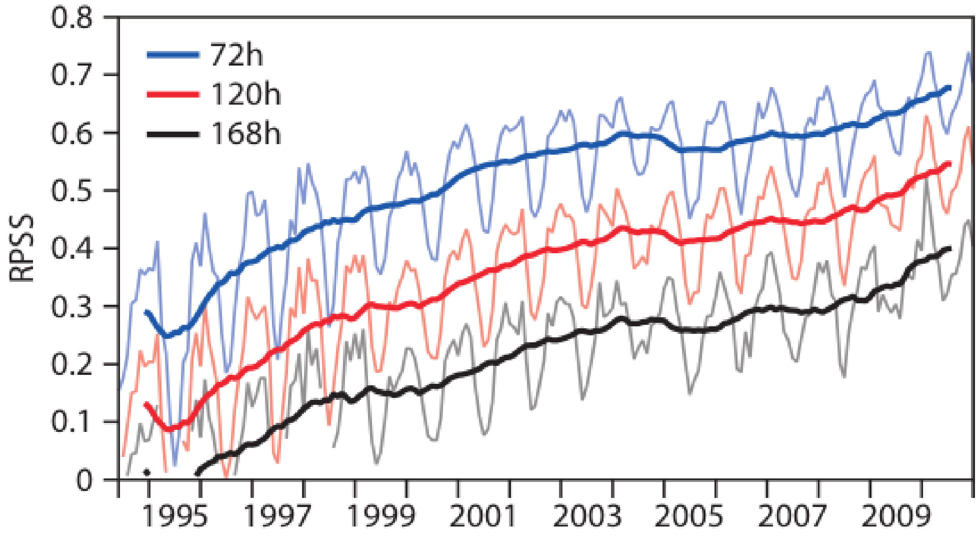


Figure 3.3: A skill measure for forecasts of the 850 hPa temperature over the northern hemisphere (20^o-90^oN) at days 3, 5 and 7. Comparing the skill measure at the three lead times demonstrates that on average the performance has improved by two days per decade. The level of skill reached by a 3-day forecast around 1998/99 (skill measure = 0.5) is reached in 2008/09 by a 5-day forecast. In other words, today a 5-day forecast is as good as a 3-day forecast 10 years ago. The skill measure used here is the Ranked Probability Skill Score (RPSS), which is 1 for a perfect forecast and 0 for a forecast no better than climatology (from ECMWF User Guide).

3.2 COSMO-based ensemble systems

The COSMO-Model is a nonhydrostatic limited-area atmospheric prediction model (www.cosmo-model.org). It has been designed for both operational numerical weather prediction (NWP) and various scientific applications on the meso-β (2-20 km) and meso-γ (20-100 km) scale. The COSMO-Model is based on the primitive thermo-hydrodynamical equations describing compressible flow in a moist atmosphere, with a variety of physical processes taken into account by parameterisation schemes (Doms et al. 2015). The basic version of the COSMO-Model (formerly known as *Lokal Modell (LM)* has been developed at the Deutscher Wetterdienst (DWD) and it has been run operationally since 1999. The subsequent developments related to the model have been organized within COSMO, the *Consortium for Small-Scale Modeling*. COSMO aims at the improvement, maintenance and operational application of a non-hydrostatic limited-area modeling system, which is now consequently called the COSMO-Model. For operational aims, COSMO model is run at horizontal resolutions ranging from 1 km to 14 km in both deterministic and ensemble mode.

3.2.1 Basic Model design and Features

The nonhydrostatic fully compressible COSMO-Model has been developed to meet high resolution regional forecast requirements of weather services and to provide a flexible tool for various scientific applications on a broad range of spatial scales. When starting with the development of the COSMO-Model, many NWP-models operated on hydrostatic scales of motion with grid spacings down to about 10 km and thus lacked the spatial resolution required to explicitly capture small-scale severe weather events (Schättler et al. 2016). The COSMO-Model has been designed for meso-β and meso-γ scales where nonhydrostatic effects begin to play an essential role in the evolution of atmospheric flows. However only by employing 1 to 3 km grid spacing for operational forecasts over a large domain, it is expected that deep moist convection and the associated feedback mechanisms to the larger scales of motion can be explicitly resolved (Doms et al. 2015). The requirements for the data assimilation system for the operational COSMO-Model are mainly determined by the very high resolution of the model and by the task of employing it also for nowcasting purposes in the future. Hence, detailed high-resolution analyses have to be able to be produced frequently and quickly and this requires a thorough use of asynoptic and high-frequency observations such as aircraft data and remote sensing data. Since both 3-dimensional and 4-dimensional variational methods tend to be less appropriate for this purpose, a scheme based on the observation nudging tecnique has been chosen for data assimilation. COSMO model is used for a wide range of applications, which imposes a number of requirements for physical, numerical and technical design of the model. The main design requirements are (Schättler et al. 2016):

- Use of nonhydrostatic, compressible dynamical equations to avoid restrictions on the spatial scales and the domain size and application of an efficient numerical method of solution;
- Provision of a comprehensive physics package to cover adequately the spatial scales of application and provision of high-resolution data sets for all external parameters required by the parametrisation schemes;
- Flexible choice of initial and boundary conditions to accommodate both real data cases and idealized initial states and use of a mesh-refinement technique to focus on regions of interest and to handle multi-scale phenomena;
- Use of a high-resolution analysis method capable of assimilating high-frequency asynoptic data and remote sensing data;
- Use of pure Fortran constructs (i.e. Fortran90) to render the code portable among a variety of computer systems and application of the standard MPI-software for message passing on distributed memory machines to accommodate broad classes of parallel computers.

The development of the COSMO-Model was organized along these basic guidelines. However, not all of the requirements are fully implemented and development work and further improvement is an ongoing task.

3.2.2 The model equations

The model equations are formulated with respect to a rotated geographical lat/lon-grid with coordinates (λ, ϕ), where λ is the latitude and ϕ is the longitude. The rotated coordinate system results from these coordinates (λ_g, ϕ_g) by tilting the north pole. In the vertical, a generalized terrein-following height coordinate ζ is used, where any unique function of geometrical height can be used for transformation. Since ζ does not depend on time, the (λ, ϕ, ζ) system represents a non-deformable coordinate system, where surfaces of constant ζ are fixed in space -in contrast to the pressure based coordinate system of most hydrostatic models, where the surfaces of constant vertical coordinate move in space with changing surface pressure.

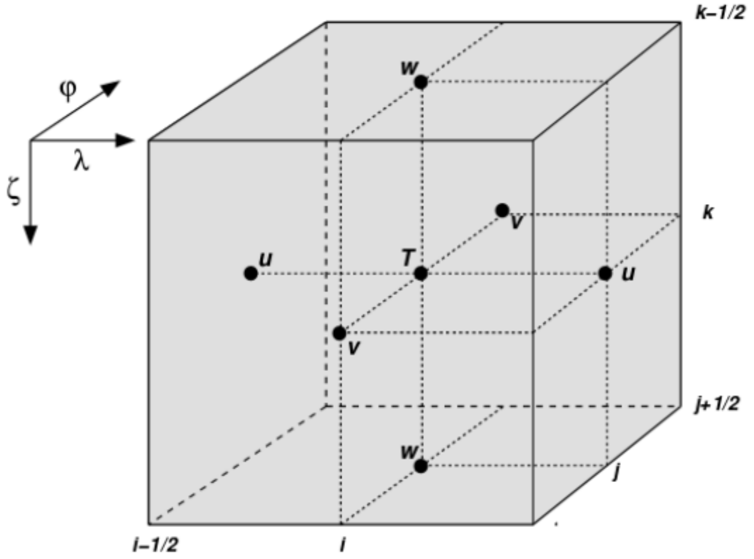


Figure 3.4: A grid box volume $\Delta V = \Delta\zeta\Delta\lambda\Delta\phi$ showing the Arakawa-C/Lorenz (Arakawa et al. 1977) staggering of the dependent model variables. ζ, λ and ϕ refer to the coordinate system.

The model variables are staggered on an Arakawa-C/Lorenz (Lorenz 1960; Arakawa et al. 1997) grid with scalars (temperatures, pressure and humidity variables) defined at the centre of a grid box and the normal velocity components defined on the corresponding box faces. For a given grid spacing, this staggering allows for a more accurate representation of differential operators than in the A-grid, where all variables are defined at the same point. The grid spacing is that relative to the different resolution at which the model is run, while in the vertical there are 40 layers from the surface up to about 24 km above ground. The set of prognostic model equations for the three components u, v and w of the wind vector, the perturbation pressure p', the temperature T and the humidity variables q is:

$$\frac{\partial u}{\partial t} + V \cdot \nabla u - \frac{uv}{a} \tan\phi - fv = -\frac{1}{\rho a \cos\phi}\left(\frac{\partial p'}{\partial \lambda} + \frac{J_\lambda}{\sqrt{G}} \frac{\partial p'}{\partial \zeta}\right) \tag{3.11}$$

$$\frac{\partial v}{\partial t} + V \cdot \nabla u - \frac{u^2}{a} \tan\phi + fu = -\frac{1}{\rho a}\left(\frac{\partial p'}{\partial \phi} + \frac{J_\phi}{\sqrt{G}} \frac{\partial p'}{\partial \zeta}\right) \tag{3.12}$$

$$\frac{\partial w}{\partial t} + V \cdot \nabla w = \frac{1}{\rho\sqrt{G}} + B + M_w \tag{3.13}$$

$$\frac{\partial p'}{\partial t} + V \cdot \nabla p' - g\rho_0 w = -(c_{pd}/c_{vd})\rho D \tag{3.14}$$

$$\frac{\partial T}{\partial t} + V \cdot \nabla T = -\frac{p}{\rho c_{vd}} D + Q_T \tag{3.15}$$

$$\frac{\partial q^v}{\partial t} + V \cdot \nabla q^v = -(S^l + S^f) + M_{q^v} \tag{3.16}$$

$$\frac{\partial q^{l,f}}{\partial t} + V \cdot \nabla q^{l,f} + \frac{1}{\rho\sqrt{G}} \frac{\partial P_{l,f}}{\partial \zeta} = S^{l,f} + M_{q^{l,f}} \tag{3.17}$$

In the previous section, it was said that ECMWF ENS is a hydrostatic ensemble, i.e. where the hydrostatic approximation is valid and the vertical velocity is a diagnostic variable not contained into the momentum equations. Instead, in COSMO-based ensembles, which are called non-hydrostatic, the hydrostatic approximation is no longer valid and therefore also the vertical velocity requires a prognostic equation, becoming a variable of the system. In the set of COSMO prognostic equations, the continuity equation has been replaced by an equation for p'. In (3.11) and (3.12) a is the radius of the earth, c_{pd} and c_{vd} are the specific heat of dry air at constant pressure and constant volume, g is the gravity acceleration, f is the Coriolis parameter, R_v and R_d are the gas constants for water vapour and dry air. Furthermore $J_\lambda = (\frac{\partial z}{\partial \lambda})_{zeta}$, $J_\phi = (\frac{\partial z}{\partial \phi})_{zeta}$ are the elements of the Jacobian matrix linked to the transformation from the zeta coordinate to the ζ. $\sqrt{G} = |\det(J^z)||\frac{\partial z}{\partial \zeta}|$ the Jacobian of the transformation from the z-to the ζ-system. ρ is the density of moist air which is calculated as a diagnostic variable from the equation of state:

$$\rho = p[R_d(1 + (R_v/R_d - 1)q^v - q^l - q^f)T]^{-1}$$

q_v is the specific humidity, q^l represents the specific water content of a category of liquid water (cloud or rain water) and q^f represents the specific water content of a category of frozen water (cloud ice, snow or graupel). The corresponding precipitation fluxes are denotated by P_l and P_f. The terms M denote contributions from subgrid-scale processes as, e.g. turbolence and convection and Q_T summarizes the diabatic heating rate due to this processes. The term B in the equation for the vertical velocity is the buoyant acceleration. The equations from (3.11) to (3.17) are solved numerically in the model using the traditional finite difference method. In this technique, spatial differential operators are simply replaced by suitable finite difference operators. The time integration is also by discrete stepping using a fixed timestep Δt depending on the horizontal resolution, in order to satisfy the CFL stability condition (that is $\Delta t \leq \Delta x/c$, where Δx is the horizontal grid spacing and c the magnitude of velocity). More details on COSMO model features can be found in the COSMO User Guide (www.cosmo-model.org).

3.2.3 The COSMO-LEPS ensemble system

As far as operational implementations are concerned, the COnsortium for Small-Scale MOdelling Limited-area Ensemble Prediction System (COSMO-LEPS) was the first mesoscale ensemble application running on a daily basis in Europe. This system, initially developed and implemented by the HydroMeteoClimate Service of Emilia-Romagna, in Bologna, Italy (ARPA-SIMC), has been running at ECMWF since November 2002 (Montani et al. 2003a). Nowadays, COSMO-LEPS is based on 20 integrations of the non-hydrostatic mesoscale model COSMO, formerly known as the Lokal Modell (Steppeler et al. 2003). The methodology (described more thoroughly in the next section) aims at combining the advantages of the probabilistic approach by global ensemble system with the high-resolution details gained in the mesoscale integrations. In the construction of COSMO-LEPS, an algorithm selects a number of members (referred to as Representative Members, RMs) from a "driving" global ensemble system (Marsigli et al. 2001). This intermediate step, referred to as "ensemble-size reduction", is required to keep the computational load operationally affordable, since it is not presently feasible to nest the limited-area model on each individual member of a global ensemble with size larger than 30 members. After the "ensemble-size reduction", the selected RMs are used to provide both initial boundary conditions to the integrations with the COSMO model, which is run once for each RM. The impact of the large ensemble-size reduction on the forecast accuracy has been studied for some case studies and it can be concluded that the accuracy of the probabilistic forecast is not noticeably improved by the increase of LEPS size. Since a large amount of computer time has spent to perform many limited-area integrations, the advantages of the clustering-selection methodology are well evident (Montani et al. 2003). Therefore, COSMO-LEPS performs a sort of dynamical downscaling of a global-model probabilistic system, limiting to a certain extent the computational cost (Tibaldi et al. 2006). Montani et al. (2011) reports on the improved skill of COSMO-LEPS throughout the years. Initial conditions are taken from the driving EPS members and interpolated on COSMO grid. Despite the reductions in number of the ensemble members, the procedure described above allows to account for the most of the variety of the scenarios represented in ECMWF EPS members, giving informations on uncertainties in initial conditions. Perturbations entering the model from the lateral boundaries are still provided by the driving EPS members and play a more and more important role in the behaviour of the limited-area system as the forecast range increases. There are due to the SPPT scheme (Stochastic Perturbations of Physical Tendencies) performed at ECMWF (Buizza et al. 1999). In general model physics scheme apply adjustments (called "tendencies") to the variables temperature, humidity and wind that are used in the equations describing atmospheric circulation. ECMWF's Integrated Forecasting System (IFS) represents uncertainty in the model physics by perturbing these physics tendencies, introducing different perturbations for each ensemble member. The tendencies are perturbed randomly within certain limits. There is a system to the randomness because the collection of ensemble members has to describe a realistic distribution of possible forecast outcomes. This is achieved by using time- and spatially-varying patterns of random numbers to provide the perturbations. Therefore the random error in parameterised forcings is assumed to be coherent between the different parameterisation modules, having a certain coherence on the space and time scales associated. Moreover, the scheme assumes that the larger the parameterised tendencies are, the larger the random error component is.

3.2.4 COSMO-2I-EPS

The higher the numerical resolution, the more accurate the calculations should become. A high spatial resolution also enables a better representation of topographical fields, such as mountains and coastlines, and the effect they have on the large-scale flow. It also produces a more accurate description of horizontal and vertical structures, which facilitates the assimilation of observations. The smallest atmospheric features which can be resolved by high-resolution forecast, have wave lengths four or five times the numerical resolution. Although these atmospheric systems have a predictability of only some hours, which is about the time it takes to disseminate the forecast, their representation is nevertheless important for energetic exchanges between different atmospheric scales. Increasing the resolution not only benefits the analyses and forecasts of the small-scale systems associated with severe weather, but also those of large-scale systems. The ability to forecast accurately the formation of large-scale blocking "omega" anticyclones and "cut-off lows" depends crucially on increasing the resolution to kilometres (Miller et al. 2010). For the above-mentioned reasons, a new high-resolution ensemble prediction system was implemented in Italy. The system, referred to as COSMO-2I-EPS, covers the domain of (Fig.3.5) with a 2.2 km horizontal resolution and 65 vertical levels.

The predictability is at convection-permitting scale and targeted products, like probability maps with thresholds, are being developed and implemented, in particular for thunderstorms and convective precipitation events. This ensemble system had 10 members in 2016, now implemented at 20 mem-

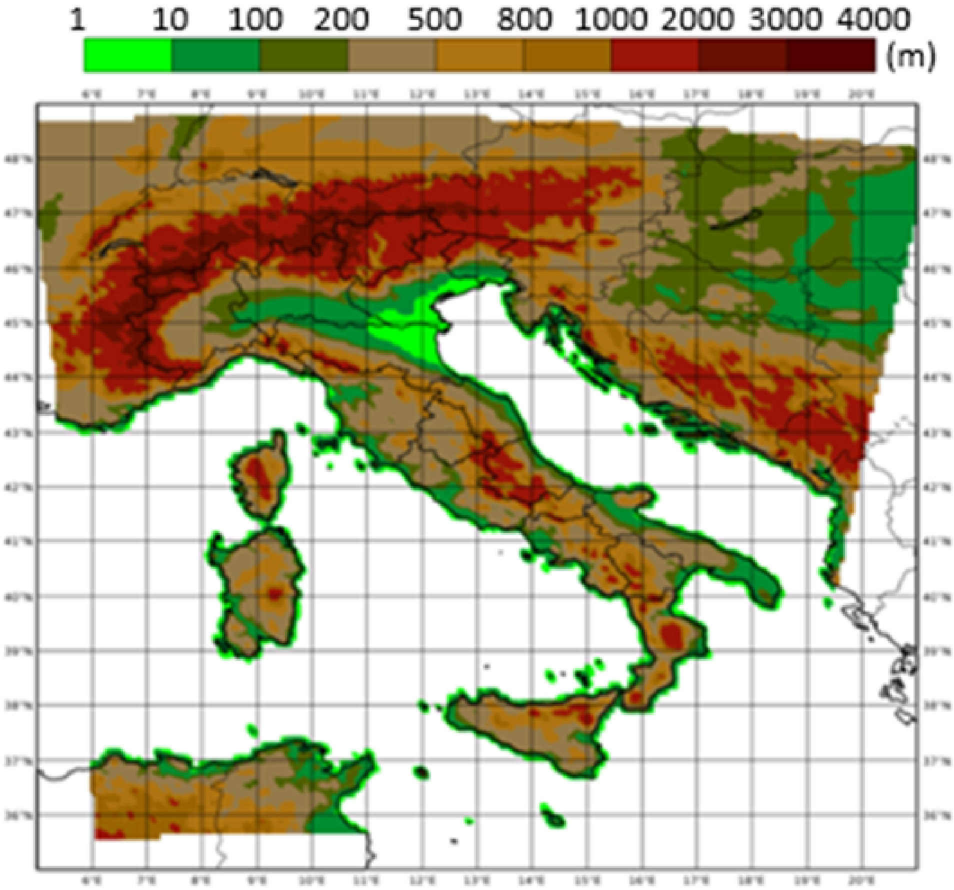


Figure 3.5: COSMO-2I-EPS integration domain.

bers that receive boundary conditions from the first 20 members of COMET ensemble system (COSMO-ME-EPS). The initial conditions, including the soil moisture and sea surface temperature perturbations, are taken from the 20 perturbed analyses of KENDA ensemble. KENDA is the data assimilation system developed by COSMO during last years. COSMO-2I-EPS runs once a day starting at 00 UTC and its forecast range covers 48 hours. COSMO-2I-EPS is not operational yet and it is only in a pre-operational fase. In the near future, the operational phase will start with an upgrade of model parameters perturbation methodology; an increasing number of parameters to perturb with their random combination will be evaluated. In addition, an operational verification of the ensemble will be implemented on the whole national territory.

3.3 Representation of orography

The different horizontal resolutions of the three ensemble systems presented in this chapter play a crucial role in the representation of the orography. It is clear that models with a higher horizontal resolution can capture smaller details of the territory; on the other hand, models with a lower resolution are able to solve only the broad features of a geographical area. In Fig 3.6 the orographies of ECMWF ENS, COSMO-LEPS and COSMO-2I-EPS (with horizontal resolutions of 18, 7 and 2.2 km respectively) are compared over part of Emilia-Romagna region.

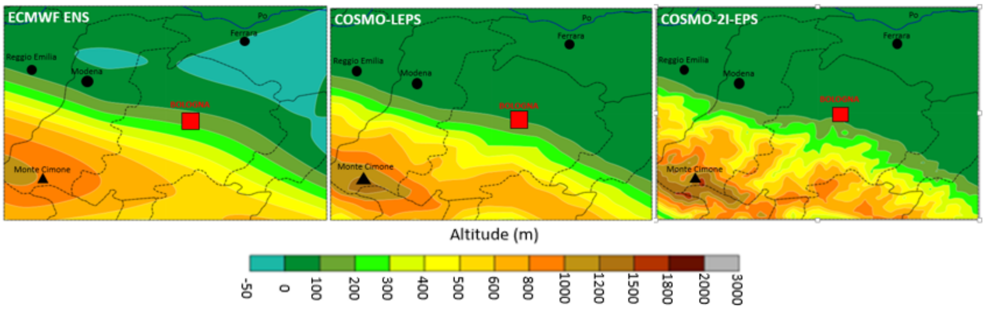


Figure 3.6: Representation of the orography (in metre) over part of Emilia-Romagna according to ECMWF ENS (left panel), COSMO-LEPS (middle panel) and COSMO-2I-EPS (right panel).

It can be noticed how the different horizontal resolutions affect the representation of the orography. In ECMWF ENS (left panel) the orography of the Apennines appears very simplified: the valleys and the highest mountain peaks (like Monte Cimone, 2138 m) are not clearly distinguishable. Furthermore, it can be observed that the maximum altitude of this part of the Apennines does not exceed 1200m, Bologna is already on the hills and some areas of the Po Valley are below the sea level. In COSMO-LEPS (middle panel), the representation of the orography is still rather approximate. However the Secchia, Panaro and Reno river valleys are just outlined and also Monte Cimone is recognizable, even if the altitude does not exceed 1500m.

On the other hand, COSMO-2I-EPS provides the best representation of the orography; in the right panel of Fig.3.6, the main valleys are described with high spatial detail and the minor ones are also visible. Monte Cimone is clearly identifiable with an altitude close to reality.

The representation of the orography described above has effects on the prediction of the atmospheric variables. COSMO-based ensemble systems, in particular COSMO-2I-EPS, are able to provide a better representation of the interation of the flow with orography and of mesoscale-related processes, that ECMWF ENS cannot solve.

