Modified shallow water models and idealised satellite data assimilation

Luca Cantarello

The University of Leeds

School of Mathematics

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

June 2021

I confirm that the work submitted is my own, except where work which has formed part of jointly authored publications has been included. My contribution and that of the other authors to this work has been explicitly indicated below. I confirm that appropriate credit has been given within the thesis where reference has been made to the work of others.

Chapter 2 Most of the content of Chapter 2 can also be also found in a paper titled "*Idealized forecast-assimilation experiments for convective-scale Numerical Weather Prediction*" co-authored by Dr Tom Kent (TK), Dr Gordon Inverarity (GI, Met Office), Professor Onno Bokhove (OB, University of Leeds), Professor Steve Tobias (ST, University of Leeds) and me, and currently available as a preprint on EarthArXiv. Table 2.2 (Table 1 in the preprint) has been created by TK, who had previously developed the model (section §2 in the paper) and the initial data assimilation scheme (cf. Kent (2016); Kent et al. (2017)). Starting from TK's work, GI suggested some modifications in order to improve the data assimilation scheme and we worked together on coding the algorithm in its current form (described in section §3 of the paper). TK and I worked together on arranging and assessing the idealised experiments (section §4 of the paper), while the numerical simulations have been run by me. Figures 3, 5, 6, 7, 8, 9, 10 and 11 of the paper appear also in this thesis and have all been generated by me. The other authors provided precious guidance, support and advice.

Chapter 3 Most of the content of Chapter 3 can also be found in two articles titled "*An idealized* $1\frac{1}{2}$*-layer isentropic model with convection and precipitation for satellite data assimilation research. Part I: model dynamics*" and "*An idealized* $1\frac{1}{2}$*-layer isentropic model with convection and precipitation for satellite data assimilation research. Part II: model derivation*", both available as preprints on EarthArXiv. These papers have been co-authored by OB, ST and me, have been submitted to the Journal of the Atmospheric Sciences on January 26th 2021 and are currently under review. The content of part I is entirely the product of research conducted by me during the PhD, with the other two authors providing guidance, support and advice. My contribution to part II is limited to the observations scaling (section §4 in the paper, here discussed in section §3.3.4), the appendices, Figures 1 and 3 and Table 1. OB conducted the mathematical derivation of the model (section 3 and 5 of the paper, not included in this thesis), which I have checked thoroughly and contributed to make its presentation clearer. ST provided guidance, support and advice.

To my family.

Acknowledgements

I want to thank all my supervisors (Professor Onno Bokhove, Professor Steve Tobias, Dr Gordon Inverarity and Dr Stefano Migliorini) for their guidance, help and support throughout the PhD. This thesis has substantially improved as a result of their feedback and suggestions.

I also want to thank the two organisations that have funded this work: the Natural Environment Research Council and the Met Office, as well as the the Research Computing Team at the University of Leeds for providing assistance while running my numerical experiments on ARC3 and ARC4.

Most of the work presented in Chapter 2 has been carried out in collaboration with Tom Kent, and my thanks go to him too. Tom has been the first person I met at the University of Leeds and his assistance was incredibly helpful at the start of my PhD, while I was familiarising with the model and the data assimilation scheme that he had developed. He conceived the idea of summarising the results of the chapter as it is done in Table 2.2 and together we worked on structuring the tuning process and assessing the results of the simulations.

I am grateful to Deputy Chief Meteorologist Nicholas Silkstone of the Met Office for his suggestion regarding the Low Level Jet flow that guided the scaling of the isentropic model.

I also want to say thanks to all the people in the School of Maths who have made Leeds feel like ~~home~~ an Italian enclave in the last four years: Andrea, Celeste, Dario, Francesco, Gabriele, Giovanni, Leonardo, Lorenzo[1], Matteo e Rosario, and to all the other PhD students, researchers and friends who contributed to make these past four years more pleasant. In this regard, a special mention goes to Liam and Robyn.

Lastly and most importantly, I don't think I would have coped with the stress of the writing up and the gloom of the pandemic without Reece. Our daily routine made of afternoon walks and PS4 nights which we developed in the course of the last year has really kept me going.

---

[1] Honorary member of the School of Maths

Abstract

Satellites constitute an essential source of observations in operational satellite data assimilation (DA). In this thesis, we investigate the impact of assimilating satellite observations at different spatial scales: is there a relative benefit in focussing on small rather than large scales (or vice versa)? In order to address this question without using complex and computationally expensive Numerical Weather Prediction (NWP) models, we conduct a series of idealised satellite DA experiments based on a modified shallow water model able to imitate convection and precipitation.

The use of an isopycnal, single-layer version of the model (modRSW) is discussed first. A series of forecast-assimilation experiments are carried out using a Deterministic Ensemble Kalman filter (DEnKF). As a result, the filter performance and the relevance of the modRSW model for convective-scale DA in Numerical Weather Prediction systems are demonstrated and a protocol to extend a similar analysis to other idealised systems is presented.

After establishing that the modRSW model is not suitable for satellite DA research, a new isentropic, $1\frac{1}{2}$-layer model (ismodRSW) is developed. The revised model is equipped with a fluid temperature definition and is therefore a better candidate for satellite DA experiments. The dynamics and the numerics of this model are discussed, and its numerical solver is verified against an analytical solution.

In order to imitate closely an operational system, an idealised observing system comprising both ground and satellite observations is created, and pseudo observations mimicking microwave radiation measured by polar-orbiting satellites are generated, with clouds and precipitation implicitly taken into account within the new (and non-linear) observation operator.

Finally, a new series of forecast-assimilation simulations is run to obtain a well-tuned system which is used as a reference in a series of data denial experiments, where satellite observations at small and large scales are selectively excluded from the assimilation to evaluate their impact on the system. Preliminary results show a degradation of both the analysis and the forecasts when only large-scale satellite observations are utilised, although further work is needed to ascertain the robustness of these findings.

All in all, this thesis shows that the idea of investigating satellite DA using a modified shallow water model is a viable strategy. By imitating closely several aspects of an operational system and by developing a more realistic model, we have demonstrated that large-scale satellite observations alone can have a negative impact on the quality of a DA system.

Abbreviations

AMV	Atmospheric Motion Vector
CRPS	Continuous Ranked Probabilty Score
DA	Data Assimilation
DEnKF	Deterministic ENsemble Kalman Filter
ECMWF	European Centre for Medium-range Weather Forecasts
EM	ElectroMagnetic
EnKF	ENsemble Kalman Filter
FOV	Field Of View
ismodRSW	ISentropic MODified Rotating Shallow Water
LW	LongWave
modRSW	MODified Rotating Shallow Water
NCEP	National Centers for Environmental Prediction
NWP	Numerical Weather Prediction
OI	Observation Influence
RMSE	Root Mean Square Error
RO	Radio Occultation
RT	Radiative Transfer
RTPP	Relaxation To Prior Perturbation
RTPS	Relaxation To Prior Spread
SPR	ensemble SPRead
SW	ShortWave
WMO	World Meteorological Organization
3D-Var	3 Dimensional VARiational data assimilation
4D-Var	4 Dimensional VARiational data assimilation

Contents

List of Figures

Chapter 1

Introduction

> Thou shalt not worship the radiosonde
>
> ---
>
> Suomi's 11th commandment

1.1 Background and motivation

On April 1st 1960, 9 seconds after 6:40 am EST, a National Aeronautics and Space Administration (NASA) rocket called *Thor-Able* was launched from Cape Canaveral, Florida (United States). Thor-Able carried on board the first-to-be weather satellite: the so-called Television and Infra-Red Observation Satellite (or TIROS-1), which entered into orbit the same day and started to return images of the Earth straight away. Since then, hundreds more missions have been carried out, operated by several national and supranational agencies, resulting in a remarkable number of satellites sent into space with the purpose of studying the Earth's atmosphere, as well as gathering more data to improve the quality of weather forecasting. The European Space Agency (ESA), for instance, launched its first satellite (called Meteosat) on November 23rd, 1977. The Japanese Space Agency had done the same with the first Himawari satellite (also known as GMS-1) in the April of the same year. China followed shortly after by sending into orbit its first Feng-Yun satellite (FY-1A) in September 1988. Currently, the website of the World Meteorological Organization (WMO) lists a total of 111 operational satellites contributing to its Integrated Global Observing System (WIGOS)[1]. The advent of satellites transformed many aspects of atmospheric sciences and meteorology. For the first time, it was possible to monitor from the skies many atmospheric phenomena, such as fronts and cyclones, which had previously been observed only from the surface. Their use also contributed to the study of the properties of clouds and precipitation, and helped to expand the real-time coverage of the weather across the planet over previously unobserved areas such as the poles and the oceans, where conventional observations were (and still are) difficult and rare.

The change brought about by satellites has been so radical that today it is difficult to imagine generating a weather forecast without their contribution. Nevertheless, the first operational forecast based on a Numerical Weather Prediction (NWP) model was issued on May 6th, 1955 by the Joint Numerical Weather Prediction Unit (JNWPU) in the United States (Harper et al., 2007), that is, five years before the first meteorological satellite was sent into orbit.

Almost seventy years later, weather forecasting can be rightly seen as one of the main scientific achievements of the Twentieth century. In a review paper published a few years ago, Bauer et al. (2015) called its success a "*quiet revolution*", as the progress made in this field was marked by small, incremental technical improvements and the gradual understanding of the underlying physical processes rather than by a series of major scientific developments. Sure enough, the inclusion of satellite observations and the gain in knowledge on how to make best use of them was part of such advances. Eventually, the emergence of increasingly complex and sophisticated NWP models has led to the ability to predict the main atmospheric variables at synoptic scales several days in advance with good accuracy (Bauer et al., 2015; Zhou et al., 2017).

Running an NWP model amounts to solving numerically a set of differential equations representing the evolution in time of various atmospheric processes. In this sense, the problem of forecasting the weather is an initial value problem and requires an initial condition representing as closely as possible the real atmospheric conditions at the validity time in which the forecast starts.

The domain of the atmospheric sciences dealing with the production of such initial conditions is called *data assimilation* (DA). Data assimilation comprises a series of

---

[1] Webpage url: `https://community.wmo.int/activity-areas/wmo-space-programme-wsp/satellite-status`. Last access on 30/03/2021.

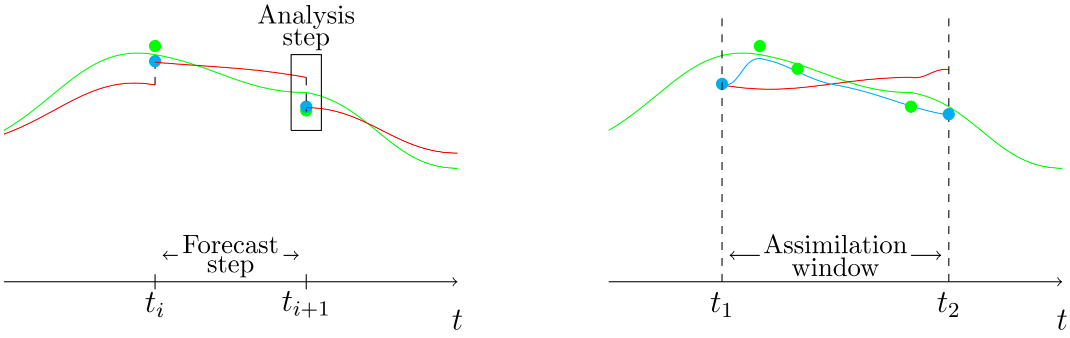


Figure 1.1: Schematic representation of a sequential (left) and a variational (right) DA scheme. In a sequential scheme, the forecast trajectory (red lines) is regularly adjusted towards the observations (green dots) at analysis times $\{t_i, t_{i+1}, ...\}$ to generate an analysis state (cyan dots) which is used as initial condition for the subsequent forecast. In a variational scheme, an analysis trajectory (cyan line) results from the minimisation of a cost function which optimises the distance from the observations (green dots) and the previous forecast (red line) over a certain assimilation window $[t_1, t_2]$; the value of the analysis at the end of the assimilation window (cyan dots) becomes the initial condition for the subsequent forecast. In both figures, the green line denotes the 'true' evolution of the atmosphere.

mathematical and statistical techniques which are used to continuously adjust recent model data with up-to-date observations in order to obtain a more accurate and precise description of the ongoing atmospheric conditions. The literature on this topic is vast, and many books cover it in detail, such as Kalnay (2003) and Reich & Cotter (2015). Among many review articles, Carrassi et al. (2018) offer an up-to-date summary on its uses and future perspectives in the geosciences.

One of the most interesting properties of the atmosphere is its intrinsic chaotic behaviour. In the context of deterministic systems, chaos is defined as a strong sensitivity to the initial conditions (Lorenz, 1963). In this sense, the ability to predict the weather in the future depends on the knowledge of the current conditions. As a result, the role of DA in generating accurate initial conditions for NWP models is particularly important and plays a crucial role in determining the accuracy of a forecast.

There exist several types of DA schemes, which can be grouped into four categories: (i) sequential, (ii) variational, (iii) hybrid and (iv) particle filters. Most of these algorithms are based on a *Bayesian* probabilistic framework, in which starting from a 'first-guess' distribution associated with the prior forecast model state (also called the *background*), and by using the observations available, a posterior distribution can be estimated, leading to the derivation of a new model state called an *analysis*. In addition, most of them (with the exception of particle filters) assume Gaussian error distributions for both the observations and the background, which allows us to formulate the DA equations only in terms of the mean state and the error covariance matrices. A brief description of each DA category mentioned above follows:

(i) Sequential schemes alternate a *forecast step* during which a model is integrated forward in time, and an *analysis step* in which the observations are assimilated and a new initial condition is generated and used to launch the subsequent forecast step (see Figure 1.1, left image). This method requires the observations to be used sequentially in time and therefore prescribes that they are assimilated only at a single time, which becomes the analysis validity time. Examples include the Optimal Interpolation (OI) method (Kalnay, 2003) and the Kalman Filter (Kalman, 1960; Kalman & Bucy, 1961), with the latter allowing a flow-dependent background error which is recomputed at every analysis step. In recent years, sequential schemes have moved significantly towards an *ensemble* approach, in which multiple model simulations are run in parallel during the forecast step, contributing to improving the representation of the forecast uncertainty. The development of the Ensemble Kalman Filter (EnKF, Evensen (1994)) was one of the first attempts made in this direction.

(ii) Variational schemes (Le Dimet & Talagrand, 1986; Talagrand & Courtier, 1987) are based on the minimisation of a cost function defined as the sum of various terms accounting for the deviation of the observations and the background from an unknown atmospheric state to be varied in the minimisation process. The model state that minimises these deviations becomes the analysis and therefore the initial condition for the subsequent forecast. A common approach based on the linear approximation of the observation penalty term leads to the so-called *incremental approach* (Courtier et al., 1994), in which the cost function is expressed in terms of the perturbation with respect to a known reference state (Veersé & Thépaut, 1998). Variational algorithms are based on the same assumptions regarding Gaussian errors as the sequential schemes, and generally assume a static background-error covariance matrix (in fact, the Kalman smoother equations and the variational cost function are equivalent, as shown in Lorenc (1986)). They can assimilate all observations comprised within an hours-long assimilation window (cf. Figure 1.1, right image) and also require the formulation of both tangent-linear operators (cf. §6.3 in Kalnay (2003)) and adjoint models (Errico, 1997). The two most common variational schemes are called 4D-Var (four-dimensional variational data assimilation) and 3D-Var (three-dimensional variational data assimilation), which differ in the way they handle the time dimension, with pure 3D-Var assuming that the observations are all valid at the same time and 3D-Var FGAT (First Guess at Appropriate Time) assuming that the observations and the previously calculated background trajectory are valid at the same time.

(iii) Hybrid schemes are in essence variational algorithms in which the ensemble approach is adopted to compute – totally or partially – the (otherwise static) background-error covariance matrix. Hybrid methods are currently in use at a number of operational weather centres (cf. Table 5 in Bannister (2017)).

(iv) Particle filters have been formulated recently (Van Leeuwen, 2009), and represent an attempt to formulate a DA technique without assuming Gaussian-distributed errors. Presently, they are not considered in any operational settings.

Bannister (2017) provides a comprehensive review of the algorithms in use at various operational weather centres.

Data assimilation strongly benefited from the advent of satellite observations, as they significantly expanded the data coverage around the planet, especially in poorly observed regions where in-situ observations were harder and more expensive to collect. This expansion led to a significant improvement in the forecast skill, particularly in the Southern Hemisphere, where for logistical reasons (i.e. a smaller population density and a high proportion of the surface occupied by water) the conventional observations were scarce (Rabier, 2005; Simmons & Hollingsworth, 2002).

With time, satellites have become the most prominent source of observations in global NWP DA in terms of both volume of data assimilated and impact on the final analysis (Kelly & Thépaut, 2007). This progress took several decades of research and was accompanied by a radical rethink of how satellite observations had to be treated. The initial approach (common in the 1980s) of treating satellite data as de-facto radiosonde measurements, and converting them into vertical profiles of temperature and humidity was later replaced by the direct assimilation of radiance, which was also made possible by a gradual shift towards the use of variational schemes in the late 1990s.

Currently, one of the main challenges is the assimilation of cloudy scenes. In fact, only a small portion of all available satellite data is routinely assimilated, and for many years most of the satellite observations contaminated by clouds and precipitation have been discarded because of the difficulties in handling the non-linearities arising from their assimilation. In practice, only satellite observations in the presence of *clear-sky* conditions used to be retained. More recently, a lot of effort has been put into so-called *all-sky* DA, that is, the assimilation of satellite data regardless of the local weather conditions (Geer et al., 2017, 2018; Migliorini & Candy, 2019). This progress is expected to have a further positive impact on the quality of weather forecasting, as the areas covered by clouds and precipitation are often those dynamically more interesting to assimilate (McNally, 2002). At present, there are already a few operational centres which have started to implement their all-sky assimilation methods, while the research in this field is still ongoing.

Satellite DA is also confronted by an ongoing expansion in the number of observations, due to the continuous addition of new instruments and the increase in both their spectral and spatial resolutions. These developments make the research into how to optimise the use of satellite observations all the more important. In this regard, one possible approach consists of applying data compression techniques that can thin the total amount of observations by taking into account their error correlations (Fowler, 2019).

In an effort to investigate more efficient ways to assimilate an expanding number of observations, in this thesis we examine the potential benefit of focussing on the assimilation of satellite observations at small rather than large spatial scales (and vice versa). In particular, we will assess the response of a forecast-assimilation system to the variation in spatial resolutions (or Field of View) of a set of synthetic satellite observations. Understanding the effect of prioritising the assimilation of satellite observations at a specific spatial scale can help orientate future strategies and decisions in the field, with more resources directed towards what is deemed most beneficial for an operational DA system.

Moreover, in an attempt to make the work presented here as relevant as possible in the context of the current research interests in satellite DA, we plan to include the effect of clouds and therefore to conduct all-sky assimilation experiments.

Unfortunately, the degree of complexity and the computational cost typical of state-of-the-art NWP models can hamper or slow down the research in DA, especially when the resources are limited. Indeed, operational NWP DA systems are built to combine routinely very large amounts of data: the number of degrees of freedom in a global weather model (i.e., the number of model variables times the number of grid points used for its numerical discretisation) can easily reach orders of magnitude of $\mathcal{O}(10^9)$, while the number of observations assimilated can be as big as $\mathcal{O}(10^7)$.
Therefore, simplified or even idealised versions of the state-of-the-art models have been developed and utilised as alternative tools. These solutions include:

- Models of intermediate complexity, based on some widely known simplifications of the equations governing the atmosphere, such as the quasi-geostrophic (QG) approximation. These models represent a compromise that retains a connection with the equations governing the atmosphere and at the same time reduce considerably the computational cost. For example, a QG model was used in Houtekamer & Mitchell (1998) to investigate the use of the Ensemble Kalman filter, and in Beck & Ehrendorfer (2005) to study the impact of a flow-dependent background-error covariance matrix in a 4DVar system. Other intermediate complexity models developed for DA purposes are: the Atmospheric Model of Intermediate Complexity (AMIC, Ehrendorfer & Errico (2008)), the Moist Atmosphere Dynamics Data Assimilation Model (MADDAM, Zaplotnik et al. (2018)), and the non-hydrostatic ABC model (Bannister, 2020; Petrie et al., 2017).

- Idealised models, based on simple sets of differential equations representing only basic processes related to the atmosphere. A common example is represented by the simplified models developed by Lorenz (e.g., Lorenz (1963) and Lorenz (1995)), which have been used in many DA studies (see, for example, Goodliff et al. (2015); Ott et al. (2004)). A different approach envisages the use of shallow water models, which have been utilised in DA research for both the atmosphere and the ocean in a series of studies, i.e. Žagar et al. (2004), Salman et al. (2006), Stewart et al. (2013) and Würsch & Craig (2014). With respect to the Lorenz models, they have the advantage of being based on actual fluid-dynamic equations. Lastly, there are single-column models, in which the atmosphere is approximated as a single, vertical column of fluid. These models have been used to investigate the role of clouds in ensemble DA (Vetra-Carvalho et al., 2011), to explore the role of satellites in variational DA (Rudd et al., 2012), and to study coupled 4DVar DA in an atmosphere-ocean model (Smith et al., 2015).

Key to the success of an idealised model is the degree to which it is able to imitate the defining characteristics of the problem at hand. As a result, a well-constructed idealised system will be able to display some of the same underlying features of the real one, therefore allowing the user to address the scientific question without having to solve or rely on the full-scale problem.

This thesis deals with two idealised models based on modified shallow water equations developed to support DA research. The first one was developed in a previous work by Tom Kent (Kent, 2016; Kent et al., 2017) and here we show its relevance in the context of convective-scale NWP DA. The second one is an upgraded version which will allow us to conduct idealised satellite DA experiments. Finally, in this thesis we focus only on sequential schemes and, in particular, we adopt the Deterministic Ensemble Kalman Filter (DEnKF) developed by Sakov & Oke (2008).

1.2 Objectives

The purpose of this thesis is to investigate the impact of satellite observations at different spatial scales on a DA system by using an idealised model of the atmosphere. In particular, we are interested in understanding whether there is a relative benefit in assimilating small scale observations over large scale ones, or vice versa.

In order to address this question, this work starts from an idealised model of the atmosphere based on modified rotating shallow water equations (modRSW), able to imitate convection and precipitation, which was presented in Kent (2016); Kent et al. (2017). Here, we show that it is possible to conduct a series of forecast-assimilation experiments with the modRSW model which are both well-tuned and relevant in the context of operational NWP models. Since the relevance of idealised models for NWP DA research is often overlooked in the literature, we present our results in the form of a protocol that can be re-applied to other idealised systems in the future.

Furthermore, we discuss the limitations of the modRSW model for satellite DA research and develop a new, revised version of the same model (in brief: ismodRSW) which is considered fit for purpose. We also present our new observing system, including our strategy to generate pseudo satellite observations, with the purpose of assimilating them in a series of new forecast-assimilation experiments. Finally, we aim to understand whether there exists a benefit in focussing on the assimilation of satellite observations at small, rather than large spatial scale. To this end, we run a series of data denial experiments, in which the scale of the observations is defined by their horizontal resolution (or field of view). Thus, by excluding small and large scale satellite observations in turn from the assimilation, it is possible to evaluate their impact on the analysis.

Overall, the objectives of this thesis can be summarised as follows:

1. Show that it is possible to obtain a well-tuned idealised experiment with the modRSW model which is also relevant for operational NWP DA research and formulate a protocol to assess other idealised systems in the same way;
2. Show that the modRSW model needs to be modified for satellite DA research, and then develop a new, revised version of the model that is fit for purpose;
3. Construct a complex (and idealised) observing system, comprising both simulated ground and satellite observations, in which the characteristics of real-world observing systems and satellite observations are imitated closely, and include the effect of clouds;
4. Conduct new forecast-assimilation experiments in which satellite observations are assimilated that are relevant for operational NWP systems;
5. Perform a series of data denial experiments – i.e. simulations in which a portion of the observations are intentionally excluded from the assimilation – to investigate the impact of satellite observations at different spatial scales.

1.3 Thesis outline

The thesis is structured as follows. Chapter 2 briefly summarises the dynamics and the numerics of the modRSW model and describes the setup and the DA scheme used in the forecast-assimilation experiments realised with the same model; moreover, it introduces a protocol to assess the performance and the relevance of an idealised system in the context of operational NWP DA and concludes with an evaluation of the experiments conducted with the modRSW model (objective 1). Chapter 3 shows why the modRSW model is unsuitable to conduct satellite DA experiments and presents the dynamics and the numerics of a revised, isentropic model (ismodRSW), together with a verification against an analytical solution and a prototype of a nature run simulation for anticipated DA experiments (objective 2). Chapter 4 presents an in-depth review of basic radiative transfer concepts, the functioning of meteorological satellites and the history of satellite DA, which is followed by a description of how all-sky pseudo satellite observations are imitated in our idealised system (objective 3). Chapter 5 describes the modifications made to the DA scheme already outlined in Chapter 2 so that it is possible to assimilate the pseudo satellite observations (objective 3); given the new configuration and the nature run introduced in Chapter 3, it then illustrates the process to obtain a well-tuned experiment and proves its relevance in the context of convective-scale NWP systems (objective 4); lastly, it shows and discusses the results of a series of data denial experiments to evaluate the impact of satellite observations at various spatial scales (objective 5). Chapter 6 reports the conclusions and discusses possible future work.

Chapter 2

The relevance of the modRSW model for NWP data assimilation

Here we discuss the use of an idealised model of the atmosphere based on modified shallow water equations to conduct a series of data assimilation experiments. In particular, we will describe the so-called 'twin-setting' configuration that is used to generate a set of pseudo observations which are subsequently combine with an ensemble of forecasts using a Deterministic Ensemble Kalman filter (DEnKF). In the final part of the chapter, we will evaluate its relevance for NWP DA research and, in doing that, we present a protocol that can be generalised to other idealised models.
The data assimilation scheme and the twin-setting configuration outlined in this chapter will be replicated in the experiments involving the assimilation of satellite observations in Chapter 5, apart from a few modifications described therein.

2.1 The *modRSW* model: dynamics and numerics

Previous work done by Kent (2016); Kent et al. (2017) led to the development of an idealised model aimed at investigating convective-scale data assimilation research. This model was based on modified rotating shallow water equations (hence the acronym *modRSW*) and was an improved modification of the simpler cumulus convection model described in Würsch & Craig (2014).

Here, we present a brief summary of the dynamics and numerical implementation of the modRSW model. Later in the chapter, we will discuss in detail its use in a series of data assimilation experiments (cf. section §2.4-2.5) thus proving its relevance for NWP data assimilation. Similar considerations can be found in Kent et al. (2020).

2.1.1 Model equations

The equations of the *modRSW* model are as follows:

$$\partial_t h + \partial_x(hu) = 0, \tag{2.1.a}$$

$$\partial_t(hu) + \partial_x(hu^2 + P) + hc_0^2\partial_x r - fhv = -Q\partial_x b, \tag{2.1.b}$$

$$\partial_t(hv) + \partial_x(huv) + fhu = 0, \tag{2.1.c}$$

$$\partial_t(hr) + \partial_x(hur) + \alpha hr + h\tilde{\beta}\partial_x u = 0; \tag{2.1.d}$$

in which: x and t are the space and the time coordinates respectively, h is the fluid depth, u and v are the zonal and meridional velocities, and r is a 'rain mass fraction' (a proxy for precipitation). All variables are defined on a one-dimensional domain, whereas the meridional components of the spatial derivative are dropped (i.e. $\partial_y(\cdot) = 0$). Other parameters in (2.1) include the Coriolis frequency f and the bottom topography b.

The model (2.1) comprises two fluid depth thresholds which act like switches and modify the model dynamics in order to generate convection and precipitation. In particular, the definition of P, Q and $\tilde{\beta}$ in (2.1) depend on the convection (H_c) and rain (H_r) thresholds (as well as convergent flow). Their expressions read:

$$P(h;b) = \begin{cases} \frac{1}{2}g(H_c - b)^2 & \text{if } h + b > H_c, \\ \frac{1}{2}gh^2 & \text{otherwise;} \end{cases} \tag{2.2.a}$$

$$Q(h;b) = \begin{cases} g(H_c - b) & \text{if } h + b > H_c, \\ gh & \text{otherwise;} \end{cases} \tag{2.2.b}$$

$$\tilde{\beta} = \begin{cases} \beta & \text{if } h + b > H_r \text{ and } \partial_x u < 0, \\ 0 & \text{otherwise;} \end{cases} \tag{2.3}$$

in which g indicates the acceleration due to gravity. The remaining model parameters c_0^2 [m^{2}s^{-2}], β (dimensionless) and α [s^{-1}] in (2.1) control: the intensity at which convection is suppressed in areas where rain is present, the rate at which rain is produced, and the velocity with which rain is removed from the system, respectively.

From equations (2.2)-(2.3), one can note that whenever $h+b<H_c<H_r$ and $r=0$, the system (2.1) reduces to the classic rotating shallow water equations.
The system (2.1) can be non-dimensionalised by introducing horizontal and vertical length scales L_0 and H, the scale velocity V_0 and the corresponding time scale L_0/V_0. By using this scaling, the model variables, parameters and coordinates can be non-dimensionalised as follows:

$$t=\widetilde{t}\frac{L_0}{V_0}, \quad x=\widetilde{x}L_0, \quad (u,v)=(\widetilde{u},\widetilde{v})V_0, \quad h=\widetilde{h}H, \quad \alpha=\widetilde{\alpha}\frac{V_0}{L_0}, \quad c_0^2=\widetilde{c}_0^2V_0^2. \tag{2.4}$$

After substituting (2.4) into (2.1) and some reordering, and once the tildes on x, t, u and v are dropped, one obtains the scaled modRSW system:

$$\partial_t h+\partial_x(hu)=0, \tag{2.5.a}$$
$$\partial_t(hu)+\partial_x(hu^2+P)+h\widetilde{c}_0^2\partial_x r-\frac{1}{\mathrm{Ro}}hv=-Q\partial_x b, \tag{2.5.b}$$
$$\partial_t(hv)+\partial_x(huv)+\frac{1}{\mathrm{Ro}}hu=0, \tag{2.5.c}$$
$$\partial_t(hr)+\partial_x(hur)+\widetilde{\alpha}hr+h\widetilde{\beta}\partial_x u=0; \tag{2.5.d}$$

in which the parameters $\widetilde{\alpha}$ and $\widetilde{c}_0^2$ are now non-dimensional, $\mathrm{Ro}=V_0/fL_0$ indicates the Rossby number, and the gravitational acceleration g within P gets redefined as $g=1/\mathrm{Fr}^2$, with Fr being the Froude number:

$$\mathrm{Fr}=\frac{V_0}{\sqrt{gH}}. \tag{2.6}$$

2.1.2 Model dynamics

In this section we summarise the dynamical mechanisms associated with the onset of convection and the production of precipitation in the *modRSW* model. We refer to section §2.2 of Kent (2016) for a more detailed description.

Onset of convection ($H_c<h+b<H_r$)

When the total fluid depth $h+b$ overcomes the convection threshold H_c, the effective pressure P is maintained at a constant value (cf. eq. (2.2a)). This modification implies that, above H_c, the fluid experiences a lower pressure value than the one it would normally have. Ultimately, this relative reduction in the effective pressure causes the fluid to rise.

From a meteorological perspective, the threshold H_c can be interpreted in the context of cumulus convection as the Level of Free Convection (LFC), which is defined as the height at which an air parcel needs to be lifted for it to become unstable and naturally more buoyant (i.e. less dense) then the surrounding environment.

Generation of precipitation ($H_c < H_r < h + b$ and $\partial_x u < 0$)

Model rain r is generated upon satisfying of two requirements: the depth of the fluid column $h + b$ overcoming the convection threshold H_r and a positive local wind convergence given by $\partial_x u < 0$. These two conditions lead to the activation of the source term $h\tilde{\beta}\partial_x u$ in the model equation (2.1d) via (2.3).
This mechanism can be interpreted in physical terms by considering that rain is generally produced at a later stage with respect to the onset of convection (hence $H_r > H_c$) and is often associated with areas of wind and moisture convergence. In order to prevent the convection from growing indefinitely in the modRSW model, the momentum equation (2.1b) is coupled to the rain equation (2.1d) via the term $hc_0^2\partial_x r$. As highlighted in Kent (2016); Kent et al. (2017), this term counteracts the fluid rise generated by the constant effective pressure P in a way that is comparable to the suppression of the updraft in a cloud when precipitation starts. Eventually, rain is removed from the system via the sink term αhr in eq. (2.1d).

2.1.3 Hyperbolicity and numerical implementation

Shallow water models are hyperbolic systems of partial differential equations (PDEs) which can be expressed as:

$$\partial_t \mathbf{U} + \partial_x \mathbf{F}(\mathbf{U}) + \mathbf{T}(\mathbf{U}) = 0. \tag{2.7}$$

The problem of how to deal with them numerically has been addressed extensively in the literature (see, for example, LeVeque (2002)). The *modRSW* model, however, represents a special case, as the addition of terms associated with convection and rain turns it into a non-conservative system in which eq. (2.7) becomes:

$$\partial_t \mathbf{U} + \partial_x \mathbf{F}(\mathbf{U}) + \mathbf{G}(\mathbf{U})\partial_x \mathbf{U} + \mathbf{T}(\mathbf{U}) = 0, \tag{2.8}$$

in which the presence of the non-conservative products (NCP) $\mathbf{G}(\mathbf{U})\partial_x \mathbf{U}$ prevents the expression above from being written in divergence form.

The numerical scheme used in Kent (2016) and Kent et al. (2017) to integrate the modRSW model was the Discontinuous Galerkin Finite Element Method (DGFEM) developed by Rhebergen et al. (2008) at the lowest order (i.e. DG0), in which the NCP terms were dealt with as per the Dal Maso, LeFloch and Murat's theory (DLM; see Dal Maso et al. (1995); Le Floch (1989)). Although this method generally offers a robust solution for non-conservative hyberbolic systems of PDEs, the presence of topography in the *modRSW* model constituted an additional challenge, as the DGFEM does not handle the preservation of the steady states at rest in a satisfactory way (Kent & Bokhove, 2020). To tackle this limitation, the NCP theory was combined with the well-balanced scheme of Audusse et al. (2004), leading to a comprehensive and robust numerical scheme for the *modRSW* model (we refer again to Kent (2016) and Kent et al. (2017) for full details).

2.2 Forecast-assimilation experiments with the modRSW model

In this section we describe the setup used in the forecast-assimilation experiments described in section §2.5. First, we give an illustration of the twin-setting configuration, which is the basis of the data assimilation simulations realised with the modRSW model. Later, we will describe the data assimilation scheme in detail, with a step-by-step description of the algorithm implementation reported in section §2.3.
Some of the parameters mentioned in this section (namely those regarding the observing system and the ensemble size) have been selected at the end of a lengthy tuning process, which for sake of simplicity is not outlined here. The tuning of further parameters related to the data assimilation scheme will be discussed in section §2.5.

2.2.1 The twin-setting configuration

In order to conduct data assimilation experiments using the modRSW model, a twin-setting configuration is adopted. This approach amounts to running the same numerical model to obtain both a *nature run* simulation – treated as the 'true' evolution of the physical system – and the forecasts. In this regard, two methods are possible: a *perfect* model scenario in which the nature run and the forecast are generated with

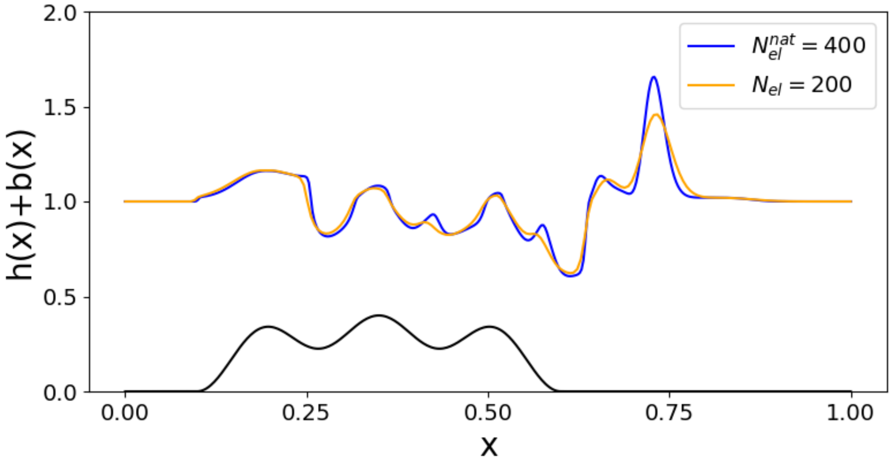


Figure 2.1: Effects of the resolution mismatch between the nature run and the forecasts. The nature run trajectory for h ($N_{el}^{nat} = 400$, blue line) is plotted alongside a single model integration with $N_{el} = 200$ (orange line) after $t = 0.144$. Both are initialised with (2.11). Topography $b(x)$ is in black.

two identical versions of the same model and an *imperfect* model scenario in which they differ in some respect.

In this thesis, we adopt a twin-setting configuration based on an imperfect model scenario, in which we utilise different spatial resolutions for the nature run and the forecasts. In particular, the single, deterministic integration of the modRSW model representing the nature run is carried out at twice the resolution of the forecasts. This resolution mismatch implies that the forecast is unable to resolve some of the features present in the nature run (this can be observed in Figure 2.1), akin to the struggles of real NWP models in representing small-scales and sub-grid phenomena, which are common at convective scales.

The nature run is used to generate a set of *pseudo-observations* which are included in the data assimilation algorithm to adjust the forecast towards the 'true' state of the system. A schematic representation of the twin-setting configuration applied to the modRSW model is shown in Fig. 2.2.

Nature run

We express the nature run solution as a time-dependent $3N_{el}^{nat}$-dimensional vector $\mathbf{x}^t(t)$ which results from a single, numerical integration of the modRSW model (2.1)

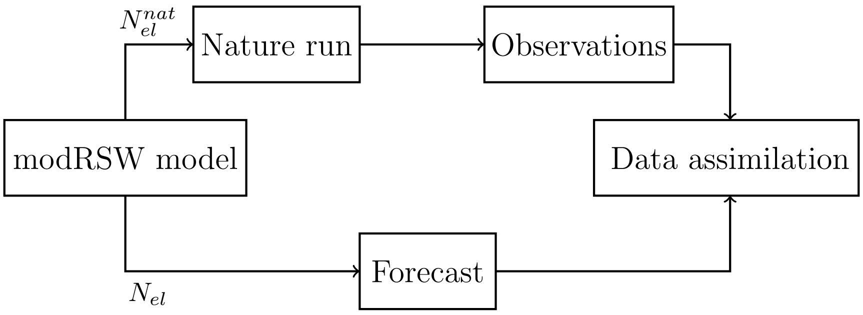


Figure 2.2: Schematic representation of the twin-setting configuration used to conduct forecast-assimilation experiments with the modRSW model. N_{el}^{nat} and N_{el} indicate the number of grid points used for the numerical discretisation of the nature run and the forecast, respectively.

(denoted by the operator $\mathcal{M}$) from time $t - dt$ to t:

$$\mathbf{x}^t(t) = \mathcal{M}(\mathbf{x}^t(t - dt)) \tag{2.9}$$

on a horizontal grid of $N_{el}^{nat} = 400$ points over a periodic domain of length L_0, in the absence of rotation (i.e. $\text{Ro} = \infty$) and under supercritical flow conditions (i.e. $\text{Fr} = 1.1$). The bottom topography $b(x)$ is formed by a series of hills covering part of the domain defined by a series of sinusoidal functions (see the black line in Fig. 2.1)), i.e.:

$$b(x) = \begin{cases} \sum_{i=1}^{3} b_i, & \text{for } x_p < x < x_p + 0.5; \\ 0, & \text{elsewhere}; \end{cases} \tag{2.10.a}$$

$$\text{with } b_i = A_i\big(1 + \cos\big(2\pi(k_i(x - x_p) - 0.5)\big)\big), \tag{2.10.b}$$

in which $x_p = 0.1$, $k = \{2, 4, 6\}$ and $A = \{0.1, 0.05, 0.1\}$. The nature run is initialised with initial conditions:

$$h(x, 0) + b(x, 0) = 1; \quad hu(x, 0) = 1; \quad hv(x, 0) = 0; \quad hr(x, 0) = 0, \tag{2.11}$$

and is integrated for a total of 48 hours, equivalent to $t = 6.912$ non-dimensional model time units. Given the absence of rotation and $hv(x, 0) = 0$ in (2.11), it follows that the meridional momentum equation in (2.1) can be neglected. The convection and rain thresholds are set to $H_c = 1.02$ and $H_r = 1.05$, respectively. A summary of the model parameters used in the nature run simulations is reported in Table 2.1.