Chapter 4

Description of the experiments

4.1 Synoptic description of the events

The short period of June 2016, considered for the verification of the three ensemble systems, was characterized by a strong atmospheric instability with numerous thunderstorms and convective precipitation events. As for the details of the weather patterns, the situation can be summarised by considering only four days (20, 23, 26, 27) which are particularly suited to represent the meteorological situation over Europe and Northern Atlantic Ocean, between 20^{th} and 29^{th} June 2016. In particular, the synoptic situation on 20^{th} June 2016 at 00 UTC is shown in Fig.4.1. This map is relative to the analysis of ERA-Interim (ECMWF): here the different colours refers to different values of 500 hPa geopotential height[1], as reported in the legend below the figure; white isolines link locations with the same values of mean sea level pressure.

The situation was characterized by a trough placed over the european central meridians, elongated from Denmark and Germany to Italy; on the other hand, an upper level ridge was located over Eastern Mediterranean Sea, Greece and Turkey, up to Ukraine and the most south-western part of Russia. The Iberian Peninsula and France were under the influence of the Azores Anticyclone stretching over western Europe. A large-scale cyclonic area was located just south of Iceland. The 300 hPa jet stream is shown in

---

[1] The geopotential Φ at a certain p-level (isobaric coordinates) is defined as $\Phi = \int_0^p g dp$; the geopotential height is computed as $GPH = \Phi / g_0$ considering g_0 (gravity at ground level) constant along the vertical direction.

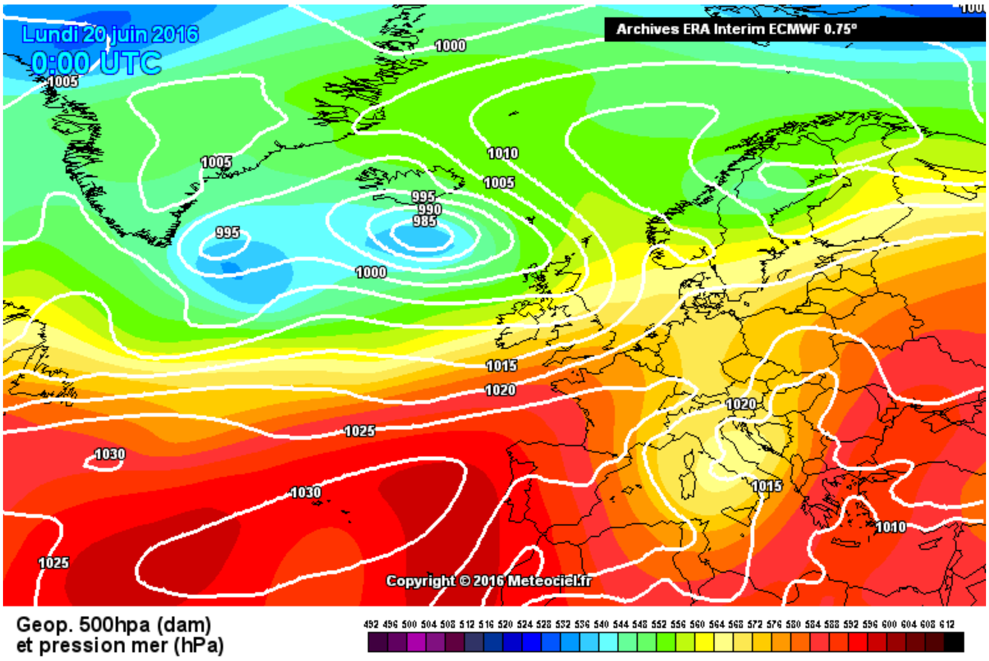


Figure 4.1: Reanalysis from ERA-Interim (ECMWF) valid at 00 UTC of 20th June: colours discriminate different value of 500 hPa height (in dam); solid white line link point with same MSLP (interpolated by Meteociel (www.meteociel.fr)).

(Fig.4.2): the highest speeds were located near the British Isles, while the anticlockwise rotation over Italy was due to the through.

In (Fig.4.3) it's possible to notice the distibution of air masses over Europe. Tipically it's a habit to consider temperature at 850 hPa isobaric height (approximately 1500 m) to evaluate the type of air mass. Infact, it's at this heigh, at the top of boundary layer, that effects of diurnal cycle can be neglected and that is suitable to establish the cold or warm advection in the low atmospheric layers. It is evident the presence of fresh air associated with the trough over Central Europe and the area of Central Mediterranean: this type of air mass originates from the Northern Atlantic Ocean. The descent of fresher air down to Italy, the surrounding Seas and the northernmost part of Algeria and Tunisia, left room for hot continental subtropical air from the Sahara desert up to the Eastern Mediterranean, the Aegean Sea, the Black Sea and all Countries facing these Seas; it was just this saharian air

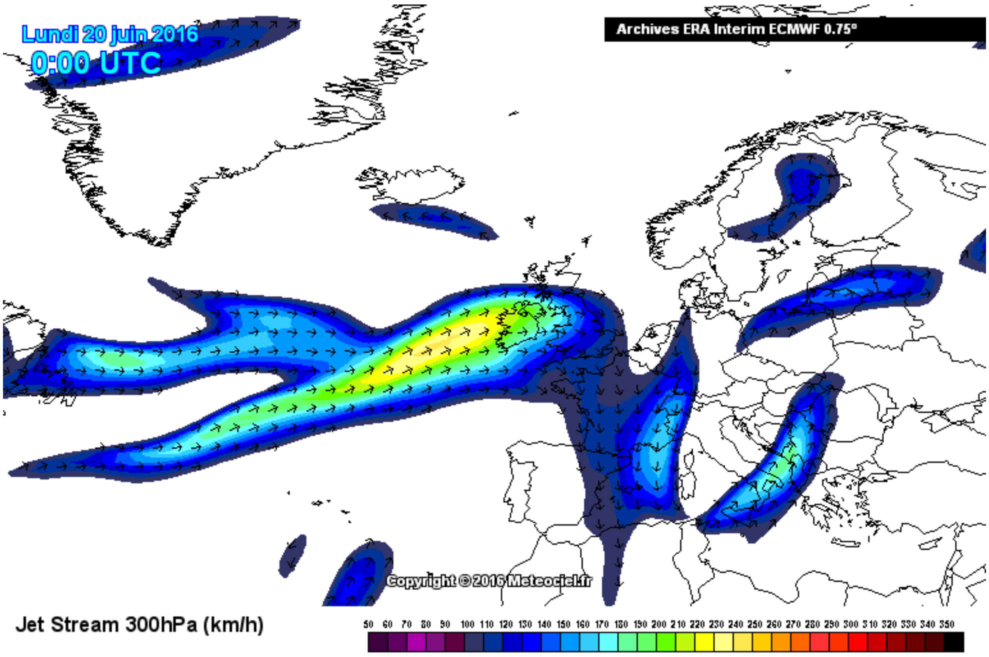


Figure 4.2: Reanalysis from ERA-Interim (ECMWF) valid at 00 UTC of 20^{th} June: colours discriminate different velocity of jet stream at 300 hPa height (in km/h), arrows show wind direction.

mass that supported the subtropical ridge over these areas. Insted, the joint action between the cyclone of Iceland and the high pressure of Azores has carried mild oceanic mid-latitude air over the British Islands and the western costlines of the continent.

Fig.4.4 shows the synoptic situation at 00 UTC of 20^{th} June. The chart allows to detect the main mesoscale features, including fronts. Grey isolines refer to points with the same value of MSLP (Mean Sea Level Pressure). At first glance, it's particulary evident the deep cyclone (982 hPa) located just south the Iceland, with the frontal bands over the British Islands. More to the south, the Anticyclone of Azores can be identified with a maximum pressure of 1033 hPa right on the archipelago and a secondary maximum in correspondence of Biscay Bay. Instead, linked to the trough over Central Europe, there was a minimum of pressure over middle Adriatic Sea. The cooler air, pushed on the sea by this minimum of pressure, had developed a cold front.

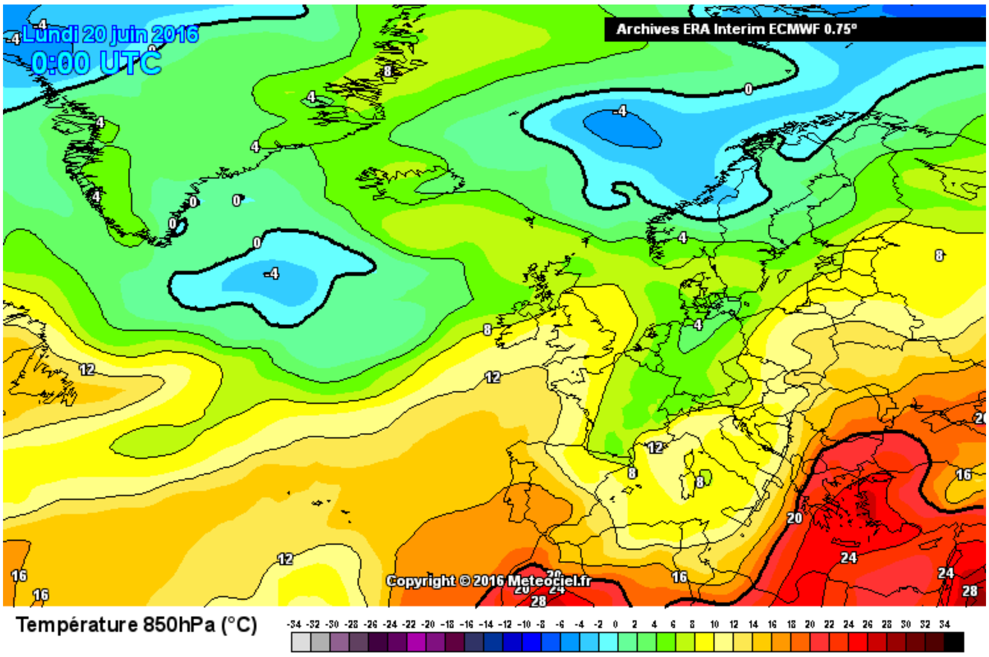


Figure 4.3: Reanalysis from ERA-Interim (ECMWF) valid at 00 UTC of 20^{th} June: colours discriminate different air temperature at 850 hPa height (in ^{o}C).

Infact, looking at the image from the satellite (Fig.4.5), it can be noticed that there was some storm cells offshore the coasts of Central Adriatic, caused by the passage of the cold front over the sea. The arch of cloud, curled around the pressure low located south of Iceland and extending southeasterly over Great Britain and Ireland, was produced by some frontal systems that were moving toward the northwestern part of the continent.

Three days later, on 23^{rd} June 2016, the trough over Central Europe evolved into a cut-off low centered in correspondence of the Sicilian Channel (as in Fig.4.6); to this geopotential minimum have been associated two lines of instability over southern-central Italy, responsible for the growth of many thunderstorms. Over the Countries of central-eastern Europe the MSLP had been increasing, thanks to a field of high pressure centered, with a maximum of 1027 hPa, over eastern Poland. Despite the Cyclone of Iceland had dissolved, a new baroclinic wave within westerlies was approaching to western Europe; infact new fronts had already reached the British Islands.

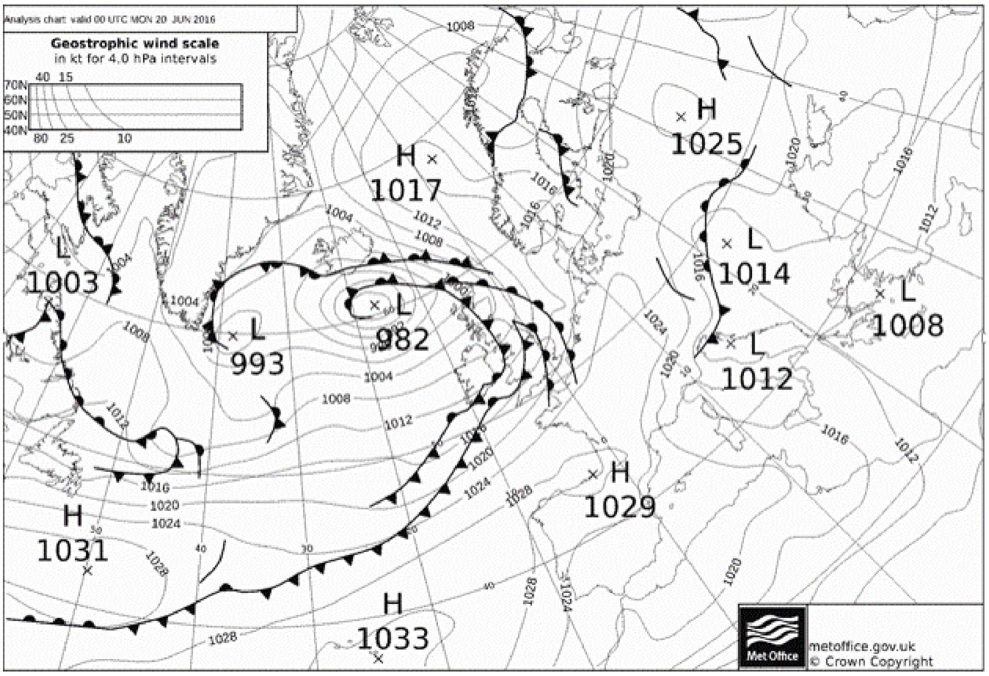


Figure 4.4: Synoptic chart valid at 00 UTC of 20^{th} June 2016, by UK Met Office.

Three days later, on 26^{th} June 2016, the cut-off low had crossed the Ionian Sea getting over Greece, with associated a small line of instability connected to a weak cyclogenesis over the Aegean Sea. Weather became worse again over central Europe, because of the new Atlantic through. The latter moved northestward with a bent axis from the North Sea to the Iberian Peninsula, through Benelux and France. So at ground it led to the development of three centers of low pressure: one over Denmark (1006 hPa), another one over the Balcanic Peninsula (1011 hPa) and the last one over the Iberian Peninsula (1011 hPa). Therefore, over a large portion of southern central Europe, there were the ideal conditions to trigger many thunderstorms, especially during the afternoon. Meanwhile, the strengthening of the jet stream coming out of Canada, was pushing again the Azores Anticyclone towards Europe (as in Fig.4.7).

On 27^{th} June 2016, the Atlantic trough had moved further eastwards, reaching the Balkans and incorporating the cut-off low over Greece. Moving

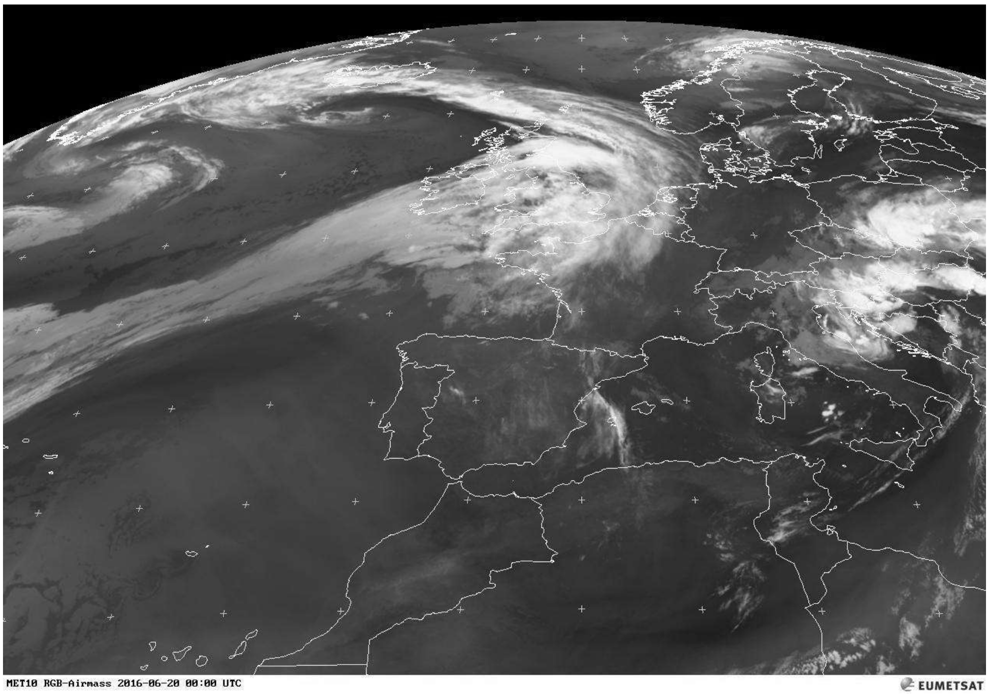
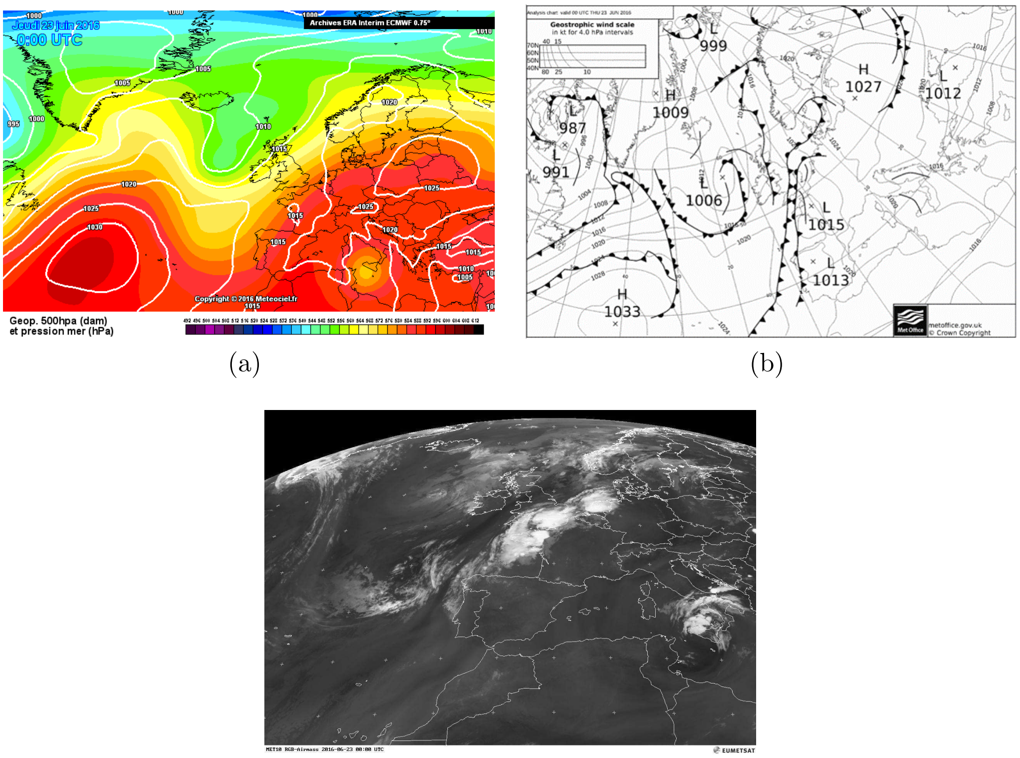


Figure 4.5: Satellite image of 20^{th} June 2016 00 UTC from EUMETSAT (European Meteorological Satellites) in the infrared channel (MET10 RGB-Airmass).

eastward the trough has only lapped Italy, where anyway it has allowed the survival of marked instability conditions. In fact, storms developed during the night on north-eastern Italy, in the late afternoon between Umbria and Marches. However the Azores Anticyclone was again extending over the western Mediterranean, where the atmosphere had became more stable, with the exception of the Alboran Sea and the surrounding areas. In fact, over the southern part of Iberian Peninsula was located a pressure low of 1016 hPa, with an instability line. Over central-northern Europe westerlies were accompanied by a series of perturbed systems (as in Fig.4.8).

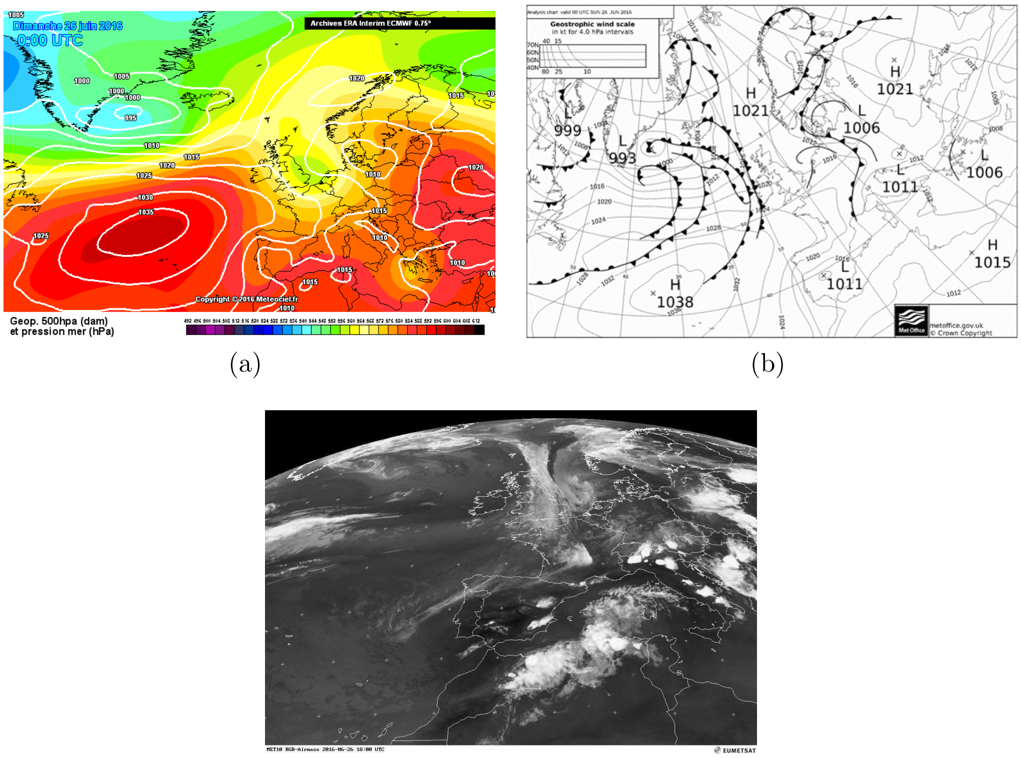
Overall, the period under investigation was characterized over Italy by marked instability conditions, with the trigger of numerous thunderstorms, especially during the central hours of the day. This could be verified due to the presence of an especially active Atlantic flow, with a rapid alternation of depressionar waves, full of fresh and unstable air. In the image reported



(a) (b)

(c)

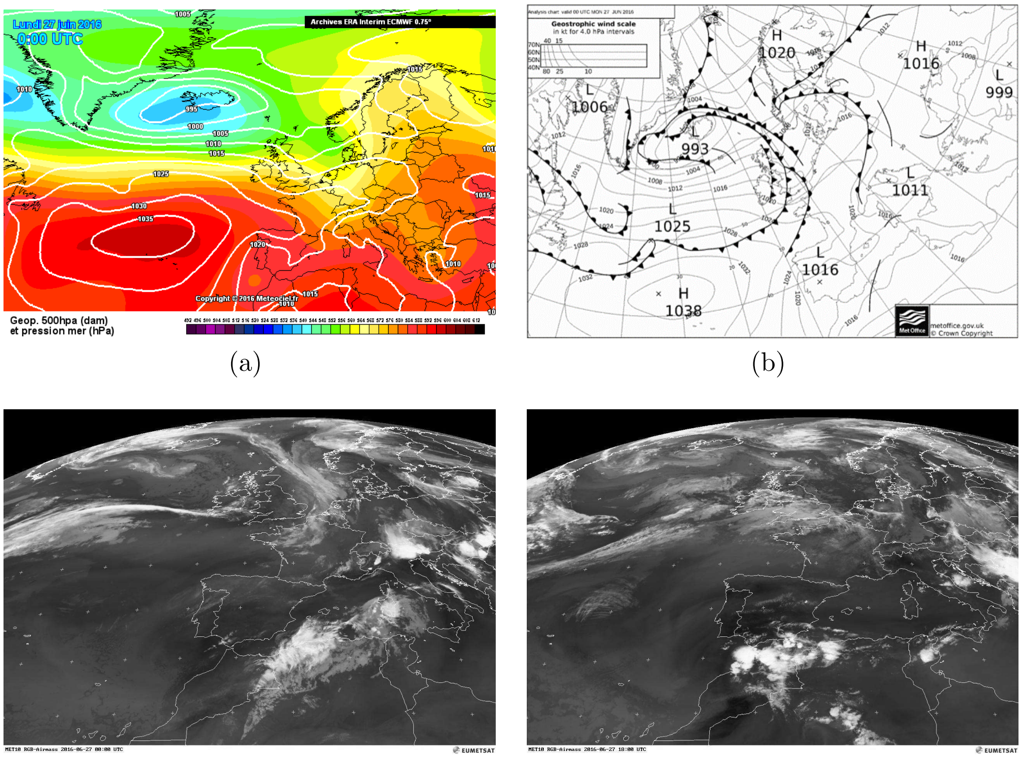
Figure 4.6: (a) Reanalysis from ERA-Interim (ECMWF) valid at 00 UTC of 23th June: colours discriminate different value of 500 hPa height (in dam); solid white line link point with same MSLP (interpolated by Meteociel (www.meteociel.fr)). (b) Synoptic charts valid at 00 UTC of 23th June 2016, by UK Met Office. (c) Satellite image of 23th June 2016 00 UTC from EUMETSAT in the infrared channel (MET10 RGB-Airmas).