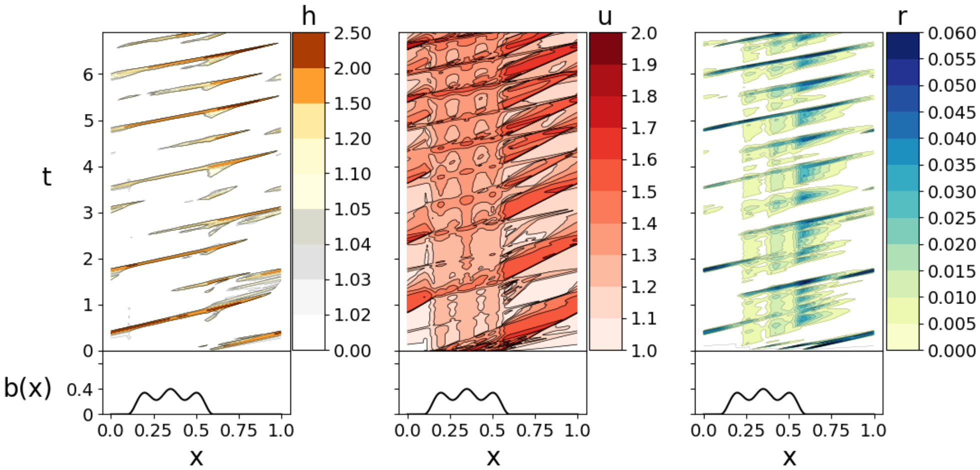


Figure 2.3: Hovmöller plot of each model variable in the nature run simulation: h (left panel), u (central panel) and r (right panel). The corresponding topography $b(x)$ is shown underneath each panel.

The presence of topography (2.10) in a periodic domain, together with the initial horizontal flow, creates the conditions for a continuous production of gravity waves, which are subsequently advected eastward and re-enter the domain on the opposite side. This process results in a self-sustained generation of convection and precipitation and contributes to keeping the simulations dynamically interesting without the need for external forcings. The model dynamics are represented in Fig. 2.3, which displays the Hovmöller plot of each model variable in the nature run simulation. Indeed, most of the convection (gray shades) and potential precipitation ranges (yellow to brown shades) in h (left panel) leading to rain r (right panel) is generated in correspondence with or downstream of the topography (bottom panels), and then travels across the domain until it reappears at $x = 0$. The underlying travelling gravity waves also have an effect on the horizontal velocity u, which experiences a regular alternation of increases and decreases in the second half of the domain, although in a way that seems out-of-phase with (and faster than) the presence of convection and precipitation.
The nature run will be used as the reference 'true' state in section §2.5 to compute various diagnostics, including the root mean squared error ($RMSE$) of the ensemble mean (as defined in section §2.4.2). This possibility figures as one of the advantages of using an idealised configuration instead of an operational NWP model, as the 'true' state of the atmosphere is not available in the real world.

Observations

A set of $p = 28$ pseudo-observations $\mathbf{y}^o$ (also called synthetic observations) are obtained from the $3N_{el}^{nat}$-dimensional nature run $\mathbf{x}^t$ every 60 minutes (i.e. every $t = 0.144$ non-dimensional model time units). To this aim, an observation generator function $\mathcal{G} : \mathbb{R}^{3N_{el}^{nat}} \longrightarrow \mathbb{R}^p$, mapping the nature run at the time t into the p-dimensional observation space can be defined:

$$\mathbf{y}^o(t) = \mathcal{G}(\mathbf{x}^t(t)) + \boldsymbol{\epsilon}^o(t), \tag{2.12}$$

in which $\boldsymbol{\epsilon}^o$ is a p-dimensional vector representing the observation error.
The observing system used in the forecast-assimilation experiments described in section §2.5 is obtained by simply sub-sampling the nature run at fixed, evenly spaced locations along the domain, with the spacing of each observation type $d_{h,u,r}$ reported in Table 2.1. This simplification makes it possible to express the observation generator function as a $(p \times N_{el}^{nat})$ linear operator $\mathbf{G}$ (a sparse matrix of 0 and 1), that is:

$$\mathbf{y}^o(t) = \mathbf{G}\mathbf{x}^t(t) + \boldsymbol{\epsilon}^o. \tag{2.13}$$

The observation error vector $\boldsymbol{\epsilon}^o$ at any assimilation time t is sampled from the distribution:

$$\boldsymbol{\epsilon}^o \sim \mathcal{N}(0, \mathbf{R}), \tag{2.14}$$

in which $\mathbf{R}$ is the static $(p \times p)$ observation error covariance matrix, defined as:

$$\mathbf{R} = \mathrm{diag}(s_h^2\mathbf{I}_h, s_u^2\mathbf{I}_u, s_r^2\mathbf{I}_r), \tag{2.15}$$

where $\mathbf{I}_{h,u,r}$ represents the identity matrix with dimensions equal to the number of h,u and r observations, and $s_{h,u,r}^2$ are the variances associated with the observation errors s_h, s_u and s_r reported in Table 2.1. In other words, the observation errors $\boldsymbol{\epsilon}^o$ are assumed to be Gaussian and unbiased (i.e. zero mean), in line with the assumptions typical of an Ensemble Kalman filter like the one described later, in section §2.2.2. Since the matrix $\mathbf{R}$ in (2.15) is taken to be diagonal, the observation errors are also uncorrelated.

Forecasts

The forecast simulations $\mathbf{x}^f$ are generated by running the modRSW model with the same parameters used in the nature run simulation (cf. Table 2.1), but on a coarser grid of $N_{el} = 200$ points. Taking into account the domain length scale of $L_0 = 500\,\text{km}$, this is equivalent to a model resolution of 2.5 km, a value in line with most operational convection-resolving NWP models currently available (Gustafsson et al., 2018).

At the initial time $t = 0$, the initial condition (2.11) is perturbed in order to obtain an ensemble of $j = 1, ..., N = 18$ initial conditions, defined as:

$$h_j(x,0) = h(x,0) + s_h^{\text{ic}} \mathbf{z}_j \tag{2.16.a}$$

$$hu_j(x,0) = hu(x,0) + s_{hu}^{\text{ic}} \mathbf{z}_j \tag{2.16.b}$$

$$hr_j(x,0) = hr(x,0) + s_{hr}^{\text{ic}} \mathbf{z}_j \tag{2.16.c}$$

in which $\mathbf{z}_j$ is an N_{el}-dimensional vector sampled independently for each variable from $\mathbf{z}_j \sim \mathcal{N}(0, \mathbf{I})$ (with $\mathbf{I}$ being an $(N_{el} \times N_{el})$ identity matrix), and $\mathbf{s}^{\text{ic}} = (s_h^{\text{ic}}, s_{hu}^{\text{ic}}, s_{hr}^{\text{ic}}) = (0.1, 0.05, 0)$ defines the standard deviation of the initial errors.

The value $s_{hr} = 0$ implies that the initial condition $hr(x,0) = 0$ in (2.11) remains unperturbed in all ensemble members (2.16c): this choice prevents the generation of nonphysical negative initial values of the mass rain fraction r. Moreover, the variable r is positively correlated with h via the rain threshold mechanism (cf. §2.1.2) and therefore the perturbation of $h(x,0)$ alone will affect $r(x,t)$ at later times, i.e. $t > 0$. The possibility of spurious, negative values of r and h arising during the application of the data assimilation scheme described in the next section cannot be excluded. Therefore, a non-negativity constraint is imposed on both variables after each analysis step. In practice, any negative values of either h or hr are set to zero (h is reset to $h = 0.001$) after the observations have been assimilated into the system. This is, of course, a very crude approach that has the benefit of solving the problem efficaciously, but does not address other issues such as the conservation of mass. More elegant alternatives have been proposed, such as in Janjić et al. (2014), in which mass conservation principles are incorporated within the data assimilation algorithm.

2.2.2 The Deterministic Ensemble Kalman Filter (DEnKF)

The data assimilation scheme used in this thesis is the Deterministic Ensemble Kalman Filter (DEnKF) developed by Sakov & Oke (2008). This algorithm is part of a broader family of DA schemes – including the so-called deterministic and square-root filters, see Tippett et al. (2003) – which, differently from the prevailing formulation of the EnKF proposed by Burgers et al. (1998), do not require the addition of random perturbations to the observations before their assimilation. This procedure constitutes a

<table>
<tr><th colspan="4">Model parameters</th></tr>
<tr><td>Rossby number, Ro</td><td>∞</td><td>Initial conditions</td><td>Eq. (2.11)</td></tr>
<tr><td>Froude number, Fr</td><td>1.1</td><td>Boundary conditions</td><td>Periodic</td></tr>
<tr><td>Convection threshold, H_c</td><td>1.02</td><td>Domain length L_0 [km]</td><td>500</td></tr>
<tr><td>Rain threshold, H_r</td><td>1.05</td><td>Velocity scale V_0 [ms^{-1}]</td><td>20</td></tr>
<tr><td>$\widetilde{\alpha}$</td><td>10</td><td>Forecast N_{el}</td><td>200</td></tr>
<tr><td>β</td><td>0.2</td><td>Nature N_{el}^{nat}</td><td>400</td></tr>
<tr><td>$\widetilde{c}_0^2$</td><td>0.085</td><td>Time-scale T_0 [h]</td><td>6.94</td></tr>
<tr><th colspan="4">Observing system</th></tr>
<tr><td colspan="2">Total number of observations, p</td><td colspan="2">28</td></tr>
<tr><td colspan="2">Observations per variable (p_h, p_u, p_r)</td><td colspan="2">(8,10,10)</td></tr>
<tr><td colspan="2">Update frequency [min]</td><td colspan="2">60</td></tr>
<tr><td colspan="2">Obs. spacing (d_h, d_u, d_r) [km]</td><td colspan="2">(62.5, 50, 50)</td></tr>
<tr><td colspan="2">Obs. error (s_h, s_u, s_r)</td><td colspan="2">(0.05, 0.02, 0.003)</td></tr>
</table>

Table 2.1: A summary of the non-dimensional model parameters and the properties of the observing system used in the idealised experiments. Units are dimensionless unless specified otherwise.

well-known source of sampling error for ensemble filters and makes the analysis error suboptimal, as discussed in Whitaker & Hamill (2002).

In our idealised system based on the modRSW model, only a small number of observations are assimilated, and they are already perturbed at the moment of their generation from the nature run, via (2.13). In this sense, the DEnKF offers a convenient setup, easy to implement and similar to the original EnKF, while at the same time it avoids the risks associated with perturbing the (few) observations twice.

Like other sequential filters, the DEnKF is characterised by a forecast step during which an ensemble of forecasts $\mathbf{x}_j^f$, with $j = 1, ..., N$ is integrated forward in time, alternated with an analysis step during which the observations $\mathbf{y}^o$ are combined with the model states $\mathbf{x}_j^f$ to obtain a new ensemble of states called analyses, $\mathbf{x}_j^a$, which are used to initialise the subsequent forecasts.

Hence, the forecast step for each ensemble member $\mathbf{x}_j^f$ between two consecutive analysis steps at times t_i and t_{i+1} can be expressed as:

$$\mathbf{x}_j^f(t_i) = \mathcal{M}(\mathbf{x}_j^a(t_{i-1})). \tag{2.17}$$

At the analysis step, the DEnKF computes the analysis mean $\overline{\mathbf{x}}^a$ according to the usual Kalman filter equation:

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^f + \mathbf{K}_e(\mathbf{y}^o - \overline{\mathcal{H}(\mathbf{x}^f)}), \tag{2.18}$$

in which $\overline{\mathbf{x}}^f$ is the forecast mean, defined as:

$$\overline{\mathbf{x}}^f = \frac{1}{N}\sum_{j=1}^{N}\mathbf{x}_j^f, \tag{2.19}$$

whereas $\mathcal{H}$ represents the observation operator, i.e. a function $\mathcal{H} : \mathbb{R}^{3N_{el}} \to \mathbb{R}^p$ able to map a model state into the observational space. Given the observing system described earlier, in our system the function $\mathcal{H}$ can be linearised and it reduces to a sparse matrix $\mathbf{H}$ of dimension $(p \times 3N_{el})$. The matrix $\mathbf{K}_e$ in (2.29) is called a *Gain matrix*, has size $(3N_{el} \times p)$ and is defined as:

$$\mathbf{K}_e = \mathbf{P}_e^f\mathbf{H}^T(\mathbf{H}\mathbf{P}_e^f\mathbf{H}^T + \mathbf{R})^{-1}, \tag{2.20}$$

in which $\mathbf{R}$ represents the observation-error covariance matrix of dimensions $(p \times p)$ defined in section §2.2.1, whereas $\mathbf{P}_e^f$ is the time-dependent forecast (or background-) error covariance matrix of size $(3N_{el} \times 3N_{el})$, which is calculated as:

$$\mathbf{P}_e^f = \frac{1}{N-1}\sum_{j=1}^{N}(\mathbf{x}_j^f - \overline{\mathbf{x}}^f)(\mathbf{x}_j^f - \overline{\mathbf{x}}^f)^T = \frac{1}{N-1}\mathbf{X}^f(\mathbf{X}^f)^T, \tag{2.21}$$

with $\mathbf{X}^f$ called a forecast perturbation matrix of dimensions $(3N_{el} \times N)$, formed by the $j = 1, ..., N$ column vectors:

$$(\mathbf{X}^f)_j = \mathbf{x}_j^f - \overline{\mathbf{x}}^f. \tag{2.22}$$

The expression (2.21) represents an approximation of the 'real' forecast-error covariance $\mathbf{P}^f$ based on the uncertainty conveyed by the ensemble $\mathbf{x}_j^f$. This is a fundamental assumption of any ensemble filter and comes with its own limitations. For example, there are sampling errors which commonly arise when using (2.21), which will be addressed in the next section.

Once the analysis mean $\overline{\mathbf{x}}^a$ is computed, the DEnKF prescribes the redefinition of the analysis perturbation matrix as:

$$\mathbf{X}^a = \mathbf{X}^f - \frac{1}{2}\mathbf{K}_e\mathbf{H}\mathbf{X}^f. \tag{2.23}$$

Finally, the analysis ensemble $\mathbf{x}_j^a$ is recalculated as:

$$\mathbf{x}_j^a = (\mathbf{X}^a)_j + \overline{\mathbf{x}}^a. \tag{2.24}$$

Our implementation of the DEnKF scheme is detailed in section §2.3 and presents some differences from what we have described so far, which will be illustrated in the next section.

2.2.3 Limitations and remedies

The DEnKF is subjected to a series of issues which are common to most ensemble Kalman filters and that make the final analysis suboptimal. In particular:

- the small, finite-size ensemble forecast leads to sampling errors: these are reflected in the forecast-error covariance matrix $\mathbf{P}_e^f$ (2.21), which underestimates the real forecast uncertainty and causes the ensemble to have too little spread (Van Leeuwen, 1999), that is, the degree of dispersion of the ensemble members (a formal definition is given in section §2.4.2).

- The impossibility for a small ensemble to span the entire model space (i.e. $N \ll 3N_{el}$) risks creating spurious correlations in the covariance matrix $\mathbf{P}_e^f$, which are not representative of real physical relationships between the variables.

- The issue of inbreeding, as identified by Houtekamer & Mitchell (1998), in which the j^{th} ensemble member is updated using a covariance matrix $\mathbf{P}^f$ computed from the perturbation matrix $\mathbf{X}^f$ based on the ensemble average $\overline{\mathbf{x}}^f$, including the j^{th} ensemble member itself.

- The imperfections and intrinsic limitations of the forecast model $\mathcal{M}$, which is inevitably a simplification of the real system (in our case, this aspect is mimicked by the resolution mismatch displayed in Fig. 2.1).

As a consequence, a series of remedies are adopted to mitigate the effects of the above issues. The solutions considered in this thesis include: adaptive multiplicative and additive inflation to correct for sampling and model errors, respectively; a self-exclusion technique to avoid inbreeding, and localisation to damp spurious correlation in the forecast covariance matrix $\mathbf{P}_e^f$. Below, we discuss each remedy applied to our system in detail.

Adaptive multiplicative inflation

In order to counteract the lack of ensemble spread resulting from the sampling error due to the finite-size ensemble forecast, an adaptive multiplicative inflation technique called Relaxation To Prior Spread (RTPS) is utilised. This method was proposed by Whitaker & Hamill (2012) and exploits the fact that the assimilation of the observations in the analysis step causes the ensemble spread of the analysis to be smaller than the spread of the forecast. Therefore, the RTPS comprises the recomputing of the analysis spread σ^a at the analysis step as:

$$\sigma^{\mathrm{a}} \leftarrow (1 - \alpha_{RTPS}) \cdot \sigma^{\mathrm{a}} + \alpha_{RTPS} \cdot \sigma^{\mathrm{f}}, \tag{2.25}$$

in which σ^f represents the forecast spread and $\alpha_{RTPS} \in [0, 1]$ is a tunable parameter. More information on the actual implementation of the RTPS in our algorithm can be found in section §2.3.

We note here that an alternative type of adaptive multiplicative inflation was proposed by Whitaker & Hamill (2012), called Relaxation To Prior Perturbations (RTPP). This technique is based on the recalculation of the analysis perturbation matrix $\mathbf{X}^a$ as:

$$\mathbf{X}^a \leftarrow (1 - \alpha_{RTPP})\mathbf{X}^a + \alpha_{RTPP}\mathbf{X}^f, \tag{2.26}$$

in which $\mathbf{X}^f$ represents the forecast perturbation matrix (cf. (2.22)) and $\alpha_{RTPP} \in [0, 1]$ is a tunable parameter. Similarly to the RTPS, also the RTPP exploits the reduction in spread of the analysis ensemble, as the perturbation matrix depends on the ensemble dispersion (see definition in (2.22)). Duc et al. (2020) presents a unifying theory about covariance inflation techniques, including RTPP, RTPS and other types of multiplicative inflation. We show in Appendix A how the DEnKF formulation is equivalent to an ensemble Kalman filter with no perturbations and a RTPP of $\alpha_{RTPP} = 1/2$. We exploit this equivalence in coding our DA algorithm (see step 2, part vi in section §2.3).

Self-exclusion

To avoid the ensemble inbreeding, each analysis member $\mathbf{x}_j^a$ is calculated using a different Kalman gain $\mathbf{K}_{e,\widehat{j}}$ computed as:

$$\mathbf{K}_{e,\widehat{j}} = \mathbf{P}^f_{e,\widehat{j}}\mathbf{H}^T(\mathbf{H}\mathbf{P}^f_{e,\widehat{j}}\mathbf{H}^T + \mathbf{R})^{-1}, \tag{2.27}$$

in which each $\mathbf{P}^f_{e,\widehat{j}}$ is defined as in (2.21) but with the exclusion of the j^{th} ensemble member. This method was first proposed by Houtekamer & Mitchell (1998) and subsequently applied by Hamill & Snyder (2000), while was first referred to as 'self-exclusion' by Bowler et al. (2017).

In order to incorporate the self-exclusion into the DEnKF, the analysis mean $\overline{\mathbf{x}}^a$ is computed as:

$$\overline{\mathbf{x}}^a = \frac{1}{N}\sum_{j=1}^{N}\mathbf{x}_j^a; \tag{2.28}$$

after the application of the Kalman filter equation:

$$\mathbf{x}^a_{\widehat{j}} = \mathbf{x}^f_j + \mathbf{K}_{e,\widehat{j}}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^f_j)). \tag{2.29}$$

We note that this represents a slight modification of the algorithm originally defined by Sakov & Oke (2008).

Localisation

Localisation techniques are employed to eliminate spurious long-range correlations in the forecast-error covariance matrix (Hamill et al., 2001). In practice, the (self-excluded) matrix $\mathbf{P}^f_{e,\widehat{j}}$ is redefined as:

$$\mathbf{P}^f_{\text{loc},\widehat{j}} = \boldsymbol{\rho} \circ \mathbf{P}^f_{e,\widehat{j}}, \tag{2.30}$$

in which $\boldsymbol{\rho}$ is a $(3N_{el} \times 3N_{el})$ localisation matrix made of coefficients $\rho_{l,m}$ representing the correlation between the elements l and m, computed according to the Gaspari-Cohn taper functions (Gaspari & Cohn, 1999), whereas the symbol '$\circ$' represents the element by element matrix multiplication $\rho_{l,m} \cdot P^f_{e,\widehat{j}\,l,m}$ (also known as the Schur product, cf. Schur (1911)). The shape of the Gaspari-Cohn taper functions depends on a parameter L_{loc} which defines the length-scale of the localisation, with $L_{loc} = 0$ representing the absence of localisation. Fig. 2.4 shows the shape of these functions for different values of L_{loc} as a function of the distance in grid points (left panel), together with a plot of the matrix $\boldsymbol{\rho}$ for $L_{loc} = 1.0$ (right panel). The mathematical expression of the Gaspari-Cohn function is given in Appendix B.

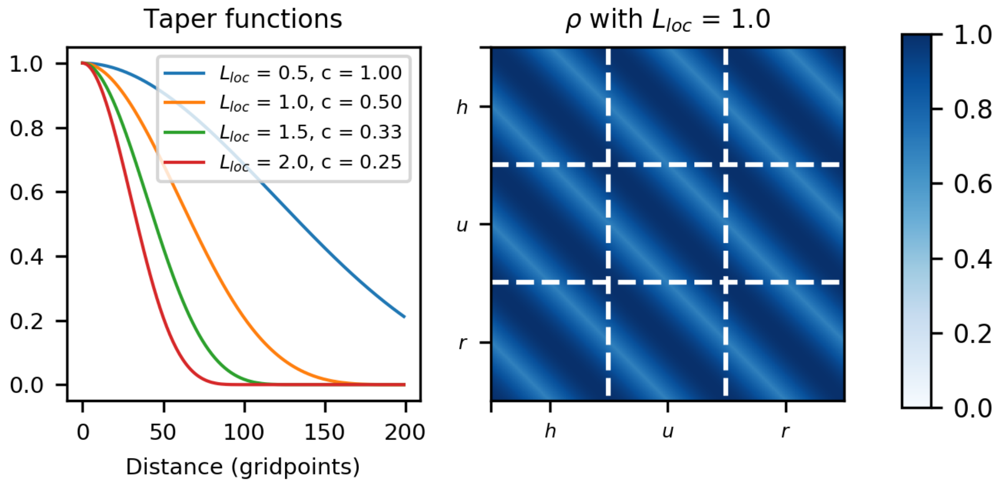


Figure 2.4: Gaspari-Cohn taper functions for different values of localisation scales and length distance ($L_{loc} = \{0.5, 1.0, 1.5, 2.0\}$, $c = \{1.0, 0.50, 0.33, 0.25\}$, cf. Appendix B) as a function of distance in number of gridpoints (left) and plot of the matrix $\boldsymbol{\rho}$ for $L_{loc} = 1.0$ (right). The values $L_{loc} = 1.0$ will be selected at the end of the tuning process presented later in this chapter.

Additive inflation

Additive inflation is implemented by injecting random perturbations $\boldsymbol{\eta}$ into the forecast model during its numerical integration, that is, by modifying (2.17) into:

$$\mathbf{x}_j^f(t_i) = \mathcal{M}(\mathbf{x}_j^a(t_{i-1})) + \boldsymbol{\eta}_j, \quad j = 1, ..., N, \tag{2.31}$$

in which the $3N_{el}$-dimensional vector $\boldsymbol{\eta}_j$ is sampled from a Gaussian distribution $\boldsymbol{\eta} \sim \mathcal{N}(0, \gamma_a^2\mathbf{Q})$, with $\mathbf{Q}$ being the model-error covariance matrix and γ_a a tunable additive inflation parameter controlling the magnitude of $\boldsymbol{\eta}$.
In general, the best way to compute the matrix $\mathbf{Q}$ for an operational system is not trivial and is still very much an active field of research. However, since the scope of this study is not to investigate the best model for $\mathbf{Q}$, we adopt a very straightforward approach, exploiting the fact that, in the idealised system that we have constructed, the only source of model error comes from the difference in spatial resolution between the forecasts and the nature run. Therefore, we launch a total of 48 deterministic one-hour forecasts (with $N_{el} = 200$) initialised at each analysis steps with the nature run so that

there is no initial condition error, and take the differences at each grid point between the forecast and the nature run (conveniently sub-sampled) at the end of each hour-long simulation, chosen to match the forecast-assimilation system's cycling period. The resulting space-dependent error distribution is subsequently used to generate a model-error covariance matrix $\mathbf{Q}$, in which the non-diagonal terms are neglected for simplicity. In addition, the hr components are set to zero in order to avoid the overinflation of the fluid depth h, which is non-linearly related to r via the system of thresholds described in section §2.1.2. The result is shown in Figure 2.5, where the diagonal terms of the matrix $\mathbf{Q}$ multiplied by different values of γ_a^2 are shown. From the graph, one can note the impact that topography has on shaping the model error, which is larger in the second half of the domain, where most convection and rain takes place (cf. Fig. 2.3). In order to avoid the introduction of biases, once $\boldsymbol{\eta}_j$ is drawn from the distribution $\boldsymbol{\eta} \sim \mathcal{N}(0, \gamma_a^2\mathbf{Q})$, it gets redefined by subtracting the ensemble average from the sampled perturbations, which leads to the unbiased additive inflation vector $\tilde{\boldsymbol{\eta}}$:

$$\tilde{\boldsymbol{\eta}}_j = \boldsymbol{\eta}_j - \overline{\boldsymbol{\eta}} \tag{2.32}$$

to be used in (2.31) instead.

Finally, the implementation of the additive inflation in (2.31) is performed via the Incremental Analysis Updates (IAU, cf. Bloom et al. (1996)). In other words, $\tilde{\boldsymbol{\eta}}_j$ is split into small increments and injected continuously throughout the forecast step, rather than added all at once at the analysis step (see forecast step, step ii and equation (2.34) in section §2.3). Again, any negative values of h and r emerging upon the injection of additive inflation are reset to $h = 0.001$ and $r = 0$, respectively. The choice of adopting the IAU method is also in line with the operational implementation of the additive inflation at the Met Office (cf. Bowler et al. (2017)), while noting that the Met Office samples from an archive of scaled and adjusted historical analysis increments rather than drawing the perturbations from a Gaussian distribution.

2.3 Step-by-step algorithm

Note that this section is an adaptation of Appendix A in Kent et al. (2020).

A compact algorithm for our implementation of one complete cycle (forecast plus analysis) of the DEnKF is summarised here. Throughout, the subscript $j = 1, ..., N$ denotes the j^{th} ensemble member and the subscript $i = 1, ..., 48$ denotes the analysis step

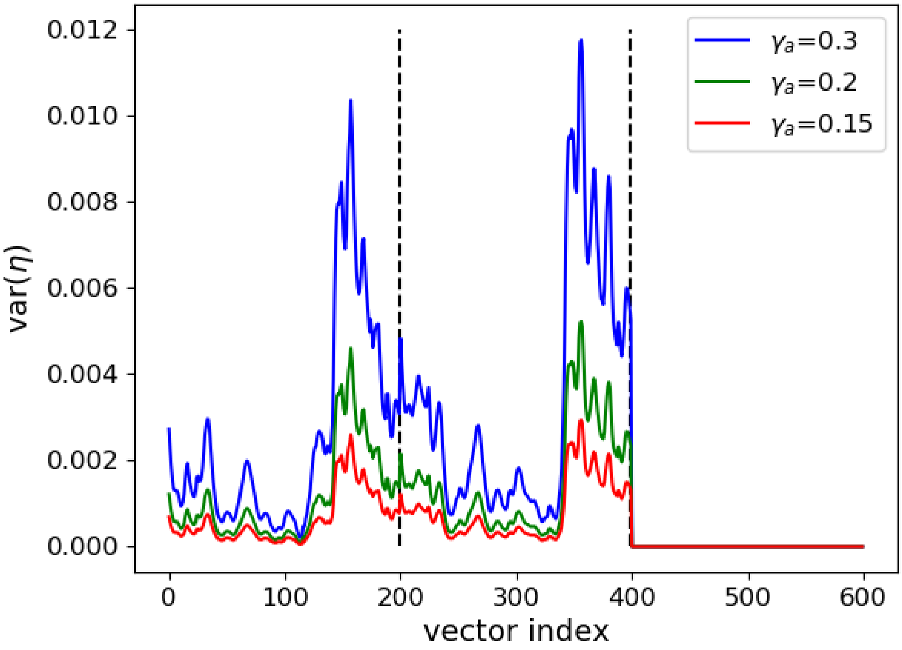


Figure 2.5: Values of the $(N_{el} \times N_{el})$ diagonal matrix $\gamma_a^2\mathbf{Q}$ for $\gamma_a = \{0.15, 0.2, 0.3\}$, representing the spatial structure of the model error used to generate the additive inflation in (2.31). The vertical dashed lines delimit the h, hu and hr components of $\gamma_a^2\mathbf{Q}$. The different values of γ_a are tested in the tuning process outlined in section §2.5.

(i.e. time). Note that prior to the start of the DA algorithm, p pseudo-observations $\mathbf{y}_i$ are generated by stochastically perturbing the nature run $\mathbf{x}^t$ valid at the observing time t_i:

$$\mathbf{y}_i = \mathcal{G}\mathbf{x}^t + \boldsymbol{\epsilon}_i^o, \qquad \text{where } \boldsymbol{\epsilon}_i^o \sim \mathcal{N}(0, \mathbf{R}), \tag{2.33}$$

in which $\mathbf{R} = \text{diag}(s_h^2\mathbf{I}_h, s_u^2\mathbf{I}_u, s_r^2\mathbf{I}_r)$ is a matrix of size $(p \times p)$, for prescribed error variances $s^2_{h,u,r}$ and identity matrices $\mathbf{I}_{h,u,r}$ with dimension equal to the number of observations of h, u, and r, respectively. Unphysical negative pseudo-observations of h and r are then reset to zero. A prescribed model-error covariance matrix $\mathbf{Q}$ of dimensions $(N^{el} \times N^{el})$ needs also to be estimated.

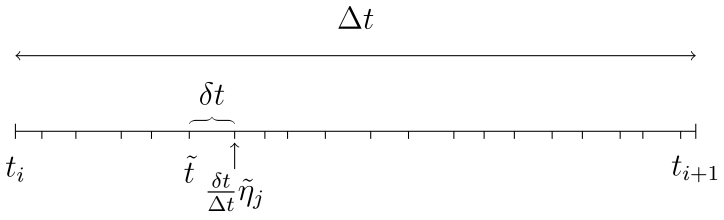
1. FORECAST STEP:

   i An ensemble of initial conditions $\mathbf{x}_j^{\text{ic}}$ is generated by taking the values from Eq. (2.11) and adding Gaussian noise for each variable according to $\mathbf{s}^{ic}$, as per Eq. (2.16). Unphysical negative initial conditions for hr are reset to zero while negative h values (very rare) are reset to 0.001.

   ii The model is integrated forward in time. Additive inflation is drawn from the model-error covariance matrix $\mathbf{Q}$ as $\eta_j \sim \mathcal{N}(0, \gamma_a^2\mathbf{Q})$ and potential biases are removed by applying equation (2.32). The unbiased model-error vector $\widetilde{\eta}_j$ is injected throughout the numerical integration by dividing it into (small) allocations proportional to the duration of each dynamical time-step δt. For time-step details, we refer to Kent (2016) and Kent et al. (2017), implemented here with a Courant-Friedrichs-Lewy (CFL) number of 0.5. Therefore, within each forecast step of duration $\Delta t = t_{i+1} - t_i$ at any time $\tilde{t} \in [t_i, t_{i+1}]$, we compute:

   $$\tilde{\mathbf{x}}_j^f(\tilde{t} + \delta t) = \mathcal{M}[\tilde{\mathbf{x}}_j^f(\tilde{t})] + \frac{\delta t}{\Delta t}\widetilde{\eta}_j, \quad j = 1, ..., N, \tag{2.34}$$

   with $\tilde{\mathbf{x}}_j^f(t_i) = \tilde{\mathbf{x}}_j^a(t_i)$. In order to ensure that the algorithm does not overshoot the time of the next forecast-assimilation cycle t_{i+1}, we take the final time-step to be the reduced value $t_{i+1} - \tilde{t}$ when this is smaller than the optimal δt determined by the CFL value.



iii At later times, the forecast uses the analysis ensemble from the previous cycle as initial condition to integrate forward in time. A 12-hour forecast is launched at the end of each analysis step, with the additive inflation re-sampled and injected hourly as in ii.

2. ANALYSIS STEP:

i Each j^{th} $T+1$ hr model state obtained from the most recently launched $T+12$ hr forecast is transformed into the state vector for assimilation: $\mathbf{x}_j^f(t_i) = \Psi(\tilde{\mathbf{x}}_j^f(t_i))$, i.e., $(h, hu, hr) \mapsto (h, u, r)$.

ii Compute the $(p \times 1)$ innovations $\mathbf{d}_j = \mathbf{y}_i - \mathcal{H}\mathbf{x}_j^f$ using the forecast states from step 1 and the pre-computed pseudo-observations.

iii Compute the (diagonal) observational-error covariance matrix $\mathbf{R}$, described after Eq. (2.33).

iv Compute the forecast perturbation matrix $\mathbf{X}$ of size $(N^{el} \times N)$ and therefore the N forecast-error covariance matrices $\mathbf{P}^f_{e,\widehat{j}}$ of dimension $(N^{el} \times N^{el})$ (see Eq. (2.21)), each of them computed excluding the j^{th} ensemble member of the forecast states from step 1, in order to avoid inbreeding.

v Apply (model-space) localisation using the Gaspari-Cohn function ρ for a given length-scale L_{loc} to each forecast-error covariance matrix $\mathbf{P}^f_{loc,\widehat{j}}$:

$$\mathbf{P}^f_{loc,\widehat{j}} \leftarrow \boldsymbol{\rho} \circ \mathbf{P}^f_{e,\widehat{j}}; \tag{2.35}$$

compute the j^{th} Kalman gain $\mathbf{K}_{e,j}$ of size $(N^{el} \times p)$ and the subsequent analysis ensemble:

$$\mathbf{x}_j^a = \mathbf{x}_j^f + \mathbf{K}_{e,j}\mathbf{d}_j. \tag{2.36}$$

vi The Deterministic Ensemble Kalman Filter is implemented. The analysis perturbation matrix $\mathbf{X}^a$ of dimensions $(N^{el} \times N^{el})$ is computed with the N members $\mathbf{x}_j^a$ (as per Eq. (2.21)); using the RTPP implementation of the DEnKF (Appendix A), $\mathbf{X}^a$ is then redefined as:

$$\mathbf{X}^a = \frac{1}{2}\mathbf{X}^a + \frac{1}{2}\mathbf{X}^f \tag{2.37}$$

vii Relaxation to Prior Spread (RTPS) is applied by recomputing the analysis ensemble spread σ^{a} (as per Eq. (2.25)) and then the analysis ensemble perturbations $\mathbf{X}^a$ are adjusted using Eq. (3) from Whitaker & Hamill (2012):

$$\mathbf{X}^a = \left(1 - \alpha_{RTPS} + \alpha_{RTPS}\frac{\sigma^{\text{f}}}{\sigma^a}\right)\mathbf{X}^a. \tag{2.38}$$

viii The final analysis ensemble is recomputed using the redefined perturbation matrix:

$$\mathbf{x}_j^a = \overline{\mathbf{x}}^a + (\mathbf{X}^a)_j, \quad j = 1, ..., N, \tag{2.39}$$

where $(\mathbf{X}^a)_j$ is the j^{th} column of $\mathbf{X}^a$. If h is negative then it is reset to 0.001 while if r is negative it is reset to 0.

ix Return to step 1: analysis states from step 2viii are transformed back $\tilde{\mathbf{x}}_j^a(t_i) = \Psi^{-1}(\mathbf{x}_j^a(t_i))$ for integration and the sequential cycle continues.

Recall that, in this implementation, the parameter L_{loc} defines the number of grid points in model space (N_{el}/L_{loc}) beyond which correlations are set to zero. In the case of the ground observations, the observation location always coincides with the model grid, and therefore $\mathbf{H}$ is a linear operator.

2.4 Relevance of the modRSW model for NWP data assimilation

As outlined in the introduction, the use of idealised models in DA research is a common strategy that presents substantial computational and logistical advantages. Nonetheless, an assessment of their relevance for operational NWP systems is often overlooked, despite being a crucial aspect impacting on their usefulness. After all, an idealised setup able to closely reproduce most of the features of a state-of-the-art DA scheme can become a precious tool in orienting future decisions and choices involving a full-scale system.

It is worth mentioning at this point that the model developed by Würsch & Craig (2014) – of which the modRSW model represents a more physically consistent version – has already been used in a number of studies to investigate various aspects of convective-scale DA, such as the use of particle filters (Haslehner et al., 2016), parameter estimation for the representation of clouds (Ruckstuhl & Janjić, 2018) and the representation of model error (Zeng et al., 2018, 2019). In this regard, we note that the aim of this study is considerably different from those applications, as we want neither to compare DA methods, nor to propose suggested changes to operational DA schemes, but instead want to provide convincing evidence that DA experiments using the modRSW model show consistency with operational forecast-assimilation systems. In doing this, we achieve two objectives. On the one hand, we develop a protocol – or at least a consistent framework – to assess the relevance of idealised forecast-assimilation systems for DA research in an operational setting. On the other hand, we can demonstrate that the modRSW model itself constitutes a useful and relevant tool for investigating DA in the presence of complex dynamics associated with convection and precipitation.

2.4.1 A protocol for assessing performance and relevance of idealised model experiments

We approach the question of whether (and to what extent) an idealised system is able to replicate the main characteristics of an operational DA scheme in a rigorous and systematic way, which is summarised in Table 2.2.

In the first column on the left, we have listed a series of parameters and properties that characterise real-world data assimilation schemes and that any idealised systems should aim to reproduce. Among these, we have also included some of the aspects and criteria that are utilised (or need to be taken into account) during the tuning and the validation processes which will be analysed in detail in the next section, §2.5. The second column reports the typical values that these aspects assume in operational convective-scale DA systems (see Gustafsson et al. (2018) for a good summary), whereas the third column displays how those are reflected in our forecast-assimilation experiments based on the modRSW model. Lastly, the fourth column presents a self-assessed evaluation of whether or not each aspect (unless inapplicable) is imitated successfully.

The values of the ‘prescribed parameters’ reported in the table are based on those reported in Table 5 of Gustafsson et al. (2018), which provides a useful overview of the operational convective-scale data assimilation systems in use at the time. Since the use of idealised models aims at reducing the computational costs, the number of observations and the size of the state vector in our system are by construction not comparable with those of a full-scale scheme. However, the rank deficiency relationship which holds in a typical operational system is maintained by choosing a number of ensemble members (i.e. $N = 18$) and observations (i.e. $p = 28$) such that $N < p < n$, with $n = 3N_{el}$.

The observing system chosen for the idealised experiments is an extremely simple one (i.e. linear observation operator, uncorrelated observations error, cf. section §2.2.1), although it replicates the network density of the ground observations assimilated at the Met Office, which is approximately 40 km[1].

The filter configuration of our idealised experiments depends on localisation and inflation parameters (such as the RTPP and RTPS) which are widely used in other operational systems. Although we do not expect the precise values of these parameters to be comparable (as they reflect the characteristics of a specific data assimilation system rather than representing a universal target value), both the RTPS and the RTPP found in our final configuration are in line with those reported in the literature for other operational implementations (Bick et al., 2016; Bowler et al., 2017; Gustafsson et al., 2018; Schraff et al., 2016).

In addition, we used common diagnostic tools such as the ensemble spread (SPR), the root mean square error of the ensemble mean (RMSE) and the Continuous Ranked Probability Score (CRPS) as tuning criteria for our Deterministic Ensemble Kalman filter (these quantities are defined in section §2.4.2). Ensemble filters are considered well–tuned when the spread of their ensemble is comparable to the RMSE of the ensemble mean[2]; we therefore set the target ratio between SPR and RMSE to unity. We also seek to minimise the values of both RMSE and CRPS for forecasts with a lead time of three hours: this reflects the objective of operational centres of optimising their data assimilation systems to produce the forecast with the smallest possible error, other than generating an accurate initial condition. Moreover, a lead time of three hours is also fully within what is considered to be the nowcasting range ($0 - 6$ hours) for a weather forecast (Ballard et al., 2012; Sun et al., 2014), which is the most relevant time-scale for convective-scale NWP.

---

[1] Source: `https://www.metoffice.gov.uk/weather/guides/observations/uk-observations-network`. Accessed on 26/05/2021.

[2] Leutbecher & Palmer (2008) show that this results is actually dependent on the ensemble size. In the case of an ensemble of $N = 18$ members, a theoretical convergence is expected for $RMS = 1.1\ SPR$, cf. equation (7) therein.

Table 2.2: Protocol to assess the relevance of idealised forecast-assimilation experiments for convective scale NWP, and results obtained with the DA simulations performed with the modRSW model.

	Operational system	Our idealised system	Relevant?
Prescribed parameters			
Forecast resolution	$\mathcal{O}(1\text{km})$	2.5km	✓
Update frequency	$\mathcal{O}(1\text{hr})$	1hr	✓
Ensemble size, N	$\mathcal{O}(10-100)$	18	✓
Number of observations, p	$\mathcal{O}(10^7)$	28	N/A
State dimension, n	$\mathcal{O}(10^9)$	600	N/A
Rank-deficiency	$N \ll p \ll n$	$N < p < n$	✓
Observation operator	Nonlinear	Linear	–
Tuning parameters			
Observing system:			
(i) observation spacing	~ 40km	~ 50km	✓
(ii) observation error	Correlated and uncorrelated	Uncorrelated	–
Filter configuration:			
(i) Localization (horiz.)	Lengthscale: $\mathcal{O}(100\,\text{km})$ (obs. space)	~ 500km (model-space)	✓
(ii) Inflation	Adaptive (RTPP; RTPS) and additive[1]	$\alpha_{RTPP} = 0.5$, $\alpha_{RTPS} = 0.7$, $\gamma_a = 0.15$	✓
Tuning criteria			
(i) $SPR/RMSE$	~ 1	~ 1	✓
(ii) $RSME$	Minimum for given lead times	Minimum for 3hrs forecast	–
(iii) $CRPS$	Minimum for given lead times	Minimum for 3hrs forecast	–
Validation criteria			
Observational influence	Global: $\sim 20\%$; high-res.: $> 20\%$	$\sim 30\%$	✓
Error-doubling time T_d	Global: $\mathcal{O}(1\text{ day})$; high-res.: $\mathcal{O}(1\text{ hr})$	$\sim 6-9$ hrs	✓

[1] Owing to the many different approaches in operational systems, it is not possible to give precise target values for inflation factors. In the literature, we find values in the region of $\alpha_{RTPP} = 0.5 - 0.75$ and $\alpha_{RTPS} = 0.7 - 0.95$ (e.g., Bick et al. (2016); Bowler et al. (2017); Gustafsson et al. (2018); Schraff et al. (2016)), for both global and convective-scale systems.

Once the tuning criteria mentioned above are met, it is possible to validate an idealised system for its relevance for NWP data assimilation. This is assessed by calculating the Observational Influence diagnostic (OID) and the error doubling time T_d and subsequently comparing them against the values found in operational systems (see section §2.4.2 for their definitions). In this regard, our well–tuned forecast-assimilation experiments produced an OID of approximately 30%, which compares with a value of 18% calculated for the global ECMWF model by Cardinali et al. (2004), but is expected to be higher (i.e. between 20% and 40%) in a convective-scale system. In addition, the value of the error-doubling time found in the experiments (i.e. $6-9$ hours) is close to those found in other operational convection-permitting models (Hohenegger & Schar, 2007).

2.4.2 Diagnostics definitions

In this section we summarise the definitions of some of the main diagnostics used in this thesis to evaluate the forecast-assimilation experiments described here and in Chapter 5.

Root mean squared error, RMSE

The root mean squared error (RMSE) is used to measure the accuracy of the ensemble mean. The value of the RMSE averaged over the state vector is computed as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(\overline{x}_k^{(\bullet)} - x_k^t)^2}, \text{where } \overline{x}_k^{(\bullet)} = \frac{1}{N}\sum_{j=1}^{N} x_{j,k}^{(\bullet)}, \tag{2.40}$$

in which the subscript k refer to each component (i.e. variable) of the state vector, the superscript t indicates the nature run $\mathbf{x}^t$, whereas the symbol $(\bullet)$ stands for either the analysis $\mathbf{x}^a$ or the forecast $\mathbf{x}^f$. All vectors are of size N_{el}, with the nature run sub-sampled on the forecast model grid). The integer N represents the size of the ensemble.

Ensemble spread, SPR

The spread (SPR) is used to measure the uncertainty (or dispersion) conveyed by an ensemble. Its average over the state vector is computed as:

$$SPR = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\frac{1}{N-1}\sum_{j=1}^{N}(x_{j,k}^{(\bullet)} - \overline{x}_{k}^{(\bullet)})^2} \equiv \sqrt{\frac{1}{n}\mathrm{Tr}(\mathbf{P}_{\mathrm{e}}{}^{(\bullet)})}, \tag{2.41}$$

in which ‘Tr’ indicates the trace of a matrix (i.e. the sum of its diagonal entries), and $\mathbf{P}_e^{(\bullet)}$ is the forecast-error covariance matrix (cf. eq. (2.21)).

Continuous Ranked Probability Score, CRPS

The continuous ranked probability score (CRPS) is used to evaluate the quality of a probabilistic forecast. In this thesis, we apply the definition for ensemble systems presented in Hersbach (2000). Therefore, after reordering the ensemble members such that $x_i \leq x_j$ (with $i < j$), we can compute the cumulative distribution $P(x)$:

$$P(x) = \frac{1}{N}\sum_{j=1}^{N}\Theta(x - x_j^{(\bullet)}), \tag{2.42}$$

and the transition values $p_j = j/N$, in which $\Theta(\cdot)$ represents a Heaviside function which is one when its argument is non-negative.
The CRPS is then computed as:

$$CRPS = \sum_{j=0}^{N} c_j, \qquad \text{with } c_j = \alpha_j p_j^2 + \beta_j (1 - p_j)^2 \tag{2.43}$$

in which α_j and β_j are coefficients defined as:

$0 < i < N$	α_j	β_j
$x_{j+1} < x^t$	$x_{j+1} - x_j$	0
$x_j < x^t < x_{j+1}$	$x^t - x_j$	$x_{j+1} - x^t$
$x_j > x^t$	0	$x_{j+1} - x_j$

The CRPS is therefore a scalar quantity included between 0 and 1, in which $CRPS = 0$ indicates a ‘perfect’ forecast. In the results reported here and in Chapter 5, the CRPS is averaged over the domain and/or each variable.

Observation Influence Diagnostic, OID

The observation influence diagnostic (OID, cf. Cardinali et al. (2004)) is used to evaluate quantitatively the sensitivity of the analysis $\mathbf{x}^a$ with respect to the observations $\mathbf{y}^o$. In order to compute this quantity, we start first by calculating the projection of the analysis in the observation space for each ensemble member, that is:

$$\hat{\mathbf{y}}_j = \mathbf{H}\mathbf{x}_j^a = \mathbf{H}\mathbf{K}_{e,j}\mathbf{y}^o + (\mathbf{I} - \mathbf{H}\mathbf{K}_{e,j})\mathbf{H}\mathbf{x}_j^f, \tag{2.44}$$

in which: $\mathbf{H}$ is the (linearised) observation operator[1], $\mathbf{K}_{e,j}$ is the Kalman gain matrix, $\mathbf{I}$ is the $p \times p$ identity matrix, and j a subscript indicating the ensemble member excluded by the self-exclusion (cf. section § 2.2.3). Therefore, the analysis sensitivity with respect to the observations for each ensemble member is defined as:

$$\mathbf{S}_j = \frac{\partial \hat{\mathbf{y}}_j}{\partial \mathbf{y}^o} = \mathbf{H}\mathbf{K}_{e,j}, \tag{2.45}$$

and the global observation influence is hence computed as:

$$OID_j = \frac{\mathrm{Tr}(\mathbf{S}_j)}{p}. \tag{2.46}$$

The values of OID reported later are averaged over the whole ensemble. Moreover, since the OID is an additive quantity (as (2.46) implies), it is possible to split the OID into contributions related to each type of observation, by considering only the diagonal entries of $\mathbf{S}$ associated with them.

Error doubling time, T_d

The error doubling time (T_d) is used to estimate the error-growth rate of the forecast-assimilation system. Here, it is computed such that:

$$\frac{E(T_d)}{E(T_0)} = 2 \tag{2.47}$$

in which $E(T_0)$ is defined as the analysis RMSE (based on the difference $\mathbf{x}_j^a - \mathbf{x}^t$) of each analysis ensemble member used to initialise a medium-range forecast during which $E(T_d)$ (i.e. the forecast RMSE, based on $\mathbf{x}_j^f - \mathbf{x}^t$) is calculated hourly.

---

[1] This is the case for the satellite observation, for which we compute $\frac{\mathrm{d}I_{sat}}{\mathrm{d}\sigma}$.

Tuning parameters	
Localisation scale, L_{loca}	$\{0.5, 1.0, 1.5, 2.0\}$
Adaptive inflation, α_{RTPS}	$\{0.1, 0.3, 0.5, 0.7, 0.9\}$
Additive inflation, γ_a	$\{0.05, 0.08, 0.1, 0.12, 0.15, 0.2, 0.3, 0.4, 0.5\}$

Table 2.3: An overview of the assimilation parameters varied during the tuning of the filter configuration.

2.5 Forecast-assimilation experiments: tuning and results

The results summarised in the third column of Table 2.2 represent the endpoint of a long tuning process that has involved the assessment of a large number of forecast-assimilation experiments until it was possible to select a single set of parameters that on the one hand delivered a well-tuned ensemble filter and on the other hand proved the relevance of our idealised configuration for convective-scale NWP data assimilation.

Despite working with an idealised system, the tuning of the system described in section §2.2.2 has required a large number of adjustments to the observing system (the number of observations p, their spacing d and their errors $s_{h,u,r}$), the cycling frequency, the resolution ratio between the forecast model and the nature run, and the filter configuration (e.g. the number of ensemble members N, the localisation and inflation parameters L_{loc}, γ_a and α_{RTPS}). For sake of simplicity, here we restrict our analysis to the variation of the filter parameters, keeping the observing system and the other aspects fixed, including the ensemble configuration illustrated in the previous section. A summary of the parameters varied in the forecast-assimilation experiments is given in Table 2.3.

The tuning process described in this section is aimed at producing the best forecast with a three-hour lead time. This time-scale lies within the definition of nowcasting for NWP forecasts (0-6 hours), that is the most relevant for convective-scale weather forecasting. We define as 'well-tuned' an experiment that satisfies three criteria:

(i) the ratio $SPR/RMSE$ is near 1,

(ii) the $RMSE$ is minimised and,

(iii) the $CRPS$ is minimised.

The definitions of spread (SPR), root mean squared error ($RMSE$) and the continuous ranked probabability score ($CRPS$) are found in section §2.4.2 and are calculated with reference to the nature run simulation.

2.5.1 Assessment of filter performance

Figure 2.6 displays a summary of the main statistics computed for the 180 simulations resulting from the combination of the assimilation parameters listed in Table 2.3. Each cell represents the average value computed over the whole domain, across all three variables[1], and for the entire duration of the experiments (after excluding the first 12 hours to remove any spin-up effects). The target values for a well-tuned simulation as set out by the criteria (i)-(iii) are highlighted by cells in white or light colors. Of course, the aim is to find an area of the parameter space that satisfies all the criteria at once.

We start by considering the first criterion, namely the ratio between the ensemble spread and the ensemble mean error, i.e. $SPR/RMSE$. Since it would be unreasonable to expect the experiments to return a value of exactly 1, we consider all simulations with a ratio within $[0.8, 1.2]$, which are highlighted in red in the top-left panel of Fig. 2.6. Among these, those showing the smallest values of RMSE and CRPS are outlined in black in the bottom panels of the same figure. As a result, we have restricted our search for a well-tuned experiment from 180 candidates to just 12.

The top-right panel of Fig. 2.6 shows the observational impact diagnostic (OID, see section §2.4.2) of each experiment. The values of OID as a function of γ_a and α_{RTPS} reflect the impact that inflating the forecast-error covariance matrix $\mathbf{P}_e^f$ has on the assimilation of the observations. Higher values of inflation (both multiplicative and additive) correspond to higher values of OID, as the Kalman filter relies more on the observations when the spread (and hence the uncertainty) in the forecast ensemble is increased. Conversely, small values of γ_a and/or α_{RTPS} lead to small values of OID as, in the absence of inflation, the ensemble spread becomes small, and so does the

---

[1] Before averaging over the variables, the SPR and $RMSE$ values of r are multiplied by a factor 100 to be comparable in magnitude with those of h and u.

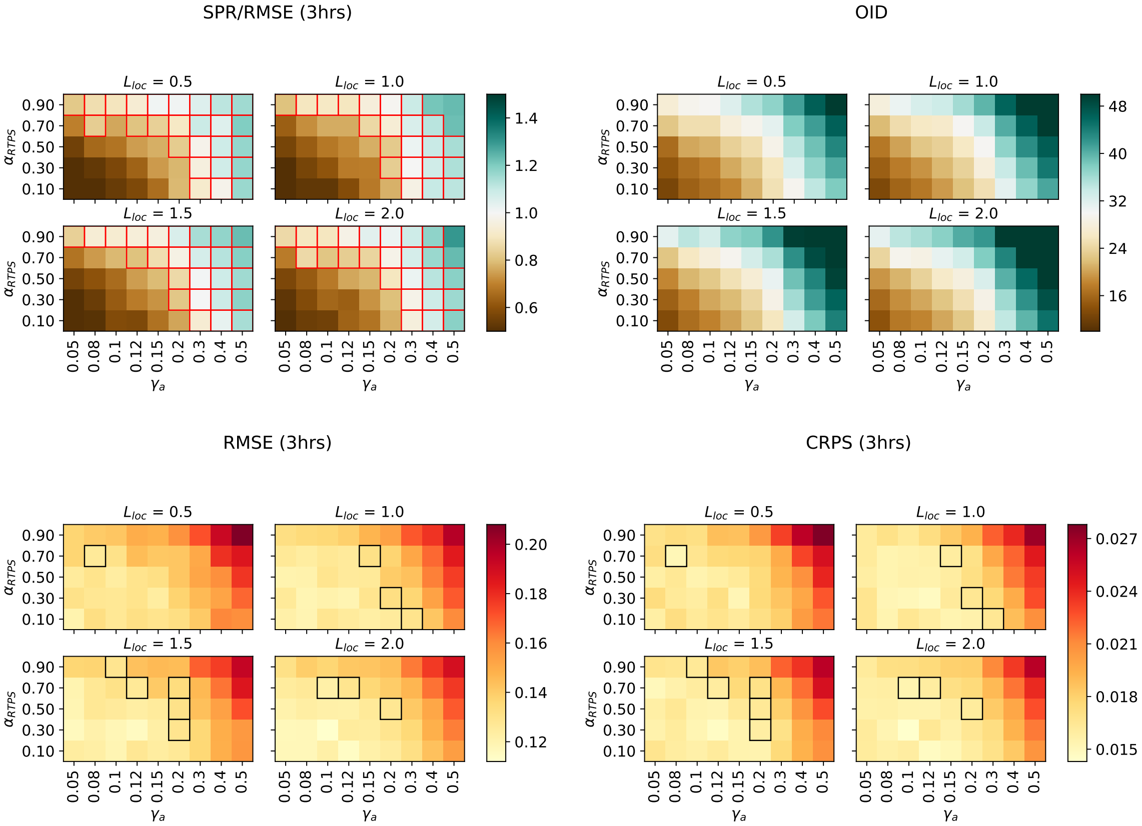
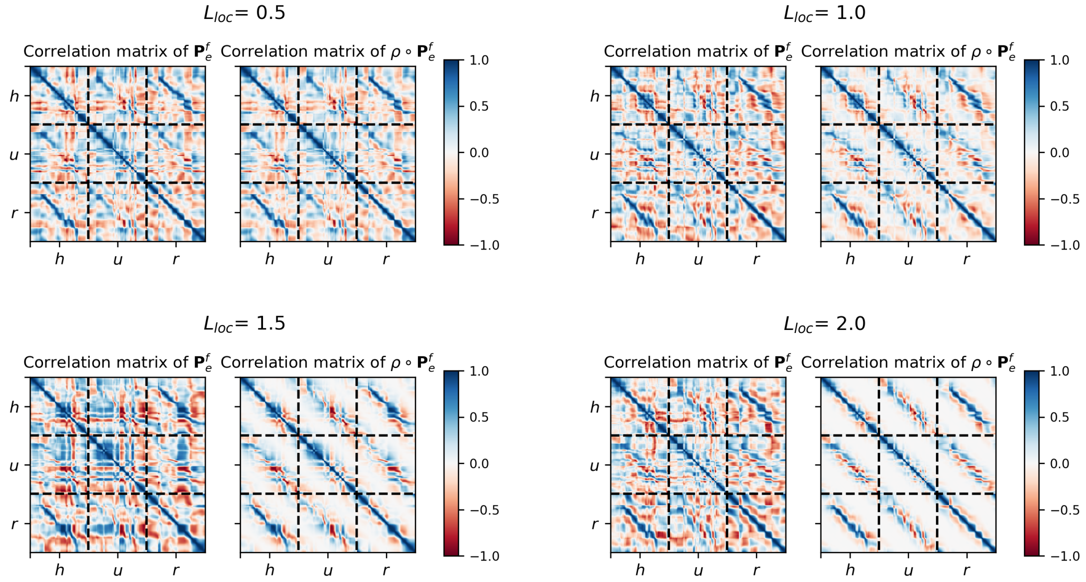


Figure 2.6: Summary of the results of the forecast-assimilation experiments for different statistics: $SPR/RMSE$ (top-left panel), OID (top-right panel), $RMSE$ (bottom-left panel) and $CRPS$ (bottom-right panel). Within a panel, each sub-panel indicates a different value of L_{loc}, and each cell represents a single experiment characterised by the pair (γ_a, α_{RTPS}) averaged over the whole duration of the experiment, domain and variables. The white/light cells indicate the target values of the corresponding statistic. The cells highlighted in red in the top-left panel indicate experiments with $0.8 < SPR/RMSE < 1.2$. Among these, the cells with low values of $RMSE$ and $CRPS$ are outlined in black in the bottom panels. Note that the OID is computed at the analysis time.



L_{loc}= 0.5

Correlation matrix of $\mathbf{P}_e^f$ Correlation matrix of $\rho \circ \mathbf{P}_e^f$

L_{loc}= 1.0

Correlation matrix of $\mathbf{P}_e^f$ Correlation matrix of $\rho \circ \mathbf{P}_e^f$

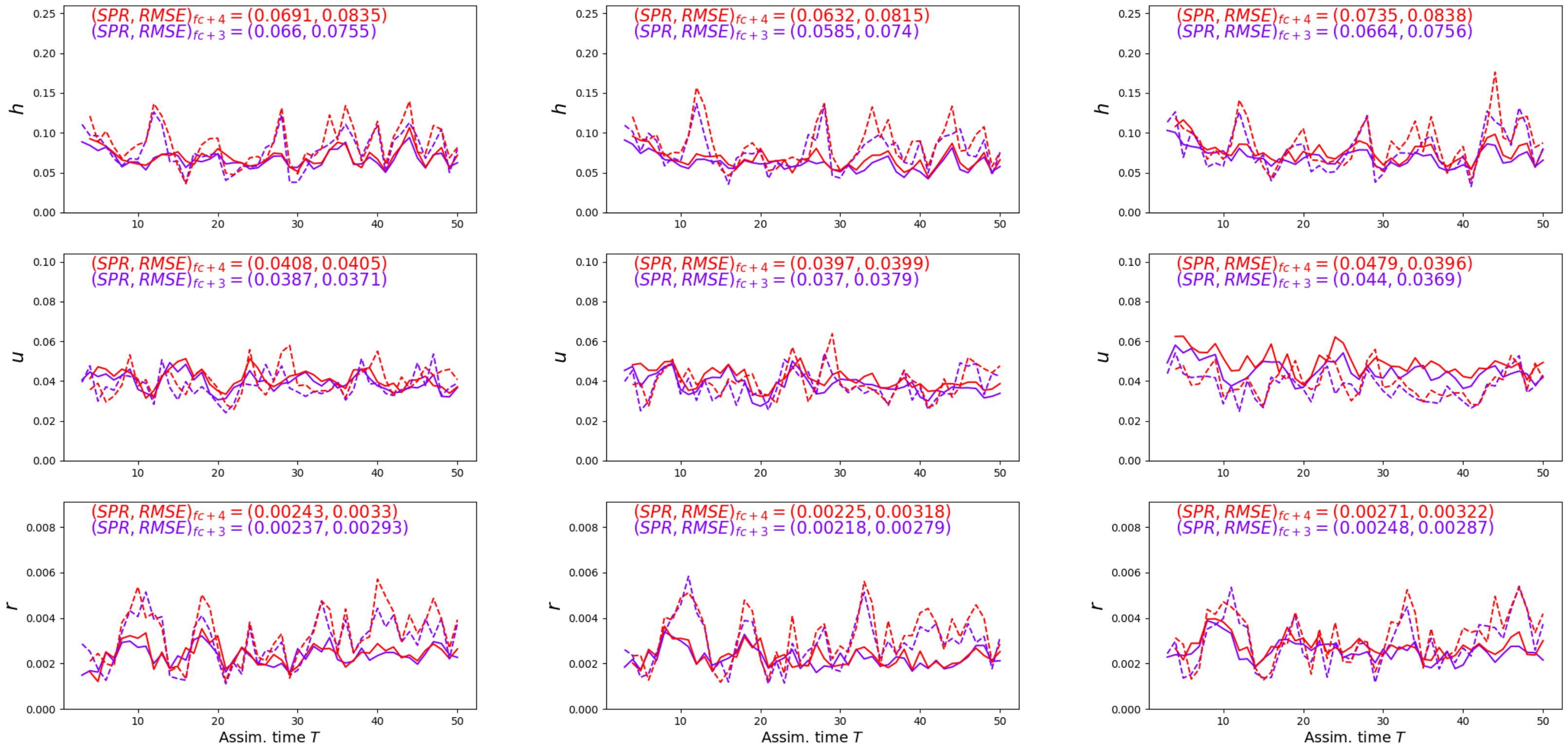
L_{loc}= 1.5

Correlation matrix of $\mathbf{P}_e^f$ Correlation matrix of $\rho \circ \mathbf{P}_e^f$

L_{loc}= 2.0

Correlation matrix of $\mathbf{P}_e^f$ Correlation matrix of $\rho \circ \mathbf{P}_e^f$

Figure 2.7: Effects of localisation (for different values of the localisation scale L_{loc}) on the forecast-error correlation matrices (derived from the covariance matrices $\mathbf{P}_e^f$ used in the DEnKF). Left: before localisation; right: after localisation. Values of γ_a and α_{RTPS}: $\gamma_a = 0.08$, $\alpha_{RTPS} = 0.7$ (top left), $\gamma_a = 0.15$, $\alpha_{RTPS} = 0.7$ (top right), $\gamma_a = 0.12$, $\alpha_{RTPS} = 0.7$ (bottom left), $\gamma_a = 0.12$, $\alpha_{RTPS} = 0.7$ (bottom right). Note that since the self-exclusion prescribes the computation of a different covariance matrix $\mathbf{P}^f_{e,\widehat{j}}$ for each j^{th} ensemble member (cf. §2.2.3, eq. (2.27)), an 'average' matrix is shown above.



(a) $L_{loc} = 1.0$, $\gamma_a = 0.15$, $\alpha_{RTPS} = 0.7$.

(b) $L_{loc} = 1.0$, $\gamma_a = 0.2$, $\alpha_{RTPS} = 0.3$.

(c) $L_{loc} = 1.0$, $\gamma_a = 0.3$, $\alpha_{RTPS} = 0.1$.

Figure 2.8: Domain-averaged time series of SPR (solid lines) and $RMSE$ (dashed lines) for 3hr forecasts (blue lines) and 4hr forecasts (red lines) of the three highlighted experiments from Fig. 2.8 with $L_{loc} = 1.0$ (parameters values reported in the subcaptions). Top panels: h; middle panels: u; bottom panels: r. Each panel reports at the top the time-averaged values (excluding the first 12 hours).

Experiment	Lead time	h	u	r	Average
$\gamma_a = 0.15$,	3hr	0.0755	0.0371	0.00293	
$\alpha_{RTPS} =$	4hr	0.0835	0.0405	0.00330	
0.7	% diff	9.6%	8.4%	11.2%	9.7%
$\gamma_a = 0.2$,	3hr	0.0740	0.0379	0.00279	
$\alpha_{RTPS} =$	4hr	0.0815	0.0399	0.00318	
0.3	% diff	9.2%	5.0%	12.2%	8.8%
$\gamma_a = 0.3$,	3hr	0.0756	0.0369	0.00287	
$\alpha_{RTPS} =$	4hr	0.0838	0.0396	0.00322	
0.1	% diff	9.8%	6.8%	10.9%	9.2%

Table 2.4: Summary of the $RMSE$ values (and percentage reduction between the 4hr and the 3hr forecasts) of the three experiments shown in Fig. 2.8.

forecast uncertainty. Although this quantity has not been used as a tuning criterion, it is worth noting how the simulations selected above display values of OID around 30%, which is fully within the range of values expected for an operational NWP system, in which $20\% \lessapprox OID \lessapprox 40\%$. This is a promising sign in terms of finding a configuration which both satisfies the criteria for a well-tuned filter and meets the conditions to be considered relevant for NWP data assimilation. We will return to these considerations in the next section.

The impact of localisation in Fig. 2.6 seems less relevant than that of inflation (although, this is not true in general and might depend on the observation spacing selected for these experiments). Therefore, in order to visualise better the impact of localisation on the experiments, Figure 2.7 shows the forecast correlation matrices[1] (at $T = 36$ hours) of four of the 12 simulations selected earlier, before and after the application of localisation, each one for a different value of the localisation length L_{loc}. The purpose of localisation should be one of damping the long-distance spurious correlations, which are an artefact produced by a finite ensemble rather than the result of

---

[1] Because of self-exclusion (cf. §2.2.3, eq. (2.27)), Fig. 2.7 reports the average of the N correlation matrices obtained by excluding one ensemble member at a time.

a true dynamical signal. At the same time, it is also important to preserve the actual intra-correlations present in the system. In this regards, Fig. 2.7 shows how values of L_{loc} bigger than one (bottom panels) might go too far in suppressing the correlations away from the diagonal bands, whereas a value of $L_{loc} = 0.5$ has a very limited effect, leaving the matrix almost unaffected. As a compromise, we decide to retain only the simulations with $L_{loc} = 1.0$. As we have seen in Fig. 2.4 (and Appendix B), this value of L_{loc} corresponds to a localisation length scale of 500 km (i.e. 200 grid points), which is also in line with the values used in NWP systems (cf. Table 2.2).

At this point, the tuning process has reduced the number of experiments to just three (see the outlined cells in Fig. 2.6 with $L_{loc} = 1.0$). However, there is one essential aspect that has not been taken into account yet, namely the impact that the data assimilation has on reducing the forecast error. Fig. 2.8 shows the time series of both SPR (solid line) and RMSE (dashed line) for each variable of the three remaining simulations, computed for lead times of 3 hours (blue lines) and 4 hours (red lines) valid at the same analysis time. Each panel reports also the average values (computed by excluding the first 12 hours) and Table 2.4 complements this information with the RMSE percentage difference between the two lead times. As one would expect, the forecast with the shorter lead time (i.e. 3 hours) is more accurate (i.e. displays a smaller RMSE), as it has been initialised more recently with observations generated from the nature run. The results show a reductions of RMSE in the range of $5-12\%$ between the variables (with rain showing the biggest improvements) and around 10% on average across the three experiments (cf. Table 2.4).

The oscillatory behaviour of both SPR and RMSE time series in Fig. 2.8 is a sign of the time-dependence of the forecast error, accentuated in our configuration by the periodic domain, which allows the continuous circulation of gravity waves. Yet, these variations look stationary in time, and this aspect makes it possible to exclude a 'filter divergence' situation in which the error increases in time while the ensemble spread collapses. This would be a sign that the data assimilation has failed, as the forecasts would then become at the same time inaccurate (large errors) and overconfident (little spread), leading the observations to be discarded entirely.

One final consideration on Fig. 2.8 can be made about the $SPR/RMSE$ ratio. While the top-left panel of Fig. 2.6 presented a single number resulting from averaging the experiments in time and space, Fig. 2.8 displays the same information in a more granular way. The match between SPR and RMSE (producing a ratio close to unity) is still visible, although there are clear differences between the variables. In particular, h

and r seem to be more underspread than u (i.e. $SPR/RMSE < 1$ for h and r, while $SPR/RMSE \sim 1$ for u); this fact makes sense from a dynamical point of view, as h and r are the most non-linear variables in the system and the Kalman filter might struggle to capture their spread correctly, especially in regions of strong convection. Overall, these results are encouraging when compared with those obtained for operational systems (see, for instance, Figures 3-8 and 13-14 in Bowler et al. (2017)).
Finally, in order to select one of the three experiments left, we choose the simulation displaying the biggest reduction in RMSE in Table 2.4, that is, the one with $\gamma_a = 0.15$ and $\alpha_{RTPS} = 0.7$ (and $L_{loc} = 1.0$). We will examine this selected experiment further in the next section, where we will also analyse in detail its relevance for NWP data assimilation.

2.5.2 Validation and relevance for convective-scale NWP

Until now, our analysis has focussed on various diagnostics and statistics that have helped us understand how well the ensemble forecast was performing and therefore which experiments could be considered as ‘well-tuned’. Here, we proceed with the next step, that is, assessing the relevance of the tuned idealised system that we have built for NWP data assimilation. To this aim, we will use two additional validation diagnostics:

- the observational influence diagnostic (OID), and
- the error-doubling times statistics.

The definitions of both diagnostics can be found in section §2.4.2.

Observational influence diagnostic (OID)

The OID for each variable and for all observations is calculated at each analysis step (cf. (2.46)) and its value as a function of the assimilation time is dispalyed in Figure 2.9. The OID is flow-dependent and varies depending on the forecast uncertainty (as the observation error is constant in time, see §2.2.2). For example, a large OID is expected in situations when the uncertainty in the forecast is large and the Kalman filter must rely heavily on the observations, which will have a stronger impact on the final analysis. On the other hand, the OID will be smaller when the forecast has a lower uncertainty (i.e. a smaller spread), leading the observations to have less influence.

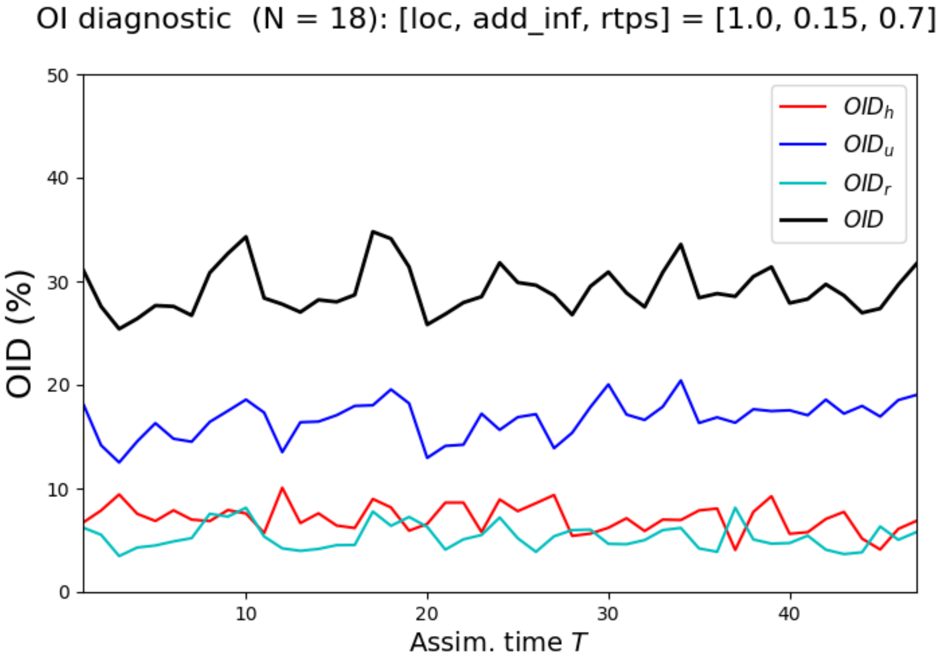


Figure 2.9: Time series of the observational influence diagnostic (OID), as a percentage (%). The total OID and the OID computed for the single variables are reported in different colours (see legend). Values on the x-axis are in hours.

Indeed, the time series of OID seen in Figure 2.9 displays these 'hourly weather' oscillations, with the total OID fluctuating between $25-35\%$. The influence of the u observations on the final analysis seems to be the dominating contribution, with values between $10-20\%$, followed by the impact given by h and r, each oscillating between $5-10\%$. The predominance of the u observations over h and r is probably a consequence of the specific observing system chosen for our experiments.
The value of total OID matches the estimates for operational systems of $20\% \lessapprox OID \lessapprox 40\%$, demonstrating the relevance of our forecast-assimilation experiments in the context of NWP data assimilation.

Error-doubling time statistics

In order to evaluate the error-doubling time statistics of the selected experiment, we have run a 24 hour forecast for each of the $N = 18$ members in the ensemble, repeated for the first 25 consecutive analysis cycles. This gives a sample of $25 \cdot 18 = 450$ forecasts, each lasting 24 hours. We then calculate the initial analysis RMSE and compute the number of hours T_d that it takes for this error to double, cf. (2.47). Figure 2.10 shows the histogram of the distribution of the doubling times T_d for each variable. We note that while h and u display very similar values (around 9 hours), the error in r grows at a faster rate (approximately every 6 hours). This is not surprising, given the high non-linearity of r.
Doubling times of 6 to 9 hours are of the same order of magnitude (i.e. $\mathcal{O}(1)$ hours) as those found for convection-resolving NWP models (Hohenegger & Schar, 2007). It is also worth noting that in a previous round of experiments with $N_{el}^{nat} = 800$ (as well as in Kent (2016)) the doubling time was approximately half of the values shown in Fig. 2.10 (i.e. 4 hours). This behaviour is not surprising given the use of a nature run simulation with a higher spatial resolution, in which the small scales features are better resolved. As the observations are generated from the nature run, the analysis will end up resolving those small scales better; however, small-scale errors grow faster than large-scale ones, causing a reduction in the error doubling times when a higher ratio between nature run and forecast resolution is adopted.

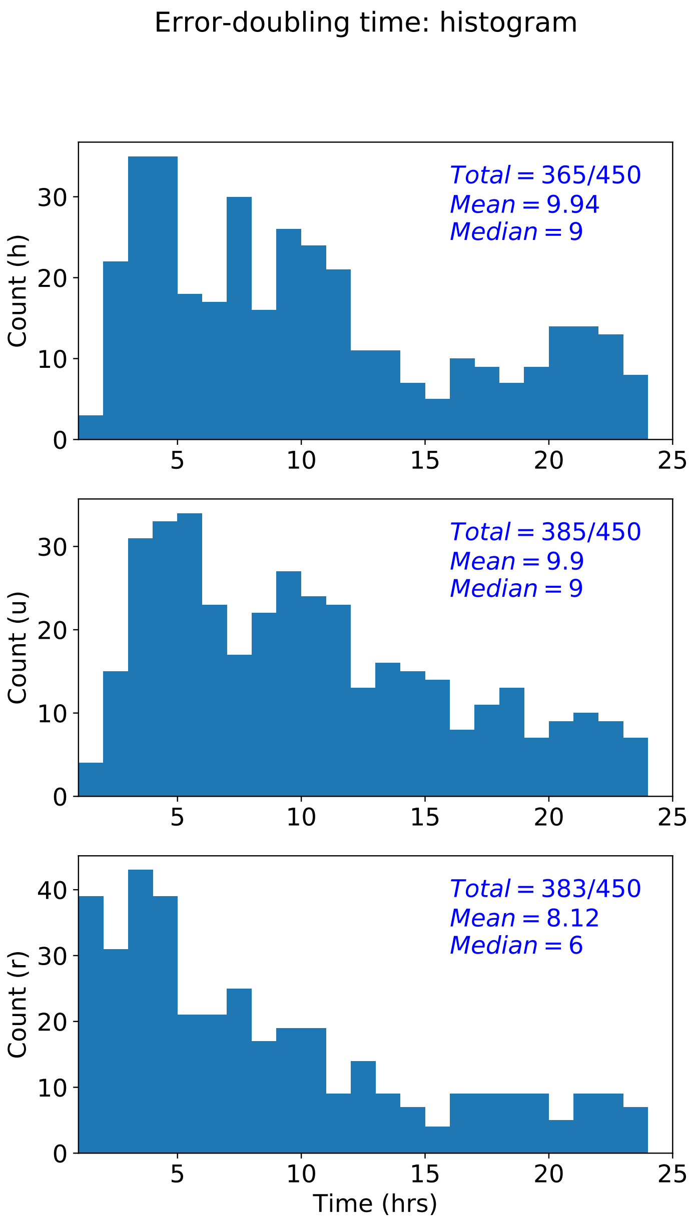


Figure 2.10: Error-doubling times distribution based on 450 24hr forecasts. Top panel: h; middle panel: u; bottom panel: r. Each panel reports the number of forecasts in which the initial time has doubled within 24 hours, the mean and the median of the distribution.

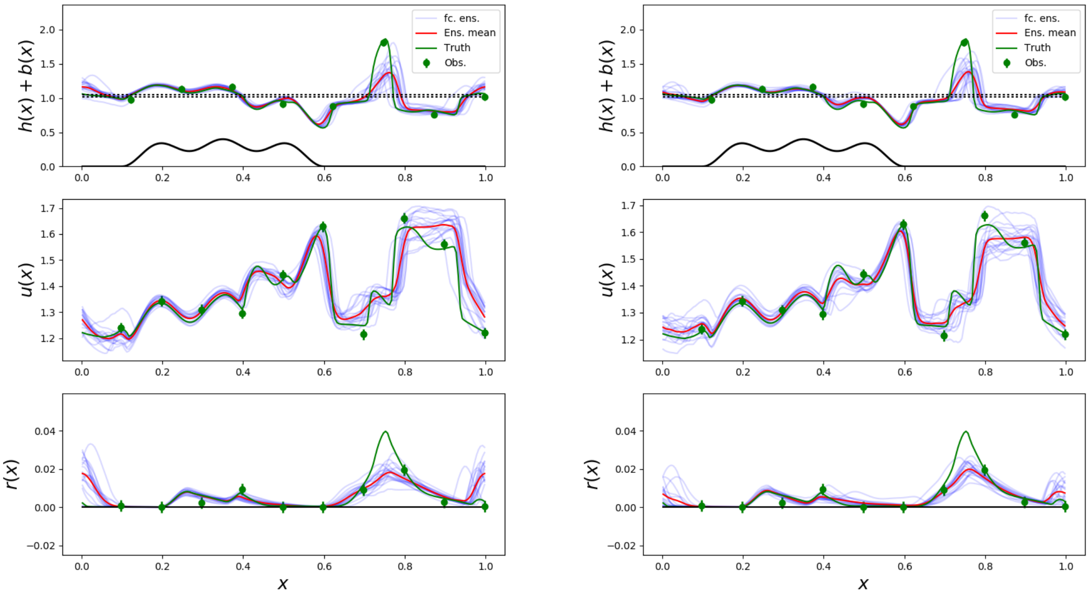


Figure 2.11: Model snapshot valid at $T = 40$ hours. Left: 4hr forecast starting at $T = 36$ hours. Right: 3hr forecast starting at $T = 37$ hours. The nature run trajectory is in green, the ensemble trajectories are in blue and the forecast mean is in red. The observations (at $T = 40$ hours) – shown in comparison only and not yet assimilated – are indicated by the green dots with error bars.

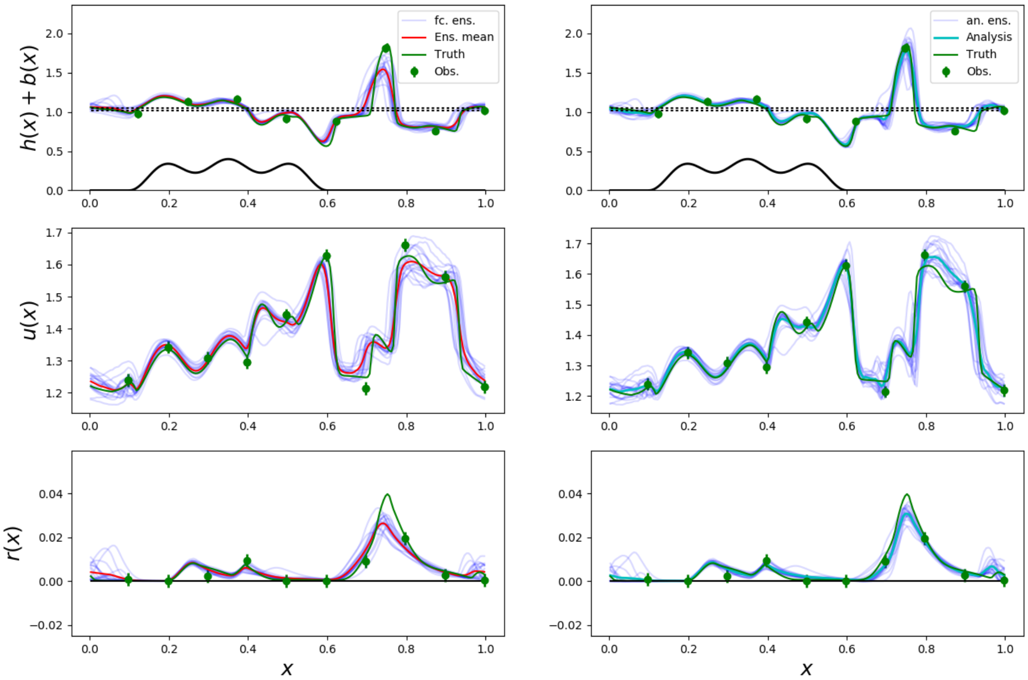


Figure 2.12: Model snapshot valid at $T = 40$ hours. Left: 1hr forecast starting at T=39 hours. Right: analysis. The nature run trajectory is in green, the ensemble trajectories are in blue, the forecast mean is in red and the analysis mean is in cyan. The observations are indicated by green dots with error bars and have been assimilated in the trajectories in the right column but not for those in the left.