(a) (b)

(c)

Figure 4.7: (a) Reanalysis from ERA-Interim (ECMWF) valid at 00 UTC of 26th June: colours discriminate different value of 500 hPa height (in dam); solid white line link point with same MSLP (interpolated by Meteociel (www.meteociel.fr)). (b) Synoptic charts valid at 00 UTC of 26th June 2016, by UK Met Office. (c) Satellite image of 26th June 2016 18 UTC from EUMETSAT in the infrared channel (MET10 RGB-Airmas).



(a) (b)

(c) (d)

Figure 4.8: (a) Reanalysis from ERA-Interim (ECMWF) valid at 00 UTC of 27^{th} June: colours discriminate different value of 500 hPa height (in dam); solid white line link point with same MSLP (interpolated by Meteociel (www.meteociel.fr)). (b) Synoptic charts valid at 00 UTC of 27^{th} June 2016, by UK Met Office. (c) Satellite image of 27^{th} June 2016 00 UTC from EUMETSAT in the infrared channel (MET10 RGB-Airmas). (d) Satellite image of 27^{th} June 2016 18 UTC from EUMETSAT in the infrared channel (MET10 RGB-Airmas).

in Fig.4.9 it is possible to see the total precipitation recorded by the pluviometric network of the National Civil Protection Department (DPCN) over the whole period considered. Rainfall affected almost the entire national territory, with the exception of Liguria, Tuscany and Apulia, where the phenomena were more sporadic. Three main rainy can be detected in Fig.4.9: the northern Piedmont and Lombardy, the Friuli Venezia Giulia and the Marches. On these areas, several rain gauges recordered value beetween 50 mm and 100 mm, with some isolated peaks over 100 mm on the northern Piedmont and Lombardy.

4.2 Methodology of evaluation

The evaluation of the performance of the model consists in the comparison of gridded model output against point observations. A number of statistical scores evaluate different aspects of model performance while the forecast "error" is simply defined as the difference between the forecast value and the observation. In a "standard" deterministic approach, the uncertainty associated with the forecast value is not estimated. An EPS aims at quantifing this uncertainty using a set of perturbed Initial Conditions (ICs) and/or perturbed model formulations. Verification methods applied to ensemble forecasts have two main objectives:

- to assess the characteristics of the ensemble distribution;
- to verify the probability forecast.

Since all perturbed ICs should be equally possible be true and all perturbed physics or varying physics schemes or alternative models be equally plausible, the performance of any ensemble member should, in principle, be equivalent to that of another member on average. If this is not the case, that is indicative of problems with the choice of ensembling the technique employed. For example, either the IC perturbations are too large or alternative models, physics schemes or perturbations are not equally plausible. In the next chapter two evaluation methods will be used:

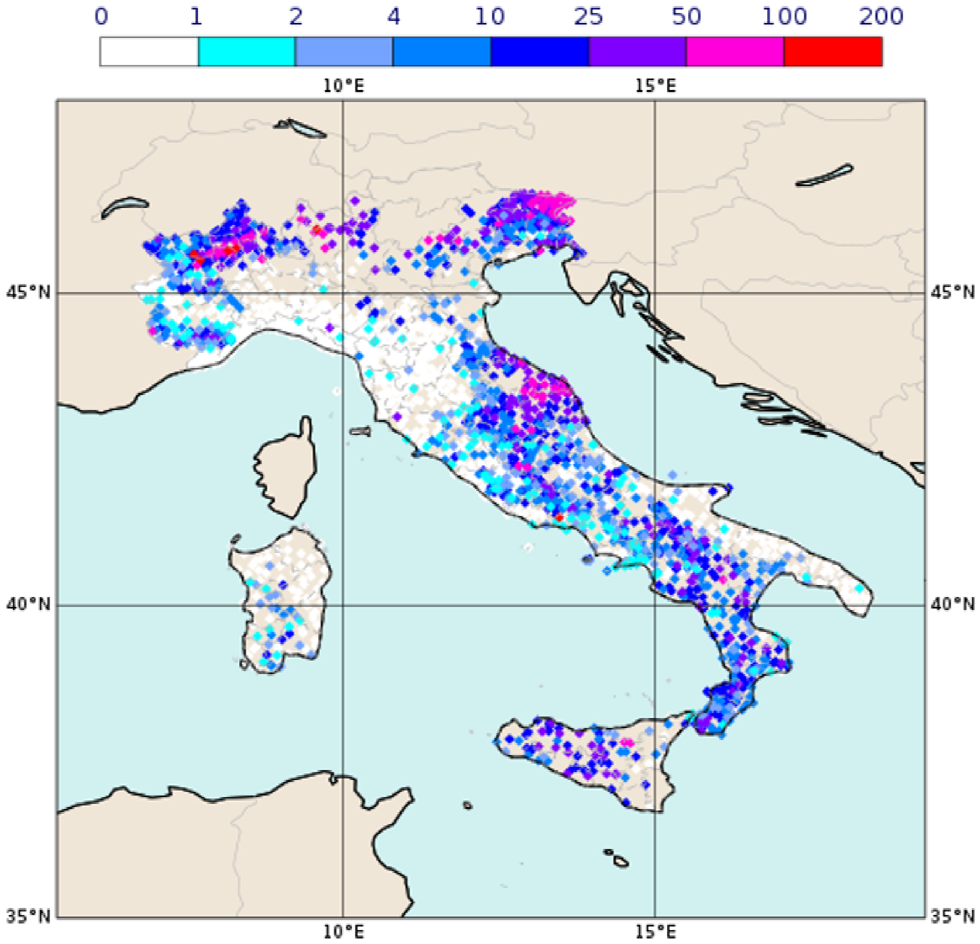


Figure 4.9: Total precipitation (mm) collected from rain gauges of DPCN (Dipartimento di Protezione Civile Nazionale) network, from 20th June 2016 at 00 UTC to 29th June 2016 at 00 UTC. The lack of some stations (e.g in Trentino Alto Adige) is due to the partial unavailability of data during the investigation period.

- **the nearest grid point**: since observations seldom occur at the precise locations represented by the grid points of one particular model, it is necessary to compare the forecast values in the grid points with those of the nearest observations (ECMWF Forecast User Guide);
- **bilinear interpolation**: it is an extension of linear interpolation for interpolating functions of two variables (e.g., x and y) on a rectilinear 2D grid. The key idea is to perform linear interpolation first in one

direction, and then again in the other direction. Although each step is linear in the sampled values and in the position, the interpolation as a whole is not linear but rather quadratic in the sample location (ECMWF Forecast User Guide). The bilinear interpolation technique gives a weighting to the effect of an observation at the grid vertices for analysis purposes and interpolates a value to a sub-grid location from values given at grid vertices for output purposes (Fig. 4.10).

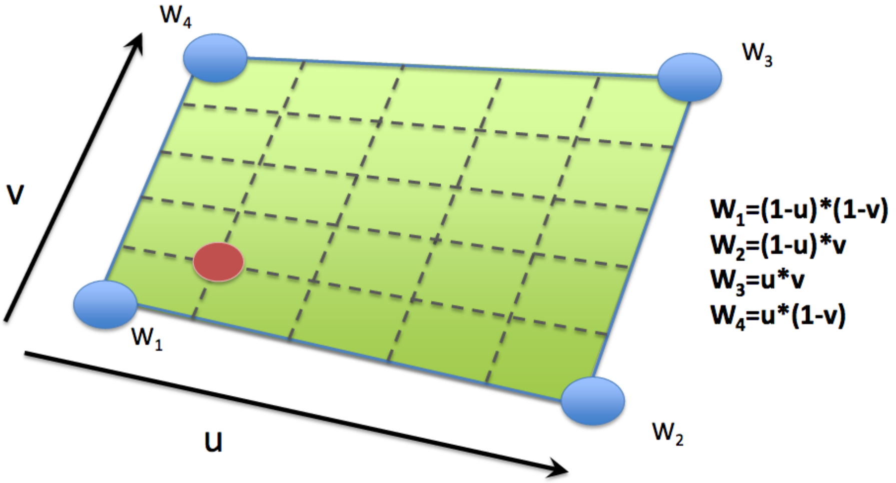


Figure 4.10: The interpolation uses the four corner points closest to the selected location and takes a weighted average to arrive at the interpolated value where u and v are non-dimensional weighting factors that vary between 0 and 1 across the blue grid (from ECMWF Forecast User Guide).

4.2.1 Deterministic scores

A number of statistical scores have been developed and are applied in order to evaluate the usefulness of an EPS forecast system; in this subsection the Mean Error and the Root Mean Square Error, that will be used in the experimental verification of the three ensemble systems, are introduced (Gofa et al., 2010).

The mean error (ME) measures the average difference between a set of forecasts and corresponding observations. It measures the average difference beetween the ensemble mean forecast and observations. The ME of the ensemble mean forecast $\overline{Y}$ given the observation, x, is given by:

$$ME = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{Y_i})$$

The Mean Square Error (MSE) measures the average square error of the forecasts. The Root Mean Square Error (RMSE) provides the square root of this value, which has the same units as the forecasts and observations. Here, the forecast corresponds to the ensemble mean value and an 'error' represents the difference between the ensemble mean $\overline{Y}$ and the observation x. The equation for the RMSE is:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{Y_i})^2}$$

RMSE of the ensemble mean measure the distance beetween forecasts and analyses (or observations). The ensemble spread (SPRD) is calculated by measuring the deviation of ensemble forecasts from their mean (Zhu, 2005). Usually, SPRD is defined as:

$$SPRD = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}(\overline{f} - f(n))^2}$$

Where $\overline{f} = \frac{1}{N}\sum_{n=1}^{N} f(n)$ is for the ensemble mean and f is for the ensemble forecast. In general, an ideal ensemble forecast will be expected to have the same size of ensemble spread as their RMSE at the same lead time in order to represent full forecast uncertainty (Zhu, 2005, Buizza et al., 2005); but most of the ensemble systems are underdispersed (lower spread) for longer lead times due to an imperfect model system (or physical parameterizations) and other factors. Anyway over a large number of ensemble forecasts, the statistical properties of the true value X_{TRUE} of any quantity X are identical to the statistical properties of a member X_j of the ensemble; in particular:

$$\overbrace{\overline{|X_j - X_{MEAN}|^2}}^{ensemblevariance} = \overbrace{\overline{|X_{TRUE} - X_{MEAN}|^2}}^{meansquarederror}$$

where X_{MEAN} is the ensemble mean. The time-mean ensemble spread around the mean equals the time-mean RMSE of the ensemble mean (Gofa et al., 2010).

4.2.2 Probabilistic Scores

In this chapter, the performance of the three ensemble systems has been evaluated by computing the following “traditional” probabilistic scores: the Brier Skill Score (BSS), the Ranked Probability Score (RPS), the debiased Ranked Probability Skill Score ($RPSS_D$), the Relative Operating Characteristic Curve (ROC) Area and the percentage of OUTLiers (OUTL).

The Brier Score (BS) measures the average square error of a probability forecast (Brier, 1950). It is analogous to the mean square error of a deterministic forecast, but the forecasts, and hence error units, are given in probabilities. The Brier Score measures the error with which a discrete event, such as “flooding” is predicted. It is given from:

$$BS = \frac{1}{N}\sum_{i=1}^{N}(f_i - o_i)^2$$

- N is the number of points in the “domain” (spatio-temporal)
- $o_i = 1$ if the event occurs, $0_i = 0$ if the event does not occur
- f_i is the probability of occurence according to the forecast system (e.g. the fraction of ensemble members forecasting the event)

it is sensitive to climatological frequency of the event (Wilks, 1995). In absence of any forecasting skill, the best strategy to optimise the Brier Score is to forecast the climatological frequency. The rarest an event, the easier is getting a good BS without having any real skill. For this reason, the Brier Score (see below) is preferred as long as it references the score to climatology (sample or long-term). The perfect score is 0 and that is possible for perfect deterministic forecast (Stanski et al., 1989; Wilks, 1995). The Brier Score can be decomposed into several components that are relevant for interpretation of the sources of errors in the forecasts (Murphy 1973) which are useful to explore the dependence of probability forecasts on ensemble characteristics:

$$BS = \overbrace{\frac{1}{N}\sum_{k=0}^{M} N_k(f_k - \overline{o_k})^2}^{reliability} - \overbrace{\frac{1}{N}\sum_{k=0}^{M} N_k(\overline{o_k} - \overline{o})^2}^{resolution} + \overbrace{\overline{o}(1-\overline{o})}^{uncertainty}$$

$\overline{o}$ is the total frequency of the event (sample climatology). For this decomposition, it is assumed that there is a discrete number of forecast possibilities, M, and the forecasts and observations have been sorted by the forecast value. Each of the terms in can be interpreted in the context of attributes of forcast quality. The first term is a *reliability* measure: for forecasts that are perfectly reliable, the sub-sample relative frequency is exactly equal to the forecast probability in each sub-sample. It measures the difference between the forecast and the mean observation associated with that forecast value, over all of the forecast. The second term is a *resolution* measure: if the forecast sorts the observations into sub-samples having substantially different relative frequencies from the overall sample climatology, the resolution term will be large. This is a desirable situation, since the resolution term is subtracted. It is large if there is resolution enough to produce very high and very low probability forecasts. The *uncertainty term* ranges from 0 to 0.25. If the event was either so common, or so rare, that it either always occurs or never occurs, then $b_{unc} = 0$. When the climatological probability is near 0.5, there is more uncertainty inherent in the forecasting situation ($b_{unc} = 0.25$).

The Brier Skill Score (BSS) measures the performance of one forecasting system relative to another in terms of the Brier Score (BS). The BS measures the average square error of a probability forecast of a dichotomous event. The BSS comprises a ratio of the BS for the forecasting system to be evaluated (the “main forecasting systems”), over the BS for the reference forecasting system BS_{REF}. Commonly, the reference forecast is the sample climatology.

$$BSS = 1 - \frac{BS}{BS_{cli}}$$

$$BS_{cli} = \overline{o}(1-\overline{o})$$

As a measure of average square error in probability, values for the BS approaching zero are preferred. It follows that a BSS closer to 1 is preferred, as this indicates a low BS of the main forecasting system relative to the BS of the reference forecasting system. This score should always be applied to a sufficiently large sample, one for which the sample climatology of the event is representative of the long term climatology. The rarer the event, the larger the number of samples needed to stabilise the score. For better results the Brier Skill Score should be computed on the whole sample, i.e., the skill should be computed for an aggregated sample, not averaged for several samples.

A widely used measure to evaluate probability forecasts of multiple categories is **the Ranked Probability Score (RPS)** (Epstein 1969; Murphy 1969,1971). This measure is analogous to the BS and has the form:

$$RPS = \frac{1}{J-1}\sum_{m=1}^{J}[(\sum_{j=1}^{m} f_i) - (\sum_{j=1}^{m} o_j)]^2$$

where

- J is the number of forecast categories
- $o_j = 1$ if the event occurs in category j, $o_j = 0$ if the event does not occur in category j
- f_j is the probability of occurrence in category j

This score is used to assess multi-category forecast, where J is the number of forecast categories (for example, rainfall bins). The RPS penalizes forecasts less severely when their probabilities are close to the true outcome and more severely when their probabilities are further from the actual outcome. For two forecast categories the RPS coincides with the Brier Score.

The **Ranked Probability Skill Score (RPSS)** measures the improvement of the multi-category probabilistic forecast with respect to a reference forecast (usually the long-term or sample climatology). It is similar to the 2-category Brier Skill Score, as for it takes climatological frequency into account. A positive value of RPSS indicates forecast benefit with respect to the climatological forecast (Wilks, 1995). Because the denominator approaches 0 for a perfect forecast, this score can be unstable when applied to small data sets. This score should always be applied to a sufficiently large sample, one for which the sample climatology of the event is representative of the long term climatology. The rarer the event, the larger the number of samples needed to stabilise the score. For better results the ranked probability skill score should be computed on the whole sample, i.e. the skill should be computed for an aggregated sample, not averaged for several samples.

$$RPSS = \frac{\overline{RPS} - \overline{RPS}_{reference}}{0 - \overline{RPS}_{reference}} = 1 - \frac{\overline{RPS}}{\overline{RPS}_{reference}}$$

where the overbar denotes the average of the scores over a given number of forecast/observation pairs.

As mentioned above, for small ensemble sizes the RPSS is negatevely biased. Müller et al. (2005) have shown that this bias can be removed if, in equation for RPSS, the climatological reference score $RPS_{reference}$ is replaced by $E[RPS_{ran}]$, which is the expectation of the scores RPS_{ran} that the ensemble prediction system under consideration would produce in the case of merely random resamples from climatology. While Müller et al. (2005) applied a Monte Carlo approach to obtain an estimate of $E[RPS_{ran}]$, Weigel et al. (2007) have derived an analytical expression for a **debiased discrete Ranked Probability Skill Score ($RPSS_D$)**. It is applicable to probabilistic uncalibrated single model ensemble forecast. The formula only depends on the classical reference score $RPS_{reference}$, the category probabilities p_i and the ensemble size M:

$$RPSS_D = 1 - \frac{\overline{RPS}}{\overline{RPS}_{reference} + D}$$

$$D = D_0/M$$

and

$$D_0 = \sum_{k=1}^{K} \sum_{i=1}^{k} [p_i(1 - p_i - 2 \sum_{j=i+1}^{k} p_j)]$$

Here D is the “intrinsic unreliability” of the EPS and is a measure for the distortion of the forecast probabilities by discretization effects due to finite ensemble size. For K equiprobable categories, that is, if $p_k = (1/k)$ for all $k \in 1, ..., K, D_0$ can be further simplified to

$$D_0 = \frac{k^2 - 1}{6k}$$

If only two dichotomous categories with probabilities p and $(1 - p)$ are considered (BS situation), D_0 becomes

$$D_{0,bin} = p(1 - p)$$

Note that this derivation of the $RPSS_D$ assumes that the ensemble members are not correlated with each other in the case of zero predictability (i.e. that ensembles are not overconfident). Otherwise, the number of indipendent ensemble members, M', would be effectively reduced with respect to M, corrsponding to an increased intrinsic unreliability and a decreased $RPSS_D$. In other words, the $RPSS_D$ penalizes overconfident ensembles, with the penalty depending on M and M'.

The Relative Operating Characteristic Curve Area (ROC Area) is defined as the area under the curve generated by plotting of the cumulative Hit Rate (H) against False Alarm Rate (F). Hit Rate and the False Alarm Rate for each probability class (threshold) are defined as:

$$H_k = \frac{a_k}{a_k + c_k}$$

$$F_k = \frac{b_k}{b_k + d_k}$$

following the contingency table (see the table below).

	Observed Y	Observed N
Forecast Y	a	b
Forecast N	c	d

Table 4.1: The contingency table

The two scores indicate, respectively, the proportion of events which were predicted by k ensemble members and actually happened, and the proportion of events forecast by k members and did not occur. It is commonly used as a probabilistic score, its maximum value being 1 and a value of 0.5 indicating a no-skill forecast system (Mason and Graham, 1999).

The Percentage of Outliers of a probabilistic forecast system is defined as the probability of the analysis (or observation) lying outside the forecast range (Buizza, 1997). Therefore this can be seen as the percentage of times the “truth” falls out of the range spanned by the forecast values. Here, it is computed as the fraction of points of the domain where the observed value lies outside the minimum or maximum forecast value.

Chapter 5

Performance of the ensemble systems

The verification of numerical models against observations has several purposes (Bougeault P., 2003). For instance:

- providing a measure of the progress of the forecast skill over the years;
- comparing the merits of two versions of a forecasting system in order to decide which one is the best for operations;
- understanding where the problems are and what aspects of the system need refinements;
- comparing the relative value of two different systems for a specific category of users.

No single verification system can be optimal for all of these tasks and there is a need to issue guidance on what methods are good for what purpose. The logical process of verification against observations can be divided in five steps (Bougeault P., 2003):

1. the choice of a set of observations for verification;
2. the technique to compare a single model forecast to a single observation;
3. the aggregation of model/observation pairs in ensembles of a convenient size;

4. the use of statistics to condense the information contained in the joint distribution of model/observation pairs;
5. the use of additional information to help the interpretation of the scores, in particular their statistical significance.

This chapter focuses on the performance of the three ensemble systems (ECMWF EPS, COSMO-LEPS, COSMO-2I-EPS) that have already been described in chapter 3. The results are reported for both upper-level and surface variables in order to provide a comprehensive description of the strengths/weakness of the different ensemble systems. Before proceeding with the analysis of data, the observational networks is presented with a description of the observation types used to evaluate the ensemble systems.