2.5.3 Subjective verification

In this final section, we examine a couple of snapshots of the model dynamics drawn from the selected experiment (i.e. $L_{loc} = 1.0$, $\gamma_a = 0.15$, $\alpha_{RTPS} = 0.7$). These snapshots are depicted in Figures 2.11-2.12 and show: the nature run solution (green lines), the observations drawn from it (green dots), the topography (black line), the forecast mean (red line), the analysis mean (cyan line) and the individual ensemble members (blue lines). The purpose of this analysis is purely qualitative, as we have already evaluated this experiment from a quantitative point of view in the previous sections.
Figure 2.11 compares the forecast launched at $T = 37$ hours with a 3-hour lead time (right panel) with the forecast launched at $T = 36$ hours with a 4-hour lead time (left panel). Clearly, both forecasts are valid at the same time (i.e. $T = 40$ hours), but the one with the shorter lead time is expected to be more accurate, thanks to its more recent initialisation.
The dynamical situation defined by the nature run simulation (green lines) at $T = 40$ hours is as follows. The flow is moving eastward ($u > 0$) and the strongest area of convection and rain ($h > H_r > H_c$) is occurring downstream of the topography, around $x = 0.75$. Convection and rain are also present in correspondence to the topography (i.e. $0.2 < x < 0.4$) and at the end of the domain (around $x = 1$). The velocity field u displays large gradients and oscillations, mostly in the second half of the domain.
Both forecasts are accurate (i.e. similar to the nature run, with small spread) away from the areas with convection and precipitation. Conversely, they seem to struggle in representing the intensity (although not the position) of the convective updraft at $x = 0.75$ and the area near the thresholds around $x = 1$. In the first case, there is almost no difference in h and r between the 3 and the 4 hour forecasts, implying little to no improvement produced by the assimilation of the observations at $T = 37$ hours; on the other hand, the impact of data assimilation is more noticeable at $x = 0/1$, where both forecasts produce spurious precipitation due to some of the ensemble members exceeding the rain threshold. This precipitation is not present in the nature run (green line), and although it is produced in both forecasts, its intensity is significantly diminished in the more recent one, which is therefore more accurate. Overall, looking back at Fig. 2.8, we notice that at $T = 40$ hours, the improvement given by the 3 hour forecast is larger in u and r than h.
Finally, we consider Figure 2.12, in which are plotted the forecasts initialised at $T = 39$ hours and valid at $T = 40$ hours (left panel) and the analysis at the same time (right

panel). This comparison is used to highlight the direct impact of data assimilation, in that the analysis on the right is the product of the combination of the observations (green dots) with the forecast on the left. First, we notice how even the 1 hour forecast struggles in representing the convection updraft at $x = 0.75$ (although this feature is greatly improved with respect to the 3 and 4 hour forecasts). In this case, however, the effect of the assimilation is strong, and the analysis captures the intensity of the peak in h very well, thanks to h being observed exactly at the same location. In addition, the benefit of adjusting h towards the nature run solution impacts positively also on the other variables; for example, the physical correlation between h and r in the covariance matrix $\mathbf{P}_e^f$ (cf. Fig. 2.7, top-right panel, although valid at a different time) means that even if r is not observed in the same position, its maximum is also revised accordingly. In general, the analysis is visually more accurate but also more precise (that is, the spread is reduced) with respect to the forecast on the left.

One final feature can be observed in the velocity field u: the observation at location $x = 0.7$ forces the analysis away from the nature run trajectory. This is not totally unexpected, as the observations are generated by perturbing the nature run with random perturbations (cf. (2.13)), and it is sometimes the case that inaccurate observations produce inaccurate analyses.

Chapter highlights and summary

- Idealised models constitute a common and practical tool used to facilitate research in data assimilation for NWP models. The use of idealised models represents a numerically inexpensive alternative to full-scale operational systems. Disappointingly, the ability of these idealised models to reproduce the features and the characteristics of operational NWP systems is rarely explored. The work summarised in this chapter has tried to filled this gap by developing a protocol in which the properties of an idealised forecast-assimilation system are systematically compared against those commonly found in an operational system (cf. Table 2.2). This protocol can be adapted and applied to other idealised models.

- The modified rotating shallow water (modRSW) model presented in Kent (2016); Kent et al. (2017) is an example of idealised models developed to facilitate data assimilation research. The modRSW model constitutes our starting point to conduct satellite data assimilation re-

search. As a preliminary step, in this chapter we have summarised the dynamics and the numerics of the modRSW model and have discussed a series of forecast-assimilation experiments realised with such model, using a Deterministic Ensemble Kalman Filter (DEnKF). By applying the protocol mentioned above, we have shown that our idealised experiments reflect a configuration which is both well-tuned and relevant for NWP convective-scale data assimilation.
Whether the modRSW model is suitable to conduct satellite data assimilation research will be discussed at the beginning of the next chapter.

Chapter 3

The ismodRSW model

In this chapter we present a revised version of the modRSW model introduced in Chapter 2. A new, isentropic $1\frac{1}{2}$-layer modified shallow water model (ismodRSW) will pave the way for further data assimilation experiments featuring idealised satellite observations. After showing that the modRSW model is not optimal for the purpose, we discuss both the dynamics and the numerics of the new model, completing the chapter with a comparison against an analytical solution and a protoype nature run simulation to be used in the subsequent experiments of chapter §5.

3.1 Motivation

One of the most common types of satellite data assimilated in operational NWP systems is passive observations of emitted thermal radiation coming from the Earth's surface and atmosphere. This radiation can be related to the temperature of the emitting source (and vice versa) by exploiting the principles of radiative transfer and blackbody radiation. In view of adopting a similar approach to generate idealised satellite observations later on (this topic will be covered extensively in Chapter 4), it is essential that either (i) the model includes temperature among the prognostic variables or (ii) one (or more) of the model variables can be readily related to temperature. In this regard, the modRSW model is not particularly well-suited, since it does not include temperature among its prognostic variables, and even though a physical relationship between fluid depth and temperature can be formulated, this leads to some scaling issues in which DA-relevant model dynamics can be achieved only for unrealistic values of the temperature. In order to visualise this limitation, we consider a straightforward diagnostic equation between the dimensionless fluid depth h and its dimensional temperature T based on the ideal gas law ($p = R\rho T$) and hydrostatic equilibrium ($p = \rho g H h$), that is:

$$T = T_0 h, \qquad \text{with} \quad T_0 = gH/R, \tag{3.1}$$

in which g is the acceleration due to gravity ($g = 9.81\,\mathrm{m\,s^{-2}}$), R is the specific gas constant for dry air ($R = 287\,\mathrm{J\,kg^{-1}\,K^{-1}}$) and H the scale height of the fluid. The scaling used in both Kent et al. (2017) and Chapter 2 – i.e. $gH = 330\,\mathrm{m^2 s^{-2}}$, or $H \approx 34\,\mathrm{m}$ – is clearly based on an unrealistic fluid depth H which leads to an unreasonable scale temperature of $T_0 \simeq 1.1\,\mathrm{K}$. This configuration is clearly not suitable for satellite data assimilation purposes and highlights the need for a different approach.

In a paper by Pan & Smith (1999), an isopycnal $1\frac{1}{2}$-layer shallow water model is used to investigate gap winds and wakes in the presence of orography. Although the purpose of the study is different from ours, the fact that the authors used a shallow water model and based their scaling on real atmospheric observations makes this study attractive and interesting. Remarkably, despite the use of an isopycnal model, they employed the *isentropic* definition of the reduced gravity g', that is:

$$g' = \frac{\theta_1 - \theta_2}{\theta_1} g, \tag{3.2}$$

to compute the Froude number utilised in their numerical simulations. On the one hand, this combined use of an isopycnal model with the isentropic definition of the reduced gravity seems to guarantee a simple but realistic testbed for their numerical experiments. On the other hand, this choice cannot be seen as an entirely consistent one. It is precisely to remove this inconsistency that we decided to develop an entirely consistent $1\frac{1}{2}$-layer isentropic model.

The newly derived model, hereafter denoted as *ismodRSW* (for isentropic modified Rotating Shallow Water), does not suffer from the scaling issues discussed above, as it is naturally equipped with a physically consistent definition of fluid temperature linked to the definition of the potential temperature θ:

$$\theta = T\left(\frac{p_r}{p}\right)^{\kappa} \Longrightarrow T = \theta\left(\frac{p}{p_r}\right)^{\kappa} = \theta\eta^{\kappa}, \tag{3.3}$$

in which p is the pressure, p_r a reference pressure and $\kappa = R/C_p$ the ratio between the specific gas constant for dry air and its specific heat capacity at constant pressure ($c_p = 1004\ \mathrm{J\,kg^{-1}K^{-1}}$), with $\eta = p/p_r$ a non-dimensional pressure. The expression $\Pi = \eta^\kappa$ is often referred in the literature as the *Exner function*.

On a further note, the transition from a single to a $1\frac{1}{2}$-layer model offers additional benefits for the modelling of idealized satellite observations. Real satellite observations are radiance measurements shaped by several processes (emission, absorption and scattering) taking place throughout the atmosphere before the radiance reaches the satellite. In this sense, the degree of vertical complexity with which the atmosphere can be modelled plays a crucial role in mimicking the most relevant features of real satellite observations.

3.2 Modifications

The ismodRSW model is obtained by performing the following modifications to the modRSW model:

i. the single-layer configuration is replaced with a $1\frac{1}{2}$-layer model;

ii. the isopycnal fluid (with its uniform density) is replaced with an isentropic fluid, in which the potential temperature is constant in each layer.

In general, $1\frac{1}{2}$-layer models are derived from two-layers models by imposing the preservation in time of a uniform total fluid depth (also called a rigid lid approximation). This applies, for example, to isopycnal models, in which this modification leads to the sole replacement of the acceleration due to gravity with the isopycnal reduced gravity g' in the single-layer equations (see, for example, Vallis (2017), §15.2.2). However, as we will show in the next section, in the case of a two-layer isentropic shallow water model this approach reveals an apparent inconsistency. Ultimately, the formal derivation of an isentropic $1\frac{1}{2}$-layer shallow water model based on principles of asymptotic and Hamiltonian fluid dynamics is able to resolve the contradiction, showing that the fluid in the upper layer needs to be at rest in order for the $1\frac{1}{2}$-layer model to emerge. This derivation is quite long and technical and goes beyond the scope of this thesis, therefore we refer to Bokhove et al. (2021) for full details.

Lastly, it is important to note here that there are no further changes to the model setup apart from the two modifications mentioned above: the essential functioning and dynamics of the model remain similar to those of the modRSW model, including the threshold mechanism used to simulate convection and precipitation (see also section §3.4.1). However, the presence of bottom topography will not be considered for the isentropic model.

3.3 $1\frac{1}{2}$layer isentropic shallow water

In this section we illustrate multiple aspects related to the isentropic $1\frac{1}{2}$-layer shallow water model. The full ismodRSW model (including convection and precipitation) will be presented in full in section §3.4.

3.3.1 A rigid-lid approximation in a 2-layer model

Starting from an isentropic two-layer model, it is possible to show how a closed $1\frac{1}{2}$-layer model emerges by taking a seemingly inconsistent rigid-lid approximation. That the final model is nonetheless consistent goes beyond the scope of this thesis and is shown in Bokhove et al. (2021).

A full, geometric derivation of an isentropic N–layer model can be found in Bokhove & Oliver (2009). Here, we take a two-layer simplification thereof, with $N = 2$ (see top image in Figure 3.1). The momentum equations of this model arise by assuming hydrostatic balance and constant entropy (i.e. constant potential temperature θ) in each layer. The continuity equations emerge once the space (x, y) and time-dependent (t) pseudo-density $\sigma_\alpha(x, y, t)$ for each layer, numbered by $\alpha = \{1, 2\}$, is defined, i.e.:

$$\sigma_\alpha = p_r\,(\eta_\alpha - \eta_{\alpha-1})/g, \tag{3.4}$$

in which g refers to the gravity acceleration, $\eta_\alpha = p_\alpha/p_r$ is the non-dimensional pressure at the bottom of the layer α and p_r indicates a reference pressure. The pseudo-density σ arises from the hydrostatic balance $\mathrm{d}p = -\rho\, g\, \mathrm{d}z$, by integrating an element of mass flux for some infinitesimal surface element $\mathrm{d}A$: $\mathrm{d}m/\mathrm{d}A = \rho\,\mathrm{d}z = -\mathrm{d}p/g$ across each layer. In Bokhove (2002) and Ripa (1993) the variational and Hamiltonian formulation of the isentropic N–layer equations are derived by simplifying the Eulerian variational principle of the compressible Euler equations.

The resulting four, isentropic two-layer (continuity and momentum) equations read as

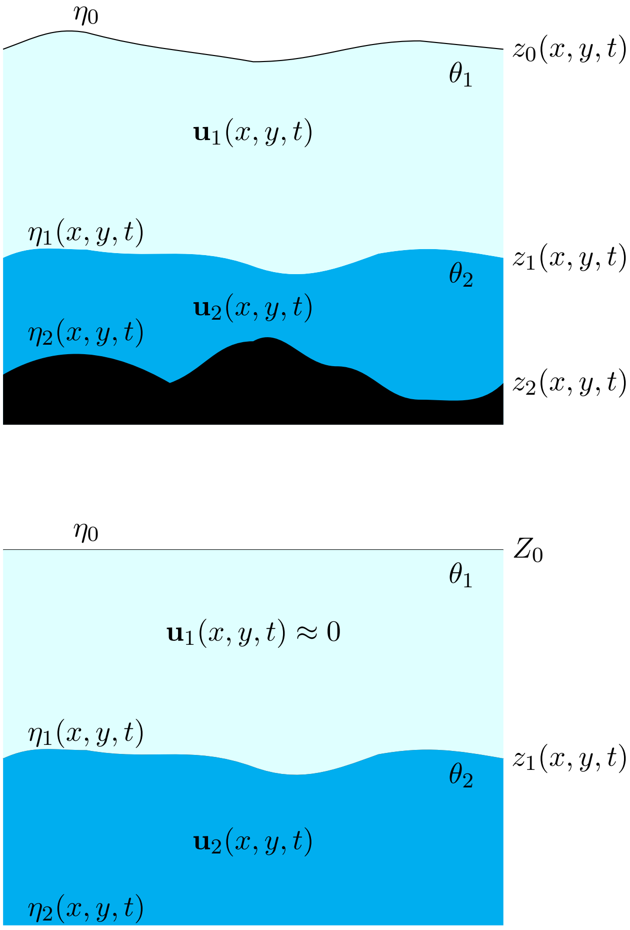


Figure 3.1: Schematic representation of a two-layer isentropic shallow water model (top) and of a $1\frac{1}{2}$-layer isentropic shallow water model (bottom) without topography ($b = 0$).

follows:

$$\partial_t \sigma_\alpha + \nabla \cdot (\sigma_\alpha \mathbf{v}_\alpha) = 0, \tag{3.5.a}$$

$$\partial_t \mathbf{v}_\alpha + (\mathbf{v}_\alpha \cdot \nabla)\mathbf{v}_\alpha + f {\mathbf{v}_\alpha}^\perp = -\nabla M_\alpha, \tag{3.5.b}$$

with $\alpha = \{1, 2\}$ and in which: ∇ is the horizontal gradient, $\mathbf{v}_\alpha = \mathbf{v}_\alpha(x, y, t) = (u_\alpha, v_\alpha)^T$ is the horizontal velocity within layer α and $\mathbf{v}_\alpha^\perp = (-v_\alpha, u_\alpha)^T$ the vector perpendicular to it, f is the Coriolis frequency, and M_α is the Montgomery potential. In order to close the system, one needs to specify the Montgomery potentials in each layer. As seen in section §3 of Bokhove & Oliver (2009), for a two-layer model these potentials can be defined as:

$$M_1 = c_p\, \theta_2\, {\eta_2}^\kappa + c_p(\theta_1 - \theta_2)\eta_1^\kappa + g z_2, \tag{3.5.c}$$

$$M_2 = c_p\, \theta_2\, {\eta_2}^\kappa + g z_2. \tag{3.5.d}$$

The hydrostatic condition for an isentropic model $\partial M/\partial z = 0$ implies that, in general, the Montgomery potential $M = c_p \theta \eta^\kappa + gz$ is independent of z within each layer. Therefore, one can evaluate M in the bottom layer (where $\theta = \theta_2$) at both $z = z_2$ and $z = z_1$, and M in the upper layer (where $\theta = \theta_1$) at both $z = z_1$ and $z = z_0$, to find:

$$g z_0 = c_p\, \theta_1\, (\eta_1^\kappa - \eta_0^\kappa) + g z_1, \tag{3.6.a}$$

$$g z_1 = c_p\, \theta_2\, (\eta_2^\kappa - \eta_1^\kappa) + g z_2, \tag{3.6.b}$$

from which is possible to express the thickness of each layer as:

$$h_1 = z_0 - z_1 = (c_p\, \theta_1/g)\, (\eta_1^\kappa - \eta_0^\kappa), \tag{3.7.a}$$

$$h_2 = z_1 - z_2 = (c_p\, \theta_2/g)\, (\eta_2^\kappa - \eta_1^\kappa). \tag{3.7.b}$$

We note here that the non-dimensional pressure η_0 is treated as a constant throughout this thesis.

Finally, the relations between layer pressure and pseudo-densities can be derived using the expressions (3.4) for σ_1 and σ_2 as follows:

$$\eta_1 = g\, \sigma_1/p_r + \eta_0 \qquad \text{and} \qquad \eta_2 = g\, (\sigma_1 + \sigma_2)/p_r + \eta_0. \tag{3.8}$$

When one takes a rigid-lid approximation, it is convenient to add a constant $K = -(c_p \theta_1 \eta_0^\kappa + g Z_0)$ to M_1 in (3.5c), leading to:

$$M_1 = c_p\, \theta_1\, (\eta_1^\kappa - \eta_0^\kappa) + c_p\, \theta_2\, (\eta_2^\kappa - \eta_1^\kappa) + g z_2 - g Z_0. \tag{3.9}$$

Therefore, by substituting (3.6.b) into (3.6.a) and subtracting gZ_0 from both sides one finds:

$$\begin{aligned} gz_0 &= c_p\,\theta_1\,(\eta_1^\kappa - \eta_0^\kappa) + c_p\,\theta_2\,(\eta_2^\kappa - \eta_1^\kappa) + gz_2, \\ gz_0 - gZ_0 &= c_p\,\theta_1\,(\eta_1^\kappa - \eta_0^\kappa) + c_p\,\theta_2\,(\eta_2^\kappa - \eta_1^\kappa) + gz_2 - gZ_0, \\ gz_0 - gZ_0 &= M_1. \end{aligned}$$

If the top surface is fixed, i.e. $z_0(x, y, t) = Z_0 = \text{const}$, then $M_1 = g\,(z_0 - Z_0) = 0$, and a closed $1\frac{1}{2}$-layer model emerges, as is illustrated in the next section. However, this condition alone produces an inconsistency, since the constraint $M_1 = 0$ is not preserved in time by the two original continuity equations. In fact, as it is shown in Bokhove et al. (2021), a closed and fully consistent $1\frac{1}{2}$-layer model results after taking $M_1 = 0$ and $\mathbf{v}_1 = 0$ in the momentum equation of the upper layer, with (3.5) still valid for the bottom layer (with $\alpha = 2$).

3.3.2 The closed system

In Bokhove et al. (2021) the authors derive an asymptotically consistent $1\frac{1}{2}$-layer shallow water model in which an isentropic shallow layer of fluid at potential temperature θ_2 lies below a second (relatively) motionless layer at potential temperature θ_1 ($\theta_1 > \theta_2$) capped by a rigid lid. The non-dimensional closed set of equations for such a system reads:

$$\partial_t \sigma_2 + \nabla\cdot(\sigma_2 \mathbf{v}_2) = 0, \tag{3.10.a}$$

$$\partial_t \mathbf{v}_2 + (\mathbf{v}_2\cdot\nabla)\mathbf{v}_2 + f{\mathbf{v}_2}^{\perp} = -\nabla M_2, \tag{3.10.b}$$

$$M_1 = 0 \qquad \text{and} \qquad \mathbf{v}_1 = 0; \tag{3.10.c}$$

in which σ_2 is the pseudo-density in the bottom layer defined as (cf. eq. (3.4)):

$$\sigma_2 = \frac{p_r}{g}\,(\eta_2 - \eta_1)\,. \tag{3.11}$$

For the purpose of this thesis, we apply two simplifications to the system above: (i) we assume flat bottom conditions (i.e. $z_2 = 0$ in (3.5d)) and (ii) flow independence in the meridional direction (i.e. $\partial_y = 0$). A sketch of this model is given in the bottom image of Figure 3.1. Given these two assumptions, the equations for (σ_2, u_2, v_2) read as:

$$\partial_t \sigma_2 + \partial_x(\sigma_2 u_2) = 0, \tag{3.12.a}$$

$$\partial_t u_2 + u_2 \partial_x u_2 - f\, v_2 = -\partial_x M_2, \tag{3.12.b}$$

$$\partial_t v_2 + u_2 \partial_x v_2 + f\, u_2 = 0. \tag{3.12.c}$$

The system of equations (3.12) is closed since the non-dimensional pressure η_2 in (3.5d) is linked to the pseudo-density σ_2 via its definition (3.11), with η_1 being

$$\eta_1 = \left[\frac{\theta_2}{\triangle\theta} \left(-\eta_2^{\kappa} + \frac{\theta_1}{\theta_2} \eta_0^{\kappa} + \frac{g}{c_p \theta_2} Z_0 \right) \right]^{\frac{1}{\kappa}}, \tag{3.13}$$

in which η_0 is the (constant) non-dimensional pressure acting on the upper lid, $\Delta\theta = \theta_1 - \theta_2$ is the difference in potential temperature between the layers and $Z_0 = h_1 + h_2$ represents the total depth of the fluid. We note that such a $1\frac{1}{2}$-layer model has the advantage over a one-layer model that the non-dimensional pressure η_1 is active and not constrained to be constant, as is η_0. The expression (3.13) is obtained by isolating η_1 after imposing $M_1 = 0$ (and $z_2 = 0$) in (3.9). A typical function relating σ_2 to η_2 is shown in Fig 3.2.

3.3.3 The conservative hyperbolic system

The numerical implementation of the nonlinear hyperbolic system (3.12) can be facilitated when it is written in the following conservative form:

$$\partial_t \mathbf{U} + \partial_x \mathbf{F}(\mathbf{U}) + \mathbf{T}\left(\mathbf{U}\right) = 0. \tag{3.14}$$

In order to write the system (3.12a-3.12c) in conservative form, we start by multiplying (3.12b) and (3.12c) by σ_2. Replacing M_2 in the momentum equation with (3.5d) and after some manipulation (herein we drop the subscripts for σ, u, v and η), we obtain:

$$\partial_t \sigma + \partial_x \left(\sigma u\right) = 0, \tag{3.15.a}$$

$$\partial_t \left(\sigma u\right) + \partial_x \left(\sigma u^2\right) - f\, \sigma v = -c_p\, \theta_2\, \sigma \partial_x(\eta^{\kappa}), \tag{3.15.b}$$

$$\partial_t \left(\sigma v\right) + \partial_x \left(\sigma u v\right) + f\, \sigma u = 0. \tag{3.15.c}$$

The right-hand-side of eq. (3.15b) can be rewritten as the composite derivative of an unknown function $\mathcal{E}(\eta(x,t))$:

$$-\, \partial_x \mathcal{E}\left(\eta(x,t)\right) = -\partial_\eta \mathcal{E} \partial_x \eta \quad \text{with} \quad \frac{\partial \mathcal{E}}{\partial \eta} = \kappa\, c_p\, \theta_2\, \sigma(\eta)\, \eta^{\kappa-1}, \tag{3.16}$$

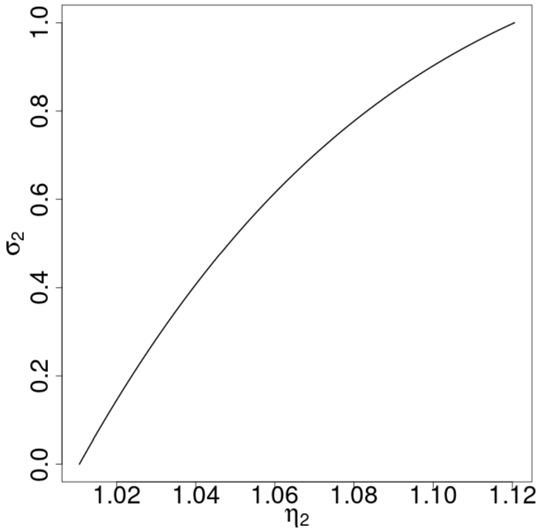


Figure 3.2: Plot of non-dimensional σ_2 as a function of η_2. The parameters used are the same as reported in Table 3.1 (see also §3.3.4): $R = 287\,\mathrm{J\,kg^{-1}\,K^{-1}}$, $c_p = 1004\,\mathrm{J\,kg^{-1}\,K^{-1}}$, $\theta_1 = 311\,\mathrm{K}$, $\theta_2 = 291.8\,\mathrm{K}$, $\eta_0 = 0.48$, $Z_0 = 6120\,\mathrm{m}$ and $g = 9.81\,\mathrm{m/s^2}$.

in which the analytical expression of the pseudo-density σ as a function of η reads (after substituting (3.13) into (3.11) and dropping the subscript in η_2):

$$\sigma(\eta) = \frac{p_r}{g}\left\{\eta - \left[\frac{\theta_2}{\triangle\theta}\left(-\eta^\kappa + \frac{\theta_1}{\theta_2}\eta_0^\kappa + \frac{g}{c_p\theta_2}Z_0\right)\right]^{\frac{1}{\kappa}}\right\}. \tag{3.17}$$

Integration of (3.16) with (3.17) yields:

$$\begin{aligned} \mathcal{E}\,(\eta) = c_p\theta_2\frac{p_r}{g}\frac{\kappa}{\kappa+1}\Bigg[\eta^{\kappa+1} + \left(\frac{\theta_2}{\triangle\theta}\right)^{\frac{1}{\kappa}}\left(\frac{\theta_1}{\theta_2}\eta_0^\kappa + \frac{g}{c_p\theta_2}Z_0 - \eta^\kappa\right)^{\frac{\kappa+1}{\kappa}} + \\ - \left(\frac{\theta_2}{\triangle\theta}\right)^{\frac{1}{\kappa}}\left(\frac{\theta_1}{\theta_2}\eta_0^\kappa + \frac{g}{c_p\theta_2}Z_0\right)^{\frac{\kappa+1}{\kappa}}\Bigg]. \end{aligned} \tag{3.18}$$

Therefore, the momentum equation can be expressed as:

$$\partial_t\,(\sigma u) + \partial_x\left(\sigma u^2 + \mathcal{E}\right) - f\,\sigma\,v = 0, \tag{3.19}$$

and system (3.15) can be written in conservative form (3.14) with $\mathbf{U}$, $\mathbf{F}(\mathbf{U})$ and $\mathbf{T}(\mathbf{U})$ defined as follows:

$$\mathbf{U} = \begin{pmatrix} \sigma \\ \sigma\,u \\ \sigma\,v \end{pmatrix}, \quad \mathbf{F}(\mathbf{U}) = \begin{pmatrix} \sigma\,u \\ \sigma\,u^2 + \mathcal{E} \\ \sigma\,u\,v \end{pmatrix}, \quad \mathbf{T}(\mathbf{U}) = \begin{pmatrix} 0 \\ -f\,\sigma\,v \\ f\,\sigma\,u \end{pmatrix}. \tag{3.20}$$

By writing $\sigma u^2 = (\sigma u)^2/\sigma$ and $\sigma uv = (\sigma u)(\sigma v)/\sigma$, the Jacobian of the system reads:

$$\mathbf{J}(\mathbf{U}) \equiv \partial_{\mathbf{U}}\mathbf{F} = \begin{pmatrix} 0 & 1 & 0 \\ -u^2 + \partial_\sigma \mathcal{E} & 2u & 0 \\ -uv & v & u \end{pmatrix}, \tag{3.21}$$

with eigenvalues:

$$\lambda_{1,2} = u \pm \sqrt{\partial_\sigma \mathcal{E}} \quad \text{and} \quad \lambda_3 = u. \tag{3.22}$$

Here, $\partial_\sigma \mathcal{E}$ is computed to be:

$$\frac{\partial}{\partial \eta}\mathcal{E}\left(\sigma\left(\eta\right)\right) = \frac{\partial \mathcal{E}}{\partial \sigma}\frac{\mathrm{d}\sigma}{\mathrm{d}\eta} \Longrightarrow \frac{\partial \mathcal{E}}{\partial \sigma} = \frac{\frac{\partial \mathcal{E}}{\partial \eta}}{\frac{\mathrm{d}\sigma}{\mathrm{d}\eta}} = \frac{\kappa c_p \theta_2 \sigma\left(\eta\right)\eta^{\kappa-1}}{\mathrm{d}\sigma/\mathrm{d}\eta}, \tag{3.23}$$

after using (3.16) in the numerator and with denominator:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\eta} = \frac{p_r}{g}\left[1 + \left(\frac{\theta_2}{\triangle\theta}\right)^{1/\kappa}\left(\frac{\theta_1}{\theta_2}\eta_0^\kappa + \frac{g}{c_p\theta_2}Z_0 - \eta^\kappa\right)^{(1-\kappa)/\kappa}\eta^{\kappa-1}\right]. \tag{3.24}$$

The plot of $\partial\mathcal{E}/\partial\sigma$ in Figure 3.3 shows that it is positive for non-negative values of σ, thus confirming the hyperbolic character of system (3.15), with real and distinct eigenvalues λ in (3.22) for $\partial_\sigma\mathcal{E} > 0$. The numerical scheme used to integrate the model can therefore be chosen to be close to the one in Kent et al. (2017) with minor adaptations. Details thereof can be found in Appendix C.

3.3.4 Observations-based scaling

The derivation of the isentropic $1\frac{1}{2}$-layer shallow water model presented in Bokhove et al. (2021) relies on a couple of assumptions concerning its physical scaling. In particular, the values of the parameters:

$$\varepsilon = \frac{U_1}{U_2}, \qquad \delta_a = \frac{H_2}{H_1}, \tag{3.25}$$

with U and H indicating each layer's scale velocity and fluid depth, are assumed to be small, i.e. $\varepsilon, \delta_a < 1$.

In view of adding convection and precipitation to the shallow water model here presented, we want our two-layer configuration to be an approximation of the troposphere (where such processes are confined), in which a thin bottom layer is surmounted by a thicker one, almost at rest, that is: $H_1 > H_2$ and $U_2 > U_1$.

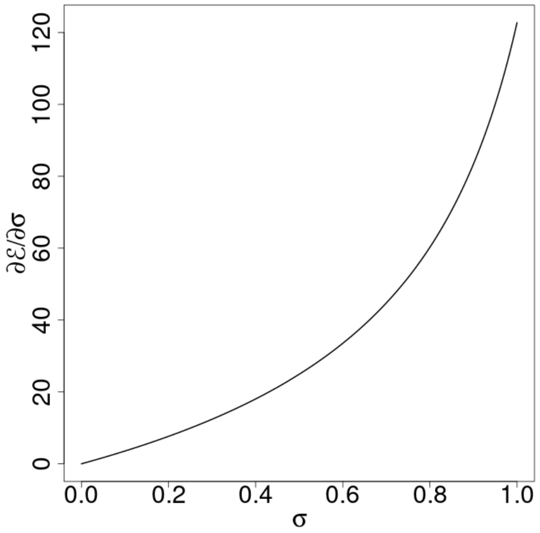


Figure 3.3: Plot of non-dimensional $\partial_\sigma \mathcal{E}$ as a function of σ. The parameters used are the same reported in Table 3.1 (see also §3.3.4): $R = 287\,\mathrm{J\,kg^{-1}\,K^{-1}}$, $c_p = 1004\,\mathrm{J\,kg^{-1}\,K^{-1}}$, $\theta_1 = 311\,\mathrm{K}$, $\theta_2 = 291.8\,\mathrm{K}$, $\eta_0 = 0.48$, $Z_0 = 6120\,\mathrm{m}$ and $g = 9.81\,\mathrm{m/s^2}$.

Low-level jets (LLJs) are recurrent meteorological features located at various locations in the world (Rife et al., 2010) and they happen to be particularly common over the Great Plains in the Southern United States (Djurić & Damiani Jr, 1980; Ladwig, 1980).

Fig. 3.4 shows vertical profiles obtained from radiosonde data of both potential temperature and wind speed during a LLJ event on 10^{th}–11^{th} December 1977 in Brownsville, Texas (US). We use this as a case study to provide a justification for the scaling chosen in Bokhove et al. (2021), which satisfies the requirements on ε and δ_a, as we are about to show. We approximate the troposphere as a two-layer fluid, exploiting the change of slope in the potential temperature profile of Fig. 3.4b as a reference. Mean potential temperature values of $\theta_1^{\mathrm{obs}} = 311.0$ K and $\theta_2^{\mathrm{obs}} = 291.8$ K follow after taking $H_1^{\mathrm{obs}} = 4.02$ km and $H_2^{\mathrm{obs}} = 2.08$ km in Fig. 3.4b. The above values of θ_1^{obs} and θ_2^{obs} are used as a constraint to compute H_1^{obs} and H_2^{obs} also in the profiles of Figs. 3.4a-c, in virtue of the isentropic assumption (i.e. constant potential temperature within each layer). Once the layer depths in each profile are established, mean wind speed values U_1^{obs} and U_2^{obs} within each layer are also computed (dashed lines

	Two-layer troposphere (Low-Level-Jet)			
	Fig. 3.4(a)	Fig. 3.4(b)	Fig. 3.4(c)	Average
H_2 [km]	2.02	2.08	1.65	1.92
H_1 [km]	3.98	4.02	4.6	4.2
p_0^{obs} [mbar]	489.6	483.8	475.4	482.9
θ_1^{obs} [K]	311.0	311.0	311.0	311.0
θ_2^{obs} [K]	291.8	291.8	291.8	291.8
U_1^{obs} [m/s]	3.6	6.6	7.0	5.7
U_2^{obs} [m/s]	13.5	12.7	11.0	12.4
δ_a	0.51	0.52	0.36	0.46
ε	0.27	0.52	0.64	0.46

Table 3.1: Summary of the values of various physical quantities obtained from the radiosonde data displayed in Fig. 3.4 and resulting values of non-dimensional scaling parameters δ_a and ε. The rightmost column reports the average values obtained from the data seen in Fig. 3.4.

in Fig. 3.4). Table 3.1 summarizes all the relevant physical parameters associated with the radiosonde data plotted in Fig. 3.4, together with the values of ε and δ_a and the values of the observed pressure on the top of the upper layer p_0. Overall, the values of ε and δ_a lie below one during the LLJ event; moreover, the rigid lid condition leading to the $1\frac{1}{2}$-layer configuration appears to be justified, as the variation in height of $Z_0 = H_1 + H_2 = \{6\,\text{km}, 6.1\,\text{km}, 6.25\,\text{km}\}$ is smaller than the change in depth of the bottom layer $H_2 = \{2.02\,\text{km}, 2.08\,\text{km}, 1.65\,\text{km}\}$.
The physical scaling summarised in Table 3.1 will be used throughout this thesis and will form the basis for the forecast-assimilation experiments reported in Chapter 5.

3.3.5 Recovery of the isopycnal model

In this section we show how the isentropic $1\frac{1}{2}$-layer shallow water model can be traced back to its isopycnal counterpart by taking $\kappa = 1$ in (3.3). Incidentally, this should help visualize the inconsistency of using the isentropic definition of the reduced gravity

(3.2) within an isopycnal model, as we argued earlier.

Starting from (3.7b) with $\kappa = 1$, a linear relation between the fluid depth h_2 and non-dimensional pressure η_2 (and thus pseudo-density σ_2) is restored:

$$h_2 = \frac{c_p\theta_2}{g}(\eta_2 - \eta_1) = \frac{c_p\theta_2}{p_r}\sigma_2. \tag{3.26}$$

Substituting the above expression back into the continuity equation (3.12a) yields:

$$\partial_t h_2 + \partial_x(h_2\, u_2) = 0, \tag{3.27}$$

equivalent to the continuity equation of the isopycnal model. In addition, we observe that for $\kappa = 1$ the Montgomery potential M_2 in (3.12b) becomes:

$$M_2 = c_p\,\theta_2\,\eta_2.$$

By using (3.7) with $\kappa = 1$, we obtain:

$$c_p\theta_2\eta_2 = gh_2 + gh_1\theta_2/\theta_1 + c_p\theta_2\eta_0. \tag{3.28}$$

After using the rigid lid constraint $H = h_1 + h_2$ and substituting back into the Montgomery potential and the momentum equation (3.12b), we obtain the usual momentum equation for an isopycnal fluid:

$$\partial_t u_2 + u_2\partial_x u_2 - f\,v_2 = -g'\partial_x h_2, \tag{3.29}$$

with reduced gravity g' defined as in (3.2). However, since $\kappa = 1$ implies $\theta = T(p_r/p)$, we note that:

$$\frac{\theta_1 - \theta_2}{\theta_1} = \frac{T_1/p_1 - T_2/p_2}{T_1/p_1} = \frac{1/\rho_1 - 1/\rho_2}{1/\rho_1} = \frac{\rho_2 - \rho_1}{\rho_2}, \tag{3.30}$$

making use of the ideal gas law ($p = \rho RT$). In other words, the isopycnal expression of the reduced gravity is recovered for $\kappa = 1$. The numerical convergence of the two models for $\kappa = 1$ will be used as a final check for the full model with convection and precipitation in section §3.4.3.