5.1 Observational networks

Following the description of the models in chapter 3, the verification domain was selected in such a way as to include the entire Italian territory, more precisely the domain having the following geographic coordinate as borders (Fig. 5.1):

- latitude: 35^oN - 48^oN
- longitude: 6^oE - 19^oE

Only direct observations within this domain were considered. This approach was preferred to the use of analysis data from ERA-Interim reanalysis [1] basically for two reasons:

- although in some cases the observational network has few station points, it nevertheless provides data that have a higher value than the reanalysis ones. Infact, data coming from stations are observed data, recorded data, thus, they are generally considered as true representation of the status of the atmosphere (Ghelli A., 2009);

---

[1] ERA-Interim is a dataset, showing the results of a global climate reanalysis from 1979 to date. ERA-Interim continues to be updated in near-real-time as new data become available. ERA stands for 'European Reanalysis' and refers to a series of research projects at ECMWF which produced various datasets (ERA-Interim, ERA-40, etcetera).

Figure 5.1: The domain, centered over Italy, considered for the verification of the three ensemble systems.

- ERA-Interim is not an indipendent analysis as it is provided by the ECMWF; therefore, it is likely to favour ECMWF ENS respect the COSMO-based ensemble systems.

The station networks, used in the evaluation procedure, are:

- the Italian *radiosoundings* and those of the neighboring Countries included within the considered domain, for a total of 17 upper level variables observations, available every 12 hours or 24 hours (Fig. 5.2);
- *the Northern-Italy non-GTS* [2] *(local) network*: it refers to about 1000 stations, over most Northern Italy and shared by the regional weather services operating in the area. These stations provide hourly data (Fig. 5.3);
- *network from National Civil Protection Department (DPCN-Dipartimento Protezione Civile Nazionale)*: this network is composed of about 5524 stations over the national territory. Also these stations provide hourly data (Fig. 5.4).

DPCN stations have been subdivided, in three groups depending on the location altitude (Fig. 5.5). For the subdivision it was decided to adopt the WMO (World Meteorologiacal Organization) directives on the subject, as follows:

- *lowland station* (under 200 m of altitude) 2311 DPCN observatories belong to this category;
- *hill stations* (between 200 m and 599 m of altitude) 1690 observatories belong to this category;
- *mountain stations* (above 600 m of altitude) 1523 observatories belong to this category.

---

[2] Global Telecommunications Systems

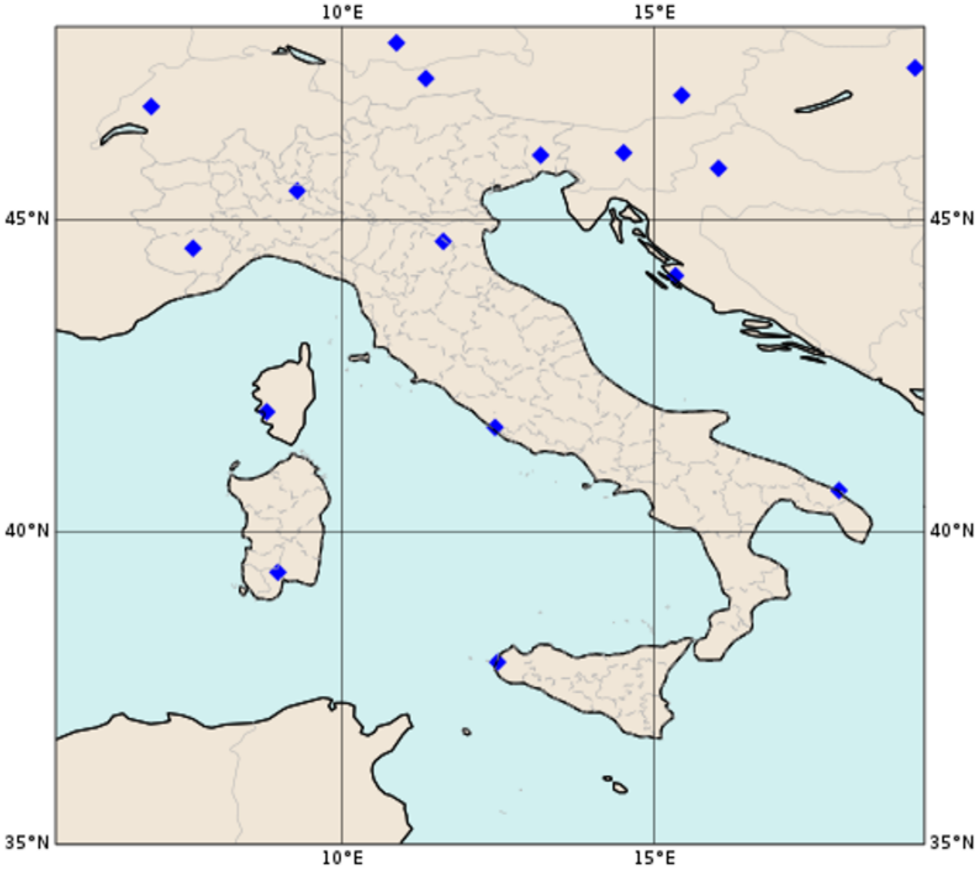


Figure 5.2: The position of radiosoundings within the domain.

5.2 Upper-level evaluation

In this section the performances of ECMWF ENS, COSMO-LEPS and COSMO-2I-EPS are evaluated taking into account the upper-level variables, i.e. geopotential height and temperature. For both variables, the verification was performed at three isobaric levels: 500hPa, 700hPa and 850hPa. The observations at these isobaric heights are obtained from radiosounding carried out every 12/24 hours at 17 different sites within the considered domain. The observations, in this case but also in all the others, are assumed without errors. In fact, an exact estimate of the errors would be extremely diffucult since they are different depending on the instrumentation manufacturer. In addition to this, the instrumental uncertainties are much lower than fore-

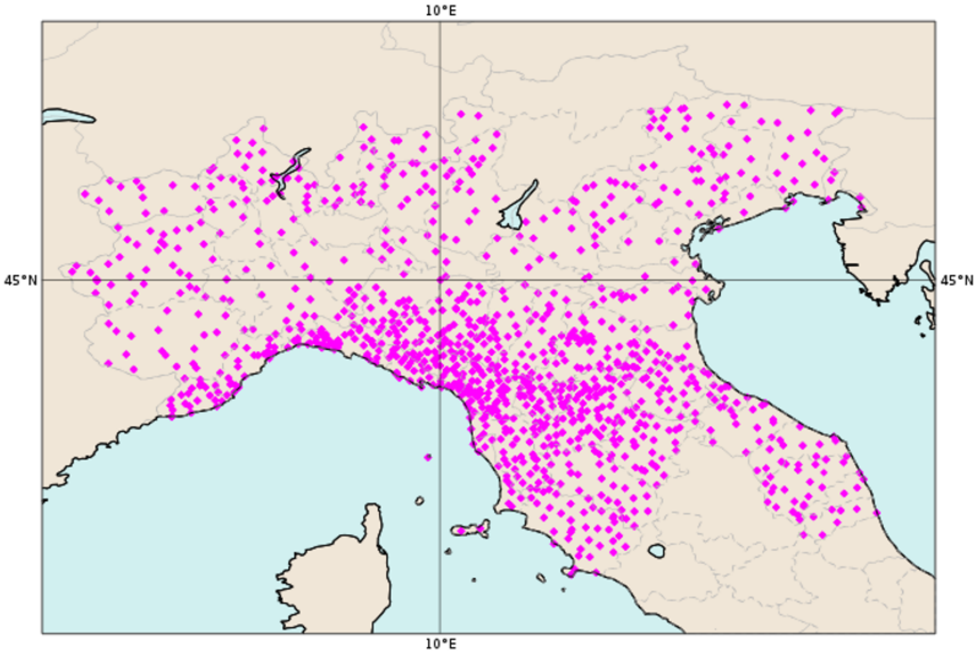


Figure 5.3: The position of the stations of the Northern-Italy non-GTS network within the domain.

cast ones as will be seen later. The verification period lasts 9 days: from 20th June 2016 at 00 UTC to 29th June 2016 at 00 UTC. The ensemble runs taken into consideration are those starting at 00 UTC from 20th June to 27th June 2016. This choice to consider only the runs of 00 UTC is due to the fact that COSMO-2I-EPS, unlike the other two ensemble systems, is available only at this time and therefore the comparison can be based only on the 00 UTC run. A similar argument applies for the forecast range: although the forecast ranges of ECMWF ENS and COSMO-LEPS extends respectively to 240 and 132 hours, that of COSMO-2I-EPS reaches 48 hours. Therefore, a comparison between the three ensemble systems can only take place within the first 48 hours of forecast. The performance of the three ensemble systems is evaluated by calculating the spread among the ensemble members and the RMSE of the ensemble mean; the verification method used is the nearest grid point. The table 5.1 summarizes the characteristics of the verification.

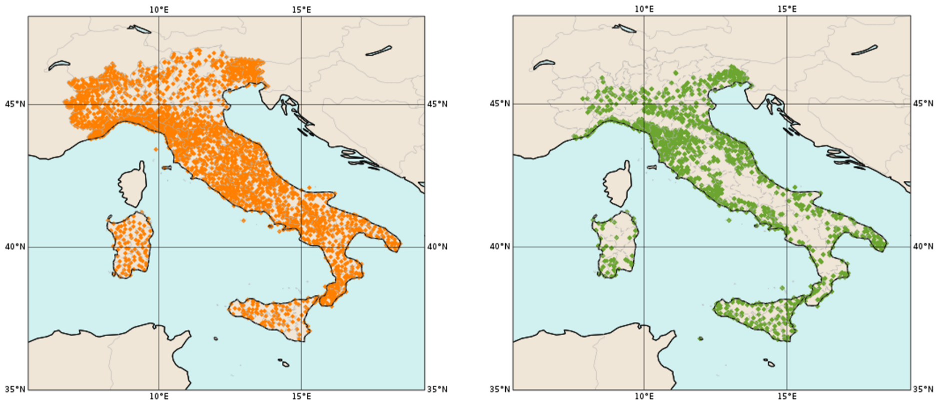
Figure 5.4: The position of DPCN stations within the domain.

Verification features	
variable:	geopotential height and temperature;
period:	from 20/06/2016 00UTC to 29/06/2016 00UTC (9 days);
region:	Italy and neighboring Countries;
method:	nearest grid point;
obs:	radiosoundings, no obs error;
fcst ranges:	0-48h (verification every 12h);
systems:	ECMWF EPS, COSMO-LEPS, COSMO-2I-EPS;
scores:	spread, RMSE;

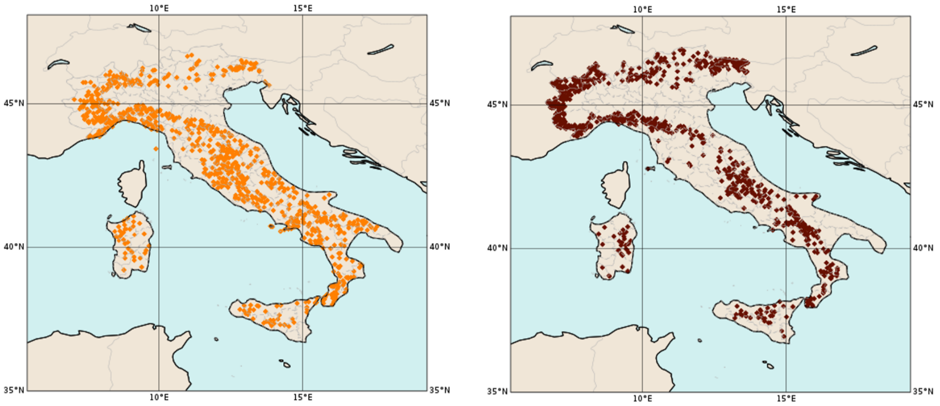
Table 5.1: geopotential height and temperature verification features

5.2.1 Geopotential height

In this subsection the verification focuses on performance of the systems in terms of geopotential height at 500hPa, 700hPa and 850hPa, according to the



(a) *All the DPCN stations* (b) *DPCN plain stations*



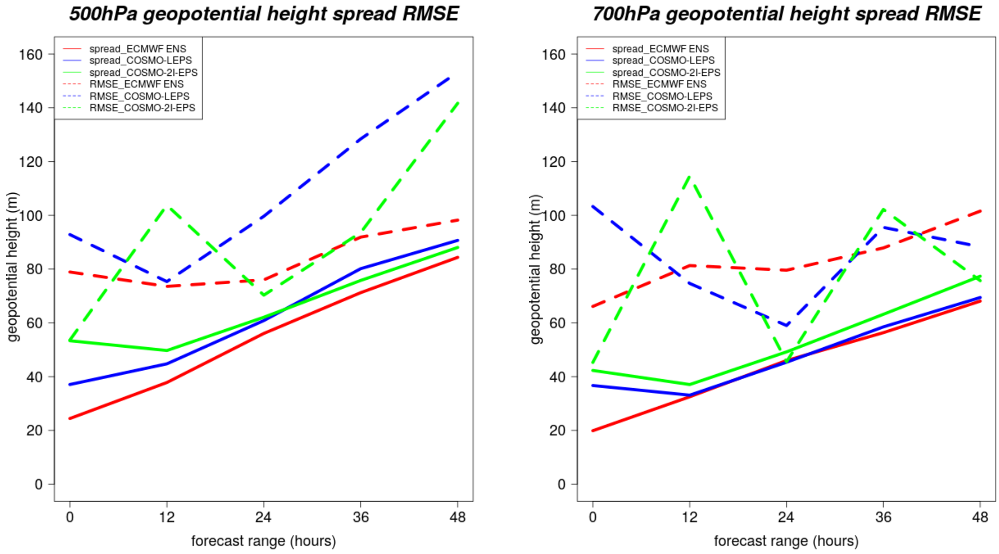
(c) *DPCN hill stations* (d) *DPCN mountain stations*

Figure 5.5: (a) The position of DPCN stations within the domain; (b) the position of DPCN plain stations within the domain; (c) the position of DPCN hill stations within the domain; (d) the position of DPCN mountain stations within the domain.

specifications explained above. The results, divided according to the three isobaric heights, are shown in (Fig. 5.6). In this graphs, the forecast range indicated in abscissa does not refer only to a single run, but to all eight run (from that of 20^{th} June 2016 at 00 UTC to that of 27^{th} June 2016 at 00 UTC). Therefore, for example, the information contained in the 12^{th} hour of the forecast range is not that of a specific run, but that of all runs at $+12 hours$ (12 UTC) from the moment of issue (00 UTC). The spread and the RMSE of the geopotential height is shown on y axis.

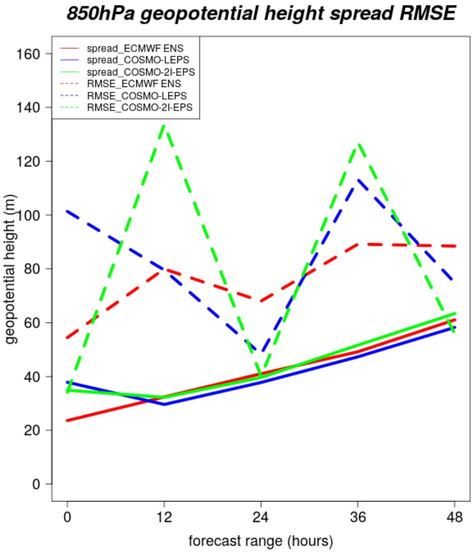
From Fig.5.6 it turns out that:

- the spread values (solid lines) are similar for all forecast systems and generally lower than the RMSE values (dashed lines);
- the spread increases almost monotonically as the time range increases;
- only at the beginning of forecast range the spreads of the three ensemble systems show greater discrepancies;
- the spread of COSMO-based model for the geopotential height at 500hPa is slightly higher (therefore better) than ECMWF ENS, in particular COSMO-2I-EPS spread is good in the first 24 hours;
- it must be highlighted that the COSMO-2I-EPS spread at 700hPa is always higher than the other ensembles ones;
- the COSMO-2I-EPS RMSE is very close to the corresponding spread during the night hours;
- in these three moments the RMSE of COSMO-2I-EPS is the lower (and thus the best) of the other two ensembles;
- during the central hours of the day, instead, COSMO-2I-EPS has the worst RMSE;
- COSMO-2I-EPS has the problem of a very marked diurnal cycle; this trend is less visible in the RMSE string of COSMO-LEPS and ECMWF ENS;



(a) *500hPa geopotential height*

(b) *700hPa geopotential height*

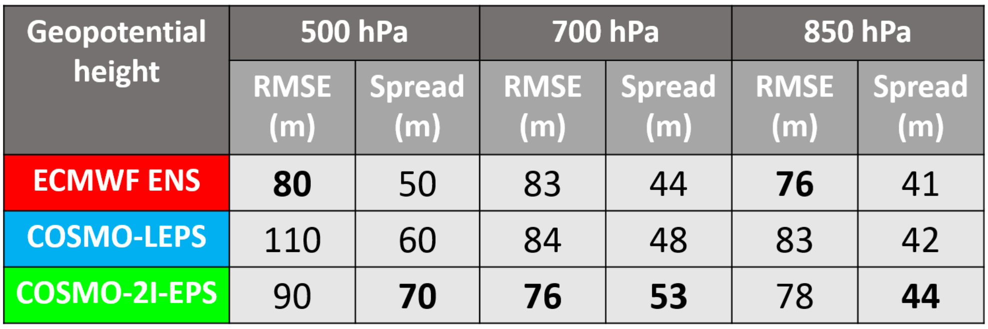


(c) *850hPa geopotential height*

Figure 5.6: The figure shows the spreads (continuous lines) and the RMSE (dotted lines) values for three pressure levels indicated in the captions. The results are obtained for the 48 hours of the forecast range every 12 hours. The ECMWF EPS scores appear in red, COSMO-LEPS in blue and COSMO-2I-EPS in green. The forecast range (in hours) is shown in the abscissa whereas in the ordinate the value of spread and RMSE (in m) is shown. All details are indicated in the legend at the top left of each figure.

- the daytime cycle of the RMSE is less visible at 500hPa, where the RMSE of the three ensemble systems tends to increase after 24 hours.

In order to summarise the informations relative to the performance of the three systems at different levels, we performed an average of the obtained values across all forecast ranges. In this way one value of spread and one of RMSE are obtained for every ensemble systems at each of the three isobaric level (500hPa, 700hPa, 850hPa). These results can be found in the table (Fig.5.7).



Geopotential height	500 hPa		700 hPa		850 hPa	
	RMSE (m)	Spread (m)	RMSE (m)	Spread (m)	RMSE (m)	Spread (m)
ECMWF ENS	**80**	50	83	44	**76**	41
COSMO-LEPS	110	60	84	48	83	42
COSMO-2I-EPS	90	**70**	**76**	**53**	78	**44**

Figure 5.7: The table shows the spread and RMSE values for the geopotential height averaged over the entire forecast range. The results are sorted by ensemble system (in rows) and by pressure levels, hence spreads and RMSE (in columns). For each class of values (spread or RMSE), it is pointed out in bold which of the three values (one for every ensemble) is the best.

Results indicate that COSMO-2I-EPS is often the best ensemble system in terms of spread/skill relation. Infact, COSMO-2I-EPS has always the highest spread values at all the isobaric height, but also the RMSE is the smallest or among the smallest. Instead, ensembles with a lower horizontal resolution are penalized by this analysis: the ECMWF ENS spread is always the lowest, while COSMO-LEPS has always the highest RMSE. In the previous chapters, it is recalled how in a perfect ensemble forecast the spread should be equal to the root mean square error (Buizza et al., 2005). This analysis shows a situation far from the ideal case, because there is a problem usually common to all ensemble systems: the underdispersion of their forecast members due to the limitation in the description of the sources of uncertainty by state-of-the-art ensemble systems. However, COSMO-2I-EPS is less underdispersed than the other ensembles and often it has a contained RMSE: for this reason it is the one that comes closest to the ideal case, while remaining far away.

5.2.2 Temperature

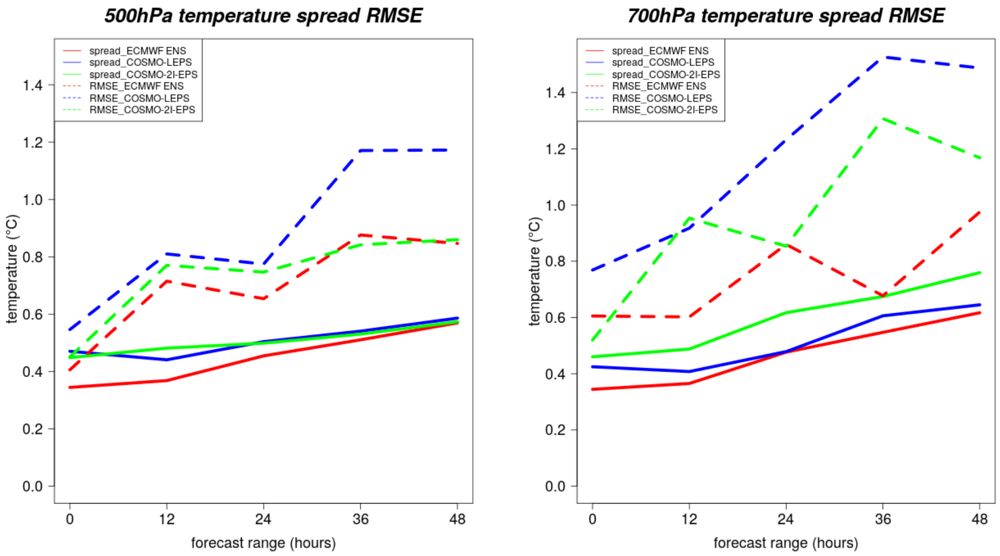
In this subsection the verification focuses on performance of the systems in terms of temperature at 500hPa, 700hPa and 850hPa, according to the specifications explained above. The results, divided according to the three isobaric heights, are shown in (Fig. 5.8).

From Fig.5.8 it turns out that:

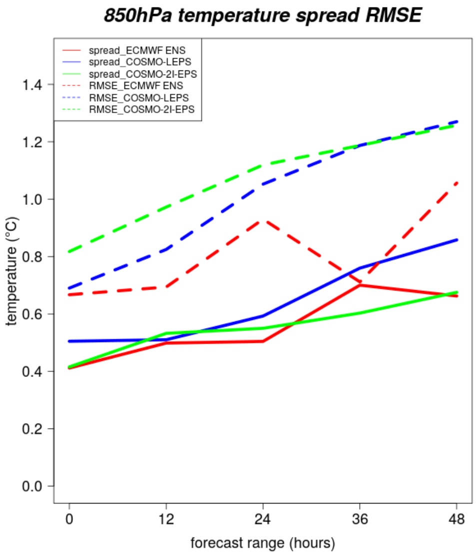
- the spread values (solid lines), calculated for the three ensemble systems, are very similar and increase slightly as the time range increases;
- the spreads of COSMO-based systems, however, are better than those of ECMWF ENS;
- COSMO-2I-EPS spread at 700hPa costantly has higher and therefore better values;
- as for the spread, also the RMSE show an upward trend as the forecast range increases;
- the RMSE is low or quite low at the beginning of the forecast range, then it tend to grow, although in a different way for the three systems, showing in few cases a diurnal cycle;
- the RMSE is generally higher for COSMO-based models.

As for the geopotential, also for the temperature, in order to summarise the informations relative to the performance of the three systems at different levels, we performed an average of the obtained values across all forecast ranges. In this way one value of spread and one of RMSE are obtained for every ensemble systems at each of the three isobaric level (500hPa, 700hPa, 850hPa). These results can be found in the table (Fig. 5.9):

The results obtained by this operation are slightly less satisfactory than in the case of the geopotential. COSMO-2I-EPS achieves good results at 500hPa

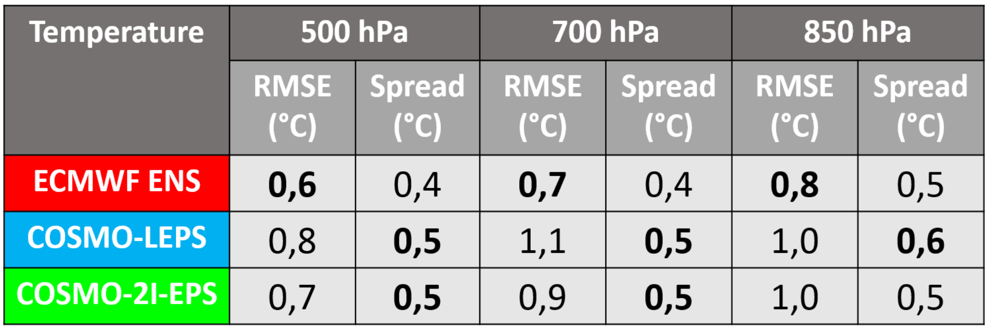


(a) *500hPa temperature* (b) *700hPa temperature*



(c) *850hPa temperature*

Figure 5.8: The figure shows the spreads (continuous lines) and the RMSE (dotted lines) values for three pressure levels indicated in the captions. The results are obtained for the 48 hours of the forecast range every 12 hours. The ECMWF EPS scores appear in red, COSMO-LEPS in blue and COSMO-2I-EPS in green. The forecast range (in hours) is shown in the abscissa whereas in the ordinate the value of spread and RMSE (in ^oC) is shown. All details are indicated in the legend at the top left of each figure.



Temperature	500 hPa		700 hPa		850 hPa	
	RMSE (°C)	Spread (°C)	RMSE (°C)	Spread (°C)	RMSE (°C)	Spread (°C)
ECMWF ENS	**0,6**	0,4	**0,7**	0,4	**0,8**	0,5
COSMO-LEPS	0,8	**0,5**	1,1	**0,5**	1,0	**0,6**
COSMO-2I-EPS	0,7	**0,5**	0,9	**0,5**	1,0	0,5

Figure 5.9: The table shows the spread and RMSE values for the temperature averaged over the entire forecast range. The results are sorted by ensemble system (in rows) and by pressure levels, hence spreads and RMSE (in columns). For each class of values (spread or RMSE), it is pointed out in bold which of the three values (one for every ensemble) is the best.

and at 700hPa, on these isobaric surfaces the spread (together with COSMO-LEPS) is barely higher than the European Centre one and the RMSE is placed on intermediate values compared to those of the other two ensembles. Instead, at 850hPa, the calculated results for the new ensemble are not as good as in the other cases, although in reality the spread and RMSE value for all the three systems are really very similar to each other. Therefore, limited to the 850hPa temperature analysis, it is perhaps more correct to say that the performance of COSMO-2I-EPS does not differ much from that of the other two models. Overall, however, even for temperature the added value of high resolution emerges quite clearly.

5.3 Surface evaluation

5.3.1 2-metre temperature

In this subsection the performance of the three ensemble systems is verified against the two-metre temperature. As already mentioned in section 4, for this verification it was decided to consider the observational dataset coming from the regional networks of the weather services on Central-Northern Italy. In this way, data coming from only one part of the Peninsula were considered. Infact, the temperature data from the national civil protection network could have been used, but these data are from time to time of low-quality in Central and Southern Italy and their use would have provided wrong evaluation on the model skill. For this reason it was decided to limit the verification only to the area shown in the Fig. 5.3. The period under investigation is from 20^{th} June 2016 at 00 UTC to 29^{th} June 2016 at 00 UTC, infact, although the last runs examined are those at 00 UTC on 27^{th} June 2016, a 48-hour forecast range must always be considered. The performance of the three ensemble systems is evaluated by calculating the spread and the RMSE of the ensemble, the verification method used is the nearest grid point. The table 5.2 summarizes the characteristics of the verification.

Verification features	
variable:	2-metre temperature;
period:	from 20/06/2016 00UTC to 29/06/2016 00UTC (9 days);
region:	Central-Northern Italy;
method:	nearest grid point;
obs:	non-GTS local fiduciary network, no obs error;
fcst ranges:	0-48h (verification every 6h);
systems:	ECMWF EPS, COSMO-LEPS, COSMO-2I-EPS;
scores:	spread, RMSE;

Table 5.2: 2-metre temperature verification features

The results are reported in (Fig. 5.10) and can be summarised as follows:

- the spread values are similar for all the three ensemble systems;
- the spread values are smaller with respect to the RMSE ones, showing a tendency of all ensembles to be underdispersive;
- with the exception of the shortest time range, COSMO-based models always show slightly higher (and therefore better) spread values than ECMWF EPS;
- RMSE values show a marked diurnal cycle, with maxima during the central hours of the day and the minimums in the night. This daytime cycle is very pronounced for ECMWF EPS and for COSMO-LEPS, less for COSMO-2I-EPS;

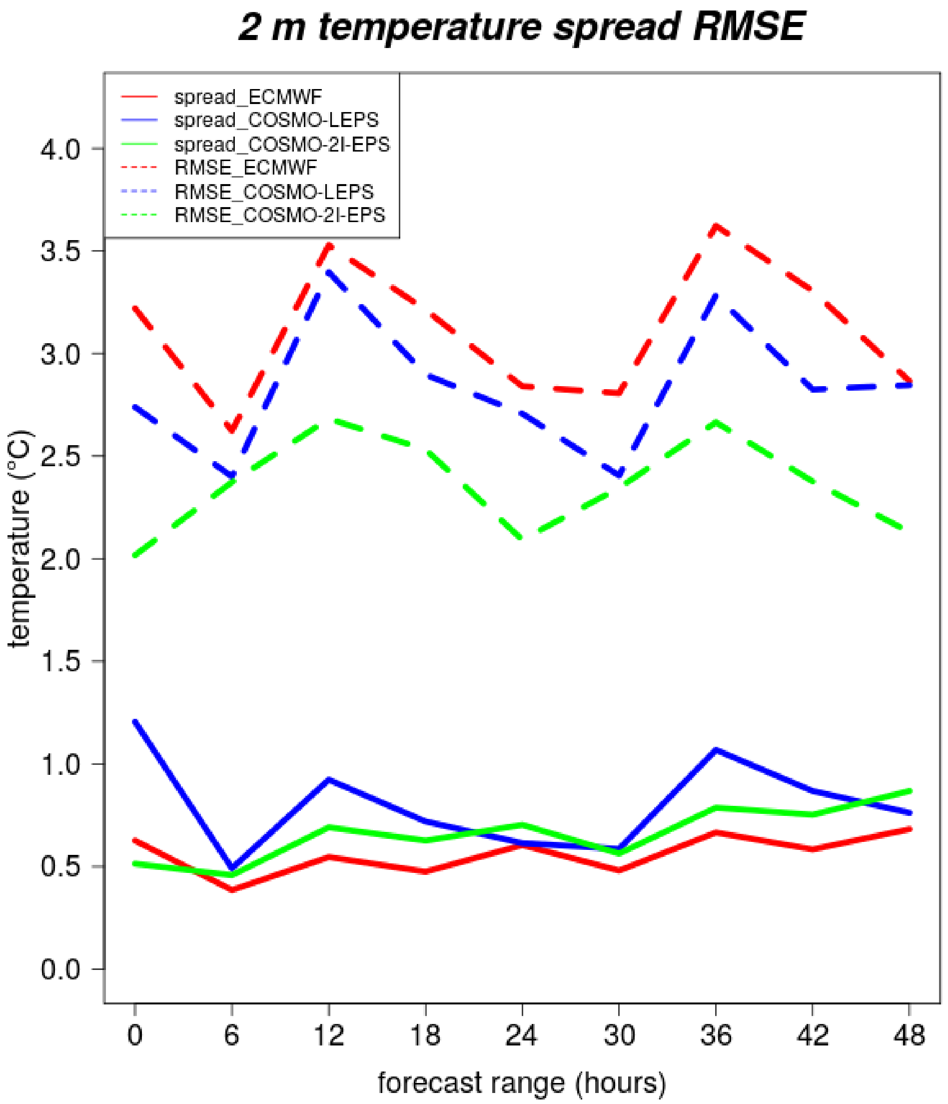


Figure 5.10: The figure shows the spreads (continuous lines) and the RMSE (dotted lines) values obtained for the 48 hours of the forecast range every 6 hours. The ECMWF EPS scores appear in red, COSMO-LEPS in blue and COSMO-2I-EPS in green. The foreacst range (in hours) is shown in the abscissa, in the ordinate the value of spread and RMSE (in °C). All details are indicated in the legend at the top left.

- the RMSE of COSMO-2I-EPS is the lowest of the three ensemble systems on the entire forecast range.

Therefore, from this 2-metre temperature verification, COSMO-based models get excellent results, especially COSMO-2I-EPS.

5.3.2 6-hourly precipitation

In this subsection the performance of ECMWF ENS, COSMO-LEPS, COSMO-2I-EPS is verified against the 6-hourly precipitation. For this verification work it was decided to use the precipitation data recorded by the rain gauges of National Civil Protection Department network. In this way, the results obtained are representative of what happened on the entire national territory between the 20^{th} June 2016 at 00 UTC and the 29^{th} June 2016 at 00 UTC. The method of the nearest grid point was used for the calculation of some probabilistic scores, among those described in chapter 4. The calculation of the BSS and the ROC Area require the chioce of thresholds that have been fixed at 1mm, 5mm, 10mm, 15mm, 25mm, 50mm in 6 hours. However, for the sake of brevity, the attention has been focused only on two thresholds (1mm/6-h and 10mm/6-h), which have a quite large number of occurrences (around a thousand for the former, some hundreds for the latter) over the verification period and represents different intensity precipitation. The other probabilistic scores (RPS, RPSSD and the percentage of outliers) are not threshold oriented. In table 5.3 are reported all the details of the verification.

Verification features	
variable:	6-hourly total precipitation ;
period:	from 20/06/2016 00UTC to 29/06/2016 00UTC (9 days);
region:	Italy;
method:	nearest grid point;
obs:	DPCN network, no obs error;
fcst ranges:	0-48h (verification every 6h);
systems:	ECMWF EPS, COSMO-LEPS, COSMO-2I-EPS;
scores:	BSS, RPS, RPSSD, ROC Area, outliers;
thresholds:	1mm, 5mm, 10mm, 15mm, 25mm, 50mm in 6 hours

Table 5.3: 6-hourly total precipitation verification features

As for Brier Skill Score, the results are presented in Fig.5.11 for the 1 and 10mm thresholds.

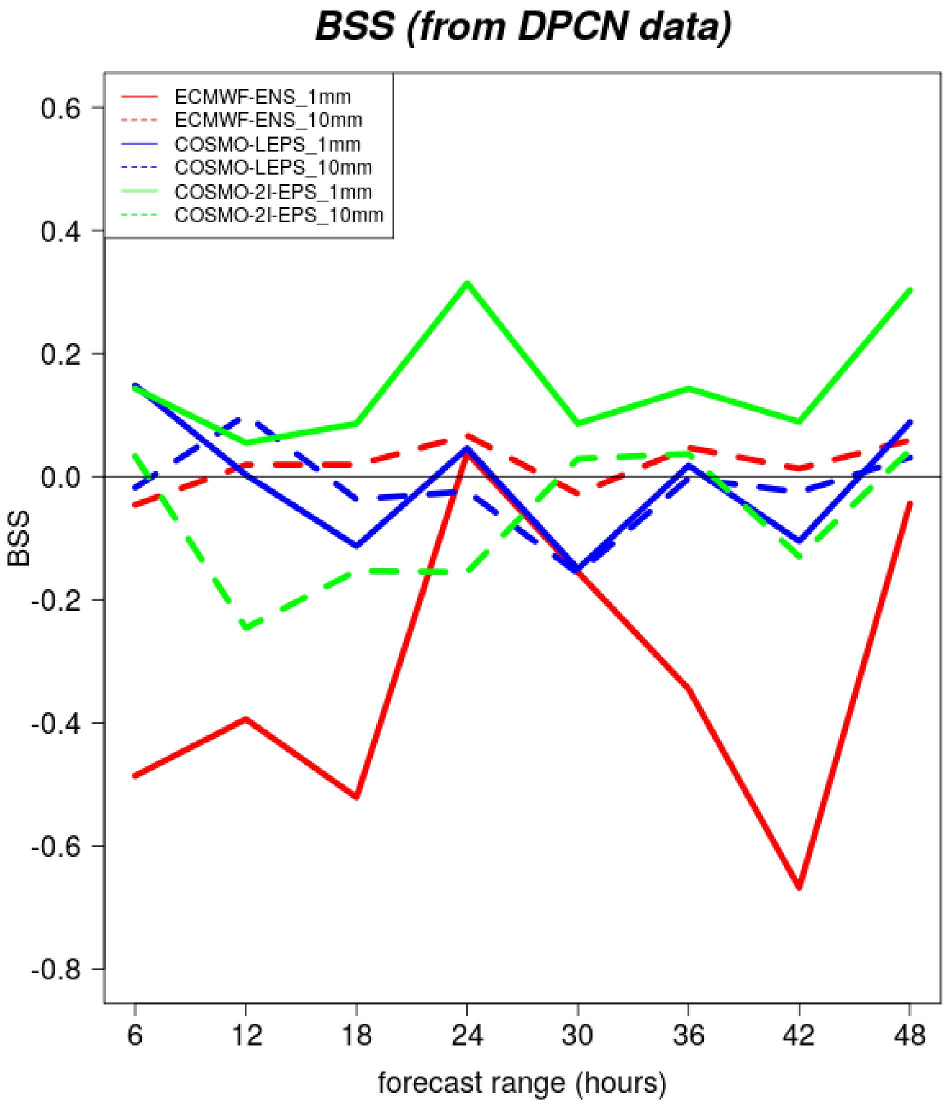


Figure 5.11: The figure shows the BSS for the three ensemble systems (ECMWF ENS in red, COSMO-LEPS in blue, COSMO-2I-EPS in green) and for the threshold of 1 mm (continuous line) and 10 mm (dotted line). The forecast range of 48 hours, in 6-hour steps, is shown in the abscissa. The dimensionless values of the BSS are marked in ordinate. In particular, a black line was placed for BSS=0. Negative values of BSS indicate a forecast skill lower than climatology.

Looking at the values plotted in the graph, it is possible to see:

- for the 1 mm threshold (solid lines) the best results are those obtained by COSMO-2I-EPS over all forecast range. Instead, the worst results (with BSS almost always negative) are those of ECMWF ENS. This could be due to the fact that ECMWF ENS sometimes predicts weak precipitation (1 mm threshold) on areas much larger than those where precipitation actually occurs;
- the BSS obtained for the 10 mm threshold (dashed lines) are similar for the three ensemble systems: there are no significant differences. Values oscillate around zero and COSMO-2I-EPS only seems to be slightly better than the others.

Therefore, it can be concluded the best BSS for the COSMO-models are for weak precipitation (threshold 1 mm/6-h).

In the Fig.5.12, the results obtained for the RPS can be consulted.

It is worth pointing out:

- the diurnal cycle of precipitation is very marked. Since it is almost exclusively afternoon convective precipitation, the highest, and therefore the worst, RPS are just in the afternoon time slots: forecast range 12-18 h, 36-42 h;
- however, the RPS of COSMO-2I-EPS, and generally the RPS of COSMO ensemble systems, is better than ECMWF ENS one over the whole forecast range.

In this case the RPS points out to the added value of COSMO-2I-EPS.

The RPS is not threshold oriented, so it lends itself to an easier interpretation of the results: for this reason, it has been done an in-depth study. The station of DPCN has been subdivided, according to their altitude, in three groups: plain, hill and mountain (see the section 5.1). Therefore, the purpose of this further study is to evaluate RPS variations according to the station altitude and see how this affects the results. RPS diagram for lowland, hill and mountain stations are shown in Fig. 5.13. In order to simplify the

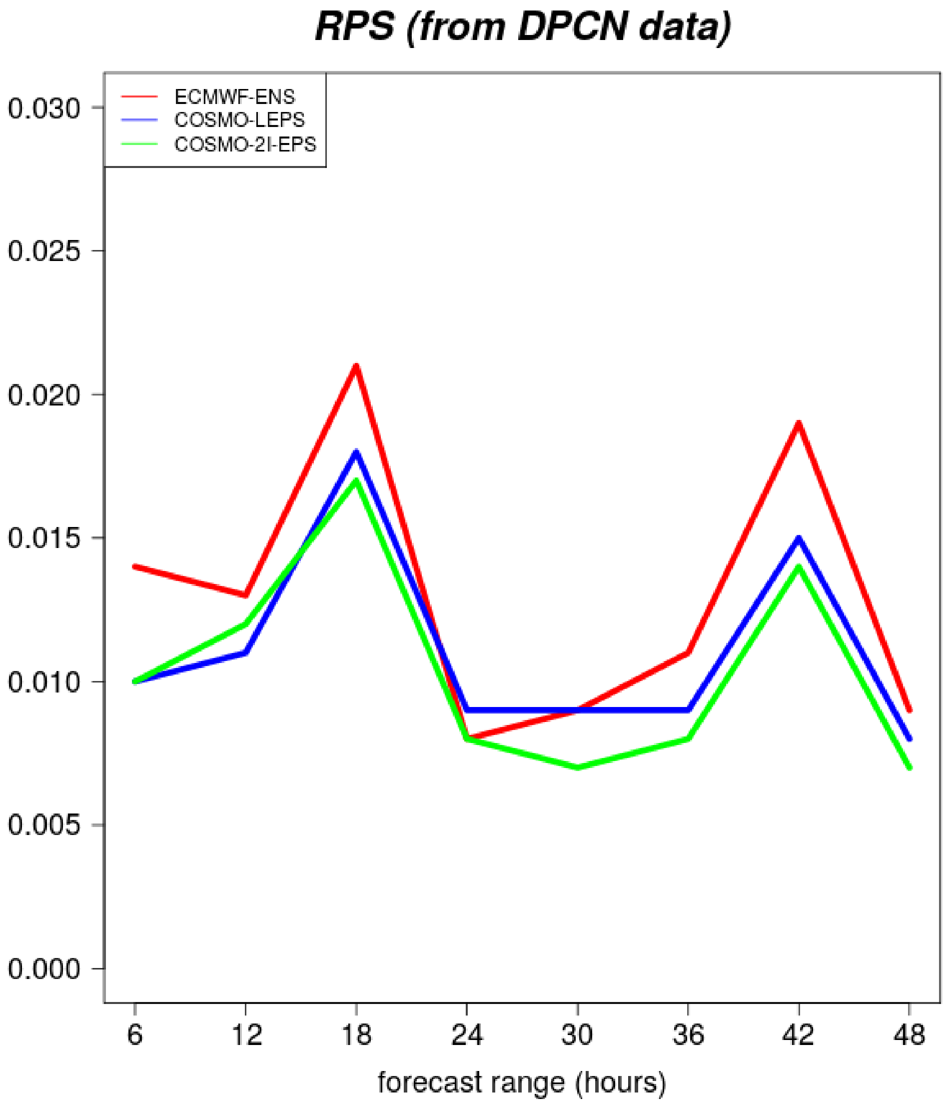
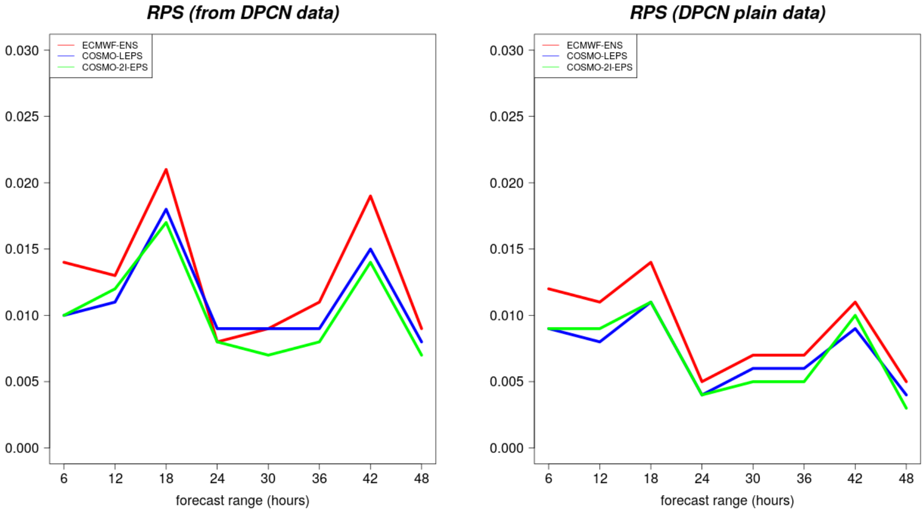


Figure 5.12: The figure shows the RPS for the three ensemble systems (ECMWF ENS in red, COSMO-LEPS in blue, COSMO-2I-EPS in green). The forecast range of 48 hours, in 6-hour steps, is shown in the abscissa; the dimensionless values of the RPS are marked in the ordinate.

comparison, the graph valid for all the DPCN stations, already commented above, has been added.