3.3.6 Shrira's solution for nonlinear waves

Here we provide an independent analytical verification of the numerical model using an extended version of Shrira's analysis of stationary nonlinear waves propagating on

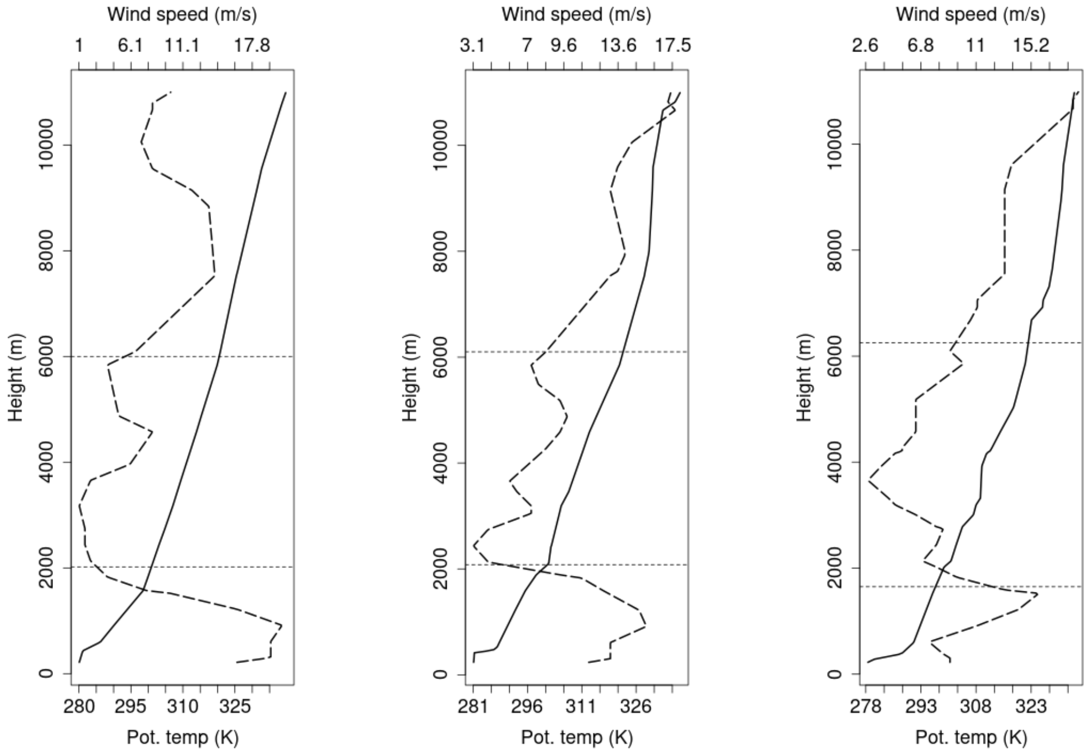


Figure 3.4: Vertical profile of potential temperature (solid line) and wind speed (dashed line) taken from radiosonde data on (a) 10/12/1977 at 00Z, (b) 10/12/1977 at 12Z and (c) 11/12/1977 at 00Z in Brownsville, Texas (US). The horizontal dotted lines indicate the depths of the two layers deduced from potential temperature data. The relevant physical parameters associated with each vertical profile are reported in Table 3.1. Source: http://weather.uwyo.edu/upperair/sounding.html

the surface of a rotating isopycnal or shallow-water layer fluid (Shrira, 1981, 1986), adapted to our isentropic model in a periodic domain. We start by splitting the pseudo-density σ into a constant state σ_0 and a perturbation $\tilde{\sigma}$:

$$\sigma = \sigma_0 + \tilde{\sigma} \quad \text{with} \quad \tilde{\sigma} = \frac{\sigma_0}{f}\frac{\partial v}{\partial x}. \tag{3.31}$$

The validity of (3.31) for $\tilde{\sigma}$ follows by substituting into the continuity equation (3.15a) and using the meridional momentum equation, obtaining the identity:

$$\begin{aligned} \frac{\sigma_0}{f}\frac{\partial^2 v}{\partial x \partial t} + \frac{\partial}{\partial x}\left[\left(\sigma_0 + \frac{\sigma_0}{f}\frac{\partial v}{\partial x}\right) u\right] &= 0, \\ \frac{\partial}{\partial x}\left[\frac{\partial v}{\partial t} + f u + u\frac{\partial v}{\partial x}\right] &= 0. \end{aligned}$$

Using (3.12c) to express u as a function of $\frac{\partial v}{\partial x}$ and $\frac{\partial v}{\partial t}$, one finds (assuming $\frac{\partial v}{\partial x} + f \neq 0$):

$$u = -\frac{\partial v}{\partial t}/(\frac{\partial v}{\partial x} + f). \tag{3.32}$$

Differentiating (3.12c) by t and using both (3.12b) and (3.32), yields:

$$\frac{\partial^2 v}{\partial t^2} + f^2 v - f\frac{\partial M}{\partial x} = \frac{\partial}{\partial t}\left(\frac{\frac{\partial v}{\partial x}\frac{\partial v}{\partial t}}{f + \frac{\partial v}{\partial x}}\right) + \frac{f}{2}\frac{\partial}{\partial x}\left(\frac{\left(\frac{\partial v}{\partial t}\right)^2}{\left(f + \frac{\partial v}{\partial x}\right)^2}\right), \tag{3.33}$$

resembling Eq. (10) in Shrira (1981) and Eq. (4) in Shrira (1986) (once y–derivatives are dropped in the older paper and the high-frequency dispersion term is neglected). The gradient of the Montgomery potential becomes:

$$\frac{\partial}{\partial x}M\left(\eta\left(\sigma\right)\right) = \frac{\partial M}{\partial \eta}\frac{\mathrm{d}\eta}{\mathrm{d}\sigma}\frac{\partial \sigma}{\partial x} = c_p\theta\kappa\eta^{\kappa-1}\left(\frac{1}{\frac{\mathrm{d}\sigma}{\mathrm{d}\eta}}\right)\frac{\sigma_0}{f}\frac{\partial^2 v}{\partial x^2}, \tag{3.34}$$

after using $\tilde{\sigma}$ from (3.31). Substituting (3.34) into (3.33) gives:

$$\frac{\partial^2 v}{\partial t^2} + f^2 v - f c_p\theta\kappa\eta^{\kappa-1}\left(\frac{1}{\frac{\mathrm{d}\sigma}{\mathrm{d}\eta}}\right)\frac{\sigma_0}{f}\frac{\partial^2 v}{\partial x^2} = \frac{\partial}{\partial t}\left(\frac{\frac{\partial v}{\partial x}\frac{\partial v}{\partial t}}{f + \frac{\partial v}{\partial x}}\right) + \frac{f}{2}\frac{\partial}{\partial x}\left(\frac{\left(\frac{\partial v}{\partial t}\right)^2}{\left(f + \frac{\partial v}{\partial x}\right)^2}\right).$$

For travelling waves of phase velocity c, we define $\zeta = x - ct$ and rewrite the equation above, with primes denoting $\frac{\partial}{\partial \zeta}$, to obtain a second order ODE in ζ:

$$c^2\, v'' + f^2\, v - f\, c_p\theta\kappa\eta^{\kappa-1}\left(\frac{1}{\frac{\mathrm{d}\sigma}{\mathrm{d}\eta}}\right)\frac{\sigma_0}{f}v'' = c^2\left(\frac{\left(v'\right)^2}{f + v'}\right)' + \frac{f\, c^2}{2}\left(\frac{\left(v'\right)^2}{\left(f + v'\right)^2}\right)'. \tag{3.35}$$

After some manipulation, one finds:

$$\upsilon'' = \frac{\frac{f^2}{c^2}\upsilon}{\frac{f}{c^2}c_p\theta\kappa\eta^{\kappa-1}\left(\frac{1}{\frac{\mathrm{d}\sigma}{\mathrm{d}\eta}}\right)\frac{\sigma_0}{f} - \frac{f^3}{(f+\upsilon')^3}},$$

non-dimensionalized as follows:

$$\mathrm{v}'' = \frac{1}{\mathrm{Ro}^2}\frac{\mathrm{v}}{\frac{c_p\theta}{c^2}\kappa\eta^{\kappa-1}\left(\frac{1}{\frac{\mathrm{d}\sigma}{\mathrm{d}\eta}}\right)\sigma_0 - \frac{1}{\mathrm{Ro}^3}\frac{1}{\left(\frac{1}{\mathrm{Ro}}+\mathrm{v}'\right)^3}}, \tag{3.36}$$

using $\upsilon = c\mathrm{v}$, $\mathrm{Ro} = c/fL$ and $(\cdot)' = \frac{\partial}{\partial\zeta} = \frac{1}{L}\frac{\partial}{\partial\zeta'} = \frac{1}{L}(\cdot)'$. This ODE is solved with a Runge-Kutta fourth order method after rewriting it as follows:

$$\chi = \mathrm{v}', \quad \chi' = F\left(\chi, \mathrm{v}, \zeta\right) = \frac{\frac{1}{\mathrm{Ro}^2}\mathrm{v}}{\left[-\frac{1}{\mathrm{Ro}^3}\frac{1}{\left(\frac{1}{\mathrm{Ro}}+\chi\right)^3} + \widetilde{c_p}\theta\kappa\eta^{\kappa-1}\left(\frac{1}{\frac{\mathrm{d}\sigma}{\mathrm{d}\eta}}\right)\sigma_0\right]}, \tag{3.37}$$

with $\widetilde{c_p} = c_p/c^2$. For comparison with the full model (3.15), we derive the expressions also for the (non-dimensional) u and $\bar{\sigma}$, using (3.32) and (3.31), and scaling u and σ_0 by c and g/p_r, respectively:

$$\mathrm{u} = \chi/(1/\mathrm{Ro} + \chi) \quad \text{and} \quad \bar{\sigma} = \sigma_0\left(1 + \mathrm{Ro}\chi\right). \tag{3.38}$$

The solution of (3.37) is stable only within a certain range of initial conditions for v_0 and χ_0. Once a stable configuration is found, the phase velocity c is tuned in order to obtain a single-wavelength wave in v, u and $\bar{\sigma}$ as solution and subsequently used as initial condition for the isentropic $1\frac{1}{2}$-layer model. A comparison between Shrira's solution and its numerical implementation at various times t and up to $t = 10\mathtt{T}$ (where $\mathtt{T}$ is the wave period), is shown in Figure 3.5. Although the dissipative character of the numerical scheme used in this paper contributes to both an amplitude and a phase error as time goes by, the numerical solution (purple to red solid lines) visually converges towards the analytical one (gray line) as the resolution increases.

3.4 Modified shallow water: the ismodRSW model

In this section, we extend the rotating isentropic $1\frac{1}{2}$-layer model described above, in order for it to mimic convection and precipitation in a similar manner to the isopcynal

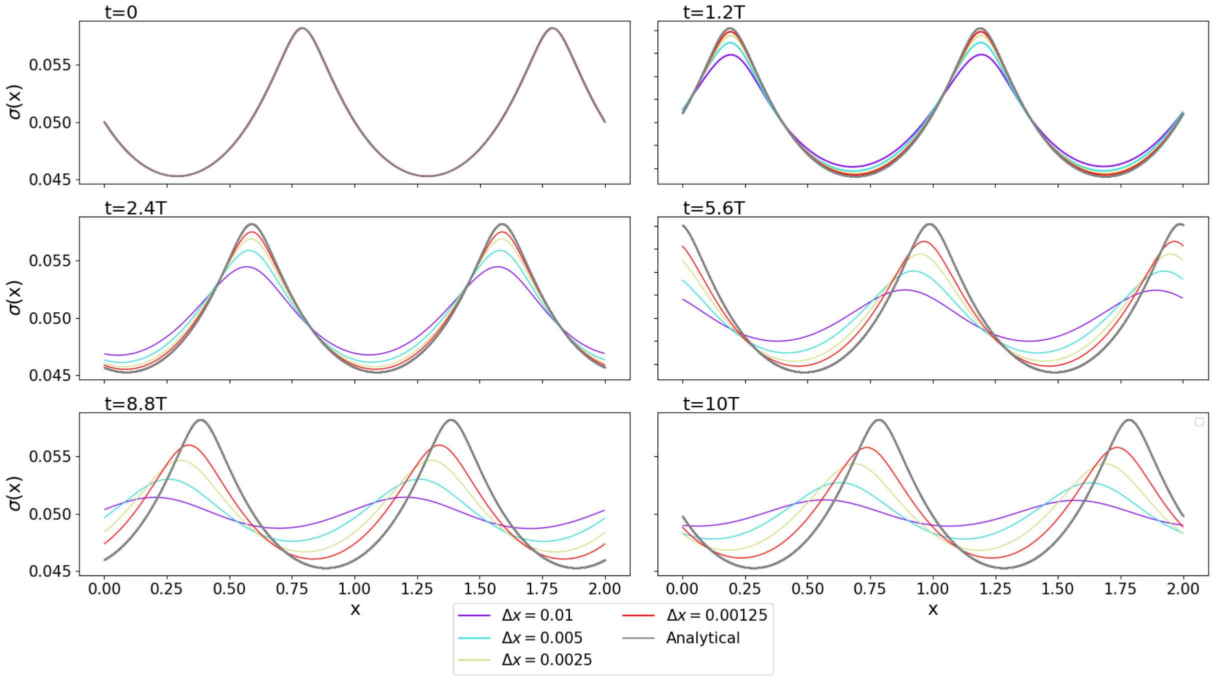


Figure 3.5: Comparison between Shrira's analytical solution for σ translated in time (gray solid line) and the evolution of the model in (3.15) at various spatial resolutions initialised with the analytical solution itself. Snapshots after $\{0, 1.2, 2.4, 5.6, 8.8, 10\}$ periods (T) are shown. Initial conditions: $\mathrm{v}_0 \equiv v(x,0) = 0.08$, $\chi_0 = 0$. Other parameters are: $\sigma_0 = 0.05$, $c = 23.05$, $\mathrm{Ro} = 0.2305$, $c_p = 1004\,\mathrm{J\,kg^{-1}\,K^{-1}}$, $\theta_1 = 311\,\mathrm{K}$, $\theta_2 = 291.8\,\mathrm{K}$, $\eta_0 = 0.48$, $p_r = 1000\,\mathrm{hPa}$, $Z_0 = 6120\,\mathrm{m}$, $\mathrm{CFL} = 0.5$.

modRSW model. In addition, we present the new nature run simulation which will be used in the forecast-assimilation experiments described in Chapter 5 and we conclude by additionally by checking the convergence of the full isentropic model towards the modRSW one by imposing $\kappa = 1$ in (3.3).

3.4.1 Model dynamics

Starting from (3.12) – and similarly to what is done in section §2.1.2 – a system of thresholds is introduced, together with an equation for the (dimensionless) rain mass fraction r. As a result, the following system of equations is obtained:

$$\partial_t \sigma + \partial_x(\sigma u) = 0, \tag{3.39.a}$$

$$\partial_t(\sigma u) + \partial_x(\sigma u^2 + \widetilde{\mathcal{E}}) + \sigma c_0^2 \partial_x r - f\sigma v = 0, \tag{3.39.b}$$

$$\partial_t(\sigma v) + \partial_x(\sigma u v) + f\sigma u = 0, \tag{3.39.c}$$

$$\partial_t(\sigma r) + \partial_x(\sigma u r) + \sigma\widetilde{\beta}\partial_x u + \alpha\sigma r = 0, \tag{3.39.d}$$

in which $\widetilde{\mathcal{E}}$ (playing the role of the effective pressure in Kent et al. (2017)) is defined as:

$$\widetilde{\mathcal{E}} = \begin{cases} \mathcal{E}(\sigma_c) & \text{for } \sigma > \sigma_c, \\ \mathcal{E}(\sigma) & \text{otherwise;} \end{cases} \tag{3.40}$$

with σ_c a convection threshold and:

$$\widetilde{\beta} = \begin{cases} \beta & \text{for } \sigma > \sigma_r \text{ and } \partial_x u < 0; \\ 0 & \text{otherwise;} \end{cases} \tag{3.41}$$

in which σ_r is a rain threshold. Again, the parameters α (s^{-1}) and β (dimensionless) control the rate at which rain is created and removed from the system. The constant speed squared c_0^2 ($\mathrm{m}^{-2}\mathrm{s}^{-2}$) converts the dimensionless rain mass fraction into a potential, introducing a coupling between the two equations and implicitly controlling the suppression of convection, cf. §2.1.2, Kent (2016); Kent et al. (2017).
Analogously to the modRSW model, the isentropic model equipped with convection and precipitation cannot be written in conservative form. Its intrinsic non-conservative vector formulation reads

$$\partial_t \mathbf{U} + \partial_x \mathbf{F}(\mathbf{U}) + \mathbf{G}(\mathbf{U})\partial_x \mathbf{U} + \mathbf{S}(\mathbf{U}) = 0, \tag{3.42}$$

where:

$$
\begin{aligned}
\mathbf{U} &= \begin{pmatrix} \sigma \\ \sigma u \\ \sigma v \\ \sigma r \end{pmatrix}, \quad \mathbf{F}(\mathbf{U}) = \begin{pmatrix} \sigma u \\ \sigma u^2 + \widetilde{\mathcal{E}} \\ \sigma u v \\ \sigma u r \end{pmatrix}, \\
\mathbf{G}(\mathbf{U}) &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ -c_0^2 r & 0 & 0 & c_0^2 \\ 0 & 0 & 0 & 0 \\ -\widetilde{\beta} u & \widetilde{\beta} & 0 & 0 \end{pmatrix}, \quad \mathbf{S}(\mathbf{U}) = \begin{pmatrix} 0 \\ -f\sigma v \\ f\sigma u \\ \alpha\sigma r \end{pmatrix}.
\end{aligned}
\tag{3.43}
$$

The Jacobian matrix $\mathbf{J} = \partial_{\mathbf{U}}\mathbf{F} + \mathbf{G}$ of the system reads:

$$
\mathbf{J} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -u^2 - c_0^2 r + \partial_\sigma \widetilde{\mathcal{E}} & 2u & 0 & c_0^2 \\ -uv & v & u & 0 \\ -u(\widetilde{\beta} + r) & \widetilde{\beta} + r & 0 & u \end{pmatrix},
$$

with eigenvalues:

$$
\lambda_{1,2} = u \pm \sqrt{\partial_\sigma \widetilde{\mathcal{E}} + c_0^2 \widetilde{\beta}} \quad \text{and} \quad \lambda_{3,4} = u. \tag{3.44}
$$

Again, the numerical discretisation is an adaptation of the one used in Kent et al. (2017). Details thereof can be found in the Appendix C.

3.4.2 The new nature run

In Chapter 2 we demonstrated how the modRSW model is a viable tool to conduct DA research for operational Numerical Weather Prediction (NWP). In particular, forecast-assimilation experiments have been conducted and subsequently analysed to show how the system, despite its idealizations, performs akin to operational DA schemes following a variety of metrics and criteria. Therein, a so-called twin-setting configuration was used, in which observations are obtained from a high-resolution nature run simulation, whereas the forecasts are generated by running the model at a lower resolution. Similarly, in order to conduct new satellite DA experiments for our novel ismodRSW model, it is essential to find a dynamically interesting nature run, characterized by continuous production of convection and precipitation. The absence of topography in the new configuration complicates the task somewhat since the presence of topography in a periodic domain with a zonal flow constituted a convenient setup to obtain self-generation of gravity waves with the modRSW model.

To compensate for the absence of topography, both rotation and a forcing term are

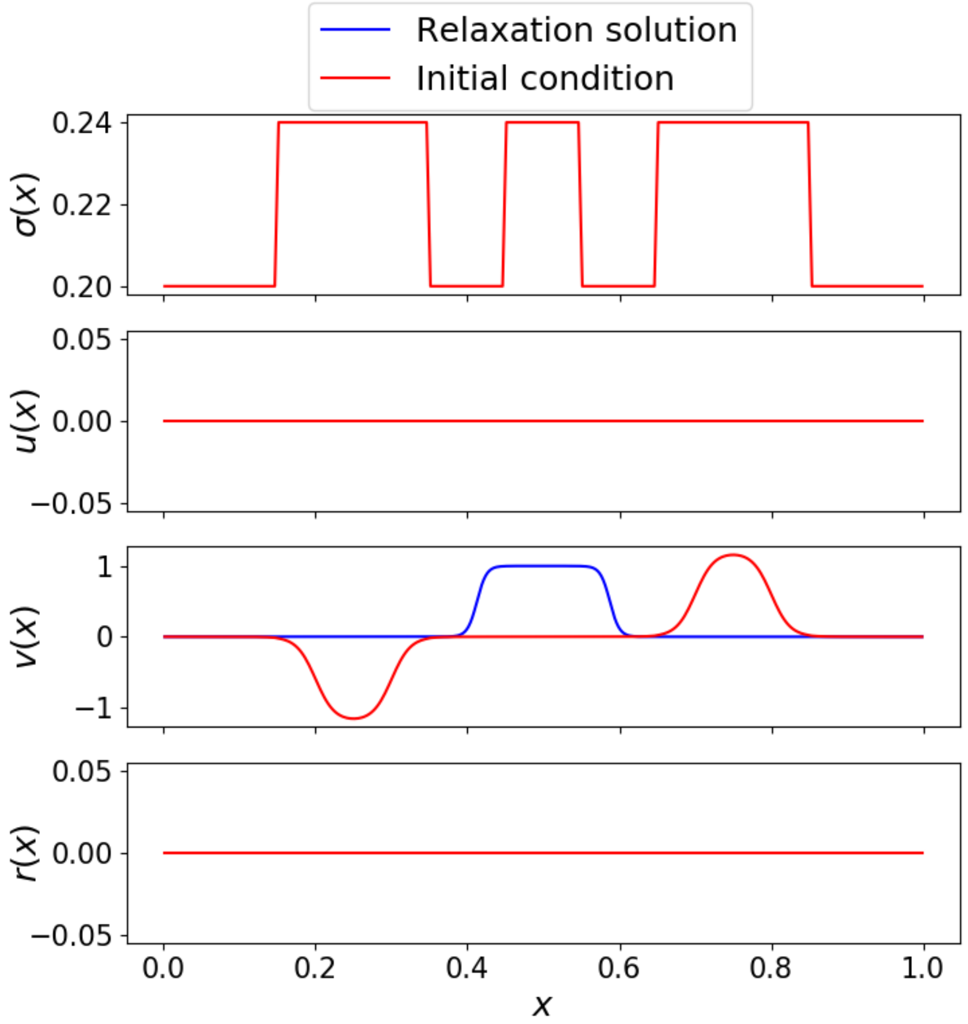


Figure 3.6: Initial condition (in red) and relaxation solution (in blue) of the nature run simulation (parameters listed in Table 3.1). The relaxation solution v_{rel} mimics a Low Level Jet (approximated by a smoothed top-hat transverse jet at the centre of the domain). All variables are non-dimensional.

Initial conditions	Fig. 3.6	η_0	0.48
Boundary conditions	Periodic	Z_0 [m]	$6.12 \cdot 10^3$
Relaxation solution	Fig. 3.6	c_0^2	1.8
τ_{rel}	4	α	6.0
CFL	0.1	β	2.0
θ_1 [K]	311	σ_c	0.21
θ_2 [K]	291.8	σ_r	0.24
Ro	0.248	L [km]	500
U [m/s]	12.4	N_{el}	400
$\mathtt{T}$ (L/U) [h]	11.2	Fr	0.36

Table 3.2: Model parameters associated with the ismodRSW nature simulation shown in Fig. 3.7. CFL indicates the Courant-Friedrichs-Lewy number. The values of σ_c, σ_r and τ_{rel} are reported as non-dimensional variables. The Froude number Fr is computed as $\mathrm{Fr} = U/\sqrt{g'h_2}$, with g' given by (3.2) and H_2 from Table 3.1.

used, with the latter consisting of a relaxation term in the meridional momentum equation:

$$\partial_t v + u\partial_x v + fu = (v_{rel} - v)/\tau_{rel}, \tag{3.45}$$

in which τ_{rel} is a relaxation time-scale defining the speed at which the meridional velocity v relaxes towards $v_{rel}(x,t)$. The relaxation profile v_{rel} is shown in Fig. 3.6 (blue line) and represents a smoothed top-hat meridional jet. This type of forcing is chosen to reflect the troposphere-based scaling used in the derivation of the ismodRSW model (cf. section §3.3.4 and Bokhove et al. (2021)), which is based on Low Level Jet (LLJ) conditions (here approximated by a transverse jet in the meridional direction) in a two-layer troposphere. In this regard, it is worth noting that v_{rel} does not represent an equilibrium solution and is not meant to be one, as its purpose is solely to excite new instability and gravity waves, counteracting the highly dissipative character of the numerical solver.

The initial condition used in the nature run is also shown in Fig. 3.6 (red lines). In this case, the shape of both the pseudo-density profile and the meridional velocity simply represents an unstable setup that maximises the production of convection and rain at the beginning of the simulation. Finally, in Table 3.2 we list the parameter set used in the ismodRSW model to generate a 48 hours-long nature run with continuous production of rain and precipitation.

The dynamical evolution of the nature run is shown in Hovmöller diagrams, one for each variable, in Fig. 3.7. The two top panels (right: σ, left: r) show how convection and rain are continuously generated across the (periodic) domain, as travelling waves move along it. Grey-shaded areas in the top-left panel indicate locations where only convection is activated ($\sigma_c < \sigma < \sigma_r$), whereas yellow-to-brown shadings denote areas of rain production ($\sigma > \sigma_r$). The fluid velocities are depicted in the bottom two panels (left: u, right: v). Areas of rain in the top right panel are spatially correlated with areas where convergence in u arises, e.g. at locations where a negative gradient of u exists. The Hovmöller diagram of the meridional velocity v (bottom-right panel) shows a much smoother time evolution than the other three variables as it shows the transition from the initial condition towards the relaxation solution.

The unsettled character of the nature run is further illustrated in Fig. 3.8, where the time series of all variables at location $x_0 = 0.5$ are shown. The peaks and troughs in the pseudo-density $\sigma(x_0)$ indicate the passage of the gravity waves at this location, with corresponding generation and removal of rain $r(x_0)$. The transit of waves is also correlated with some irregular oscillations in the horizontal velocity $u(x_0)$, while the meridional velocity $v(x_0)$, initially zero, gradually settles towards its relaxation solution v_{rel}, that is, a jet at the centre of the domain (cf. Fig. 3.6).

3.4.3 Recovery of the *modRSW* model in presence of convection and rain

As a final check performed on the full model, we compare the evolution in time of the modRSW against the ismodRSW model with $\kappa = 1$. The two models are initialized with the same initial condition and the parameters reported in Table 3.2. The initial values of the fluid depth H_0, and the related thresholds H_c and H_r are scaled to $H_0 = 1$, $H_c = \sigma_c/\sigma_0 = 1.05$ and $H_r = \sigma_r/\sigma_0 = 1.2$ with $\sigma_0 = 0.2$ (see Table 3.2 and Fig. 3.6). The value of Froude number in the bottom layer Fr_2 (which needs to be specified in the modRSW model) is reported in Table 3.2 and is computed as:

$$\mathrm{Fr}_2 = \frac{U_2}{\sqrt{g' h_2}} = \frac{U_2}{\sqrt{g \frac{\Delta\theta}{\theta_1} \frac{R\theta_2}{g} \frac{g}{p_r} \sigma}}, \tag{3.46}$$

in which we have used the definition of the isentropic reduced gravity g' in (3.2), the expression of the bottom layer depth h_2 in (3.7b) and the definition of σ in (3.11), noting that $c_p = R$ when $\kappa = 1$.

We run the two models side-by-side in three different configurations: (i) classic rotating shallow water ($\sigma_c, \sigma_r \to \infty$), (ii) convection-only regime ($\sigma_r \to \infty$) and (iii) fully modified shallow water. The results (limited to h, σ and r) are shown in Figure 3.9. While in the first two cases (left and central panels) the two models behave identically (we notice the good agreement between the solid black for h and the dashed gray line for σ in the left and central panels), we observe a gradual divergence in time between the two once both convection and rain are turned on (right panels, after $t = 1.0$). The differences are particularly noticeable in the rain solutions (blue solid and cyan dashed lines). This divergence seems to originate (and grow thereafter) at various locations in the domain where σ (and similarly h) decreases until it falls below the rain threshold. We believe this is an indirect effect of rounding errors generated by the computation of the different flux expressions in the two models (cf. (3.43) in this thesis with (6) in Kent et al. (2017)), which in turn has an impact on the values of the non-conservative products. This type of behaviour is not surprising given the non-linear character of the two models.

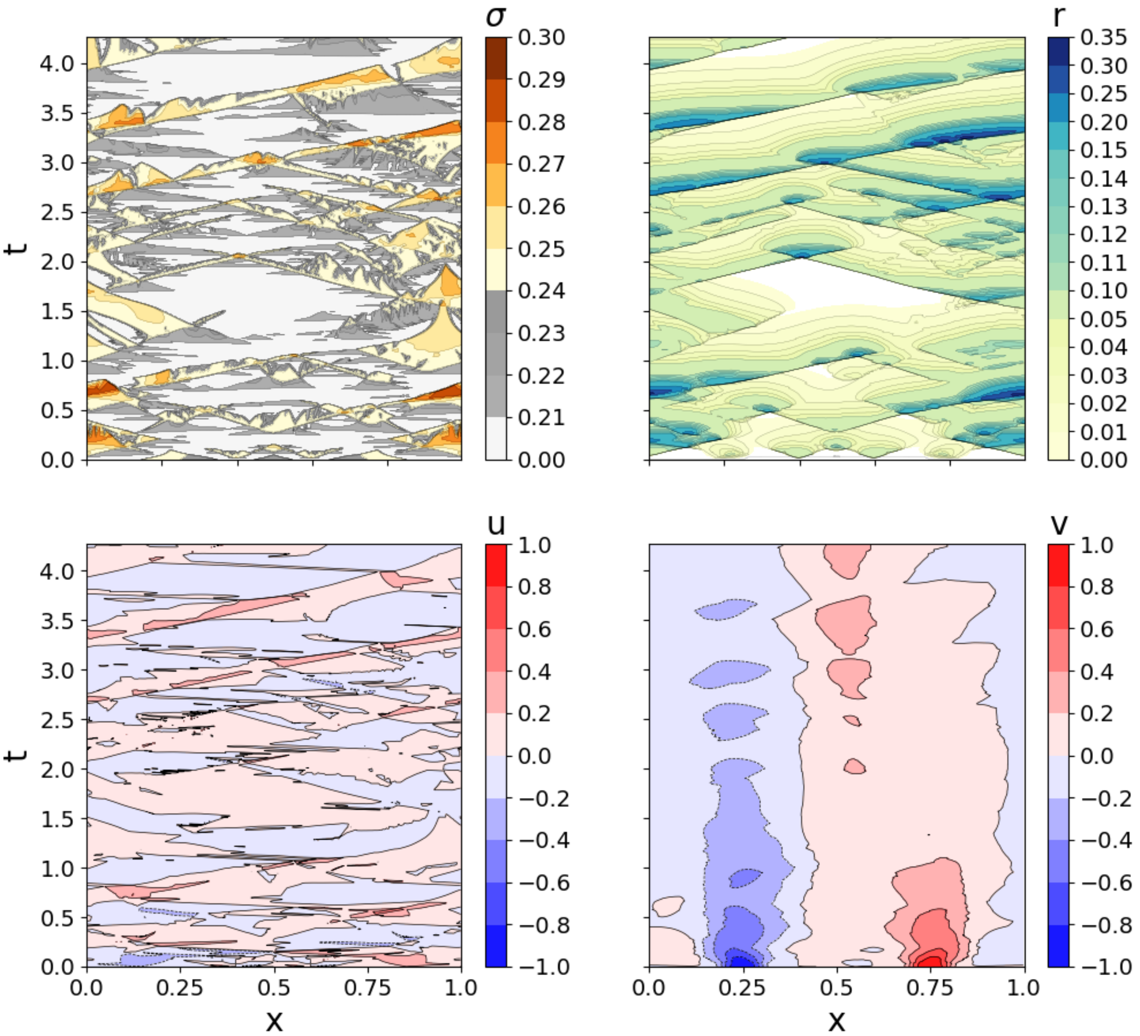


Figure 3.7: Hovmöller diagrams of the nature run simulation (parameters listed in Table 3.2). Non-dimensional variables: σ (top-left panel), r (top-right panel), u (bottom-left panel) and v (bottom-right panel). The gray and yellow shading in the top left panel indicates the areas where σ is above the convection (e.g. $\sigma_c < \sigma < \sigma_r$) and the rain ($\sigma > \sigma_r$) thresholds, respectively. Note that the length of the y axis ($t = [0, 4.272]$) is the non-dimensional equivalent of a 48 hours period, given the scaling in Table 3.2. The values of σ_c and σ_r are also reported in Table 3.2.

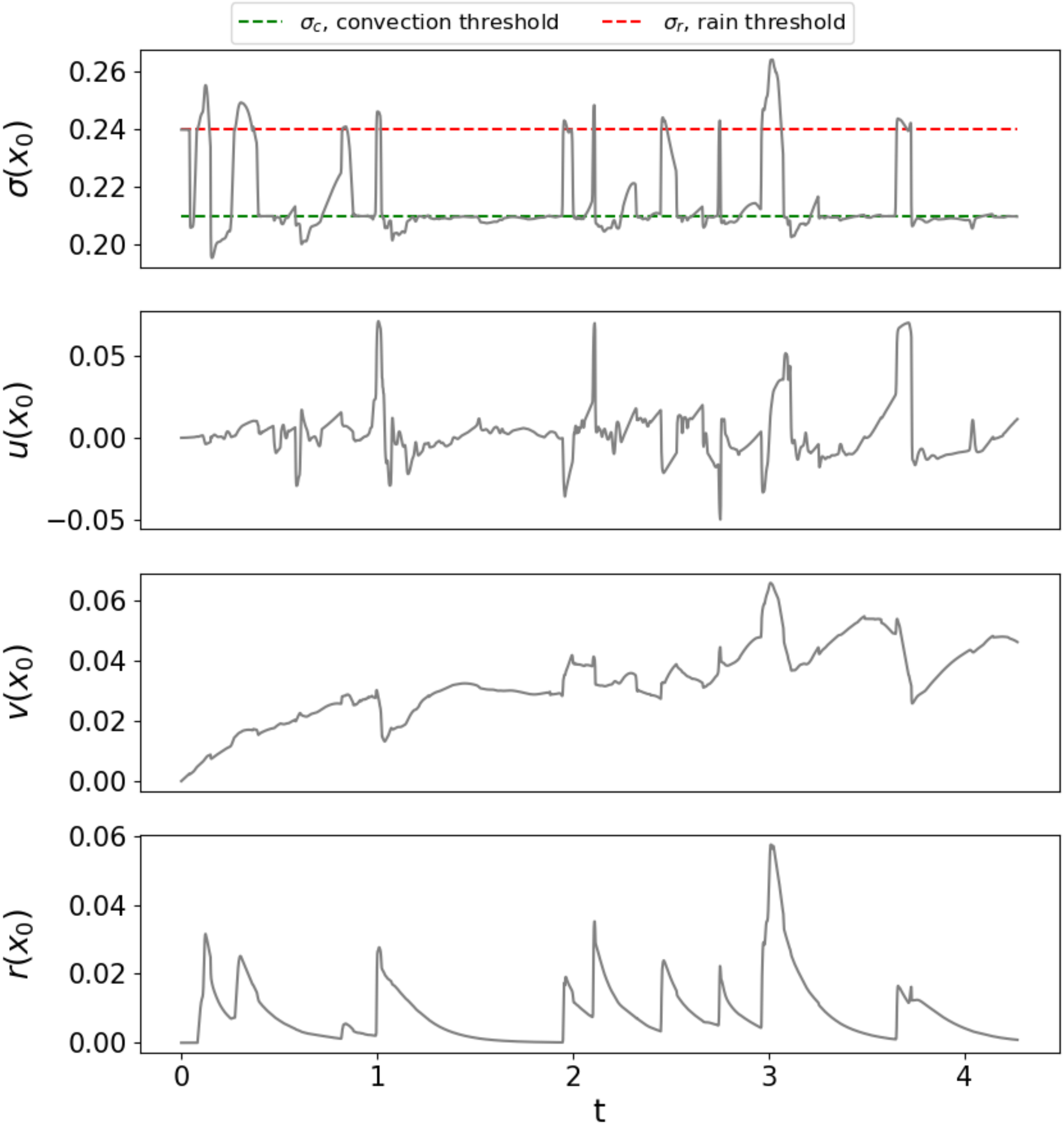


Figure 3.8: Time series of the nature run variables at location $x_0 = 0.5$ (parameters listed in Table 3.2).

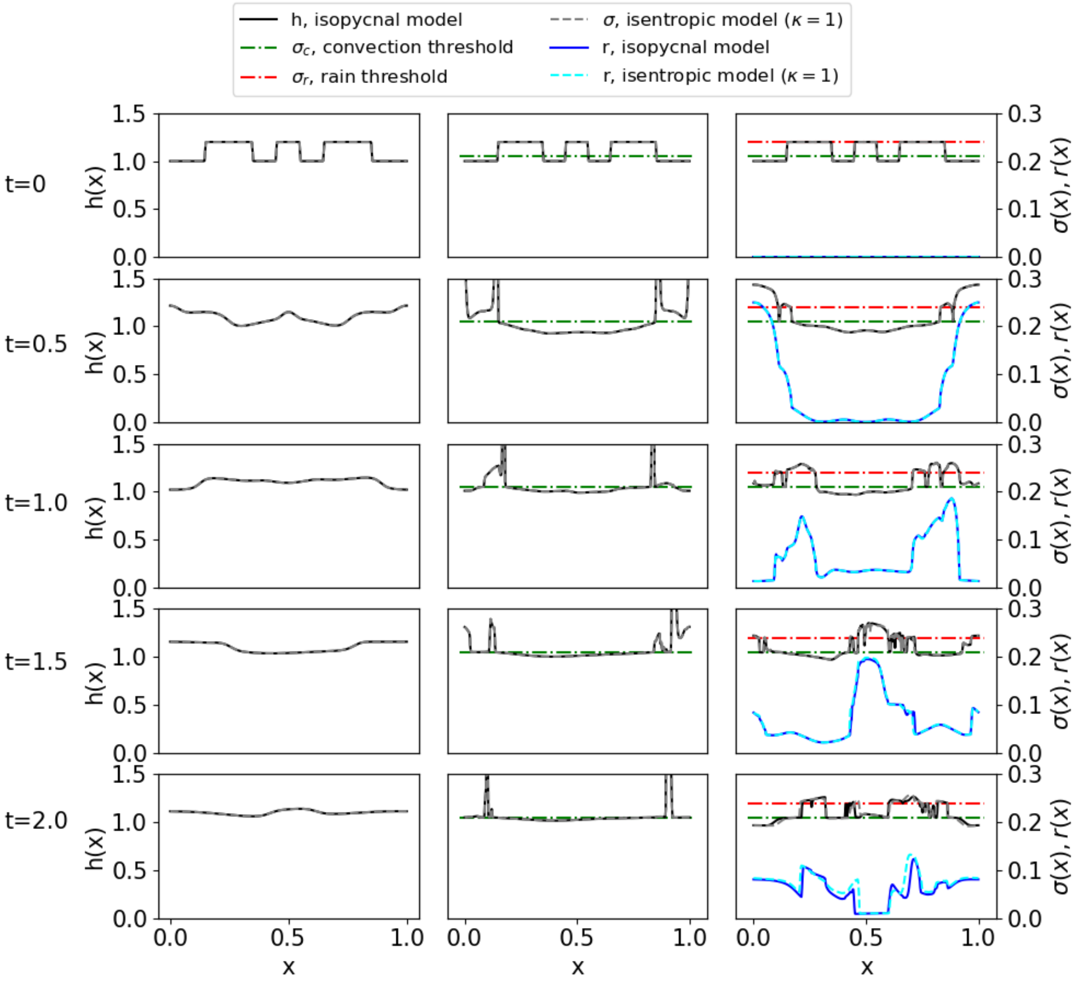


Figure 3.9: Comparison between the isopycnal and the isentropic model with $\kappa = 1$ at times (from top to bottom panels) $t = \{0, 0.5, 1.0, 1.5, 2.0\}$. Left panels: classic rotating shallow water with $\sigma_c, \sigma_r, H_c, H_r \to \infty$; central panels: convection-only regime with $\sigma_c = 0.21,\ H_c = 1.05$ (green dashed-dotted line) and $\sigma_r, H_r \to \infty$; right panels: fully modified shallow water with $\sigma_c = 0.21,\ H_c = 1.05$ and $\sigma_r = 0.24,\ H_r = 1.2$ (dashed-dotted red line). Variables: isopycnal fluid depth h (black solid line) and mass rain fraction r (blue solid line), isentropic pseudo-density σ (dashed gray line) and mass rain fraction r (cyan dashed line).

Chapter highlights and summary

- At the start of the chapter we have discussed the limitations of the modRSW model that hamper its use for satellite data assimilation research. In particular, we have showed how the lack of a robust fluid temperature definition leads to some scaling issues when a simple prognostic relationship based on the ideal gas law and the hydrostatic equilibrium is formulated.

- A revised version of the modRSW model is therefore presented, namely the isentropic modified rotating shallow water (ismodRSW) model. The ismodRSW model is naturally equipped with a physically consistent temperature definition and includes an additional layer for increased vertical complexity. At the same time, the revised model retains most of the dynamical and numerical characteristics of the modRSW model, including a threshold mechanism to imitate convection and precipitation. The revised model (without convection and precipitation) is checked numerically against an analytical solution based on Shrira (1981, 1986).

- The ismodRSW model equipped with convection and rain is used to create a new nature run simulation for the anticipated satellite data assimilation experiments. A combination of fluid rotation and relaxation to a meridional jet creates an interesting dynamics characterised by continuous production of convection and precipitation via the generation of traveling gravity waves.

Chapter 4

Idealised satellite observations

This chapter describes how satellite observations are generated using the ismodRSW model introduced in Chapter 3 and subsequently used in the forecast-assimilation experiments discussed in Chapter 5.

Since the aim of this study is to replicate the characteristics of real satellite observations in an idealised fashion, some context regarding the theoretical background and the historical perspective on the use of meteorological satellites in data assimilation is also provided.

Therefore, this chapter will start with an overview of the main concepts related to radiative transfer (RT) and atmospheric radiation, and will continue with a review of the historical developments of satellite DA, before moving to the actual description of how these aspects have been replicated in our idealised configuration.

4.1 Weather satellites: a theoretical overview

The main task of meteorological satellites is to measure electromagnetic (EM) radiation to extrapolate information about the earth's surface and atmosphere. In order to understand how this is achieved in practice, a basic understanding of the main RT processes is required. To this aim, we start this section by introducing some fundamental concepts and definitions about EM radiation and its interaction with matter. Later on, we will introduce the full RT equation, discuss the main features of atmospheric radiation and how they are exploited by satellites for meteorological purposes. Towards the end, we will summarise the main characteristics and properties of satellites and the instruments they are equipped with.

We refer to Petty (2006); Salby (1996); Stamnes et al. (2017) and Kidder et al. (1995) for a more extensive discussion of the content of this section.

4.1.1 Radiative transfer: definitions and processes

In this thesis, we will primarily describe EM radiation in terms of its wave properties, that is: its *frequency* ν, measured in Hertz (s^{-1}, [Hz]), and *wavelength* λ [m], which are related by the formula:

$$\lambda = \frac{c}{\nu}, \tag{4.1}$$

in which c represents the speed of light. Indeed, this is the speed at which the EM radiation travels, and its value depends on the medium in which it is propagating. The speed of light in vacuum is $c = 2.998 \cdot 10^8\ \mathrm{m\,s^{-1}}$. Of course, EM radiation does not propagate in vacuum when atmospheric processes are concerned, and the way it interacts with various components of the atmosphere will be at the centre of this section. However, the reduction of c caused by these processes is marginal and it appears safe to neglect these effects for the rest of the thesis.

In the context of RT applications, it is also useful to introduce a quantity called the *wavenumber* [cm^{-1}], defined as:

$$\tilde{\nu} = \frac{1}{100 \cdot \lambda}. \tag{4.2}$$

Exactly like all waves, EM radiation has the important property of carrying energy. This property helps define the quantities that are generally used to measure EM radiation quantitatively. To this aim, two physical quantities typically used in RT are defined below: *radiance* and *irradiance*.

Irradiance By considering the oriented surface element dA with unit vector $\hat{\mathbf{n}}$ in Fig. 4.1, we define the spectral (or monochromatic) *irradiance*[1] F_λ as the net amount of radiant energy dE in the wavelength interval $\lambda, \lambda + \mathrm{d}\lambda$ flowing through dA from all

---

[1] Irradiance is sometimes called *flux density* in the literature, hence the symbol F.

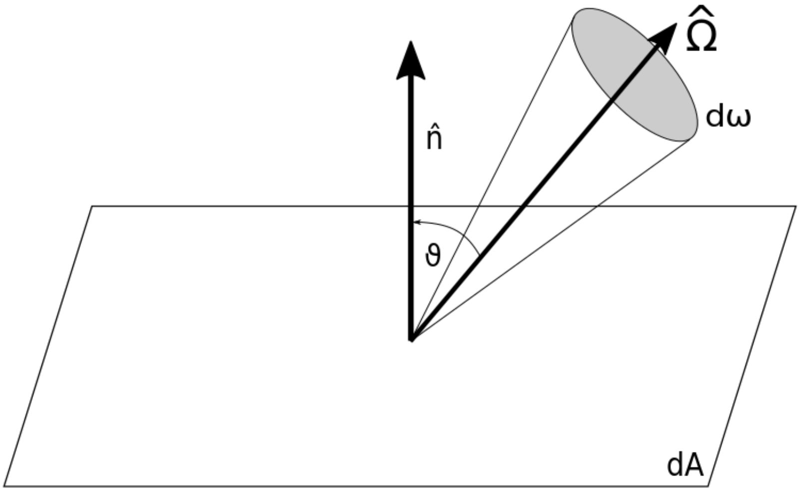


Figure 4.1: Geometrical representation of an oriented surface $\mathrm{d}A$ with unit vector $\hat{\mathbf{n}}$. A solid angle $\mathrm{d}\omega$ oriented in the $\hat{\mathbf{\Omega}}$ direction forms an angle ϑ with $\hat{\mathbf{n}}$.

directions during the time interval $t, t + dt$, that is:

$$F_\lambda = \frac{dE}{\mathrm{d}A\mathrm{d}\lambda\mathrm{d}t} \qquad \left[\frac{\mathrm{W}}{\mathrm{m}^2 \cdot \mathrm{m}}\right]. \tag{4.3}$$

Radiance By considering again the oriented surface element $\mathrm{d}A$ in Fig. 4.1, we call spectral (or monochromatic) *radiance*[1] I_λ the amount of radiant energy dE in the wavelength interval $\lambda, \lambda + \mathrm{d}\lambda$ coming from the direction $\hat{\mathbf{\Omega}}$ (forming an angle ϑ with $\hat{\mathbf{n}}$) within the solid angle $\mathrm{d}\omega$ flowing through dA during the time interval $t, t + dt$, that is:

$$I_\lambda = \frac{dE}{\mathrm{d}A\mathrm{d}\lambda\mathrm{d}t\mathrm{d}\omega} \frac{1}{\hat{\mathbf{n}} \cdot \hat{\mathbf{\Omega}}} \qquad \left[\frac{\mathrm{W}}{\mathrm{m}^2 \cdot \mathrm{m} \cdot \mathrm{sr}}\right]. \tag{4.4}$$

From the definitions above, it can be noted that the radiance, unlike the irradiance, is a directional quantity and provides information on the direction of propagation of the radiation. In practice, the (spectral) radiance and irradiance are related by the integrals:

$$F_\lambda^\uparrow = \int_{2\pi} I_\lambda^\uparrow(\hat{\mathbf{\Omega}})\hat{\mathbf{n}} \cdot \hat{\mathbf{\Omega}}\,\mathrm{d}\omega, \qquad F_\lambda^\downarrow = \int_{2\pi} I_\lambda^\downarrow(\hat{\mathbf{\Omega}})\hat{\mathbf{n}} \cdot \hat{\mathbf{\Omega}}\,\mathrm{d}\omega, \tag{4.5}$$

---

[1] Radiance is sometimes called *intensity* in the literature, hence the symbol I.

in which the arrows indicate upward (↑) or downward (↓) propagation of radiation with respect to the oriented surface in Fig. 4.1 and the integrals are performed over the upper (↑) and lower (↓) hemisphere. In those cases when the spectral radiance is isotropic (i.e. independent of the direction of propagation, $I_\lambda(\hat{\mathbf{\Omega}}) = I_\lambda = \text{const}$), the relations (4.5) simplify to[1]:

$$F_\lambda^\uparrow = \pi \cdot I_\lambda^\uparrow, \qquad F_\lambda^\downarrow = \pi \cdot I_\lambda^\downarrow. \tag{4.6}$$

It is very common for natural EM radiation (for example solar radiation) to be a broadband (or incoherent) quantity, hence spanning a wide range of non-interfering wavelengths or frequencies. However, the use of spectral quantities is a practical theoretical approximation that will be applied throughout the chapter.

The nature of EM radiation and its interactions with a medium can be characterised by three fundamental processes: *thermal emission*, *absorption* and *scattering*, which we will examine briefly below.

Thermal emission

Every object or substance with a temperature above absolute zero (i.e. $T = -273.15°\text{C}$) emits EM radiation. This process can be seen as the conversion of part of its internal energy into radiant energy and can be explained in terms of the decay of quantum states, which release energy in the form of photons.

In order to quantitatively describe the emission of thermal radiation, we introduce the concept of a *blackbody*. A blackbody is an idealised object able to absorb all radiation at any wavelengths, and as we will discuss in a later paragraph on absorption, a perfect absorber behaves also like a perfect emitter. A common thought experiment to visualise the concept of blackbody is that of a cavity with a hole (see, for example, section §6.1 in Petty (2006)). The (isotropic) spectral radiance $B_\lambda(T)\,\mathrm{d}\lambda$ emitted in the interval $\lambda, \lambda + \mathrm{d}\lambda$ by a blackbody at temperature T is defined by the Planck function:

$$B_\lambda(T) = \frac{2hc^2}{\lambda^5 \left(e^{\frac{hc}{k_B \lambda T}} - 1 \right)}, \tag{4.7}$$

---

[1] Note that the infinitesimal solid angle $\mathrm{d}\omega$ can be expressed in spherical coordinates as $\mathrm{d}\omega = \sin(\vartheta)\mathrm{d}\vartheta\mathrm{d}\phi$, with ϕ being the azimuthal angle.

in which: c is the speed of light, h is Planck's constant ($h = 6.626 \cdot 10^{-34}\,\mathrm{m^2\,kg\,s^{-1}}$) and k_B is Boltzmann's constant ($k_B = 1.381 \cdot 10^{-23}\,\mathrm{m^2\,kg\,s^{-2}\,K^{-1}}$). Given a blackbody at temperature T, its emitted spectral radiance determined by (4.7) peaks at a specific wavelength $\lambda_{\max}$ determined by the *Wien Displacement law*:

$$\lambda_{\max} = \frac{2897.8}{T} \qquad \left[\frac{\mu\mathrm{m} \cdot \mathrm{K}}{\mathrm{K}}\right]. \tag{4.8}$$

For certain RT applications, the broadband radiation emitted by a source can also be of interest. In particular, the broadband energy flux density emitted by a blackbody F_{BB} follows the *Stefan-Boltzmann law*, which is obtained by substituting (4.7) into (4.5) and integrate the result over the whole EM spectrum. Its final expression reads:

$$F_{BB}(T) = \sigma_B T^4, \tag{4.9}$$

in which σ_B is called the Stefan-Boltzmann constant:

$$\sigma_B = \frac{2\pi^5 k_B^4}{15c^2h^3} = 5.67 \cdot 10^{-8} \quad \left[\frac{\mathrm{W}}{\mathrm{m^2} \cdot \mathrm{K^4}}\right]. \tag{4.10}$$

The blackbody theory becomes useful when the thermal emission of a real-world source (for example a certain surface) is considered. Since a blackbody is a perfect emitter, no real surface will emit as much radiation as the Planck function (4.7) prescribes. Therefore, the thermal emission of a surface at temperature T at any given wavelength λ can be characterised in terms of its spectral *emissivity* ϵ_λ, defined as the ratio between the radiation I_λ emitted by the surface in a certain direction $\hat{\mathbf{\Omega}}$ over the radiation $B_\lambda(T)$ that a blackbody would emit at the same temperature, that is:

$$\epsilon_\lambda = \frac{I_\lambda(\hat{\mathbf{\Omega}})}{B_\lambda(T)}. \tag{4.11}$$

Clearly, ϵ_λ takes values comprised in the interval $[0, 1]$ and $\epsilon_\lambda = 1$ in the case of a blackbody. The spectral emissivity of a surface as defined in (4.11) can also depend on other physical properties, such as its temperature T and the direction considered $\hat{\mathbf{\Omega}}$.

An object (or surface) whose emissivity ϵ_λ is constant throughout the EM spectrum is called a *graybody*. Therefore, a coefficient ϵ called a graybody emissivity (i.e. $\epsilon_\lambda = \epsilon$) can be defined as the ratio:

$$\epsilon = \frac{F}{F_{BB}}, \tag{4.12}$$

in which F is the total broadband irradiance emitted by the object (or surface).

A useful concept related to these considerations regarding thermal emission and commonly used in many satellite applications is that of *brightness temperature*. Given a source of (monochromatic) thermal radiation, this quantity is defined as the temperature T of an equivalent blackbody emitting the same amount of radiation according to the Planck function (4.7).

Finally, as this thesis will later focus on microwave radiation, it is useful to introduce an approximation of the Planck function (4.7) that is applicable in this part of the EM spectrum (i.e. with $\lambda > 1$mm). For large enough values of the wavelength λ, the argument of the exponential in (4.7) becomes small, that is:

$$\frac{hc}{k_B \lambda T} \ll 1;$$

a Taylor expansion of $e^{\frac{hc}{k_B \lambda T}}$ in (4.7) leads to the *Rayleigh-Jeans law*:

$$B_\lambda(T) \approx \frac{2ck_B}{\lambda^4} T. \tag{4.13}$$

The validity of (4.13) depends on the value of λ at a given T. We will discuss this approximation in detail in the section regarding the radiative scheme developed for the ismodRSW model (see section §4.3.1).

Absorption

The absorption of EM radiation involves the conversion of radiant energy into the thermal or chemical energy of an object. Hence, the process can be seen as the inverse of thermal emission, in which the quantum states of atoms and molecules are promoted to higher energy levels by the absorption of photons.

The ability of a surface to absorb EM radiation is inherently linked to its ability to emit it. In fact, according to *Kirchhoff's law* of thermal radiation, the absorptivity a_λ of a surface for radiation at a certain wavelength λ coming from a certain direction $\hat{\mathbf{\Omega}}$ equals the value of emissivity for the same surface:

$$a_\lambda(\hat{\mathbf{\Omega}}) = \epsilon_\lambda(\hat{\mathbf{\Omega}}). \tag{4.14}$$

Theoretically, the validity of Kirchhoff's law is restricted to the spectral and directional dependence present in (4.14) and depends on the fulfilment of *Local Thermodynamic Equilibrium* (LTE) conditions, in which the energy exchange between the molecules

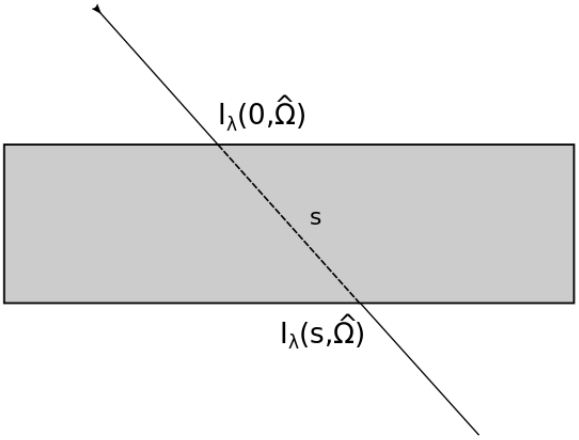


Figure 4.2: Layer of medium crossed by a beam of EM radiation of wavelength λ along the path s.

of a medium is dominated by their collisions rather than by radiative processes, such as the emission and the absorption of photons. In this thesis, we will often generalise this concept more broadly to the emission and absorption of surfaces and objects that can be treated as graybodies.

The absorption of EM radiation propagating through a medium is described by *Beer's Law* (or *extinction law*). Given a layer of medium like the one in Fig. 4.2 and a beam of EM radiation at a certain wavelength λ travelling through it along the path s in the direction $\hat{\mathbf{\Omega}}$, the radiance attenuated by absorption at any point along the path s can be defined as:

$$I_\lambda(s,\hat{\mathbf{\Omega}}) = I_\lambda(0,\hat{\mathbf{\Omega}})e^{-\tau_s(\lambda)}, \tag{4.15}$$

in which $\tau_s(\lambda)$ is a dimensionless quantity called *optical depth*, reading:

$$\tau_s(\lambda) = \int_0^s \beta_a(\lambda,s')\mathrm{d}s', \tag{4.16}$$

with $\beta_a(\lambda,s)$ [m^{-1}] being the *absorption coefficient* along the path s for radiation of wavelength λ. In other words, because of absorption processes, the radiance travelling through a medium decreases exponentially.

Equation (4.15) can also be rewritten as:

$$I_\lambda(s,\hat{\mathbf{\Omega}}) = t_\lambda(0,s)I_\lambda(0,\hat{\mathbf{\Omega}}),$$

in which $t_\lambda(0,s)$ is called a *transmittance* along the path s, that is:

$$t_\lambda(0,s) = e^{-\tau_s(\lambda)}. \tag{4.17}$$

Scattering

We call *scattering* the process by which the radiation travelling through a medium is redirected from the original direction in which it was travelling. In certain applications, scattering and absorption are considered jointly as a single process called *extinction* (or attenuation). To this aim, the (absorption) optical depth of Eq. (4.16) can be redefined as an extinction optical depth $\tau_e(\lambda)$, that is:

$$\tau_e(\lambda) = \int_0^s \left(\beta_a(\lambda,s') + \beta_s(\lambda,s')\right) \mathrm{d}s' = \int_0^s \beta_e(\lambda,s')\mathrm{d}s', \tag{4.18}$$

in which $\beta_e(\lambda,s)$ [m^{-1}] and $\beta_s(\lambda,s)$ [m^{-1}] represent the extinction and the scattering coefficients, respectively. Hence, both scattering and absorption will contribute to the attenuation of a beam of EM radiation travelling through a layer of medium like the one in Fig. 4.2. Beer's Law can also be modified in order to include scattering processes by redefining (4.15) as:

$$I_\lambda(s,\hat{\Omega}) = I_\lambda(0,\hat{\Omega})e^{-\tau_e(\lambda)}. \tag{4.19}$$

The properties of scattering processes depend on many factors, but two are of particular importance: i) the size of the particles which the radiation is scattering from, and ii) the wavelength λ of the radiation itself. A *size parameter* χ can be defined as the ratio:

$$\chi = \frac{2\pi r}{\lambda}, \tag{4.20}$$

in which r indicates the radius of a sphere approximating the particles in the medium. The size parameter χ can therefore be used to define three different regimes (see, for example, Kidder et al. (1995)):

- **Geometric optics** ($\chi > 50$): the laws of geometric optics apply whenever EM radiation scatters from particles or objects much greater than its wavelength; in this regime, the shape of the particles subject to the scattering is generally very important. The interaction between solar radiation and most hydrometeors can be described in terms of geometric optics.

- **Mie scattering** ($0.1 < \chi < 50$): this is the regime in which the size of the particles and the wavelength of the radiation are comparable. Mie scattering is used in radar meteorology to describe the scattering of radio waves generated by radars on snow and rain droplets (approximated to spherical particles).
- **Rayleigh scattering** ($\chi < 0.1$): in this regime, the particles are much smaller than the wavelength of the EM radiation and the shape of the particles themselves is generally unimportant. One area of application is the interaction of solar radiation with atmospheric molecules of O_2 and N_2. Rayleigh scattering is strongly dependent on the wavelength of the incident radiation and becomes less relevant for processes involving infrared and microwave radiation (cf. sections §4.1.3).

4.1.2 Radiative Transfer equations

In the previous section we have defined several concepts related to the propagation of EM radiation in a generic medium. It is now time to apply some of those notions to the transmission of radiation in the atmosphere.

Figure 4.3 presents a schematic representation of a satellite located at altitude z_{sat}, zenith angle θ and azimuthal angle ϕ in a *plane-parallel atmosphere*: a geometrical approximation in which the radiative properties of the atmosphere are considered as uniquely dependent on the vertical coordinate z, while all horizontal variations are neglected. Under these assumptions, the infinitesimal optical path $\mathrm{d}s$ defined in the previous section can be rewritten in terms of an infinitesimal atmospheric layer of thickness $\mathrm{d}z$ as $\mathrm{d}s = \mathrm{d}z/\mu$, where μ is the cosine of the zenith angle θ, i.e. $\mu = \cos(\theta)$.

The amount of radiation reaching the satellite in Fig. 4.3 is the result of various processes (thermal emission, absorption, scattering) occurring between the surface and the sensor, within any infinitesimal layer of depth $\mathrm{d}z$. Formally, they are summarised in the *Radiative Transfer equation* (cf., for example, equation (3.32) in Kidder et al. (1995)):

$$\mu \frac{\mathrm{d}I_\lambda(\mu,\phi)}{\mathrm{d}\tau} = (1-\tilde{\omega})B_\lambda(T) - I_\lambda(\mu,\phi) + \frac{\tilde{\omega}}{4\pi}\int_{4\pi} p(\mu,\phi;\mu',\phi')I_\lambda(\mu',\phi')\mathrm{d}\mu'\mathrm{d}\phi', \quad (4.21)$$

which describes the variation of the spectral radiance I_λ with respect to the extinction vertical optical depth, defined as:

$$\tau = \int_{z_1}^{z_2} \beta_e(\lambda, z)\mathrm{d}z. \quad (4.22)$$

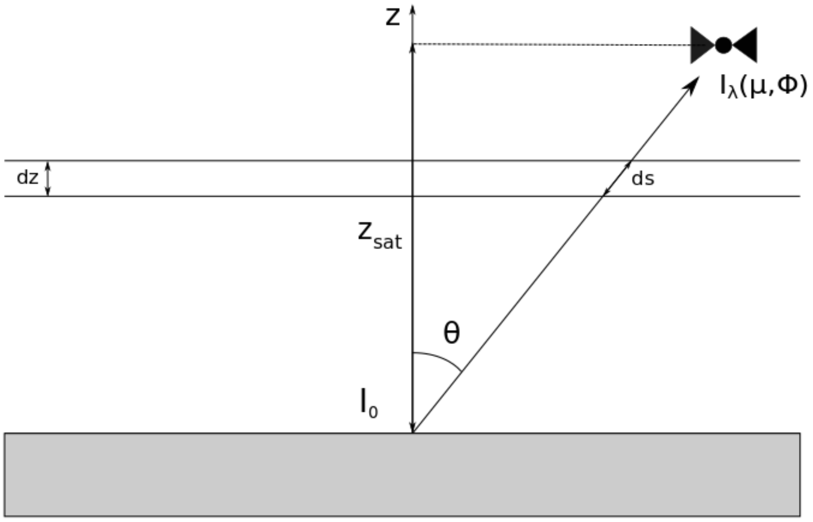


Figure 4.3: Schematic representation of a plane-parallel geometry with a satellite located at altitude z_{sat}, zenith angle θ and azimuthal angle ϕ (indicating the rotation around the axis z). The infinitesimal layer depth dz is related to the infinitesimal optical path ds by $\mathrm{d}z = \mathrm{d}s \cdot \cos(\theta) = \mathrm{d}s \cdot \mu$. $I_\lambda(\mu, \phi)$ and I_0 indicate the radiance reaching the satellite and leaving the surface, respectively.

In particular, the terms on the right-hand side of equation (4.21) represent (from the left to the right): the emitted thermal radiation emerging from the layer (and possibly attenuated by scattering); the extinction of radiation passing through the layer; the radiation redirected by scattering processes within the layer towards the direction of propagation (μ, ϕ). In the third term, $p(\mu, \phi; \mu', \phi')$ represents the *scattering phase function*, which describes the likelihood of the radiation coming from any direction (μ', ϕ') of being redirected towards (μ, ϕ). Finally, $\tilde{\omega} = \beta_s/\beta_e$ is the so-called *single-scatter albedo* defined as the ratio between the scattering and the extinction coefficients. Depending on the situation, equation (4.21) can be simplified by neglecting the terms which are deemed unimportant. For example, by ignoring all scattering processes (i.e. $\beta_s = \tilde{\omega} = 0$), one obtains the *Schwarzchild's equation*, that is:

$$\mu \frac{\mathrm{d}I_\lambda(\mu, \phi)}{\mathrm{d}\tau} = B_\lambda(T) - I_\lambda(\mu, \phi). \tag{4.23}$$

We will turn to this equation and its solution at the end of the next section.

4.1.3 Atmospheric radiation

The EM radiation that pervades the atmosphere and is measured by satellites comes primarily from two sources of emission: the Sun and the Earth. By approximating them to blackbodies, and using their surface temperature as brightness temperature, we can utilise the Planck function in Eq. (4.7) to calculate their emission spectra, which are shown in Fig. 4.4. Clearly, their different surface temperatures (here we take $T_s = 5700$ K for the Sun, in red, and $T_e = 300$ K for the Earth, in blue) lead to very different results. As predicted by the Wien-displacement law (4.8), the intensity of the solar radiation peaks at $\lambda^s_{\max} \approx 0.51\ \mu$m, which is the typical wavelength of green visible light, whereas the radiation emitted by the Earth peaks in the infrared spectrum at $\lambda^e_{\max} \approx 9.7\ \mu$m (see the black dashed lines in Fig. 4.4). Moreover, as highlighted by the two different vertical axes, the radiation emitted by the Sun is approximately 10^6 times more intense than that produced by the Earth. However, their spectra barely overlap, and this aspect makes it possible to separate clearly the two sources of radiation in most satellite-related applications.

As a consequence, atmospheric radiation can be split into two intervals: shortwave (SW) radiation emitted by the Sun and longwave (LW) radiation emitted by the Earth's surface and atmosphere, including clouds, water vapour, and other gases. The SW radiation spans a broad range of wavelengths which includes ultraviolet light (UV,

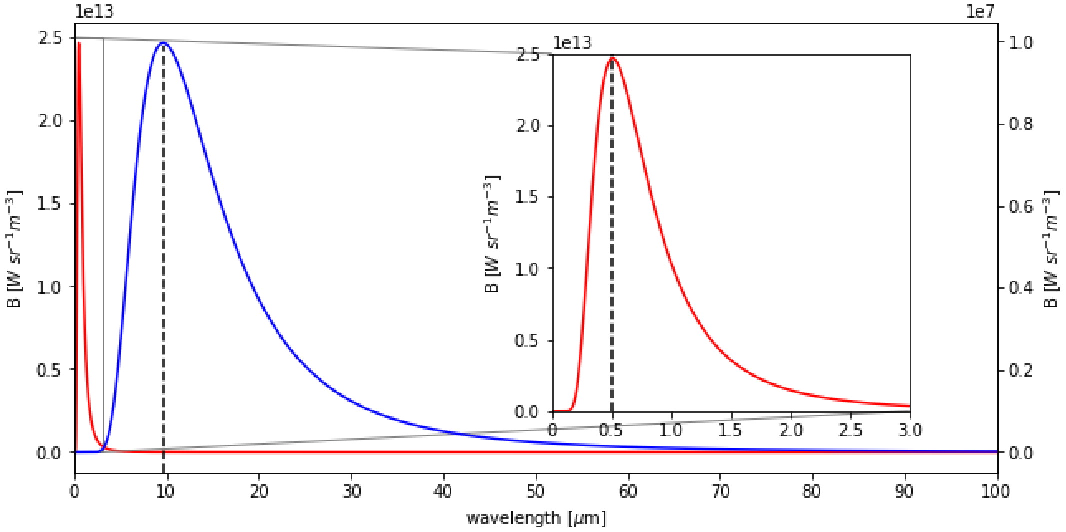


Figure 4.4: Emission spectra computed via the Planck function (4.7) for a blackbody at $T = 5700$ K (red line) and one at $T = 300$ K (blue line). These functions are meant to approximate the emission spectra of the Sun and the Earth, respectively. The black dashed lines indicate the wavelength of maximum emission estimated with the Wien-displacement law (4.8).

$0.1 < \lambda < 0.4\,\mu$m), the visible spectrum ($0.4 < \lambda < 0.7\,\mu$m) and near-infrared radiation (NIR, $0.7 < \lambda < 4.0\,\mu$m). The LW radiation also includes a wide spectrum of wavelengths, which covers the whole infrared range ($0.7\,\mu\text{m} < \lambda < 1\,\text{mm}$), microwaves ($1\text{ mm} < \lambda < 1\text{ m}$) and radio waves ($\lambda > 1$ m).

This separation between SW and LW radiation is particularly useful when their interaction with various gaseous components of the atmosphere (e.g. water vapour, carbon dioxide (CO_2), Ozone (O_3)) is concerned, as these effects vary strongly across the EM spectrum. In addition, there are other atmospheric components that perturb the transmission of EM radiation, such as clouds, precipitation, and aerosol particles.

The impact of atmospheric gases on EM radiation can be analysed by comparing the theoretical emission spectra shown in Fig. 4.4 with a measured spectrum. Many books on radiative transfer and atmospheric sciences report such comparisons and one example is shown in Figure 4.5 (this appears in Fig. 8.2 of Salby (1996)), where the EM spectra of solar radiation measured at sea level and at the top of the atmosphere, as well as the theoretical emission spectrum of a blackbody at $T = 5900\,$K, are shown. In general, the intensity of solar radiation reaching the surface (grey shaded area) is only a fraction of its theoretical estimated maximum. This reduction in intensity is the result of various processes that contribute to the extinction of solar radiation while it travels through the atmosphere. As an example, a broad reduction in intensity of UV light can be explained by the thermochemical reactions involving ozone taking place in the stratosphere (the layer of the atmosphere between 10 and 50 km above the surface), leading to a smaller portion of UV radiation reaching the ground. Moreover, the molecules of O_2 and N_2 play a big role in the scattering of solar radiation at small wavelengths (see, for example, section §9.4 of Salby (1996) on Rayleigh scattering), resulting in further extinction of UV and visible solar radiation. In addition, Fig. 4.5 shows a significant reduction in the transmitted radiance in various bands (both in the visible and the infrared spectrum) centred on specific wavelengths which are associated with absorption from various gases, such as water vapour, CO_2 and O_2 (each absorption band is indicated and its respective absorber are indicated in the figure). The nature of these bands and their position within the spectrum can be explained in terms of the interaction between the molecules of a gas and the EM radiation. Many books have covered this subject in detail: see for example section §8.3 of Salby (1996) and section §4.3 of Stamnes et al. (2017). In the context of satellite observations, it is more interesting to compare the LW radiation measured by an instrument located at the top of the atmosphere against various blackbody emission spectra, representing the

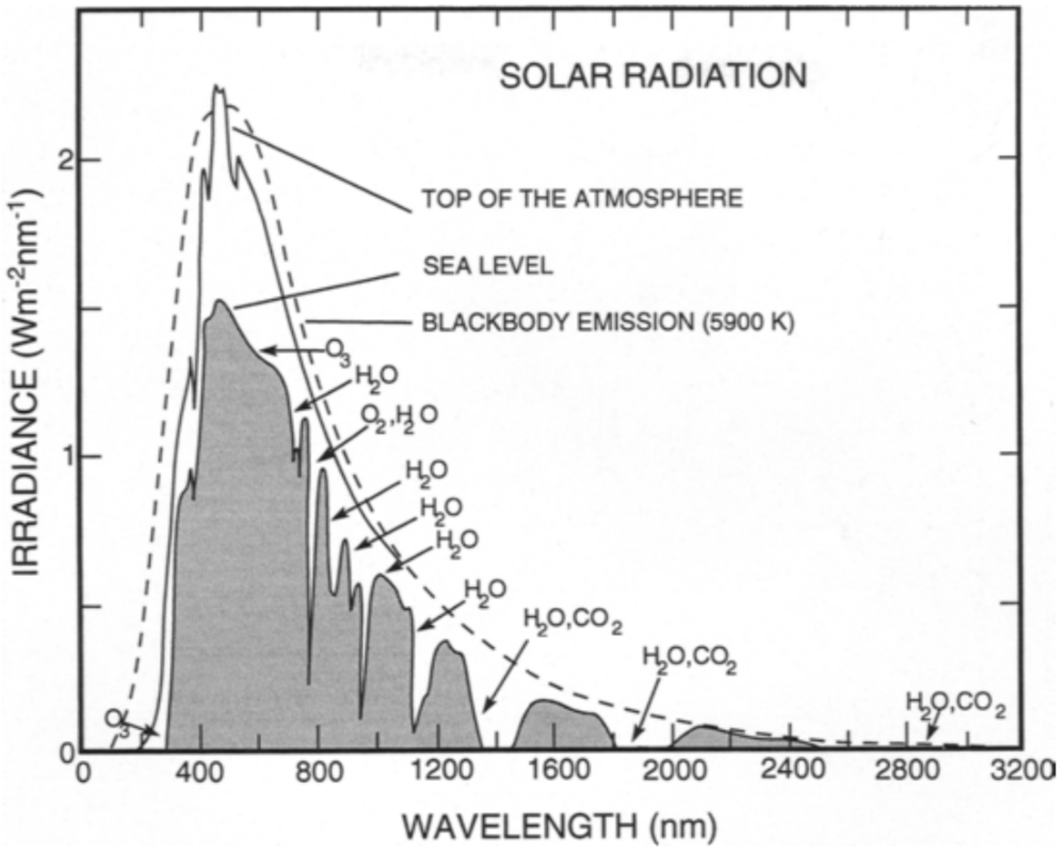


Figure 4.5: Spectrum of solar radiation measured (in irradiance) at the surface (grey shaded area) and at the top of the atmosphere (white area) against the theoretical emission spectrum of a blackbody at $T = 5900$ K. This figure is shared with the permission of Elsevier and was published in *Fundamentals of Atmospheric Physics*, Murry L. Salby, Chapter 8, page 201, Copyright 1996 by ACADEMIC PRESS and is an adaptation of Figure 3.1 in *Solar and Terrestrial Radiation*, Kinsell Coulson, Chapter 3, page 40, Copyright 1975 by ACADEMIC PRESS.

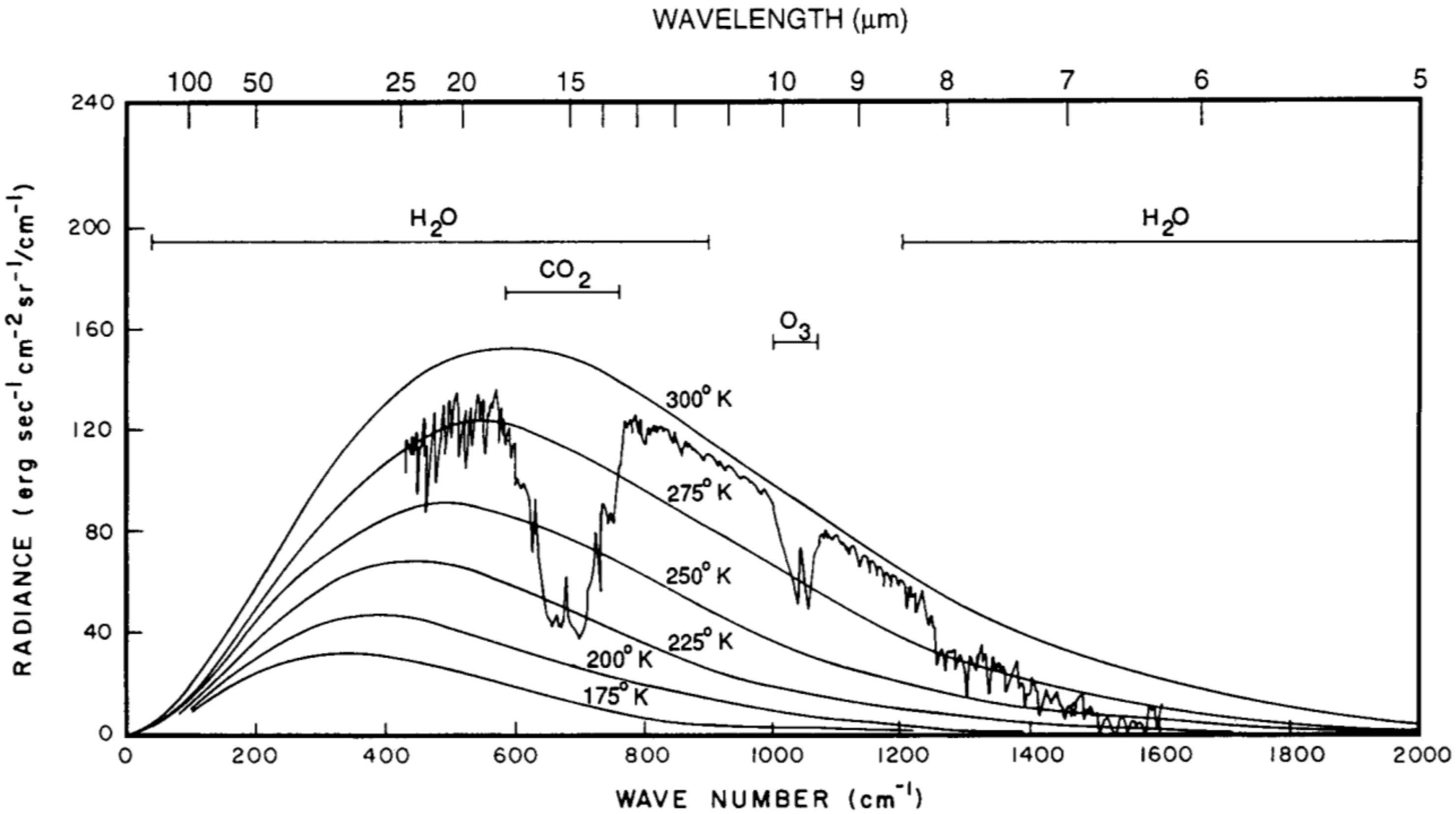


Figure 4.6: Spectrum of LW radiation measured at the top of the atmosphere by Nimbus-4 IRIS at coordinates 215 deg W, 15 deg N. This figure is shared with the permission of Elsevier and was published as Figure 8.5 in *Fundamentals of Atmospheric Physics*, Murry L. Salby, Chapter 8, page 204, Copyright 1996 by ACADEMIC PRESS and is an adaptation of Figure 4.1 in *An introduction to Atmospheric Radiation*, Kuo-Nan Liou, Chapter 3, page 89, Copyright 1980 by ACADEMIC PRESS.

theoretical emission of the Earth'surface (i.e. $T = 300\,\mathrm{K}$) and that of the atmospheric layers above it ($T = 275$ K, $T = 250$ K, etc.). This is shown, for example, in Fig. 4.6 (this appears in Fig. 8.5 of Salby (1996)). In this case, since the scattering of LW radiation by gas molecules is negligible, most of the differences between the predicted and the measured spectrum can be explained in terms of the absorption and emission properties of various gases, as well as their concentration profiles. As a result, since the temperature in the troposphere decreases with height, one observes a reduction in the measured radiance (or equivalently brightness temperature) within the absorption bands of various optically active gases (for example around $\lambda = 15\mu$m for the CO$_2$, or at $\lambda = 9.5\,\mu$m for O$_3$). In other words, the majority of the radiation reaching the satellite at these wavelengths comes from higher (and colder) layers in the atmosphere, where these gases reside. Overall, the values of brightness temperature associated with these absorption bands will be strongly influenced by the temperature of the atmospheric layer where most of the absorption has taken place.

In order to extend these considerations quantitatively, one can consider the solution to Schwarzchild's equation (4.23) for the LW radiation reaching a satellite located at top of the atmosphere (like the one in Fig. 4.3). By choosing $\tau_\lambda = \int_0^z \beta_a(\lambda, z)\mathrm{d}z$ as the vertical coordinate, one obtains (see section §3.3.1 in Kidder et al. (1995) for a proof):

$$I_\lambda(\mu, \phi) = I_0 \cdot e^{-\frac{\tau_0}{\mu}} + \int_0^{\tau_0} e^{-\frac{\tau_0 - \tau_\lambda}{\mu}} B_\lambda(T) \frac{\mathrm{d}\tau_\lambda}{\mu}, \tag{4.24}$$

in which I_0 is the radiance leaving the surface and τ_0 indicates the optical depth computed between the surface ($z = 0$) and the satellite height ($z = z_{sat}$), that is $\tau_0 = \int_0^{z_{sat}} \beta_a(\lambda, z)\mathrm{d}z$.

Equation (4.24) can be manipulated and rewritten to be a function of z (noting that from (4.22), and without scattering, $\mathrm{d}\tau = \beta_a \mathrm{d}z$) and exploiting the definitions of the transmittance functions, that is:

$$I_\lambda(\mu, \phi) = I_0 \cdot t_\lambda(0, z_{sat})^{\frac{1}{\mu}} + \int_0^{z_{sat}} B_\lambda(T(z))\, W(z)\, \mathrm{d}z, \tag{4.25}$$

in which $t_\lambda(0, z_{sat})$ indicates the transmittance of the whole atmosphere comprised between the surface ($z = 0$) and the satellite height ($z = z_{sat}$):

$$t_\lambda(0, z_{sat}) = e^{-\tau_0}. \tag{4.26}$$

The function $W(z)$ multiplying the emission term $B_\lambda(T(z))$ is called a *weighting function* and is defined as:

$$W(z) = \frac{\mathrm{d}}{\mathrm{d}z}\left(t_\lambda(z, z_{sat})^{\frac{1}{\mu}}\right) = \frac{\rho_a(z)\, k_a}{\mu} \cdot t_\lambda(z, z_{sat})^{\frac{1}{\mu}}, \tag{4.27}$$

in which we have expanded the absorption coefficient β_a as the product of the gas density $\rho_a(z)$ [$\mathrm{kg\,m^{-3}}$] and a mass absorption coefficient k_a [$\mathrm{kg^{-1}\,m^2}$]. In this case, $t_\lambda(z, z_{sat})$ indicates the transmittance of the atmosphere between the altitude z and the height of the satellite z_{sat}, i.e.:

$$t_\lambda(z, z_{sat}) = e^{-(\tau_0 - \tau)}. \tag{4.28}$$

The solution to Schwarzchild's equation (4.25) illustrates that, in the absence of scattering, the radiation reaching a satellite is the sum of two contributions. In particular, the first term in (4.25) represents the radiation coming from the surface (either reflected or emitted), attenuated by the whole atmospheric column via the transmittance $t_\lambda(0, z_{sat})$, whereas the integral term represents the emission contribution of each infinitesimal atmospheric layer dz, possibly attenuated by the atmospheric layers above. In this regard, we explore the relationship between the transmittance and the weighting function in Fig. 4.7, where both functions are plotted for various absorptivity profiles determined by the mass absorption coefficient k_a. For the same (exponentially decaying) density profile $\rho_a(z)$, higher values of k_a lead to a stronger reduction in the atmospheric transmittance, and the corresponding weighting functions peak higher in the atmosphere, where the contribution of the emitted radiance $B(T(z))$ to the total radiance $I_\lambda(\mu, \phi)$ in (4.25) will be maximum. In a situation like the one in the troposphere where the temperature decreases with height, the equivalent brightness temperature resulting from (4.25) will be lower for a strongly absorbing gas (e.g. $k_a = 5.0\ \mathrm{kg^{-1}\,m^2}$, red curve in Fig. 4.7) than for a weakly absorbing one (e.g. $k_a = 0.1\ \mathrm{kg^{-1}\,m^2}$, blue curve in Fig. 4.7), in which most of the radiation comes from the lower troposphere and the surface.

Of course, both solar and terrestrial radiation are subjected to scattering and absorption processes when they interact with rain, clouds and other hydrometeors. In this regard, it is possible to model the effects of these interactions by adding additional terms to (4.25). We will cover this topic in a later section, focussing specifically on the transmission of microwave radiation through clouds and precipitation, which will help us set up the cloud model for the *ismodRSW* model.