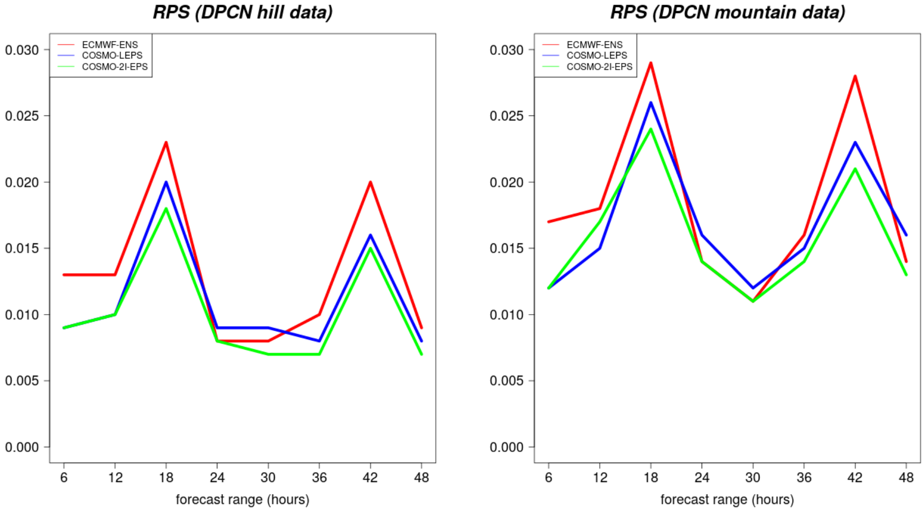
Looking at the plots it can be concluded that:

- the RPS values obtained for the lowland stations are lower (therefore



(a) *All the DPCN stations*

(b) *DPCN plain stations*



(c) *DPCN hill stations*

(d) *DPCN mountain stations*

Figure 5.13: The figure shows the RPS for four different observational dataset, indicated in the caption under each image. The ensemble systems are ECMWF ENS in red, COSMO-LEPS in blue, COSMO-2I-EPS in green. The forecast range of 48 hours, in 6-hour steps, is shown in the abscissa; the dimensionless values of the RPS are marked in the ordinate.

better) than those obtained for hill and mountain ones, in particular the results of mountain stations are the highest;

- in most cases, regardless of altitude, the RPS obtained for COSMO-2I-EPS is always lower (therefore better) than for COSMO-LEPS and ECMWF ENS;
- in the plain stations (top-right panel), there is a good gap beetween COSMO-2I-EPS and ECMWF ENS in the first day of forecast range. For the other stations this gap extends no longer than the first 18 hours, then the RPS tend to be similar for the three ensembles, except for the precipitation cumulated beetween the 36^{th} and the 42^{nd} hour of the forecast range;
- in the graph for hill and mountain stations (bottom left and bottom right panel respectively), the RPS follows a very strong daytime cycle, that is definitely less visible on the plain: this is could be due to the pluviometric regime of those days, with rainfall concentrated almost always in the afternoon hours and on the internal areas of hills and mountains.

So all the observational networks, built on altitude, confirm that the RPS of COSMO-based ensembles, but in particular COSMO-2I-EPS, are better than the global ensemble of Reading.

The results for the $RPSS_D$ are reported in Fig. 5.14, where the influence on the scores by the different ensemble size was removed.

The main findings can be summarised as follows:

- the best performance by COSMO-2I-EPS compared to other ensemble systems is confirmed;
- COSMO-LEPS, provides good results than the global ensemble over the whole forecast range;
- ECMWF ENS is always the worst of the three ensemble systems, with the $RPSS_D$ exceeding zero only for one forecast range.

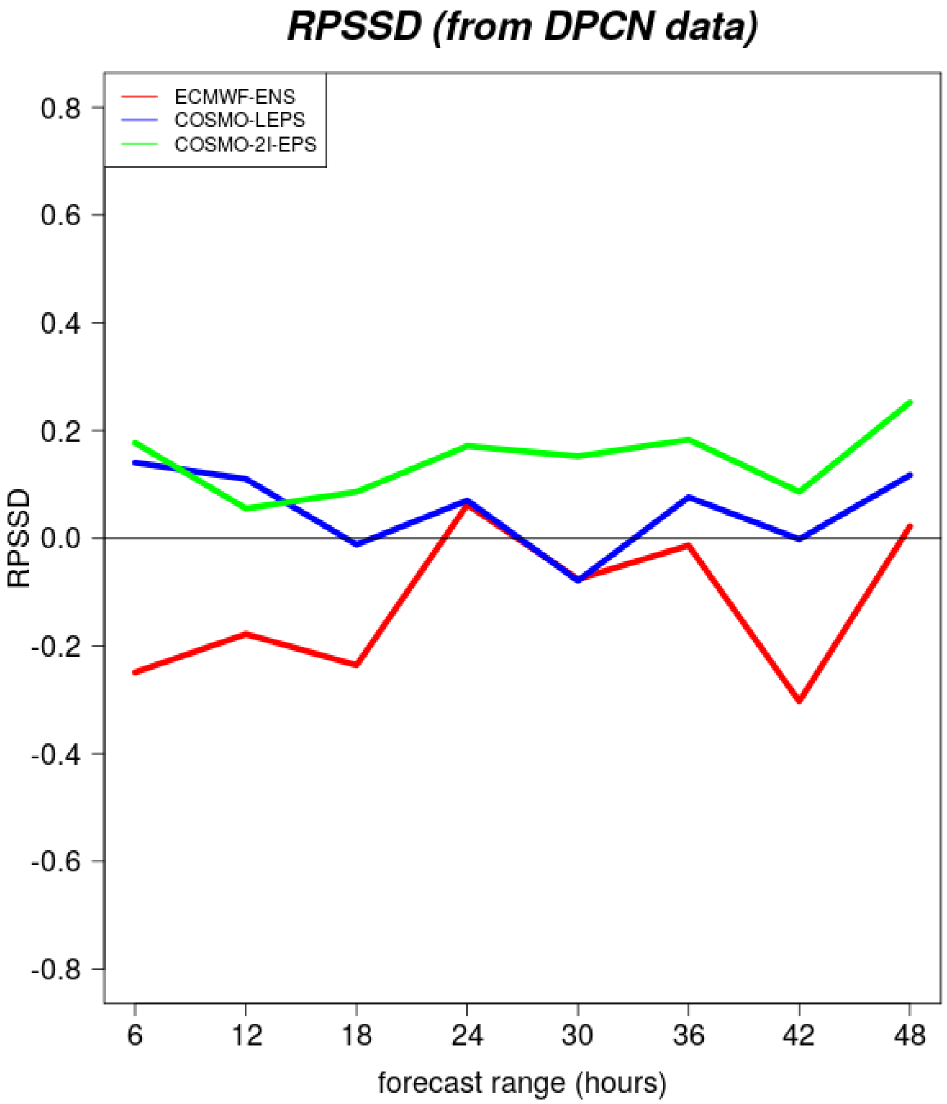


Figure 5.14: The figure shows the $RPSS_D$ for the three ensemble systems (ECMWF ENS in red, COSMO-LEPS in blue, COSMO-2I-EPS in green). The forecast range of 48 hours, in 6-hour steps, is shown in the abscissa. The dimensionless values of the BSS are marked in the ordinate. In particular, a black line was placed for $RPSS_D = 0$. Negative values of $RPSS_D$ indicate a forecast skill lower than climatology.

Also for the $RPSS_D$ it was decided to perform verifications dividing the DPCN station network according to their altitude, as already done for the

RPS. Also in this case the objective is to assess the performance of the systems for different altitudes of the stations used for verification. $RPSS_D$ diagram for lowland, hill and mountain stations are shown in Fig. 5.15. In order to simplify the comparison, the graph valid for all the DPCN stations, already commented a little above, has been added.

Observing the four panels of Fig. 5.15 it is possible to notice that:

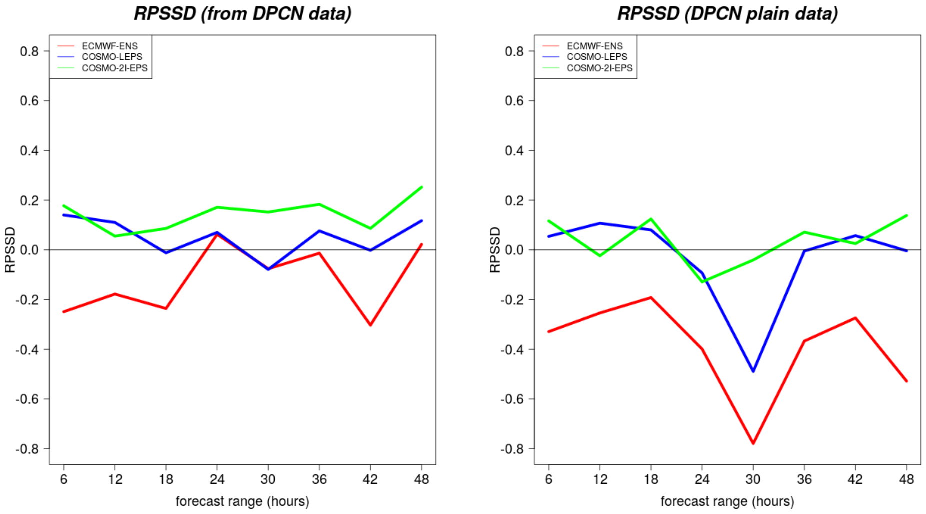
- on average, a slight increase in values with the quota is visible;
- in the case of the plain station dataset (top right panel), $RPSS_D$ is below zero at forecast range +30h for both ECMWF ENS and COSMO-LEPS: unfortunately it is not clear what could have caused two values so low and why they are found only for lowland stations;
- $RPSS_D$ values obtained with mountain stations (bottom right panels) show a slight diurnal cycle, probably due to the convective precipitation, which particulary affected the mountainous areas during the afternoon hours;
- the ECMWF ENS $RPSS_D$ always has values smaller than zero in plains and only occasionally positive for hill and mountain areas: so it is often worse than climatology;
- COSMO-2I-EPS almost has always higher RPSSD than other ensembles, especially for hill and mountain networks;
- COSMO-LEPS has intermediate results.

Also this second verification, conducted taking into account the station altitude, confirms the best trend of COSMO ensembles, in particular of COSMO-2I-EPS.

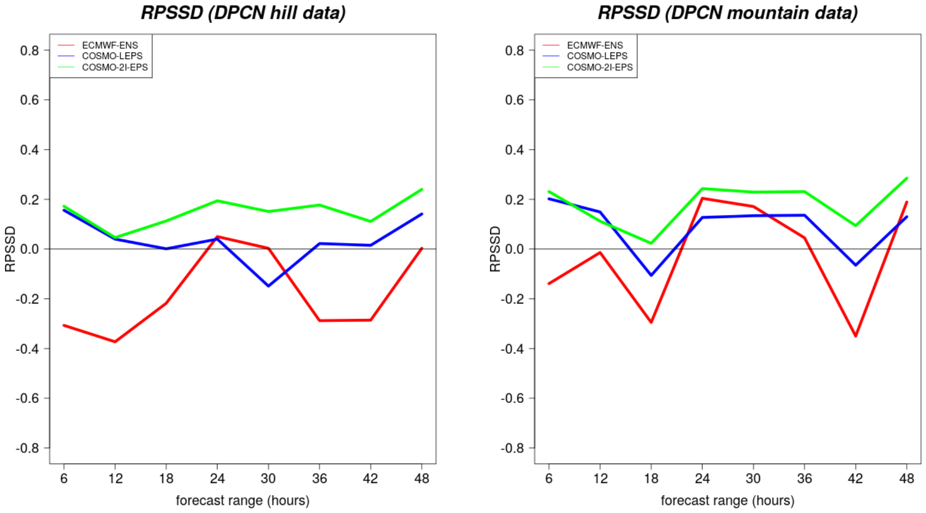
As for the ROC Area, results are reported in Fig. 5.16.

It is possible to notice that:

- considering the 1 mm threshold the best ROC Area is in favour of ECMWF ENS over the whole forecast range, while COSMO-2I-EPS is a bit below, on values overall better than COSMO-LEPS;



(a) *All the DPCN stations* (b) *DPCN plain stations*



(c) *DPCN hill stations* (d) *DPCN mountain stations*

Figure 5.15: The figure shows the $RPSS_D$ for four different observational dataset, indicated in the caption under each image. The ensemble systems are ECMWF ENS in red, COSMO-LEPS in blue, COSMO-2I-EPS in green. The forecast range of 48 hours, in 6-hour steps, is shown in the abscissa; the dimensionless values of the $RPSS_D$ are marked in the ordinate. In particular, a black line was placed for $RPSS_D = 0$. Negative values of $RPSS_D$ indicate a forecast skill lower than climatology.

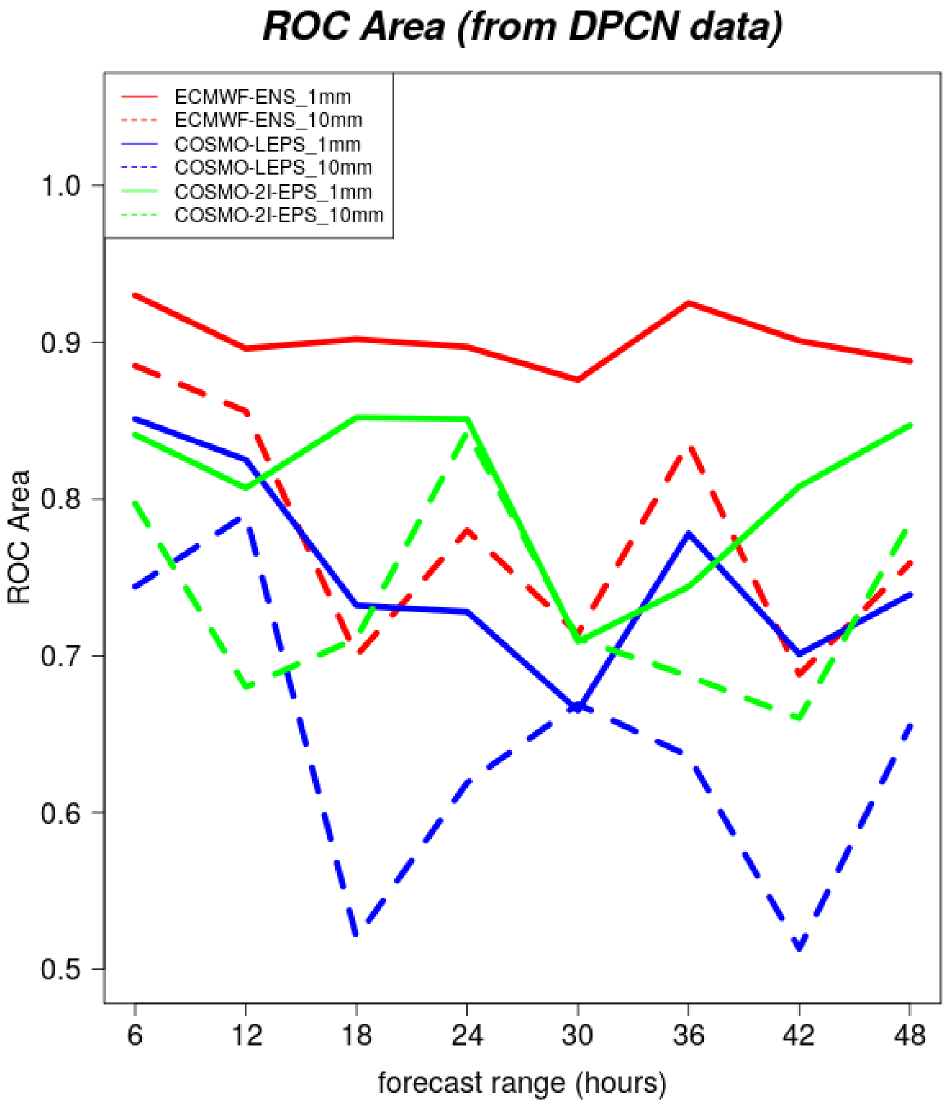


Figure 5.16: The figure shows the ROC Area for the three ensemble systems (ECMWF ENS in red, COSMO-LEPS in blue, COSMO-2I-EPS in green) and for the threshold of 1 mm (continuous line) and 10 mm (dotted line). The forecast range of 48 hours, in 6-hour steps, is shown in the abscissa. The dimensionless values of the ROC Area are marked in the ordinate.

- COSMO-2I-EPS obtains in some cases higher values than other two ensemble systems for precipitations over 10 mm.

However, as it already appears at first sight looking at the figure 5.16, the results are not very clear and it does not come out unequivocally which ensemble systems has the best ROC Area.

The percentages of outliers for the ensemble system considered as a function of the forecast range are shown in Fig. 5.17.

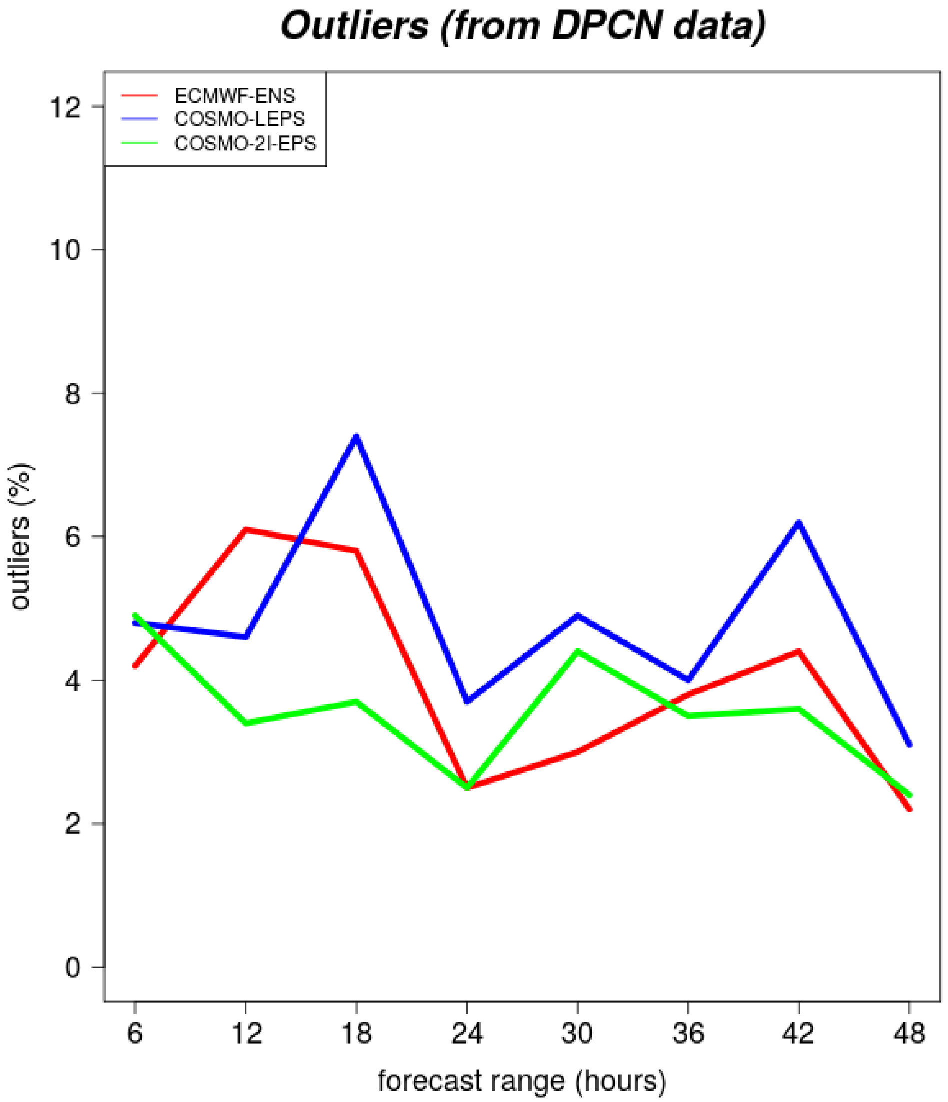


Figure 5.17: The figure shows the percentage of outliers for the three ensemble systems (ECMWF ENS in red, COSMO-LEPS in blue, COSMO-2I-EPS in green). The forecast range of 48 hours, in 6-hour steps, is shown in the abscissa. The percentage of outliers is marked in the ordinate.

Despite the lower ensemble size, COSMO-2I-EPS has often the lowest values, compared to the other two ensemble systems with a lower horizontal resolution. So, it can be stated that in this case too, the results obtained by COSMO-2I-EPS are satisfactory.

Finally, the verification according to the station altitude also involved the percentage of outliers; also in this case the experiment proceeded similarly to what was done before with the RPS and with the $RPSS_D$. Percentage of outliers diagram for lowland, hill and mountain stations are shown in Fig. 5.18. In order to simplify the comparison, the graph valid for all the DPCN stations, already commented a little above, has been added.

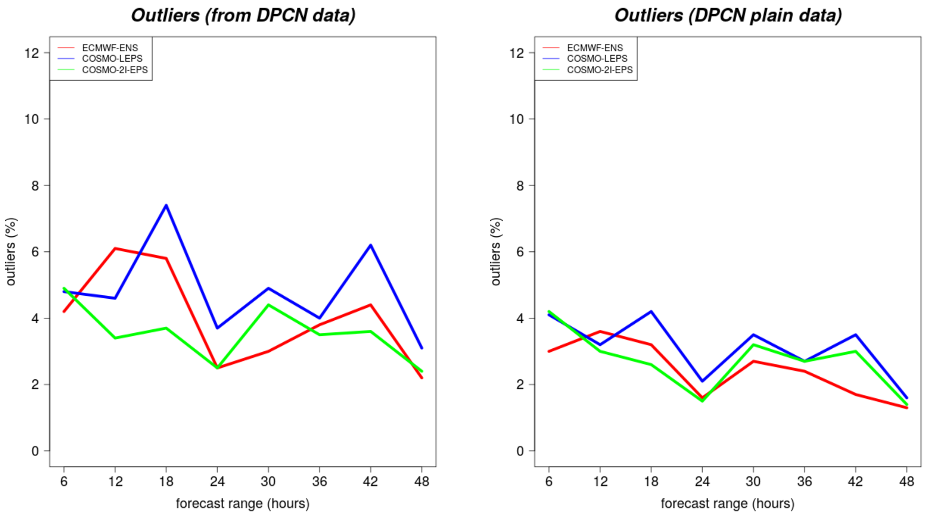
Looking at the four panels of the Fig. 5.18 it can be stated that:

- the percentage of outliers increases according to the station altitude: there are less outliers in the plains than in the mountains;
- in the plain there is little difference between the three ensemble systems; these differences increase with the altitude, infact the percentage of outliers obtained with the only mountain stations shows considerable dissimilarity beetween the ensembles;
- for hill and mountain observation datasets, a diurnal cycle is visible only in systems with parametrized convection (ECMWF ENS, COSMO-LEPS); instead, the diurnal cycle is hardly identifiable for the lowland stations;
- for almost all forecast ranges COSMO-2I-EPS has the lowest percentage of outliers.

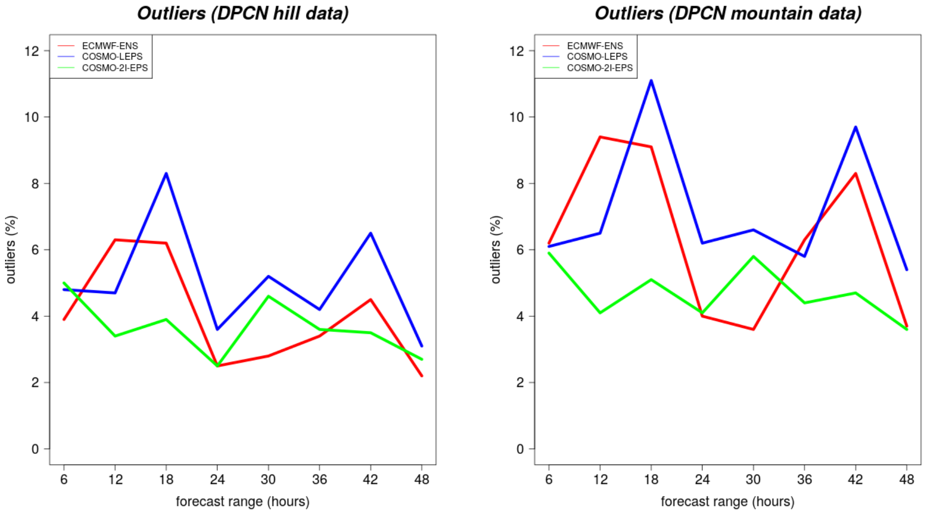
Therefore also the percentages of outliers, studied according to the altitude of DPCN stations, indicate the good skill of COSMO-2I-EPS.

5.4 Sensitivity of the scores to the verification methodology

Until now, for all the experiments carried out so far, the model grid point closest to the observation station was selected. Instead, in this paragraph a



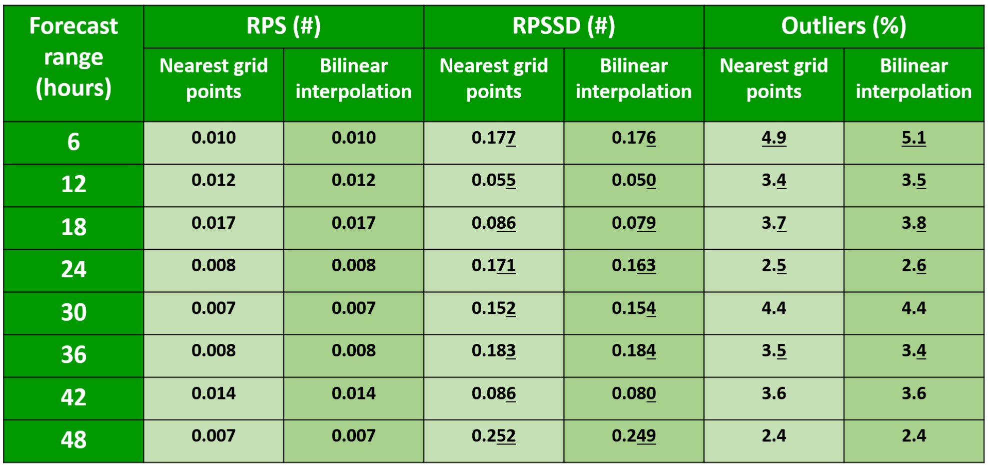
(a) *All the DPCN stations* (b) *DPCN plain stations*



(c) *DPCN hill stations* (d) *DPCN mountain stations*

Figure 5.18: The figure shows the percentage of outliers for four different observational dataset, indicated in the caption under each image. The ensemble systems are ECMWF ENS in red, COSMO-LEPS in blue, COSMO-2I-EPS in green. The forecast range of 48 hours, in 6-hour steps, is shown in the abscissa; the percentage of outliers is marked in the ordinate.

new type of model-observation pairs is introduced: the bilinear interpolation. The aim is to compare these two methods of verification. For simplicity, it has been decided to make a comparison only with COSMO-2I-EPS and only probabilistic scores without thresholds (RPS, $RPSS_D$, percentage of outliers) were considered. The results obtained for the two methods, for each of the three scores are shown in the table (Fig. 5.19).



Forecast range (hours)	RPS (#)		RPSSD (#)		Outliers (%)	
	Nearest grid points	Bilinear interpolation	Nearest grid points	Bilinear interpolation	Nearest grid points	Bilinear interpolation
6	0.010	0.010	0.177	0.176	4.9	5.1
12	0.012	0.012	0.055	0.050	3.4	3.5
18	0.017	0.017	0.086	0.079	3.7	3.8
24	0.008	0.008	0.171	0.163	2.5	2.6
30	0.007	0.007	0.152	0.154	4.4	4.4
36	0.008	0.008	0.183	0.184	3.5	3.4
42	0.014	0.014	0.086	0.080	3.6	3.6
48	0.007	0.007	0.252	0.249	2.4	2.4

Figure 5.19: In the table for each step of the forecast range, the values for RPS, $RPSS_D$ and percentage of outliers, for both verification methods(nearest grid point and bilinear interpolation) are written. In order to help the reader in the comparison, the digits that change from one method to another have been underlined.

The bilinear interpolation has already been discussed, even briefly, in the previous chapter; here it is said only that, being a more elaborate verification method than the nearest grid point, the results are expected to be better. However, looking at the results written in the table, there are no substantial improvements using the bilinear interpolation instead of the nearest grid points. Infact, the RPS values are even identical for all forecast ranges, while for the $RPSS_D$ and the percentage of outliers there are variations at decimal positions at most, therefore not really significant. Moreover, these small variations often have an opposite sign compared to those expected. For this reason it has been decided not to extend this verification method to all cases already dealt with in the previous paragraphs, since there seems to be no noteworthy impact on the so-far obtained results.

5.5 Deterministic evaluation of the ensemble systems

After verifing the three systems in ensemble mode, a deterministic approach is discussed in this section. As for the ensembles starting at 00 UTC of 26^{th} June 2016, the total precipitation cumulated over 24-hours, predicted by member 1 at the forecast range +24h/+48h, is considered in order to be compared with the 24-hours total precipitation collected on 27^{th} June 2016 by the rain gauges of the DPCN network. The comparison has been made according to the nearest grid point method. The results of this comparison are shown in the boxplots in the figure 5.20. In order to obtain better boxplots, the zeros have been eliminated by considering only 391 precipitative events greater than 0.01mm. This operation prevents the boxplots from being crushed on values that are too small and therefore not significant. For a particular data set, the boxplots presented here show the range of data falling between the 25^{th} and 75^{th} percentiles and the median value (i.e. the value for which half of the total number of values is higher/lower than it, the horizontal line inside the box). The whiskers define all the values inside 1.5 times the interquartile range above the upper quartile and below the lower quartile (supposed to be the complete range of the data). Black circles refer to external values, commonly called "outliers".

Looking at the figure it's possible to see that the boxplots are quite similar to each other. The ECMWF ENS distribution does not adequately cover the observed data. On the one hand, the top whisker and the median of COSMO-LEPS are similar to those of the observations, while the top whisker of COSMO-2I-EPS is placed on precipitation value of about 10mm higher and also the median is slightly higher. On the other hand the latter models has the limit of outliers almost coinciding with that of the observations. So, although not very clearly, the rain distribution of COSMO-2I-EPS seems to be the best because the limits of the distribution, considering also the outliers, reflect those of the observations.

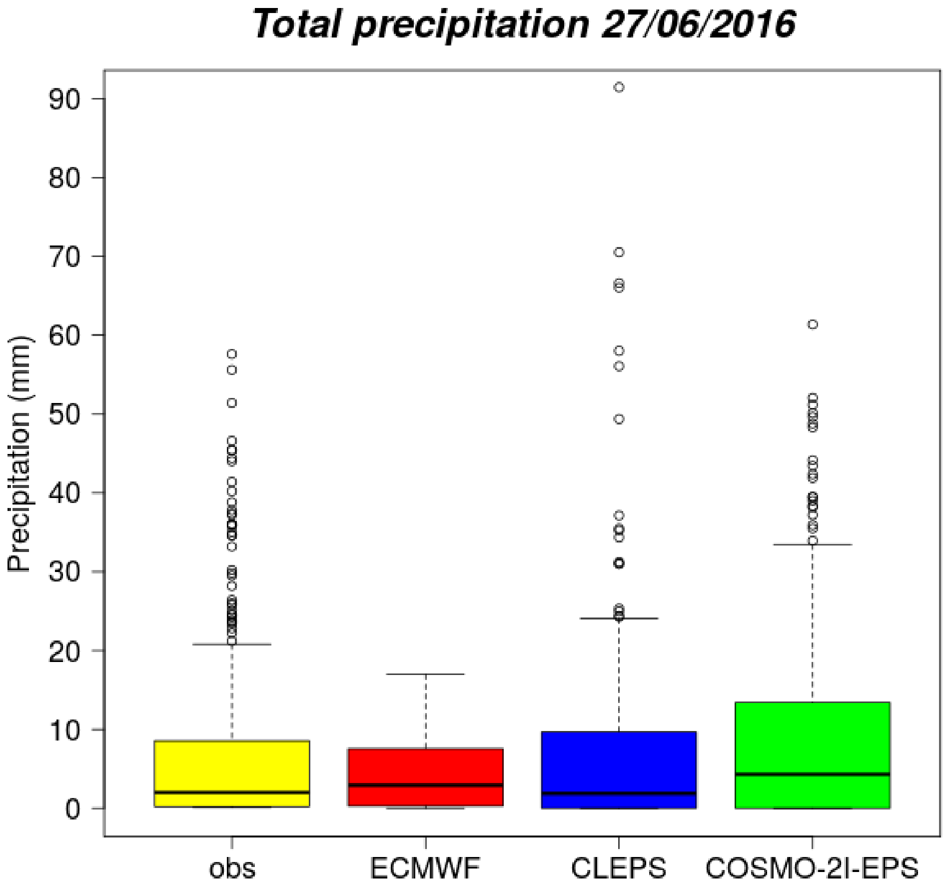


Figure 5.20: Boxplots for the total precipitation cumulated over 24-hours on 27th June 2016: in yellow the observations from the rain gauges of the DPCN, in red as expected by the member 1 of ECMWF ENS with the forecast range +24-48h, in blu and in green the same but for COSMO-LEPS and COSMO-2I-EPS respectively.

Subsequently, again in relation to runs starting at 00 UTC of 26th June 2016, the total precipitation cumulated over 6-hours has been considered (forecast range +24h/+48h). For the reasons explained above, even in this case the zeros have not been taken into consideration. The day has been divided into four time slots:

- (00-06)UTC with 220 events;

- (06-12)UTC with 166 events;
- (12-18)UTC with 154 events;
- (18-24)UTC with 23 events.

For each of these time intervals have been made boxplots to compare the data forecast by the member 1 of each ensemble system with the observations, in the same way as what has been done for the whole day. The results are shown in Fig. 5.21.

The boxplot representing the data from 00UTC to 06UTC are very similar to each other. As in the previous case also here the boxplot and the median of COSMO-LEPS are those closer to observations. However, considering also the outliers, in some cases COSMO-LEPS is too more rainy than in reality, while the COSMO-2I-EPS provides a better representation of the precipitation peaks. The same statements can in large part be made also for the cumulative precipitation between 06UTC and 12UTC. This time the only difference is given by the COSMO-LEPS outliers, which are now roughly comparable to those of the observations. The boxplot relating to precipitation beetween 12UTC and 18UTC show a substantial understimation of the precipitation by all the members 1 of the ensembles. However, COSMO-2I-EPS has the outliers closer to the observed values. Finally, the results obtained from 18UTC to 24UTC show how all the models have forecasted the absence of noteworthy precipitations.

The last deterministic verification concerns the total precipitation cumulated over 24 hours again. This time the observations have been divided into four intensity categories, in order to have a good statistical sample also for the class with more intense precipitation. The four categories of precipitation are composed in this way:

- *drizzle* (from 0.0mm to 0.9mm), with 2504 events ;
- *light* (from 1.0mm to 9.9mm) with 161 events;
- *moderate* (from 10.0mm to 19.9mm) with 40 events;

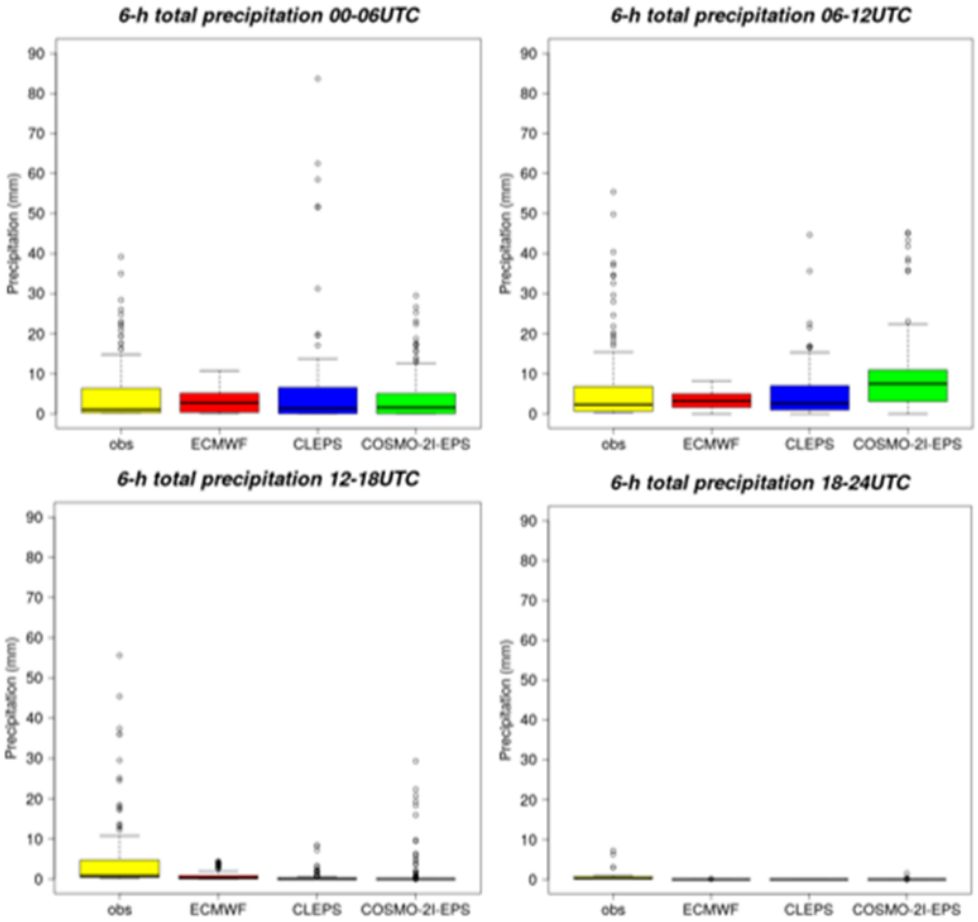


Figure 5.21: Boxplots for the total precipitation cumulated over 6-hours on 27^{th} June 2016: in yellow the observations from the rain gauges of the DPCN, in red as expected by the member 1 of ECMWF ENS with the forecast range +24-48h, in blu and in green the same but for COSMO-LEPS and COSMO-2I-EPS respectively. Time slots are indicated in the titles of the figures.

- *heavy* (over 20.0mm) with 41 events.

Therefore, the technique of the nearest grid point has been used to compare the observations of each class with the predictions of member 1 of the three ensemble systems; the forecast range is still (24-48 hours). The results are illustrated in Fig. 5.22.

On grid points close to the stations that have recorded the lower rainfall amounts, there have been numerous overstimations by the models; in

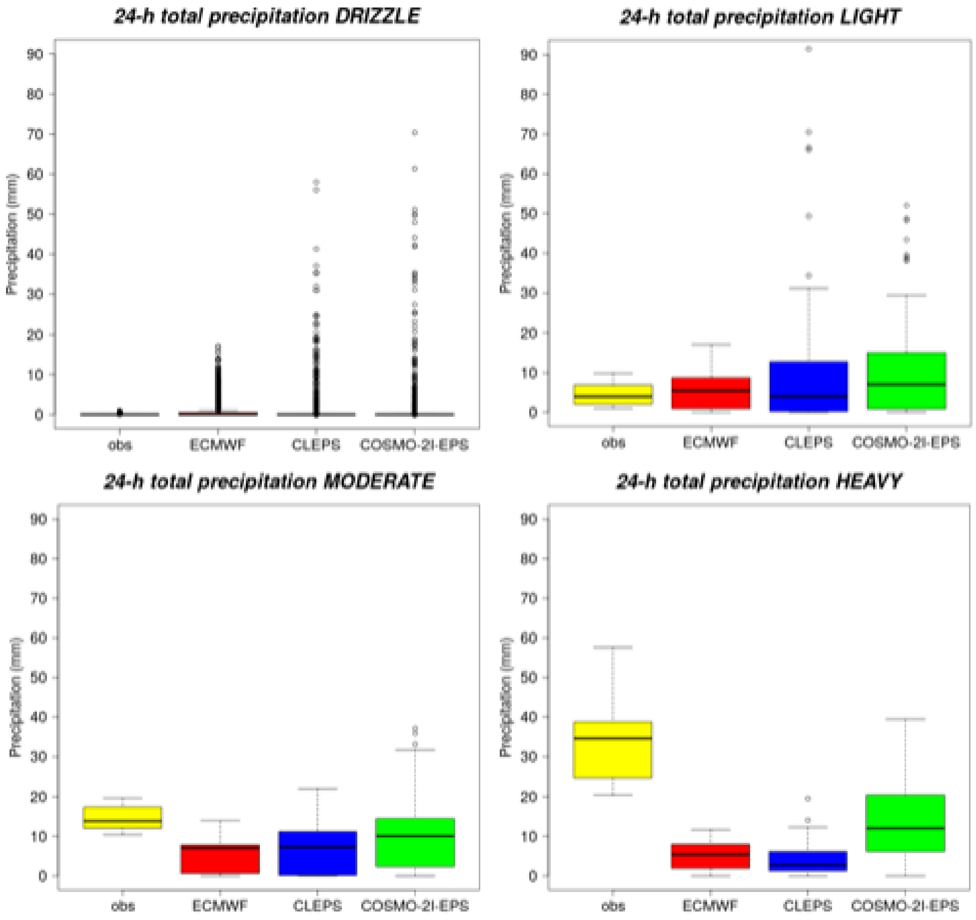


Figure 5.22: Boxplots for the total precipitation cumulated over 24-hours on 27^{th} June 2016: in yellow the observations from the rain gauges of the DPCN, in red as expected by the member 1 of ECMWF ENS with the forecast range +24-48h, in blu and in green the same but for COSMO-LEPS and COSMO-2I-EPS respectively. The intensity categories are indicated in the titles of the figures.

particular, the larger outliers belong to the models with a higher horizontal resolution. Also the light precipitation class rewards the global model, although COSMO-LEPS has the median closest to the observations. The situation changes with the most intense precipitation categories. In fact, in these cases the distributions of COSMO-2I-EPS are the ones that best suit the observations.

In conclusion, it is not easy to find a synthesis from all the results obtained with the deterministic verification. However it can be said that:

- the results of COSMO-2I-EPS are quite satisfactory accumulating the precipitation over 24 hours, without making distinction in intensity;
- COSMO-2I-EPS seems to be better than the other ensembles in cases of intense precipitation.
- the unclear results of this deterministic verification impose to reason in terms of ensemble systems, where all members contribute to more solid weather predictions.

Chapter 6

Conclusions

The present work aims to establish the performance of three ensemble systems with different characteristics, but in particular with a different horizontal resolution. The three ensemble systems involved in the verifications are ECMWF ENS, with 51 members and a horizontal resolution of about 18 km, COSMO-LEPS (16 members and 7 km of resolution) and COSMO-2I-EPS (10 members and 2.2 km of resolution). While ECMWF ENS and COSMO-LEPS run on an daily basis, COSMO-2I-EPS is still on an experimental phase with a full operational implementation planned towards the end of 2018.

Therefore, particular attention has been paid to this new ensemble, especially because it provides new types of numerical modeling products which needs to be assessed and because the best performances were expected from it.

A systematic comparison between the three ensemble systems was undertaken during a "pilot period" from 20^{th} to 27^{th} June 2016. During this period, characterised by particularly unstable weather situation over the Italian Peninsula, the performances of the three systems were compared in terms of both upper-level and surface variables. As for upper-level, both temperature and the geopotential height at 500hPa, 700hPa and 850hPa were considered; the ensemble spread and the root mean square error (RMSE) of the ensemble mean were computed using the available Italian radiosounding data every 12/24 hours for verification. The obtained results indicate that the ensemble spread of the three systems show very similar behaviour, increasing almost monotonically with the forecast range. In addition to this, there is a marked under-dispersion of the different ensemble systems, the ensemble spread being smaller than the RMSE of the ensemble mean at all forecast ranges. However, the spreads obtained for the COSMO-based systems are slightly higher than those of ECMWF ENS, this suggesting a higher ability of the limited-area ensembles to describe more sources of uncertainty. In particular at 700 hPa, both for geopotential and temperature, the results obtained by COSMO-2I-EPS are satisfactory. As for the RMSE, it is more difficult to identify the same trend for each ensemble system regardless of the variable and the isobaric level. We noticed a marked diurnal cycle, especially for the results in the lower troposphere, in addition to the tendency of the RMSE to increase with the forecast range (as expected). Results averaged over the different forecast ranges indicate a better performance by COSMO-2I-EPS, especially for the geopotential.

As for the surface, the forecasts in terms of 2-metre temperature and 6-hourly cumulated precipitation were verified against the Northern-Italy non-GTS network and the National Civil Protection Department network respectively. The results for 2-metre temperature indicate the under-dispersion issue for the different ensemble systems, although it can be noticed that the performance obtained by COSMO-2I-EPS (and in general by the COSMO-based ensembles) is quite satisfactory.

Several probabilistic scores were considered to evaluate the skill of the three ensemble systems in terms of precipitation. In most cases, the scores indicate COSMO-2I-EPS having the best performance. In order to provide more insight on the obtained results and to assess the dependence of the scores on the altitude, it was decided to divide the stations of the National Civil Protection Department into three groups: plain, hill and mountain stations.

With this division, it turns out that the performance of the systems tends to worsen with the altitude, also accentuating the diurnal cycle. This happens because it has rained more over mountain areas and during the afternoon. Anyway the scores obtained by COSMO-2I-EPS remain the best in most cases.

After veryfing the three ensemble systems in ensemble mode, a deterministic evaluation is performed. This consists in comparing the 24-hours total precipitation forecasted by member 1 of the three ensemble systems starting at 00 UTC of 26^{th} June 2016 at the forecast range +24h/+48h, with that collected on 27^{th} June 2016 by the rain gauges of the DPCN network. This verification has been carried out in two different ways: first considering all the precipitation values without distinction and then subdividing them into four classes according to the intensity. A further analysis has been made considering the 6-hours total precipitation. Also this deterministic evaluation shows encouraging results for COSMO-based ensembles. In particular, among the three systems, COSMO-2I-EPS is the best in case of heavy or moderate precipitation. Instead, models with a higher horizontal resolution tend to overstimate weak precipitation.

This thesis can be seen as a pilot study, there is no claim to consider it a complete and exhaustive work, but rather a starting point for further developments and investigations or a "modus operandi" for similar studies. In fact, the period examined, having only one week of data from 20^{th} to 27^{th} June 2016, is too short to have solid results from a statistical point of view. This would take a longer evaluation time, comparing the three ensembles for different atmospheric phenomena and weather types.

All the results shown in this thesis have been obtained with the verification method of the nearest grid point. In the case of 6-h precipitation analysis, the use of the bilinear interpolation has been attempted, however without obtaining the expected improvements. In fact it has been seen that this verification method affects at most the first decimal place and not always improving the results. So a further idea for future works may be to use the bilinear interpolation and the method of boxes to calculate the probabilistic scores in other cases; it will be interesting to see if the results will be better or worse than those obtained with the nearest grid point.