4.1.4 Types of satellites and instruments

Satellites used for meteorological purposes can be classified according to various properties related to their orbit, the type of instruments they carry on board and the

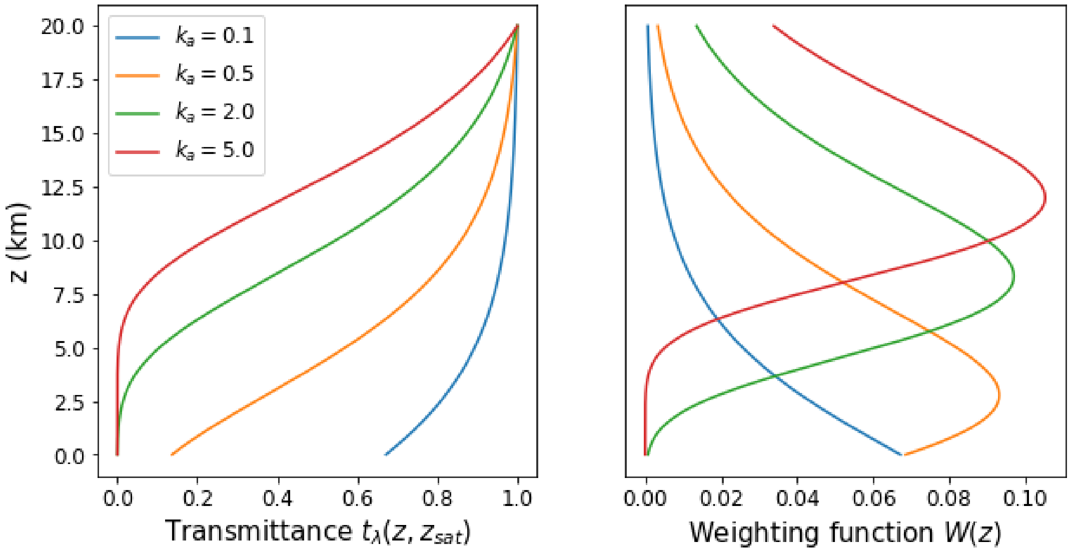


Figure 4.7: Transmittance profiles ($t_\lambda(z, z_{sat})$, left) and associated weighting functions ($W(z)$, right) for four arbitrary values of the mass absorption coefficient k_a (assuming $\mu = 1$). The density of the absorbing gas is defined as $\rho_a(z) = e^{-z/H}$ with height scale $H = 4\,\text{km}$. The values of k_a are reported in the legend.

way they perform their measurements. Here we present some of the most common terms and definitions used in the literature, which will recur later in the chapter while summarising past and present developments of satellite DA.

Polar-orbiting and geostationary satellites

Meteorological satellites are classified according to the orbit they travel on. *Geostationary* satellites are situated on a *geosynchronous* orbit over the equator and possess the same angular velocity as the Earth's surface. In practice, they hold a constant view on the same portion of the globe from an altitude of around 36,000 km. They provide good coverage of most equatorial and extra-tropical areas, but they cannot view over the poles. *Polar*-orbiting satellites are situated on a *sunsynchronous* orbit. This trajectory is chosen such that the satellite crosses the equator multiple times a day and at the same local time every day. The satellites travelling on this orbit are much closer to the surface – that is, they gravitate at altitudes in the range of 600-800 km – and complete a full orbit in about 1 to 2 hours. Therefore, they have a much narrower view than most geostationary satellites, but their spatial coverage is more homogeneous and they also cover the polar regions.

Channels

Most of the definitions and the considerations made in the earlier sections §4.1.1-4.1.2 apply to monochromatic radiation. Unfortunately, no instrument is able to measure radiation at a single wavelength. Instead, satellites are generally equipped with devices able to detect radiation within an interval of wavelengths $[\lambda_1, \lambda_2]$ called a *channel*. In the case of a radiometer, for example, the radiation measured in a certain channel is defined as (see also Chapter 3 of Kidder et al. (1995)):

$$I_{\mathrm{channel}} = \frac{\int_{\lambda_1}^{\lambda_2} I_\lambda f(\lambda) \mathrm{d}\lambda}{\int_{\lambda_1}^{\lambda_2} f(\lambda) \mathrm{d}\lambda}, \tag{4.29}$$

in which $f(\lambda)$ is the response function of the instrument, determining what portion of the radiation I_λ is detected at a certain wavelength λ within the interval $[\lambda_1, \lambda_2]$. Their number, position and width are chosen so that the most important spectral features are covered. In this regard, a compromise is required between the desire of building very narrow channels able to capture specific absorption lines and, at the same time, minimising the amount of noise, which in very narrow bands can hamper the quality of the measurements.

Field of view

The horizontal resolution of a satellite instrument is generally associated with its Field of View (FOV). The FOV is defined as the solid angle within which the detector of a device is able to receive EM radiation. The intersection between the solid angle defining the FOV and the Earth's surface is called ground FOV and varies according to the satellite's height.

Imagers and sounders

Meteorological satellites have been historically divided into *imagers* and *sounders*, depending on whether their instruments were able to observe the Earth's surface or not. Satellite *sounders*, operating channels that are opaque to the surface, are commonly utilised to reconstruct the vertical profiles of temperature and water vapour. This approach was very common in the early days of satellite DA as we will see in the next section.

Satellite *imagers* are particularly useful in the development of the AMVs (cf. section §4.2.3). They also constituted the first generation of weather satellites.

Passive and active satellites

Another common criterion to classify meteorological satellites is whether they are *active* or *passive*. Active satellites emit EM radiation towards a target and subsequently measure the returned signal. Passive instruments measure radiation emitted by another sources (e.g. the Earth or the Sun).

Most satellite applications discussed in this thesis are based on passive instruments, but active satellites endowed with lidar or radar instruments have gradually occupied an important place in the global observing system.

4.2 The use of satellites in data assimilation

In this section we will provide an historical overview of the role played by satellite observations in data assimilation. Of course, this is a very large topic and it would be impossible to provide a complete and satisfactory review on the subject in just a few pages. As a consequence, we will present only a summary of the main developments and advances that can help collocate the work done in this thesis in a broader perspective. For more in-depth and detailed reviews we refer the reader to Derber (2014); Eyre (2007); Eyre et al. (2020); Menzel (2001); Menzel et al. (2018); Saunders (2021).

4.2.1 The early imagers of the 1960s

The era of weather satellites started in 1960 when NASA sent TIROS-1 into orbit (see Allison & Neil (1962) for a detailed report of the mission). TIROS-1 became the first meteorological satellite to return images of the Earth and paved the way for other missions, such as (cf. Table 1 in Eyre et al. (2020)): Nimbus-1 (the first satellite designated for meteorological research, 1964), the Applications Technology Satellite, or ATS-1 (the first geostationary weather satellite, 1966) and the Environmental Science Services Administration satellite, or ESSA-1 (the first operational satellite to be used for meteorological purposes, 1966). All these satellites carried imaging instruments on board.

4.2.2 From the sounder retrievals to the assimilation of radiances

The first satellite sounder, Nimbus-3, was also launched by NASA in 1969. This satellite was equipped with the Satellite Infra-Red Spectrometer (SIRS) and the Infrared Interferometer Spectrometer (IRIS), and aimed at reconstructing vertical profiles of temperature, water vapour and ozone. Wark (1970) showed that the retrieved temperature profiles were in agreement with radiosonde measurements, suggesting that the technique could be used for operational and monitoring purposes. In a similar study, Smith et al. (1970) illustrated how the operational system in use at the National Meteorological Center (NMC) of the United States could benefit from the inclusion of such retrieved profiles into the objective analysis.

In 1972, the National Oceanic and Atmospheric Administration 2 (NOAA-2) satellite was launched and started its operations, carrying on board the Vertical Temperature Profile Radiometer (VTPR). Unlike Nimbus-3, NOAA-2 was a fully fledged operational polar-orbiting satellite (in fact, it was the first ever operational sounder) and was at the centre of a number assimilation experiments during the 1970s. Two other satellite sounders became operational around the same time: Nimbus-6 (launched in 1975) – which carried on board both the High-resolution Infrared Radiation Sounder (HIRS) and the SCAnning Microwave Spectrometer (SCAMS) – and TIROS-N (launched in 1978), equipped with the Microwave Sounding Unit (MSU), the Stratospheric Sounding Unit (SSU) and the HIRS (this combination of instruments is historically known as TIROS-N Operational Vertical Sounder, or TOVS). Ohring (1979) provides a nice overview of some of the early assimilation experiments realised with these instruments during the late 1970s and concluded that they provided a small benefit to weather forecasts.

Between December 1978 and November 1979, the WMO conducted a large campaign of atmospheric measurements called the First Global Atmospheric Research Program (GARP) Global Experiment (FGGE). By that time, a variety of satellite observations were already available and could subsequently be included in the experiment. As a consequence, several Observing System Experiments (OSEs) focussing on the assimilation of satellite soundings were carried out in the early 1980s. As an example, Halem et al. (1982) conducted a series of experiments showing that the assimilation of satellite data over regions with sparse conventional observations led to a significant reduction in 6-hour forecast errors downstream of those areas, with a larger improvement in the Southern Hemisphere (e.g. over Australia) than in Europe and North America.

This widely accepted idea of using satellite sounders to reconstruct vertical profiles of temperature and other atmospheric variables through inversion techniques (see Rodgers (1976)) reflected an attempt to treat satellite observations like radiosonde measurements, which before the advent of satellites constituted the main source of information on the vertical structure of the atmosphere. This approach, however, showed its limitations reasonably soon, as the vertical resolution of satellite observations was poor and could easily lead to inaccurate profiles. Moreover, the generation of retrievals relied on prior information (such as climatological data) which, whenever distant from the real atmospheric conditions, could lead to significant errors with complex characteristics, hence complicating their assimilation. A more detailed discussion of the challenges posed by the assimilation of retrievals is given in section §3.5 of Eyre et al. (2020).

As hinted already by Halem et al. (1982), since the number of observations available in the Southern Hemisphere was limited in the first place, the inclusion of satellite retrievals was initially helpful and contributed positively to the quality of weather forecasts (Kelly et al., 1978). Conversely, in the observation-rich Northern Hemisphere, the impact of satellite observations had been much smaller since the beginning, and the assimilation of retrieved vertical profiles was producing little or no benefit by the late 1980s, as both NWP models and assimilation techniques had meanwhile improved significantly. In other words, it was time for a paradigm shift.

In the early 1990s, the interest of the DA community towards variational schemes increased, and many operational centres moved towards the direct assimilation of radiances into their DA schemes. The first, intermediate step in this direction was taken by the European Centre for Medium-Range Weather Forecasts (ECMWF) in 1992, which implemented operationally a 1D-Var algorithm exploiting TOVS data (Eyre et al., 1993). With this method, satellite observations were converted first into a vertical temperature profile that was subsequently assimilated within the Optimal Interpolation scheme (OI, cf. for example Kalnay (2003), section §5.4.1) that was in use at that time as a DA algorithm. A more radical change was introduced in January 1996, when a full 3D-Var scheme that assimilated radiances directly was adopted (Andersson et al., 1994). In this regard, Eyre (1997) discussed the challenges and the advantages of this new, emerging approach. Several other operational weather centres switched to variational assimilation schemes around the same time. The Met Office followed a similar path, starting with some experiments on the assimilation of TOVS data through a 1D-Var scheme (Renshaw et al., 1997) and moving to a later implementation of a 3D-Var DA scheme which became operational in 1999 and could assimilate radiances directly (Lorenc et al., 2000). In the meantime, the National Centers for Environmental Prediction (NCEP) in the United States had already adopted a similar algorithm in 1995 (Derber & Wu, 1998).

4.2.3 Wind observations: atmospheric motion vectors (AMVs) and scatterometers

The possibility of carrying out wind observations using satellites represented a positive development in the history of satellite DA, as conventional wind measurements were scarce in large portions of the planet (i.e. over the oceans and the Southern Hemisphere). Being located outside the troposphere, satellites estimate the velocity and the direction of the wind indirectly, resorting to two different techniques: the *Atmospheric Motion Vectors* (AMVs) and the use of scatterometers.

Atmospheric Motion Vectors

Atmospheric Motion Vectors (AMVs) are used to estimate the speed and the direction of the wind in the atmosphere. Menzel (2001) provides an extensive review on the subject, including a detailed overview of the work done by pioneer Tetsuya Fujita, whose contribution to this topic goes as back as the first TIROS satellite mission in 1960.

AMVs are obtained by identifying the same clouds or water vapour features in a sequence of consecutive satellite images. Once the height of the selected tracer is known, its displacement in time can be used to estimate both the direction and the speed of the wind at a certain altitude. Therefore, the use of geostationary satellites represents a more natural choice for the production of the AMVs, as they have a constant view on the same portion of the Earth's atmosphere. To achieve this aim, several channels of various satellite imagers have been used over time, both in the visible and in the infrared spectrum. AMV data are produced by operational satellite agencies such as the National Environmental Satellite, Data, and Information Service (NESDIS) in the United States and the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT) in Europe. Schmetz et al. (1993) and Nieman et al. (1997) described the automated algorithms developed for the Meteosat and GOES 8/9 satellites, respectively.

AMVs started to be produced routinely in the 1970s and were already included in the FGGE experiment of 1979, during which cloud-based motion vectors were derived from five different geostationary satellites twice a day. An assessment of these early observations was made at the ECMWF by Kållberg et al. (1982); in their report, the authors commented positively on their impact, both on the analysis at the tropics, and on the global medium-range forecasts. In the meantime, as the new water vapour imager mounted on Meteosat (1978) became operational, Eigenwillig & Fischer (1982) started to explore the possibility of tracking areas and gradients of water vapour to produce AMVs. Later on, other studies on the impact of assimilating AMV-derived wind observations were conducted at the ECMWF: Tomassini et al. (1999) demonstrated that the use of such data had a positive impact on the 3D-Var system in use at the time, especially in the tropics and in the Southern Hemisphere; subsequently, Rohn et al. (2001) described the inclusion of the new generation of AMV data (with an improved time sampling of 90 minutes, down from 6 hours) into the 4D-Var scheme and found neutral impact in the Northern Hemisphere and a positive impact in the Southern Hemisphere. Similar impact studies were carried out at the Met Office, where Butterworth & Ingleby (2000) reported a degrading effect on the overall quality of the NWP forecasts. In a later report, Butterworth et al. (2002) showed that a revised satellite wind observation error variance in their DA system conveyed better results. A small positive impact from the same type of data was also reported at the NCEP by Zapotocny et al. (2008).

Scatterometers

While AMVs are used to reconstruct the wind vector in the atmosphere, scatterometers are used to estimate the speed and direction of winds on the surface of the ocean. Scatterometers are polar-orbiting, active satellites able to measure the radiation backscatter produced by the roughness of the ocean, which in turn is shaped by the intensity and the direction of the wind. These instruments generally emit pulses of radiation in two or more directions and produce two different estimates of wind direction and speed, an ambiguity that needs to be handled by the DA scheme.

After a couple of early experiments in the 1970s (the Microwave Radiometer Scatterometer and Altimeter, MRSA, on board Skylab and the Seasat-A Scatterometer System, SASS, on board Seasat), the first significant contribution to operational DA systems was given by the Active Microwave Instrument (AMI) mounted on the 1st European Remote-Sensing Satellite (ERS-1). A series of papers by Stoffelen & Anderson (1997a,b,c) describe in details the functioning and the instrumental properties of AMI, together with the methods used to interpret, disambiguate and actively assimilate its measurements at the ECMWF. Overall, they reported a positive impact on both the analysis and the short-range forecast. In addition, Andrews & Bell (1998) at the Met Office showed that the assimilation of ERS-1 measurements led to a benefit to the forecast system at all lead times in the Southern Hemisphere. The operational implementation of these data started at the Met Office in 1992, at NCEP (Caplan et al., 1997) in 1994, and at the ECMWF in 1996 (Eyre et al., 2020). Later on, in 1999, a new NASA satellite called QuikSCAT, equipped with a wind scatterometer called Seawinds was sent into orbit (see Hoffman & Leidner (2005) for more details). The data from QuikSCAT have been assimilated at both the ECMWF and the NCEP with good results (Chelton & Freilich, 2005; Chelton et al., 2006).

4.2.4 Radio occultation

Another (and more recent) type of satellite observations is represented by Radio Occultation (RO) measurements, which provide an estimate of the vertical gradient of the refractive index of the atmosphere[1]. This quantity can be related to the vertical profiles of air density and water vapour concentration, from which further information on various atmospheric quantities, such as temperature and humidity, can be inferred. The RO technique is based on the communication between two satellites (for example a navigation system satellite and another satellite with a Global Navigation Satellite System, or GNSS, receiver) that are separated by the Earth's limb and exchange a radio signal. While the two satellites are not in line of sight, the refractivity field bends

---

[1] The refractive index n of a medium is defined as the ratio between the speed of light in vacuum c and the phase velocity in the medium v, that is: $n = \frac{c}{v}$.

the emitted signal while crossing the atmosphere, which therefore reaches the receiver. By knowing the position of the two satellites, it is possible to infer the amount by which the radio signal has been deflected, which can therefore be related to the gradient of the refractivity index. Melbourne et al. (1994) provides an extensive description of the topic.

The assimilation of RO data was discussed initially by Eyre (1994). The first operational mission started in 2006 with the launch of the GNSS Receiver for Atmospheric Sounding (GRAS) instrument aboard the MetOp-A satellite. Earlier, other instruments were used for RO campaigns such as the Challenging Mini-satellite Payload (CHAMP) and the Gravity Recovery And Climate Experiment satellite (GRACE-A). The former is the focus of a couple of RO assimilation impact studies made at the ECMWF (Healy & Thépaut, 2006; Healy et al., 2007). The impact of RO data on the assimilation system at the Met Office is discussed in Rennie (2010).

4.2.5 Recent developments

Until recently, only a fraction of the available satellite observations have been assimilated into operational NWP systems. For example, Bauer et al. (2010) estimated that more than 75% of satellite observations were discarded at the ECMWF in 2010. One reason for this large rejection rate was the difficulty in directly assimilating clouds, as NWP models have been struggling in representing accurately physical and radiative processes related to them. Obviously, since clouds and precipitation affect continuously a significant portion of the Earth, the rejection of the observations contaminated by them led to the discard of a great deal of information.

The problems with assimilating cloudy scenes and precipitation are summarised in Errico et al. (2007), where issues related to both the nature of satellite observations and their assimilation in NWP models are discussed. Among the aspects that seemed particularly challenging to the authors, they noted: the importance of selecting the most appropriate observable (for example between the direct assimilation of radiance and that of other – indirect – retrieved parameters); the complexity of the satellite observations' error properties, featuring biases, representativeness issues and non-Gaussianity; the limitations of forward operator models converting the model state into simulated radiances, which are affected by a myriad of issues including the accuracy of parametrisation schemes for clouds and precipitation, the modeling of ice hydrometeors and the poor representation of convection processes; lastly, the high degree of nonlinearity in both observations and processes involved that can jeopardize some of the assumptions commonly made in many DA algorithms (for example Gaussian-shaped error distributions) and can lead to issues with the minimisation of the cost functions in a variational scheme.

Despite the numerous challenges, a great deal of progress has been achieved in assimilating clouds directly in DA algorithms over the past decade. The attempts in this direction are usually referred to as *all-sky* assimilation, as opposed to *clear-sky*, in which all satellite observations contaminated by clouds are discarded. The idea behind all-sky DA is to convert the output of a forecast model able to represent clouds and precipitation into modelled values of radiance via the observation operator (comprising, among others, a RT scheme), thus allowing the assimilation of satellite observations in all sky conditions.

Assimilating clouds directly is expected to be beneficial to the quality of a weather forecast, as their presence is typically concentrated over regions of high dynamical and meteorological interest (e.g. cyclones, fronts), where the assimilation of observations can be the most impactful. Therefore, their assimilation can contribute to reducing the uncertainty on the initial conditions in areas which would be otherwise scarcely observed, leading to a reduction in the forecast error (McNally, 2002).

The early advances in the assimilation of cloudy satellite observations were discussed in Bauer et al. (2011) and an additional update on all-sky DA specifically is given in Geer et al. (2018). An overview of observation errors in all-sky DA is given by Geer & Bauer (2011). As suggested earlier (cf §4.1.3), the properties of radiation and its interaction with clouds and precipitation are not homogeneous across the EM spectrum, and it is therefore convenient to treat the assimilation of all-sky infrared and microwave radiation separately.

All-sky infrared radiation

For a long time, measurements of infrared radiation made by polar-orbiting sounders have been an important source of clear-sky observations. Unfortunately, their current use in all-sky assimilation methods is made difficult by the substantial nonlinear effects that clouds introduce at these wavelengths. In particular, since clouds are strong absorbers (and therefore strong emitters) of infrared radiation, their presence results in an abrupt reduction in brightness temperature, generally associated with the temperature of the cloud top. This issue becomes problematic in a DA system when model and observations disagree on the location (or the presence) of a cloud. The large difference in brightness temperature between the top of a cloud and the surface create large '$o - b$' (observations minus background) differences that can hamper the quality of the analysis, especially when the errors are difficult to characterised.

As a result, there is no operational weather centre currently assimilating all-sky infrared radiances into their NWP system, although research is ongoing and a few methods are under development at Météo-France, NCEP and the ECWMF (cf. Table 3 in Geer et al. (2018)).

Nonetheless, a few weather centres use infrared cloudy observations in a more restricted or indirect way. The ECMWF and Météo-France, for example, have extended their operational scheme to allow for the assimilation of infrared radiances in overcast conditions (Bauer et al., 2011; Lupu & McNally, 2012; Pangaud et al., 2009), whereas at the Met Office certain cloud parameters (such as cloud-top pressure and effective cloud fraction) are estimated from infrared sounder instruments (the Atmospheric Infrared Sounder, or AIRS, the Infrared Atmospheric Sounding Interferometer, or IASI, and the Cross-track Infrared Sounder, or CrIS) using a 1D-Var algorithm and subsequently passed on to the operational 4D-Var scheme (Pavelin et al., 2008).

All-sky microwave radiation

When it comes to the interaction with clouds and precipitation, the behaviour of microwave radiation is less nonlinear than infrared radiation. In fact, microwave radiation is attenuated only partially by the presence of atmospheric components such as water vapour, liquid water, ice and precipitation, and, as a result, tends to be sensitive to the whole atmospheric column rather than just the cloud-top height or the cloud fraction. Not surprisingly, more progress has been made with the operational assimilation of microwave radiation in the presence of clouds.

One of the first attempts to assimilate cloudy microwave observations directly was made by the ECWMF in 2010 (Bauer et al., 2010; Geer et al., 2010), when data from the Special Sensor Microwave/Imager (SSM/I) and the Advanced Microwave Scanning Radiometer - Earth Obseving System (AMSR-E) over the ocean started to be assimilated operationally. Since then, the relevance of this type of observations in the ECMWF assimilation system has been growing and reached a Forecast Sensitivity to Observation Impact (FSOI, cf. Langland & Baker (2004)) of about $\sim 20\%$, also thanks to new instruments added to the observing system (Geer et al., 2017). Since 2016, an

all-sky assimilation programme has started at the NCEP with the assimilation of the Advanced Microwave Sounding Unit (AMSU-A) data, although precipitating clouds are excluded (Zhu et al., 2016).
Geer et al. (2018) reports the details of other microwave all-sky assimilation schemes under development at various operational centres. The Met Office, for example, is already assimilating all non-precipitating scenes from AMSU-A and the Microwave Humidity Sounder (MHS) (Migliorini & Candy, 2019).

4.3 Idealised satellite observations with the ismodRSW model

We now turn to the critical task of applying the theoretical concepts defined in the earlier sections to the modelling of idealised satellite observations. In this thesis, we focus on the mimicking of passive observations of microwave radiation measured by a polar-orbiting satellite. The reasons for choosing this type of observations are motivated by an attempt to maximise the relevance of this work in the context of operational DA research. In particular:

- passive microwave observations are commonly used in most operational weather services. In this regard, a list of operational satellites providing data of the same type are listed in Table 4.1;
- they have a significant impact on operational DA systems. For example, the ECMWF estimated that the overall impact of passive microwave observations on their system was around 40% in 2017 (cf. Fig. 1 in Geer et al. (2017));
- they are commonly used in most all-sky DA methods, which we aim to replicate in this work, too.

An important step in mimicking real satellite observations using an idealised model is establishing the essential characteristics that we want to imitate. Hence, we aim to reproduce the following aspects:

(i) a radiative scheme able to convert the model state into radiance;

(ii) the spatially varying character of the observations carried out by polar-orbiting satellites;

(iii) the horizontal resolution of the observations, which depends on the properties of the instrument (i.e. the FOV).

In practice, the elements (i)-(iii) will have to be included in an *observation generator* function $\mathcal{G}_s$ that can be applied to the nature run simulation $\mathbf{x}^t$ (cf. §3.4.2) to generate a number p_{sat} of satellite observations $\mathbf{y}^o_{sat}$, that is:

$$\mathbf{y}^o_{sat} = \mathcal{G}_s(\mathbf{x}^t) + \boldsymbol{\epsilon}^o_{sat}, \tag{4.30}$$

with $\boldsymbol{\epsilon}^o_{sat}$ being the satellite observation error. A discussion on how the properties (i)-(iii) are mimicked in the context of the ismodRSW model follows next.

4.3.1 Radiative scheme

Radiative schemes are complex algorithms able to convert the state of an NWP model into values of radiance that can be directly compared with satellite observations. In the case of the ismodRSW model, we decided to formulate a simple radiative scheme based on the isentropic definition of fluid temperature valid in each layer (cf. (3.3)), and the Rayleigh-Jeans law (4.13).

Approximating the fluid temperature (3.3) as a brightness temperature, and substituting it into (4.13), one obtains an expression for the spectral radiance emitted by the ith layer (herefrom, we drop the subscript λ):

$$B_i(T) = 2\frac{ck_B}{\lambda^4}T_i = 2\frac{ck_B}{\lambda^4}\theta_i\eta_i^\kappa, \tag{4.31}$$

which can be conveniently non-dimensionalised as:

$$B_i'(T) = \frac{B_i(T)}{B_{0,i}} = \eta_i^\kappa, \qquad \text{with } B_{0,i} = 2\frac{ck_B}{\lambda^4}\theta_i. \tag{4.32}$$

We note here that given their linear relationship in (4.13), the non-dimensional model temperature T_i' coincides with the non-dimensional radiance $B_i'(T)$:

$$T_i' = \frac{T_i}{\theta_i} = \eta_i^\kappa. \tag{4.33}$$

As shown in section §4.1.1, the validity of the Rayleigh-Jeans law depends on the validity of the condition:

$$\lambda \gg \frac{hc}{k_B T}.$$

Instrument	Satellite	Launch	Assimilated by
Advanced Microwave Sounder Unit A (AMSU-A)	NOAA 15–19 Aqua MetOp-A MetOp-B	1998 2002 2006 2012	ECMWF, Japanese Meteorological Agency (JMA), NCEP, Met Office, Météo-France, Deutscher Wetterdienst (DWD)
Microwave Humidity Sounder(MHS)	NOAA 18–19 MetOp-A MetOp-B	2005 2006 2012	ECMWF, JMA, NCEP, Met Office, Météo-France, DWD
Advanced Technology Microwave Sounder (ATMS)	Suomi-NPP NOAA-20	2011 2017	ECMWF, JMA, NCEP, Met Office, Météo-France
Special Sensor Microwave Imager / Sounder (SSMIS)	Def. Met. Sat. Program F-16 F-17 F-18	2005 2008 2009	ECMWF, JMA, NCEP, Met Office, Météo-France
Sounder for Probing Vertical Profiles of Humidity (SAPHIR)	Megha-Tropiques	2011	ECMWF, JMA, Met Office, Météo-France
Advanced Microwave Scanning Radiometer 2 (AMSR2)	Global Change Observation Mission – Water “Shizuku” (GCOM-W)	2012	ECMWF, JMA, Met Office, Météo-France
Global Precipitation Measurement (GPM) Microwave Imager (GMI)	Global Precipitation Measurement (GPM) Core Observatory	2014	ECMWF, JMA, Météo-France
Micro-Wave Humidity Sounder (MWHS) 1,2	Fengyun-3B (FY-3B) FY-3C FY–3D	2010 2013 2017	ECMWF, Met Office

Table 4.1: List of instruments mounted on polar-orbiting satellites able to measure microwave radiation. For each instrument, the satellites that are equipped with it are indicated, together with the year it was sent into orbit. The rightmost column gives a non-exhaustive list of the operational centres that assimilate the observations from a given instrument. See also Geer et al. (2018).

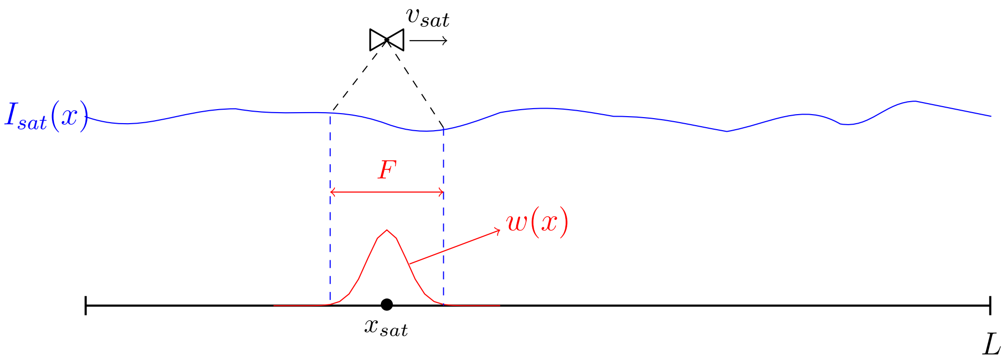


Figure 4.8: Schematic representation of an idealised satellite observation generated via (4.30). The satellite is indicated by a bowtie symbol and its position is determined by (4.34). The way $I(x)$ is calculated is described in section §4.4.4.

Using the bottom layer potential temperature θ_2 used in the nature run ($\theta_2 = 291.8$ K, cf. §3.4.2), one obtains:

$$\lambda \gtrsim 49\ \mu\text{m}.$$

Since we are interested in modelling microwave radiation (for which $\lambda > 1$ mm), the use of the Rayleigh-Jeans is an acceptable approximation.

4.3.2 Spatially varying observations

As discussed in section §4.1.4, polar-orbiting satellites transit periodically over the same geographical location. In our idealised one-dimensional setup, this can be approximated by a satellite travelling across the horizontal periodic domain at constant velocity v_{sat} and regularly re-entering it. The position of the satellite at any given time $x_{\text{sat}}(t)$ can be expressed as:

$$x_{\text{sat}}(t) = (x_0 + v_{\text{sat}}\, t) \bmod L, \tag{4.34}$$

in which x_0 is its location at the initial time $t = 0$, L represents the domain length and *mod* is the modulo function.

4.3.3 Horizontal resolution

The main parameter to evaluate the horizontal resolution of a satellite instrument is its FOV (see section §4.1.4). In order to apply this concept to our idealised one-dimensional system, we compute each satellite observation y^o_{sat} at any given location x_{sat} as the weighted average:

$$y^o_{\text{sat}}(x_{\text{sat}}) = \frac{\displaystyle\int_{x_{\text{sat}}-F/2}^{x_{\text{sat}}+F/2} w(x)I(x)\mathrm{d}x}{\displaystyle\int_{x_{\text{sat}}-F/2}^{x_{\text{sat}}+F/2} w(x)\mathrm{d}x}, \tag{4.35}$$

in which $I(x)$ is the radiance as it will be defined in section §4.4.4 (cf. (4.37)), F is the one-dimensional FOV determining the portion of the domain over which the weighted average is computed, and $w(x)$ is a Gaussian function defined as:

$$w(x) = e^{-\frac{1}{2}\frac{(x-x_{\text{sat}})^2}{(F/6)^2}}. \tag{4.36}$$

The role of $w(x)$ is one of imitating the angular response function of a real satellite instrument. Hence, the radiation measured within the FOV F is weighted by $w(x)$. As a result, the signal comes primarily from x_{sat} (i.e. $w(x_{\text{sat}}) = 1$) and rapidly decays away from it (i.e. $w(x_{\text{sat}} + F/2) \approx 0.01$). Fig. 4.8 shows a schematic representation of an idealised satellite observation defined according to (4.34) and (4.35).

4.4 Idealised observations of clouds and precipitation

In view of the recent developments in satellite DA (cf. § 4.2.5), we want our idealised system to account for the assimilation of clouds in an all-sky approach.

The ismodRSW model does not explicitly include clouds, but rather than modifying further its equations and dynamics, we can include their effects indirectly via the observation generator (4.30). Therefore, we construct the function $\mathcal{G}_s$ in a way that mirrors the interaction between microwave radiation and clouds (both precipitating and non-precipitating ones).

A review of the knowledge on this topic and a description of both the physical and radiative processes involved in it are discussed in the subsequent sections, followed by their implementation in the observation generator of the ismodRSW model.

4.4.1 Retrival of clouds and precipitation using microwave radiation: a review

Other than for DA purposes, the retrieval of various atmospheric quantities via satellite observations has helped understand many aspects of cloud physics and has contributed to better NWP model parametrisation schemes. In addition, many studies have focussed on reconstructing the spatio-temporal distribution of certain meteorological variables for climatological purposes - such as global precipitation - which had limited coverage before the advent of satellites.

The use of microwave radiation to investigate cloud properties, precipitation and other atmospheric parameters from satellites dates back to the late 1960s, stemming from studies based on the Russian satellites *Cosmos* (Akvilonova et al., 1973; Basharinov et al., 1969). Around the same time, many other satellites equipped with instruments able to measure passive microwave radiation started to be launched into orbit, leading to further research in the field. Staelin et al. (1975) was one of the first studies focussing on detecting clouds using microwave radiation through estimates of vapour and liquid water content in the atmosphere, although the paper was only interested in the impact of clouds on the retrieval of vertical temperature profiles using the Nimbus-E Microwave Spectrometer (NEMS); shortly after, and relying on the same instrument, both Grody (1976) and Staelin et al. (1976) developed algorithms to convert brightness temperature measurements at $\nu = 22.235$ GHz ($\lambda = 11$ mm) and $\nu = 31.4$ GHz ($\lambda = 9.5$ mm) into estimates of integrated water vapour, in an attempt to use microwave radiation to infer atmospheric parameters other than just vertical profiles of temperature. Similar algorithms were developed afterwards, for a variety of instruments and applications: Rosenkranz et al. (1978) estimated various atmospheric parameters such as water vapour, liquid water contents and surface wind, to characterise the typhoon June, using the SCAMS mounted on Nimbus 6; Grody et al. (1980) adapted the algorithm from the 1976 paper to retrieve water vapour and liquid water content in the Tropical Pacific environment using the SCAMS; Chang & Wilheit (1979) developed a three-channel retrieval method for water vapour and liquid water content based on the Electrically Scanning Microwave Radiometer (ESMR) onboard Nimbus 5; Wilheit & Chang (1980) and Wentz (1983) developed an algorithm to retrieve a number of ocean surface and atmospheric parameters using the Scanning Multichannel Microwave Radiometer (SMMR) orbiting aboard Nimbus 7; based on the same instrument, Prabhakara et al. (1982) proposed a retrieval algorithm for precipitable water over the ocean and Takeda & Liu (1987) presented their method to retrieve vertically-integrated liquid water amount in the atmosphere whilst identifying precipitating clouds; a series of papers published in the 1990s (Greenwald et al., 1993; Jones & Vonder Haar, 1990; Tjemkes et al., 1991; Wentz, 1997) developed similar algorithms for the SSM/I. Back in the 1970s, a number of studies were also trying to infer information about the intensity of precipitation using satellites. After a preliminary investigation by Savage & Weinman (1975), Wilheit et al. (1977) proposed a theoretical model to relate brightness temperature measurements over the oceans to rain rates based on the 1.55 cm channel of the ESMR. Later on, other studies tried to achieve the same (Huang & Liou, 1983; Liou et al., 1980; Spencer, 1986; Spencer et al., 1989), and at the same time the impact of frozen hydrometeors in precipitating clouds started to be acknowledged and incorporated in both theoretical studies and retrieval algorithms. One of the first papers to highlight the issue of frozen hydrometeors was Spencer et al. (1983), in which very low values of brightness temperature measured by the SMMR over Kansas during a summer thunderstorm were reported and attributed to the presence of ice. Soon after, a paper by Wu & Weinman (1984) developed a comprehensive radiative model for precipitating clouds that included liquid, frozen and combined-phase hydrometeors; Mugnai & Smith (1988) and Smith & Mugnai (1988) extended the investigation of Wilheit et al. (1977) by developing a model in which cloud microphysics was dealt with in detail.

Finally, several other papers (Curry et al., 1990; Greenwald et al., 1995; Njoku & Swanson, 1983; O'Dell et al., 2008; Prabhakara et al., 1983; Spencer, 1984) used similar techniques and algorithms for climatological purposes, trying to reconstruct timeseries of precipitation, clouds and other meteorological variables over different areas, surfaces and time periods.

4.4.2 Microwave radiation and non-precipitating clouds

In order to describe the interaction between microwave radiation and non-precipitating clouds, two aspects need to be considered: the difference in size between the wavelength of microwave radiation and the average dimension of the cloud droplets ($r \leq 0.1\,\mathrm{mm}$), and the fact that liquid water is a strong absorber of radiation in the microwave spectrum.

In general, for microwave radiation of wavelength larger than $\lambda > 5\,\mathrm{mm}$, scattering

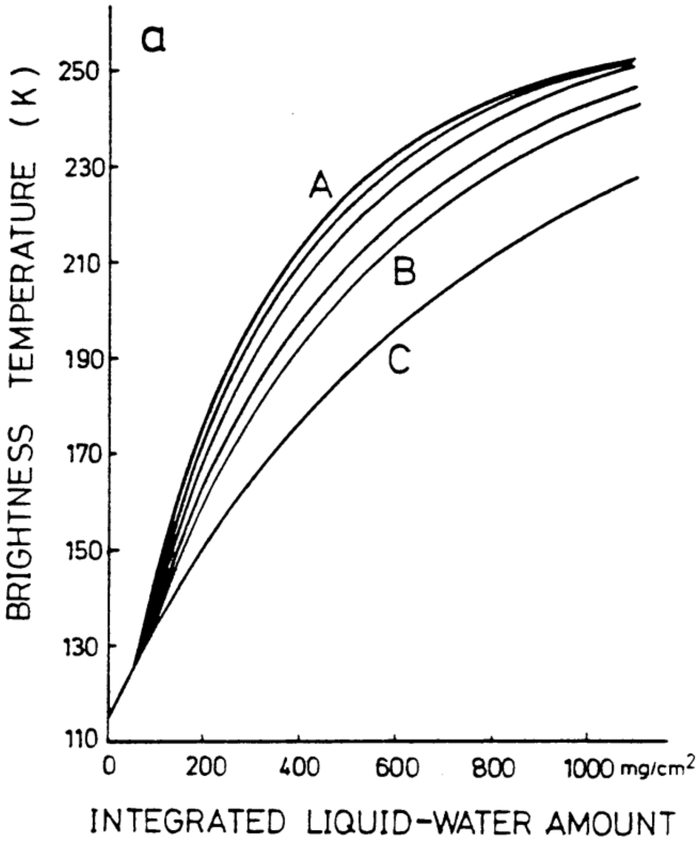


Figure 4.9: Relationship between brightness temperature and integrated liquid water concentration at 18GHz (i.e. $\lambda \approx$ 16mm) in the presence of a clouds above a sea surface at $T = 293$ K, for various size droplet distributions, as modelled by Takeda & Liu (1987). This is Figure 1 in Takao Takeda, Guosheng Liu, *Estimation of Atmospheric Liquid-Water Amount by Nimbus 7 SMMR Data*, Journal of the Meteorological Society of Japan. Ser. II, 1987, Volume 65, Issue 6, Pages 931-947, Released October 19, 2007, https://doi.org/10.2151/jmsj1965.65.6_931.

processes have a negligible impact and non-precipitating clouds figure as an efficient absorber of microwave radiation. In other words, the thermal radiation emitted by the surface gets absorbed by the small liquid water droplets in the cloud, which subsequently re-emit radiation towards the satellite.

The detection of clouds based on these principles is particularly effective over the oceans. The reason behind this lies in the low emissivity of water surfaces (Petty & Katsaros, 1994), which display very small values of brightness temperature in the microwave spectrum. As a result, even the thermal radiation emitted by small water droplets at high altitudes appear warmer against the ocean, making the cloudy pixels easy to identify.

We highlight the effect of these processes in Figure 4.9, where we report the relationship between brightness temperature and integrated liquid water contents resulting from a model built by Takeda & Liu (1987), in which a cloud layer at $T = 273\,\mathrm{K}$ lies above a (sea) surface at $T = 293\,\mathrm{K}$. Clearly, an increase in the amount of liquid water content (a proxy for the thickness of a cloud) results in an increase in brightness temperature against a cold background.

4.4.3 Microwave radiation and precipitating clouds

Precipitating clouds differ from non-precipitating ones in two respects: they are made of substantially larger water droplets and contain a greater quantity of ice and frozen hydrometeors. While for microwave radiation of $\lambda > 5\,\mathrm{mm}$ (considered in the previous section regarding non-precipitating cloud) absorption continues to represent the dominant interaction process, at smaller wavelengths (such as $\lambda < 2\,\mathrm{mm}$), the presence of ice and larger droplets contribute to increase the amount of microwave radiation that gets scattered by a precipitating cloud.