Finally, as for the weather event of 27^{th} June 2016 already presented in the introduction, the panels of Fig. 6.1 report again the precipitation observed and forecast for the 27^{th} June 2016. Differently from Fig. 1.1, we also report the forecast by member 1 of COSMO-2I-EPS at the top right panel.

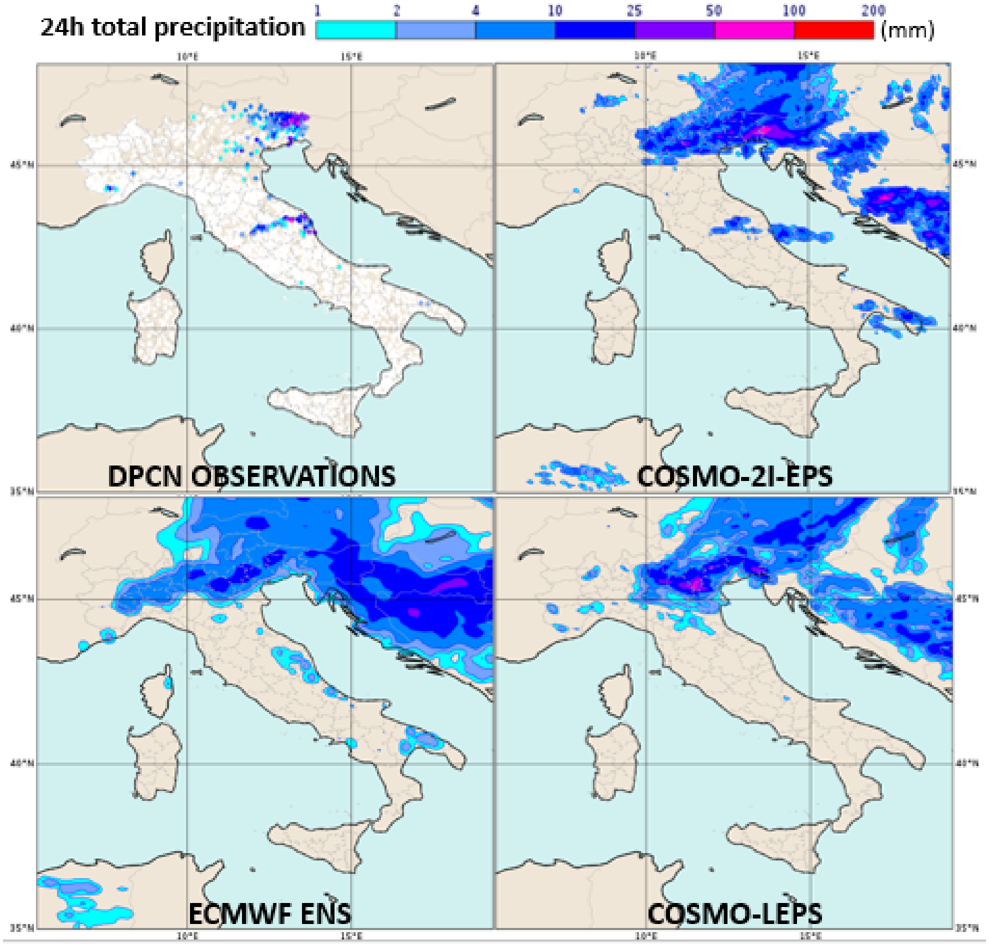


Figure 6.1: Maps of total precipitation cumulated over 24 hours for the 27th June 2016: at the top left observations from rain gauges of DPCN, at the top right as was foreseen by the run of 26th June 00 UTC from the member 1 of COSMO-2I-EPS, at the bottom left and bottom right the same but for ECMWF ENS and COSMO-LEPS respectively.

It can be clearly noticed that the precipitation forecast by COSMO-2I-EPS is the one closest to the observations both for distribution and intensity. Despite this is only a pilot study, the added value of high resolution in mesoscale ensembles seems to play a crucial role in the probabilistic prediction of atmospheric fields at all levels. In particular, in the case of 27th June 2016, the higher forecast skill by COSMO-2I-EPS is highlighted, especially as concerns the prediction of smalle-scale phenomena like summer convective instability.

The main outcomes of this work will be presented to the scientific community at the Annual Meeting of the European Meteorological Society (EMS), which will take place in Budapest from 3 to 7 September 2018 and at the 1^{st} AISAM (Associazione Italiana di Scienze dell'Atmosfera e del Clima) National Congress, that will be held in Bologna from 10 to 13 September 2018.

References

Andersson, E., Thépaut, J.-N., 2008. ECMWF's 4D-Var data assimilation system -the genesis and ten years in operations. *ECMWF Newsletter*, **115**, 8-12.

Arakawa and Lamb, 1977. Computational design of the basic dynamical processes of UCLA general circulation model. Chapter 17, Methods of Computational Physics. New York Academic Press, 173-265.

Balsamo, G., Bousetta, S., Dutra, E., Beljaars, A., Viterbo, P., Van den Hurk, B., 2011. Evolution of land surface processes in the IFS. *ECMWF Newsletter*, **127**, 17-22.

Bechtold, P., M. Köler, T. Jung, Doblas-Reyes, M. Leutbecher, M. J. Rodwell , F. Vitart, G. Balsamo, 2008. Advances in simulating atmospheric variability with the ECMWF model:from synoptic to decadal time-scales. *Q.J.R.Meterol. Soc.*, **134**, 1337-1351, also available as *ECMWF Tech. Memo.* **55**.

Bechtold, P., Semane, N., Lopez, P., Chaboureau, J.-P., Beljaars, A. and Bormann, N., 2014. Representing Equilibrium and Nonequilibrium Convection in Large-Scale Models. *J. Atmosph. Sc.*, **71**, 734-753.

Bougeault P., 2003. WGNE recommendations on verification methods for numerical weather prediction of weather elements and severe weather events. *CAS/JSG WGNE Report*, **18**.

Brier, 1950. Verification of Forecast Expressed in Terms of Probability. *Mon. Wea. Rev.*, **78**, 1-3.

Buizza, 1997. Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF Ensemble Prediction Systems. *Mon. Wea. Rev.*, **125**, 99-119.

Buizza, 2001. Chaos and weather prediction - A review of recent advances in Numerical Weather Prediction: ensemble forecasting and adaptive observation targeting. *Il nuovo cimento*, **24C**, 273-301.

Buizza, R., Milleer, M. and Palmer, T. N., 1999. Stochastic Representation of model uncertainties in the ECMWF Ensemble Presdiction System. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887-2908.

Buizza, R., & Montani, A., 1999. Targeting observations using singular vectors. *J. Atmos. Sci.*, **56**, 2965-2985.

Buizza, R., & Palmer, T. N., 1995. The singular-vector structure of the atmospheric general circulation. *J. Atmos. Sci.*, **52, 9**, 1434-1456.

Buizza, R., P. L. Houtekamer, Z. Toth, G. Pellerin, M. Wei, Y. Zhu, 2005: A comparison of the ECMWF, MSC and NCEP global ensemble prediction systems. *Monthly Weather Review*, **133**, 1076-1097.

Charney, J. G., 1947. The dynamics of long waves in a baroclinic westerly current. *J. Meteor.*, **4**, 135-162.

Charney, J.G., 1948. On the scale of atmospheric motions. *Geofys. Publ.*, **17**, 1-17.

Doms, Förstner, Heise, Herzog, Mironov, Raschendorfer, Reinhardt, Ritter, Schrodin, Schulz and Vogel, 2001. A descriptio of the Nonhydrostatic Regional COSMO Model. Part II: Physical Parameterization. (www.cosmo-model.org).

Ehrendorfer, M., 1994. The Lioville equation and its potential usefulness for the prediction of forecast skill, Part I. *Theory. Mon. Wea. Rev.*, **122**, 703-713.

Epstein, E. S., 1969. Stochastic dynamic prediction. *Tellus*, **21**, 739-759.

Fleming, R. J., 1971a. On stochastic dynamic prediction. I: the energetics of uncertainty and the question of closure. *Mon. Wea. Rev.*, **99**, 851-872.

Fleming, R. J., 1971b. On stochastic dynamic prediction. II: predictability and utility. *Mon. Wea. Rev.*, **99**, 927-938.

Ghelli, A., 2009. Observations and their importance in the verification process: View of the Joint Working Group on Forecast Verification Method Research. *ECMWF, Twelfth Workshop on Meteorological Operational Systems, 2-6 November 2009.*

Gleeson, T. A., 1970. Statistical-dynamical predictions. *J. Apll. Meteorol.*, **9**, 333-344.

Gofa, F., Tzaferi, D. and Charantonis, T., 2010. Application and verification of ECMWF products. *Hellenica National Meteorological Service (HNMS)*

Hersbach, H. P., Janssen, 2007. Operational assimilation of surface wind data from the MetOpASCAT scatterometer at ECMWF. *ECMWF Newsletter*, **113**, 6-8.

Holton, J. R.,1992. An Introduction to Dynamic Meteorology. *Academic Press Inc.*, 511 pp.

Jung, T., Balsamo, G., Bechtold, P., Beljaars, A., Köhler, M., Miller, M., Morcrette, J.-J., Orr, A., Rodwell, M., Tompkins, A., 2010: The ECMWF model climate: Recent progress through improved physical parametrizations. *Quart. J. Roy. Meteor. Soc.*, **136(650)**. 1145-1160, doi: 10.1002/qj.63 also available as *ECMWF Tech. memo*, **623**.

Leith, C. E., 1974. Theoretical skill of Monte Carlo forecast. *Mon. Wea. Rev.*, **102**, 409-418.

Leutbecher M. and Palmer T. N., 2007. Ensemble Forecasting *Research Department.*

Lorenz, E. N., 1960. Energy and Nunerical Weather Prediction. *Tellus*, **12**, 364-373.

Lorenz, E. N., 1963. Deterministic non-periodic flow. *J. Atmos. Sci.*, **20**, 130-141.

Lorenz, E. N., 1965. A study of the ptredictability of a 28-variable atmospheric model. *Tellus*, **17**, 321-333.

Lynch, Giard and Ivanovici, 1997. Improving the Efficiency of a Digital Filtering Scheme for Diabatic Initialization. *Mon. Wea. Rev.*, **125**, 1976-1982.

Marsigli, C., Montani, A., Nerozzi, F., Paccagnella, T., Tibaldi, S. and co-authors, 2001. A strategy for high-resolution ensemble prediction. Part II: limited-area experiments in four Alpine floods events. *Quart. J. Meteor. Soc.*, **127**, 2095-2115.

Marsigli, C., Boccanera, F., Montani, A. and Paccagnella, T., 2005a. The COSMO-LEPS ensemble system: validation of the methodology and verification. *Nonlin. Proc. in Geophys.*, **12**, 527-536.

Mason and Graham, 1999. Conditional probabilities, Relative Operating Characteristics, and relative operating levels. *Wea. For.*, **14**, 713-725.

Miller, M., Buizza, R., Haseler, J., Hortal, M., Janessen, P., Untch, A., 2010. Increased resolution in the ECMWF deterministic and ensemble prediction systems. *ECMWF Newsletter*, **124**, 10-16.

Mironov and Ritter, 2004. A New Sea Ice Model for GME. Technical Note, Deutscher Wetterdienst, Offenbach am Main, Germany.

Molteni, F., Buizza, R., Marsigli, C., Montani, A., Nerozzi F. and co-authors, 2001. A strategy for high-resolution ensemble prediction. Part I: definition of representative members and global-model experiments. *Quart. J. Roy. Meteor. Soc.*, **127** 2069-2094.

Montani, A., Capaldo, M., Cesari, D., Marsigli, C., Modigliani, U. and co-authors, 2003a. Operational limited-area ensemble forecast based on the Lokal Modell. ECMWF Newsletter 98, 2-7. Available at: http://www.ecmwf.int/publications/.

Montani, A., Marsigli, C., Nerozzi, F., Paccagnella, T., Tibaldi S. and Buizza, R., 2003b. The Soverato flood in Southern Italy: performance of global and limited-area ensemble forecasts. *Nonlin. Proc. Geophys.*, **10**, 261-274.

Montani, A., Cesari, D., Marsigli, C. and Paccagnella, T., 2011. Seven years of activity in field of mesoscale ensemble forecasting by the COSMO-LEPS system: main achievements and open challenges. *Tellus*, **63**, 605-624.

Müller, W. A., C. Appenzeller, F. J. Doblas-Reyes, and M. A. Liniger, 2005. A debiased ranked probability skill score to evaluate probabilistic ensemble forecasts with small ensemble sizes. *J. Climate*, **18**, 1513-1523.

Murphy, 1969. On the "Ranked Probability Score". *J. Appl. Meteor.*, **8**, 988-989.

Murphy, 1971. A note on the Ranked Probability Score. *J. Appl. Meteor.*, **10**, 155-156.

Murphy, A. H., 1973. A new vector partition of the probability score. *Journal of Applied Meteorology*, **12**, 595-600.

Murphy, J. M., Sexton, D. M. H., Barnett, D. N., Jones, G. S., Webb, M. J., Collins, M., Stainforth, D. A., 2004. Quantifying uncertainties in climate change using a large ensemble of global climate model predictions. *Nature*, **430**, 768-772.

Palmer, T. N., 2001. A nonlinear dynamical perspective on model error: a proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Quart. J. Roy. Meteor. Soc.*, **127** 279-304.

Palmer, T. N., Alessandri, A., Andersen, U., Cantelaube, P., Davey, M., Delecluse, P., Deque, M., Diez, E., Doblas-Reyes, F. J., Feddersen, H., Graham, R., Gualdi, S., Gueremy, J., Hagedorn, R., Hoshen, M., Keenlyside N., Latif, M., Lazar, A., Maisonnave, E., Marletto, V., Morse, A. P., Orfila, B., Rogel, P., Terres, J., Thomson, M. C., 2004. Development of a European multimodel ensemble system for seasonal-to-interannual prediction (Demeter). *Bull. Am. Meteor. Soc.*, **85**, 853-872.

Palmer, T. N., Molteni, F., Mureau, R., Buizza, R., Chapelet, P., & Tribbia, J., 1993. Ensemble prediction. *Proceedings of the ECMWF Seminar on Validation of models over Europe: vol. I, ECMWF, Shinfield Park, Reading, RG2 9AX, UK.*

Richardson, L. F.,1922. Weather Prediction by Numerical Process. *Cambridge University Press (reprt. Dover, New York).*

Ritchie, H., 1998. Application of the semi-Lagrangian method to a spectral model of the shallow water equations. *Mon. Wea. Rev.*, **116**, 1587-1598.

Ritchie, H., 1991. Application of the semi-Lagrangian method to a multilevel spectral primitive equations model. *Q. J.R. Meteorol. Soc.*, **117**, 91-106.

Ritter and Geleyn, 1992. A comprehensive radiation scheme for numerical weather prediction models with potential applications in climate simulations. *Mon. Wea. Rev.* **120** 303-325.

Schraff, 1996. Data assimilation and mesoscale weather prediction: A study with a forecast model for the Alpine region. Publication 56, Swiss Meteorological Institute, Zürich.

Schraff, 1997. Mesoscale data assimilation and prediction of low stratus in the Alpine region. *Meteor. Atmos. Phys.*, **64**, 21-50.

Schaer, Leuenberger, Fuhrer, L'uthi and Girard, 2002. A New Terrain Following Vertical Coordinate Formulation for Atmospheric Prediction Models. *Mon. Wea. Rev.*, **130**, 2459-2480.

Schättler, Doms and Schraff, 2016. A Description of the Nonhydrostatic Regional COSMO-Model. Part VII: User Guide. (www.cosmo-model.org).

Simmons, A. J., Mureau, R., & Petroliagis, T., 1995. Error growth and predictability estimates for the ECMWF forecasting system. *Q. J. R. Meteorol. Soc.*, **121**, 1739-1771.

Skamarock and Klemp, 1992. The Stability of time-split numerical methods for the hydrostatic and the nonhydrostatic elastic equations. *Mon. Wea. Rev.*, **120**, 2109-2127.

Somerville, R. C. J., 1979. Predictability and prediction of ultra-long planetary waves. *Pre-prints of the Amer. Meteor. Soc. Fourth Conference on Numerical Weather Prediction, Silver Spring, MD*, 182-185.

Stanski et al., 1989. Survey of common verification methods in meteorology. *Atmospheric Environmental Service, Canada.*

Stephan, K., Klink, S. and Schraff, C., 2008. Assimilation of radar-derived rain rates into the convective-scale model COSMO-DE at DWD. *Quart. J. Roy. Meteor. Soc.*, **134**, 1315-1326.

Steppeler, Doms, Schättler, Bitzer, Gassmann and co-authors, 2003. Meso-gamma scale forecast using the nonhydrostatic model LM. *Meteor. Atmos. Phys.*, **82**, 75-96.

Strainforth, D. A., Aina, T., Christensen, C., Collins, M., Faull, N., Frame, D. J., Kettleborough, J. A., Knight, S., Martin, A., Murphy, J. M., Piani, C., Sexton, D., Smith, L. A., Spicer, R. A., Thorpe. A. J., Allen, M. R., 2005. Uncertainty in predictions of the climate response to rising levels of greenhouse gases, *Nature*, **433**, 403-406.

Tennekes, Baede and Opsteegh, 1986. Forecasting forecast skill. *Proceedings of the ECMWF Workshop on Predictability in the Medium and Extended range*, Reading, England.

Thomas, Girard, Doms and Schättler, 2000. Semi-implicit scheme for the DWD Lokal-Modell. *Meteor. Atmos. Phys.*, **75**, 105-125.

Thorpe, A. J., Buizza, R., Montani, A., & Palmer, T. N., 1998. Chaotic-control for weather prediction. *Royal Society New frontiers in Science Exhibition, June 1998, The Royal Society, 6 Carltn House Terrace, London SW1Y-5AG, UK.*

Tibaldi, S., Paccagnella, T., Marsigli, C., Montani, A. and Nerozzi, F., 2006. Limited area ensemble forecasting: the COSMO model. Predictability of Weather and Climate. *Cambridge University Press*, 489-513.

Toth, Z., & Kalnay, E., 1993. Ensemble forecasting at NMC: the generation of perturbations. *Bull. Am. Met. Soc.*, **74**, 2317-2330.

Tracton, M. S., & Kalnay, E., 1993. Operational ensemble prediction at the National Meteorological Centre: pratical aspects. *Weather and Forecasting*, **8**, 379-398.

Von Neumann, J., & Richtmeyer, R. D., 1950. A method for the numerical calculation of hydrodynamical shocks. *J. Appl. Phys.*, **21**, 232.

Weigel, A. P., M. A. Liniger, and C. Appenzeller, 2007. The discrete Brier and ranked probability skill scores. *Mon. Wea. Rev.*, **135**, 118-124.

Wieker and Skamarock, 2002. Time-splitting methods for elastic models using forward time schemes. *Mon. Wea. Rev.*, **130**, 2088-2097.

Wilks, 1995. Statistical methods in the atmospheric sciences. *International Geophysics Series, Academic Press*, **59**.

Zhu, Y., 2005: Ensemble forecast: A new approach to uncertainty and predictability, *Advance in Atmospheric Sciences*, **22**, No. 6, 781-788.

Acnowledgements

Al termine di questa tesi trovo doveroso spendere ancora un po' di inchiostro per alcuni ringraziamenti. Un grazie va anzittutto alla Prof.ssa Silvana Di Sabatino che ha accettato con entusiamo l'invito ad essere mia relatrice, guidandomi con i suoi consigli in questi mesi di lavoro. Un grazie particolare va al Dott. Andrea Montani e alla Dott.ssa Tiziana Paccagnella che durante questo periodo con grande pazienza non si sono mai stancati di insegnarmi cose nuove; da loro ho appreso moltissimo e mi scuso se non sono stato secondo desiderio. Un sincero graziamento rivolgo anche alla Dott.ssa Valentina Pavan, mia compagna di stanza...ehm sgabuzzino, per avermi sopportato in questo periodo. Un pensiero va anche a Francesca che si é sacrificata con il mio "bad english".

Questa tesi la voglio dedicare a nonno Umberto e nonna Nerina che sono venuti a mancare nel corso di questi anni, lasciando un grande vuoto, pur pensandoli sempre al mio fianco, con me in ogni momento. Uno speciale ringraziamento va ai miei genitori, senza di loro non avrei mai tagliato questo traguardo. Durante i difficili anni di studio mi hanno sostenuto moralmente ed economicamente, spronandomi ad andare avanti anche nei momenti piú bui e difficili. Grazie a nonna Anna, anche lei non mi ha fatto mai mancare il suo supporto (specie quello alimentare!). Ringrazio anche mia sorella Valentina; dopo 13 anni di scuola dove non ha avuto da fare niente o quasi, con gli anni dell'universitá é diventata per me un esempio vivente di quanto possa essere nocivo uno studio eccessivo. Dopotutto, nonostante il suo impegno sui libri non conosca alba né tramonto, in questi anni a Bologna mi ha sempre aiutato e ben consigliato...ringrazio Tommaso che la tiene calma, altrimenti non l'avrei retta (in particolare in quest'ultimo anno vissuto assieme in pochi metri quadrati).

Vorrei dire grazie di cuore a tutti gli amici, senza i quali le giornate trascorrerebbero meno spensieratamente e con loro, ieri come oggi, ho condiviso molto. Purtroppo mi tocca ricordarli per gruppi, con annessi e connessi, altrimenti se dovessi scrivere qualcosa per ognuno... sarebbero piú le pagine dei ringraziamenti che quelle della tesi. Quindi ringrazio il magico terzetto (con me quartetto) dei mediocri, unito in tante avventure, forse meglio dire dis-avventure! Ricordo il gruppo della 5A, classe antropologicamente molto eterogenea che spaziava dalle timide ragazze acqua e sapone alle ammalianti megere..."gli amici di Girelli", tuttavia, sono sopravvissuti alla selezione del tempo. Poi non posso che ringraziare infinitamente tutti gli amici che, come me, odorano di cloro: sono i ragazzi della Vela Nuoto Ancona e della De Akker Team, oltre a tutto il gruppo del GUG che da 11 anni ritrovo tutte le domenica (e non solo) in vasca.....GRAZIE A TUTTI!

Le serate, trascorse a fare avanti e indietro lungo la riga nera, a volte servono anche per pensare. In una di queste sere mi sono chiesto a che cosa avrei potuto paragonare questi 10 anni lunghi e difficili. Forse ció che rende meglio l'idea é una lunga traversata in barca. Partito con il favore dei venti, mi sono fin da subito imbattuto in molte tempeste, piú volte la barca ha rischiato di affondare, piú volte mi sono domandato se abbandonarla. Poi, peró, ho intravisto l'altra sponda del mare e, non senza vento contrario, la luce del faro mi ha guidato ad un porto sicuro.