Scattering processes contribute to deflect the radiation away from its path towards the satellite, attenuating the signal and reducing the brightness temperature. Therefore, the presence of precipitating clouds results in a decrease in brightness temperature when microwave radiation at small wavelength is considered. This process has been used to estimate precipitation (or rain rates) above both water and land surfaces.

The effect of precipitation on microwave radiation is displayed in Figure 4.10, where we show the relationship between brightness temperature and rain rate above ocean (continuous line) land (dashed line) as generated by Spencer et al. (1989) (Fig. 3 therein), using the RT model of Wu & Weinman (1984). The reduction in brightness

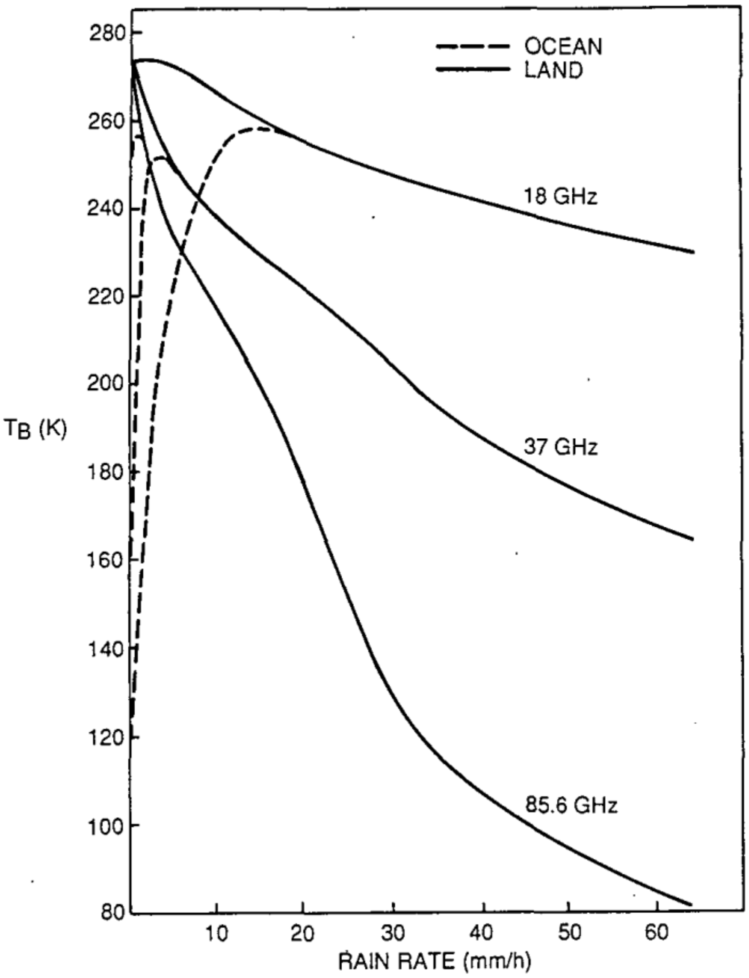


Figure 4.10: Relationship between brightness temperature and rain rate above land (continuous line) and ocean (dashed line), generated by Spencer et al. (1989) using the RT model of Wu & Weinman (1984). This figure appears in Spencer, R. W., Goodman, H. M., & Hood, R. E. (1989). *Precipitation retrieval over land and ocean with the SSM/I: Identification and characteristics of the scattering signal.* Journal of Atmospheric and Oceanic Technology, 6(2), 254-273. © American Meteorological Society. Used with permission.

Clear-sky	$\sigma < \sigma_c$
Non-precipitating clouds	$\sigma_c < \sigma < \sigma_r$
Precipitating clouds	$\sigma > \sigma_r$

Table 4.2: Range of values of the (non-dimensional) pseudo-density σ associated with different types of satellite observations. σ_c and σ_r indicate the values of the convection and rain thresholds as illustrated in section §3.4.1.

temperature is clear at all frequencies, but is stronger at higher frequencies (that is, lower wavelengths[1]), as the scattering processes become more and more important.

4.4.4 Model implementation

In order to incorporate the processes described above into our idealised system, we start by defining what constitutes a cloud in the ismodRSW model. Since the model already includes a 'rain' variable (associated with the threshold σ_r), it is reasonable to assume the presence of precipitating clouds whenever the condition $\sigma > \sigma_r$ is met (for simplicity, we will ignore here the condition on flow convergence). In addition, we make the assumption that a non-precipitating cloud is present whenever the condition on the generation of convection is satisfied, i.e. $\sigma_c < \sigma < \sigma_r$.

The latter condition is, of course, a simplification, as the presence of convection does not always translate into the formation of clouds. However, as we discussed in section §3.4.1, the convection threshold σ_c can be related to the Level of Free Convection (LFC), that is, the altitude at which a saturated air parcel in a conditionally unstable atmosphere becomes warmer than the surrounding environment and starts to rise without the need for an external force. A parcel reaching the LFC has already saturated and it is therefore reasonable to expect that the condensation of water vapour into cloud droplets has started by then.

Overall, we have defined three regimes for the generation of satellite observations associated with values of the pseudo-density σ, as summarised in Table 4.2.

It is worth noting that the dynamics of the model as described in Chapter 3 is not affected by the presence of precipitating and non-precipitating clouds. Rather, only

---

[1] Note that $\nu = 37$ GHz equals to $\lambda \approx 8$ mm and that $\nu = 85.6$ GHz equals to $\lambda \approx 3.5$ mm.

the observation generator $\mathcal{G}_s$ in (4.30) is impacted by it.

In practice, the satellite observations generated by $\mathcal{G}_s$ are equivalent to vertically integrated radiance measurements, with contributions to the 'measured' radiance I_{sat} coming from both layers and defined as:

$$I_{\mathrm{sat}} = \alpha_1\, \alpha_3\, B_1 + (\alpha_2 + \alpha_4)\, B_2, \tag{4.37}$$

in which α_1, α_2, α_3 and α_4 are non-dimensional σ-dependent functions determining the contribution of various radiative processes in each layer (see also Table 4.3), whereas B_1 and B_2 are the (σ-dependent) values of radiance emitted by each layer according to (4.31).

We note here that while I_{sat} is a function of both layers' brightness temperature B_1 and B_2, one of the limitations of this idealised configuration is the lack of a surface emission term in (4.37), which cannot be included in the absence of a 'bottom surface' temperature in the ismodRSW model.

Each coefficient α_k is defined as a function of the pseudo-density in the bottom layer σ:

$$\alpha_k = a_k + b_k \cdot \mathrm{erf}(-c_k \cdot \sigma + d_k), \quad k = 1, ..., 4, \tag{4.38}$$

in which erf indicates the so-called 'error function':

$$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-x^2} \mathrm{d}x. \tag{4.39}$$

Table 4.3 lists the values of a_k, b_k, c_k and d_k associated with each α_k. These coefficients have been obtained by manually calibrating the functions α_k so that the measured radiance I_{sat} in (4.37) could imitate the interaction of microwave radiation in the presence of precipitating clouds at small wavelengths ($\lambda < 2\,\mathrm{mm}$) and non-precipitating clouds at larger wavelength ($\lambda > 5\,\mathrm{mm}$), as described in sections §4.4.2-4.4.3. Plots of the coefficients $\alpha_k(\sigma)$ in each satellite observation regime (cf. Table 4.2) are reported in Fig. 4.11. A detailed description of the processes involved in each regime, together with their impact on I_{sat}, follows.

Clear-sky

Under clear-sky conditions ($\sigma < \sigma_c$, cf. left panel of Fig. 4.11), the non-negligible terms contributing to I_{sat} in (4.37) are just α_2, α_3 and α_4. However, since α_1 is very small,

	a	b	c	d	Process
α_1	0.5	-0.5	95	21.5	Emission of the top layer
α_2	0.425	0.425	95	21.5	Emission of the bottom layer
α_3	0.5	0.5	5	3	Scattering in the top layer
α_4	0.5	0.5	3	-1.16	Extinction in the bottom layer

Table 4.3: Values of a, b, c and d used in the calculation of α_k according to (4.38). These coefficients have been manually calibrated to obtain a profile of I_{sat} (cf. (4.37), Fig. 4.12) that reflects the response of passive microwave radiation in the presence of clouds and precipitation, as outlined in sections §4.4.2-4.4.3.

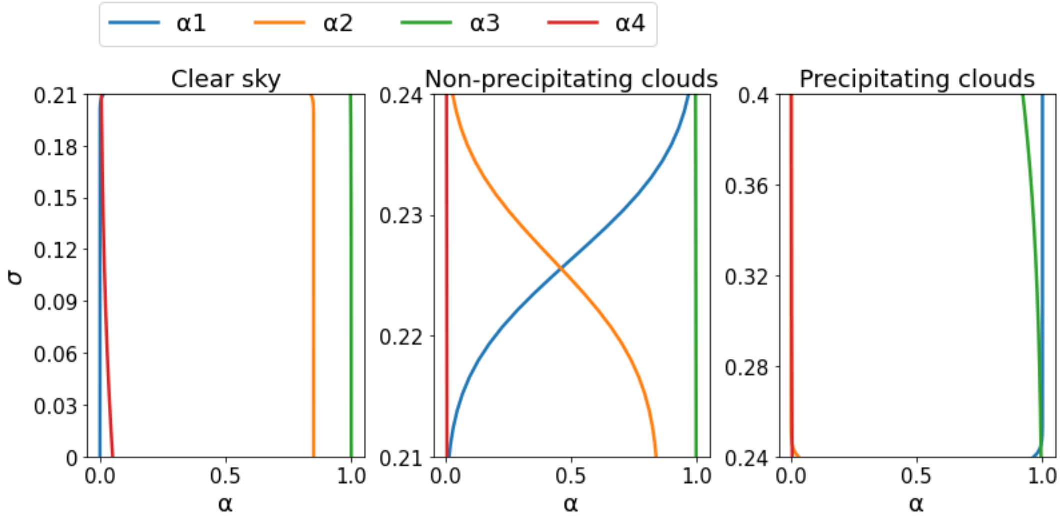


Figure 4.11: Values of α_k as a function of (non-dimensional) σ according to (4.38) for each regime defined in Table 4.2. Left panel: clear-sky conditions ($\sigma < \sigma_c$); central panel: non-precipitating clouds ($\sigma_c < \sigma < \sigma_r$); right panel: precipitating clouds ($\sigma > \sigma_r$). Note the different y-axis scale in each panel. The values of σ_c and σ_r are those already used in the nature run described in section §3.4.2, i.e. $\sigma_c = 0.21$, $\sigma_r = 0.24$.

the term associated with B_1 in (4.37) vanishes. Therefore, only radiation coming from the bottom layer (B_2) will contribute to I_{sat}, that is:

$$I_{\text{sat}}^{\text{clear-sky}} \approx (\alpha_2 + \alpha_4) \cdot B_2. \tag{4.40}$$

In the expression above, the coefficients α_2 and α_4 control the thermal emission of the lower layer and the extinction of radiation in clear-sky conditions. Hence, the radiation reaching the satellite would be maximum for $\sigma = 0$ (where $\alpha_2 + \alpha_4 = 0.9$) and decreases slightly as the bottom layer becomes thicker (i.e. $\alpha_2 + \alpha_4 \approx 0.84$ at $\sigma = \sigma_c$). This small decrease in the radiation reaching the satellite is only qualitatively in line with Beer's law (4.19), which predicts the radiation to decay exponentially as the optical depth increases.

By fixing the maximum value of $\alpha_2 + \alpha_4$ to 0.9, we can imitate the effect of the reduced surface emissivity exploited in many cloud detection algorithms at large microwaves (i.e. $\lambda > 5\,\text{mm}$, cf. §4.4.2). As we will see in Fig. 4.12, this reduction will be enough to obtain an increase in the radiance (or brightness temperature) reaching the satellite in the presence of non-precipitating clouds, despite a lower fluid temperature in the upper layer.

Non-precipitating clouds

For values of σ compatible with the presence of non-precipitating clouds ($\sigma_c < \sigma < \sigma_r$, cf. central panel of Fig. 4.11), the expression of I_{sat} can be approximated with:

$$I_{\text{sat}}^{\text{non-prec}} \approx \alpha_1\, \alpha_3\, B_1 + \alpha_2\, B_2. \tag{4.41}$$

In this regime, the radiation reaching the satellite consists of contributions from both layers. In practice, as σ reaches the convection threshold σ_c and a cloud starts to form, the coefficient α_2 associated with the emission of the bottom layer starts to decay, while α_1 (associated with the emission from the cloud) increases. As this process continues, the thickness of the cloud increases until $\sigma = \sigma_r$; at this stage the contribution of B_2 has almost vanished (as $\alpha_2 \to 0$), and is replaced almost entirely by the thermal emission of the upper layer B_1. In this regime the value of α_3 is constant and approximately equal to 1, as this coefficient is used to model the impact of scattering, which we consider negligible in these circumstances.

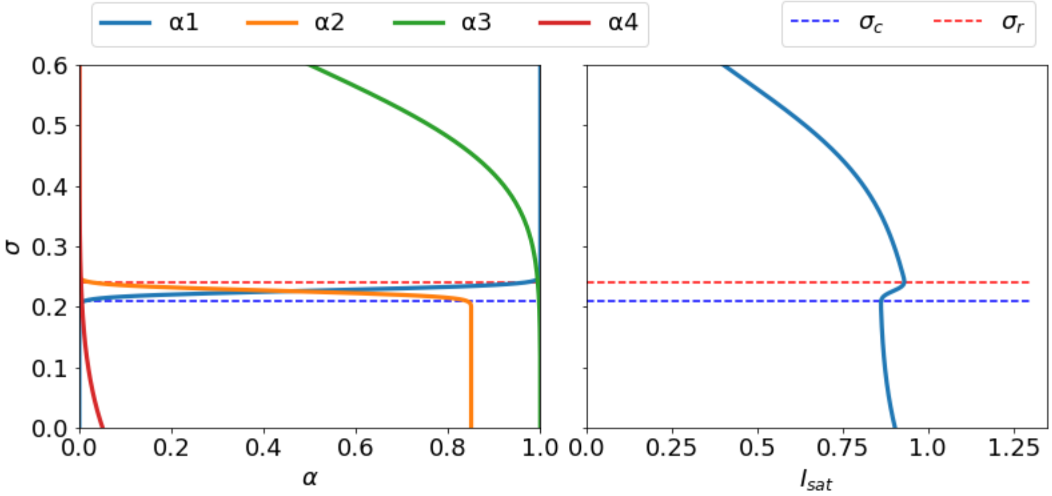


Figure 4.12: Values of α_k as a function of (non-dimensional) σ (left panel) and resulting values of (non-dimensional) radiance I_{sat} (right panel). The values of σ_c and σ_r are those already used in the nature run described in section §3.4.2, i.e. $\sigma_c = 0.21$, $\sigma_r = 0.24$. The left panel combines the three plots shown in Fig. 4.11.

Precipitating clouds

In the precipitating clouds regime ($\sigma > \sigma_r$, cf. right panel of Fig. 4.11), the sum of the coefficients $\alpha_2 + \alpha_4$ become rapidly negligible, making it possible to approximate the total radiance with the contribution coming only from the top layer, that is:

$$I_{\text{sat}}^{\text{prec}} \approx \alpha_1\, \alpha_3\, B_1. \tag{4.42}$$

In this region, the thermal emission of the cloud has reached is maximum (i.e. $\alpha_1 \to 1$). However, as the pseudo-density σ increases beyond σ_r, more precipitation is produced in the ismodRSW model, leading to an intensification of the scattering processes at small microwaves ($\lambda < 2\,\text{mm}$). This phenomenon is modelled via the coefficient α_3, which decreases gradually with σ (see Fig. 4.12).

Finally, the plots of all coefficient α_k and the radiation reaching the satellite I_{sat} as a function of the pseudo-density σ are shown in Fig. 4.12 (left panel). Crucially, the value of I_{sat} matches the response of microwave radiation in interacting with clouds and precipitation as outlined earlier, and in particular:

- it increases in the presence of non-precipitating clouds ($\sigma_c < \sigma < \sigma_r$) at large microwaves ($\lambda > 5\,\mathrm{mm}$), as the emission from liquid water droplets stands out against a cold background;
- it decreases in the presence of precipitating clouds ($\sigma > \sigma_r$) at small microwaves (($\lambda < 2\,\mathrm{mm}$)), as scattering processes become important in attenuating the radiation emitted by the cloud.

In addition, in agreement with the Beer's law (4.19), I_{sat} decreases slightly in clear-sky conditions ($\sigma < \sigma_c$), as the value of the pseudo-density σ (and therefore the depth of the bottom layer h_2) increases.

Chapter highlights and summary

- At the beginning of the chapter we have given an overview of basic radiative transfer principles behind the use of meteorological satellites, followed by a brief historical review of satellite data assimilation.
- By exploiting the principles outlined in the first part of the chapter, we have subsequently described the generation of pseudo satellite observations using the ismodRSW model. We chose to recreate measurements of microwave radiation carried out by polar orbiting satellites, which constitute a common... and formulated a simple radiative transfer model based on the Rayleigh-Jeans equation.
- Because of the current interest of the data assimilation community in all-sky assimilation, we have tried to incorporate the effects of clouds in our idealised system. To this aim, by exploiting the convection and the rain thresholds already present in the model, we could define an observable that imitates the physical response of microwave radiation to the presence of clouds and precipitation without having to modify the underlying model equations.

Chapter 5

Forecast-assimilation experiments with the ismodRSW model

In this chapter we summarise the modifications made to the DA system described in Chapter 2, in order to carry out a new series of forecast-assimilation experiments with the ismodRSW model, including the assimilation of idealised satellite observations. In addition, we will describe the process undertaken to obtain a well-tuned simulation and use such a simulation as a control in a series of data denial experiments (in which different groups of observations are excluded in turn from the assimilation) to investigate the impact of satellite observations at different spatial scales.

5.1 Modifications to the data assimilation scheme

A new series of experiments have been carried out, based on the twin-setting configuration described in section §2.2.1, and the nature run simulation introduced in section §3.4.2 (the step-by-step algorithm described in section §2.3 will be revised in section §5.1.5). However, in order to include the assimilation of the pseudo satellite observations described in Chapter 4, some aspects related to the DA scheme and the observing system have to be revisited.

In this section we describe some of these modifications, such as the properties of the new observing system (i.e. the number and type of observations) and the introduction

Observing system	
Total number of observations, p	38
Obs. per variable (p_{sat}, p_u, p_v, p_r)	(8,10,10,10)
Update frequency [min]	60
Ground obs. density (d_u, d_v, d_r) [km]	(50, 50, 50)
Satellite obs. FOV [km]	(20, 20, 40, 40, 60, 60, 80, 80)
Satellite obs. initial position [L]	(0.1, 0.64, 0.17, 0.75, 0.35, 0.88, 0.41, 0.9)
Satellite velocity v_{sat} [L/hour]	(0.2, -0.12, -0.25, 0.18, 0.15, -0.15, 0.1, -0.1)
Obs. error $(I_{sat}, \sigma_h, \sigma_u, \sigma_r)$	(0.01, 0.04, 0.04, 0.01)

Table 5.1: A summary of the observing system used in the idealised experiments. Units are dimensionless unless specified otherwise.

of a nonlinear observation operator. Moreover, some technical features are discussed, such as the implementation of a *modulated* ensemble technique to preserve the direct localisation of the forecast covariance matrix $\mathbf{P}_e^f$ in model space, and a revision of the model-error covariance matrix $\mathbf{Q}$.

5.1.1 The new observing system

A new and more complex observing system is created, comprising p_{g} ground observations $\mathbf{y}_{\text{g}}^o$ of fluid velocity (u and v) and rain (r), and p_{sat} pseudo satellite observations $\mathbf{y}_{\text{sat}}^o$. This distinction between ground (sometimes also called *conventional*) and satellite observations is an important one, as it is common in real-world observing systems and contributes to making our idealised configuration more relevant to that of a typical NWP DA system.

As a result, the observation generator function $\mathcal{G}$ can be split into two components. On the one hand, similarly to what was described in section §2.2.1, the ground observations of u, v and r are generated from the nature run $\mathbf{x}^t$ by sub-sampling their values at fixed locations along the domain at every analysis step. In other words, a

$(p_g \times 4N_{el}^{nat})$[1] linear operator $\mathbf{G}$ (i.e. a sparse matrix) is defined, such that:

$$\mathbf{y}^o_{\mathrm{g}} = \mathbf{G}\mathbf{x}^t + \boldsymbol{\epsilon}^o_{\mathrm{g}}. \tag{5.1}$$

On the other hand, a non-linear operator is used to generate pseudo satellite observations (as extensively described in Chapter 4), that is:

$$\mathbf{y}^o_{\mathrm{sat}} = \mathcal{G}_{sat}(\mathbf{x}^t) + \boldsymbol{\epsilon}^o_{\mathrm{sat}}. \tag{5.2}$$

Together, the vectors $\mathbf{y}^o_{\mathrm{sat}}$ and $\mathbf{y}^o_{\mathrm{g}}$ form the p-dimensional observation vector $\mathbf{y}^o$.
In order for our experiments to be relevant for satellite DA research, the relative numbers of ground observations p_g and satellite observations p_{sat} should reflect the characteristics of a real observing system. In the case of satellite observations, aspects such as their frequency and availability in time and space vary widely depending on the type of instrument and the orbit of the satellite it travels on.

Ground observations

Analogously to what was done in Chapter 2, we use the surface density of ground observations over the UK (around 40 km) as an upper bound for p_g. Hence, we generate 10 evenly spaced observations (one every 50 km) for each ground variable (u, v and r), to give a total of $p_g = 30$ ground observations at each assimilation time.
In this regard, it is worth noting that the meridional velocity v is included among the observations despite being actively forced towards the relaxation solution v_{rel} (cf. section §3.4.2) in both the nature run and the forecast. A further exploration of whether it is possible to achieve a well-tuned experiment in the absence of v observations (exploiting the constraint given by v_{rel}) is left for future examination.

Satellite observations

In this study we focus on a very specific type of observation, namely measurements of microwave radiation carried out by polar-orbiting satellites. Therefore, we are interested in quantifying how many observations of this type are available hourly over the domain of a high-resolution (and convection resolving) limited area model, as this

---

[1] Note that although the number of grid point in the nature run N_{el}^{nat} has not changed, the model state here includes the meridional velocity v, other than σ, u and r.

represents the closest type of configuration to the ismodRSW model with a domain of $L = 500\,\text{km}$. As an example, we consider the UKV model run at the Met Office, which covers the British Isles and the surrounding area and has a grid spacing of 1.5 km in its interior and 4 km at its extremities (Tang et al., 2013). Milan et al. (2020) describe in detail the operational 4DVar scheme used for this model, and provide information on the observing system. In particular, Figures 11 and 13 of their paper report the distributions of the number of microwave and conventional observations assimilated every hour. The mean values are also shown, with an average of 1900 microwave observations from sounder instruments and 6830 conventional observations assimilated every hour. These numbers give a ratio of satellite to ground observations of approximately 0.27 which, given $p_g = 30$ ground (conventional) observations, translates into a value of $p_{sat} = 8$ in our idealised system. In order to investigate the impact of observations at various spatial scales, we create four sets of observations with different fields of view, that is $\{20, 40, 60, 80\}$ km. The initial position and velocity v_{sat} of each satellite (cf. (4.34)), which determines the location of the observation at the analysis step, are reported in Table 5.1.

5.1.2 Nonlinear observation operator

The modifications to the observing system through the generation of pseudo satellite observations requires a revision of the observation operator, that is, the function that maps a model state into the observational space (cf. section §2.2.2, eq. (2.18)). In particular, the relationship between the pseudo-density σ and the radiance $B_{i,\lambda}(T)$ in the Rayleigh-Jeans law (4.31) makes the observation operator nonlinear.

The development of nonlinear observation operators represents a common challenge in most operational systems, especially in the context of satellite DA, which comprises the use of complex radiative transfer schemes able to convert typical model variables such as temperature and humidity into simulated radiances. In this sense, this modification contributes further to making our idealised configuration more similar and relevant for NWP DA research.

In a similar manner to the observation generator, the observation operator can also be split into a ground and a satellite component. For ground variables (i.e. u, v and r), it consists of a linear operator $\mathbf{H}_{\text{g}}$ (a sparse matrix of 0 and 1) sub-sampling the model state $\mathbf{x}^f$ at the positions of the corresponding observations. Conversely, the observation operator for satellite observations is a nonlinear function $\mathcal{H}_{\text{sat}}$ acting on the model state $\mathbf{x}^f$ such that the predicted radiance I_{sat} is computed at the satellite observations positions x_{sat} (cf. (4.34)). It is worth noting that the observation operator $\mathcal{H}_{\text{sat}}$ differs from the observation generator $\mathcal{G}_s$ in that it does not apply the weighted horizontal average to the values of the forecast $\mathbf{x}^f$ near the satellite position $\mathbf{x}_{\text{sat}}$.

5.1.3 Modulated ensemble for model-space localisation

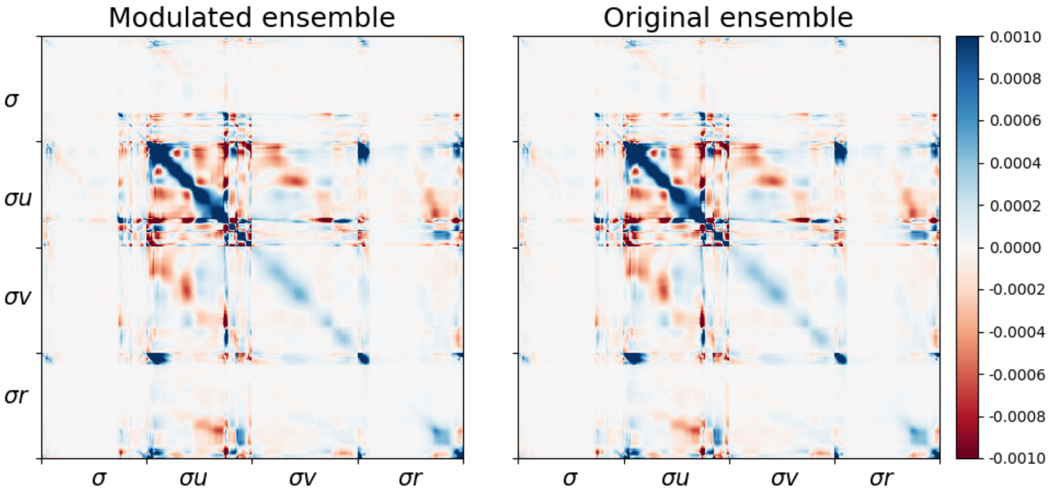


Figure 5.1: Localised covariance matrix $\mathbf{P}^f_{\text{loc}}$ computed with the modulated ensemble (cf. (5.9), left) and with the original one (cf. (2.30), right). Both matrices are computed with the model ensemble forecast valid at $T = 49$ hours with a lead time of one hour. Visually, the differences between the two matrices are very small.

In a deterministic ensemble Kalman filter, the observation operator appears twice: first, in the innovation vector $\mathbf{y}^o - \mathcal{H}(\mathbf{x}^f)$ within the analysis equation (2.18) and furthermore in the expression of the Kalman gain $\mathbf{K}$, cf. (2.20). Although Kalman's original approach and its deterministic nonlinear extensions normally require $\mathcal{H}$ to be a linear operator (or the computation of its Jacobian if it is nonlinear), alternative ensemble formulations of the Kalman filter in the presence of nonlinear observation operators have also been developed that take a different approach. One of the most popular approach is adopted in a series of papers from Houtekamer & Mitchell (1998, 2001, 2005), in which the expressions involving the nonlinear observation operator within the Kalman gain matrix $\mathbf{K}$, i.e. $\mathbf{P}^f\mathcal{H}^T$ and $\mathcal{H}\mathbf{P}^f\mathcal{H}^T$, are calculated exploiting the definition of $\mathbf{P}^f_e$ used in an ensemble Kalman filter (cf. (2.21)), that is:

$$\mathbf{P}^f_e\mathcal{H}^T = \frac{1}{N-1}\sum_{i=1}^{N}\left(\mathbf{x}^f_i - \overline{\mathbf{x}}^f\right)\left(\mathcal{H}(\mathbf{x}^f_i) - \overline{\mathcal{H}(\mathbf{x}^f_i)}\right)^T, \tag{5.3.a}$$

$$\mathcal{H}\mathbf{P}^f_e\mathcal{H}^T = \frac{1}{N-1}\sum_{i=1}^{N}\left(\mathcal{H}(\mathbf{x}^f_i) - \overline{\mathcal{H}(\mathbf{x}^f_i)}\right)\left(\mathcal{H}(\mathbf{x}^f_i) - \overline{\mathcal{H}(\mathbf{x}^f_i)}\right)^T. \tag{5.3.b}$$

A possible difficulty in using the expressions above to compute the gain matrix $\mathbf{K}_{e,\widehat{j}}$ in our DA scheme is due to the fact that we apply the spatial localisation directly to the (self-excluded) ensemble covariance matrix $\mathbf{P}^f_{e,\widehat{j}}$, as discussed in section §2.2.3. In practice, we use a redefinition of (5.3) in terms of $\mathbf{P}^f_{\text{loc},\widehat{j}}\mathcal{H}^T$ and $\mathcal{H}\mathbf{P}^f_{\text{loc},\widehat{j}}\mathcal{H}^T$, adopting the modulated ensemble technique outlined in Bishop et al. (2017). The objective is to find the square root matrix $\mathbf{Z}$ of $\mathbf{P}^f_{\text{loc},\widehat{j}}$, that is:

$$\mathbf{P}^f_{\text{loc},\widehat{j}} = \boldsymbol{\rho} \circ \mathbf{P}^f_{e,\widehat{j}} = \mathbf{Z}\mathbf{Z}^T. \tag{5.4}$$

First, we consider the square root of the localisation matrix $\boldsymbol{\rho}$:

$$\boldsymbol{\rho} = \mathbf{W}\mathbf{W}^T, \tag{5.5}$$

in which $\mathbf{W}$ is a $(4N_{el} \times K)$ matrix, with $K \leq 4N_{el}$ being its rank. Here we compute $\mathbf{W}$ exploiting a low-rank approximation of the spectral theorem discussed in Appendix D. Therefore, we form a new matrix by multiplying each column of $\mathbf{W}$ with the $N-1$ columns of the (self-excluded) perturbation matrix $\mathbf{X}^f_{\widehat{j}}$ to obtain $\mathbf{Z}_{\widehat{j}}$:

$$\begin{aligned}\mathbf{Z}_{\widehat{j}} = \Big[&(\mathbf{w}_1 \circ \mathbf{X}^f_1, \mathbf{w}_1 \circ \mathbf{X}^f_2, ..., \mathbf{w}_1 \circ \mathbf{X}^f_{\widehat{j}-1}, \mathbf{w}_1 \circ \mathbf{X}^f_{\widehat{j}+1} ..., \mathbf{w}_1 \circ \mathbf{X}^f_N), ...,\\ &(\mathbf{w}_K \circ \mathbf{X}^f_1, \mathbf{w}_K \circ \mathbf{X}^f_2, ..., \mathbf{w}_K \circ \mathbf{X}^f_{\widehat{j}-1}, \mathbf{w}_K \circ \mathbf{X}^f_{\widehat{j}+1}, ..., \mathbf{w}_K \circ \mathbf{X}^f_N)\Big].\end{aligned} \tag{5.6}$$

The matrix above has dimension $3N_{el} \times M$, in which $M = (N-1)K$ (with $N = 20$ being the size of the original ensemble, cf. Table 5.3, and $K = 8$ being the rank of $\mathbf{W}$). Therefore, a new expanded (or modulated) ensemble of size $M = 152$ can be defined exploiting the column vectors of $\mathbf{Z}_{\widehat{j}}$ (with components $\mathbf{z}_{k,\widehat{j}}$) namely:

$$\mathbf{v}^f_{k,\widehat{j}} = \left(\frac{1}{N-1}\sum_{\substack{i=1\\ i\neq\widehat{j}}}^{N}\mathbf{x}^f_i\right) + (\sqrt{M})\mathbf{z}_{k,\widehat{j}}. \tag{5.7}$$

Equation (5) in Bishop et al. (2017) shows that the column vectors $\mathbf{z}_{k,\widehat{j}}$ have zero mean, provided that this applies also to the ensemble perturbations $\mathbf{X}^f_{\widehat{j}}$. As a result, the modulated ensemble $\mathbf{v}^f_{k,\widehat{j}}$ has the same mean value as the original forecast ensemble without the j^{th} member, that is:

$$\overline{\mathbf{v}}^f_{\widehat{j}} = \frac{1}{M}\sum_{i=1}^{M}\mathbf{v}^f_{i,\widehat{j}} = \frac{1}{N-1}\sum_{\substack{i=1\\ i\neq\widehat{j}}}^{N}\mathbf{x}^f_i = \overline{\mathbf{x}}^f_{\widehat{j}}. \tag{5.8}$$

Finally, the covariance matrix obtained from the ensemble $\mathbf{v}^f_{k,\widehat{j}}$ is equal to $\mathbf{P}^f_{\text{loc},\widehat{j}}$:

$$\mathbf{P}^f_{\text{loc},\widehat{j}} = \frac{\sum_{i=1}^{M}(\mathbf{v}^f_{i,\widehat{j}} - \overline{\mathbf{v}}^f_{\widehat{j}})(\mathbf{v}^f_{i,\widehat{j}} - \overline{\mathbf{v}}^f_{\widehat{j}})^T}{M}. \tag{5.9}$$

Overall, what Bishop et al. (2017) have obtained is a new, expanded ensemble of size M with the same mean as the original and a covariance matrix that is equal to the localised matrix $\mathbf{P}^f_{\text{loc},\widehat{j}}$. Using the new modulated ensemble, we can redefine (5.3) as:

$$\mathbf{P}^f_{\text{loc},\widehat{j}}\mathcal{H}^T = \frac{1}{M}\sum_{i=1}^{M}\left(\mathbf{v}^f_{i,\widehat{j}} - \overline{\mathbf{v}}^f_{\widehat{j}}\right)\left(\mathcal{H}(\mathbf{v}^f_{i,\widehat{j}}) - \overline{\mathcal{H}(\mathbf{v}^f_{i,\widehat{j}})}\right)^T, \tag{5.10.a}$$

$$\mathcal{H}\mathbf{P}^f_{\text{loc},\widehat{j}}\mathcal{H}^T = \frac{1}{M}\sum_{i=1}^{M}\left(\mathcal{H}(\mathbf{v}^f_{i,\widehat{j}}) - \overline{\mathcal{H}(\mathbf{v}^f_{i,\widehat{j}})}\right)\left(\mathcal{H}(\mathbf{v}^f_{i,\widehat{j}}) - \overline{\mathcal{H}(\mathbf{v}^f_{i,\widehat{j}})}\right)^T. \tag{5.10.b}$$

Figure 5.1 shows a comparison between $\mathbf{P}^f_{\text{loc}}$ computed with (2.30) and (5.9). The two matrices are visually equal, demonstrating that the two methods lead to the same result.

5.1.4 Model-error covariance matrix

The use of a new nature run means that the model-error covariance matrix $\mathbf{Q}$ used to generate the additive inflation perturbations $\boldsymbol{\eta}$ (cf. section §2.2.3, Fig. 2.5) needs to be recomputed.

Figure 5.2 (left panel) shows the new elements of diagonal matrix $\mathbf{Q}$, obtained with the same procedure outlined in Chapter 2, that is, by running a series of 96 hourly forecasts (re-initialised hourly with the nature run) and generating a covariance matrix from the distribution of the differences $\mathbf{x}^t - \mathbf{x}^f$ computed at the end of each forecast; the non-diagonal terms of the $\mathbf{Q}$ matrix are set to zero for simplicity. The vertical dashed lines in Fig. 5.2 separate the σ, σu, σv and σr components of $\mathbf{Q}$.

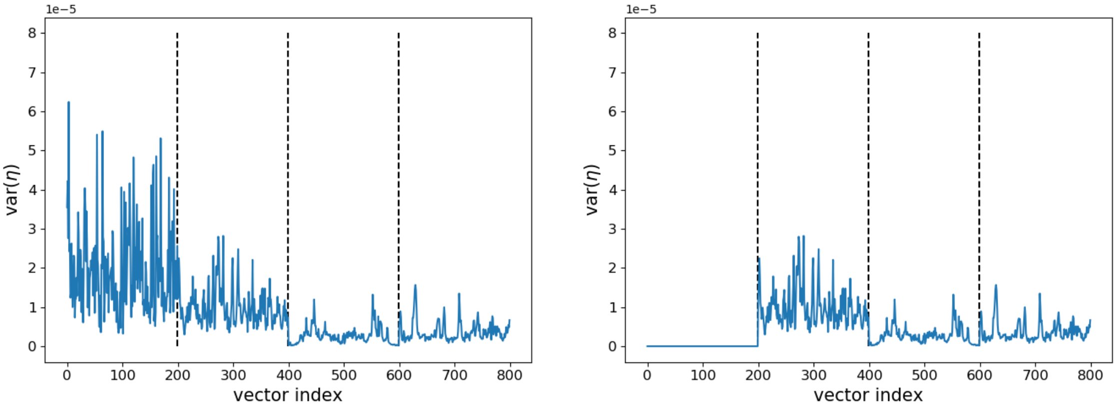


Figure 5.2: Entries of the (diagonal) model-error covariance matrix $\mathbf{Q}$. The black dashed lines separate the vector components σ, σu, σv, σr. The matrix on the right has the σ components set to zero and is used to compute the additive inflation $\boldsymbol{\eta}$ as per (2.31).

In Chapter 2, the hr component of $\mathbf{Q}$ was set to zero to avoid twice perturbing twice two positively correlated variables (h and r). Here, we decide to set to zero the σ component instead. This decision was taken in the course of the tuning process, during which the pseudo-density σ has shown the smallest benefit from the DA in terms of reduction of error (see also Table 5.4). In particular, the σ component of $\mathbf{Q}$ (left panel of Fig. 5.2) displays a high degree of noise, which would be reflected in the spatial pattern of the additive inflation perturbations. Hence, setting the σ component to zero (and therefore limiting the additive inflation to the σr component, right panel of Fig. 5.2) has the effect of both mitigating the possible negative impact of adding excessive noise to σ (which would degrade the analysis), and avoiding the double perturbation of two correlated variables, that is σ and r.

5.1.5 Algorithm modifications

Given the modifications described above and the model revision discussed in Chapter 3, the algorithm described in section §2.3 is revised accordingly:

- The matrix $\mathbf{R}$ used to generated the observations $\mathbf{y}_i$ in 2.33 is redefined as $\mathbf{R} = \text{diag}(s^2_{I_{sat}}\mathbf{I}_{I_{sat}}, s^2_u\mathbf{I}_u, s^2_v\mathbf{I}_v, s^2_r\mathbf{I}_r)$, with error variances $s^2_{I_{sat},u,v,r}$ and identity matrices $\mathbf{I}_{I_{sat},u,v,r}$ with dimension equal to the number of observations I_{sat}, u, v and r, respectively.

- In step i of the forecast step, σr is reset to zero and σ is reset to 0.001.
- In step i of the analysis step, the transformation before the assimilation reads $(\sigma, \sigma u, \sigma v, \sigma r) \mapsto (\sigma, u, v, r)$.
- Step iv and v of the analysis step are merged together via the calculation of the modulated ensemble technique discussed in section §5.1.3.

5.2 A well-tuned experiment

Tuning parameters	
Localisation scale, L_{loca}	$\{0.5,\ 1.0,\ 1.5,\ 2.0\}$
Adaptive inflation, α_{RTPS}	$\{0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.5,\ 0.6,\ 0.7,\ 0.8,\ 0.9\}$
Additive inflation, γ_a	$\{0.3,\ 0.4,\ 0.5,\ 0.6,\ 0.7,\ 0.8\}$

Table 5.2: An overview of the filter parameters used during the tuning of the forecast-assimilation experiments.

In this section we summarise the process undertaken to obtain a well-tuned experiment to be used as control in the data-denial simulations described in the next section. The procedure follows the same approach outlined in Chapter 2, conveniently modified to infer the relevance of the new idealised system for satellite DA research. We report in Table 5.3 a revised version of Table 2.2, in which the main modifications associated with the new system and the assimilation of idealised satellite observations are highlighted in red.

We start with the simultaneous evaluation of 216 experiments in which the observing system is kept fixed (cf. Table 5.1) and only the filter-related parameters are varied, that is, the localisation scale L_{loc}, the RTPS coefficient α_{RTPS} and the additive inflation scaling factor γ_a. The initial ensemble perturbation applied to all experiments is injected as in (2.16), here with $(s^{\text{ic}}_{\sigma}, s^{\text{ic}}_{\sigma u}, s^{\text{ic}}_{\sigma v}, s^{\text{ic}}_{\sigma r}) = (0.02, 0.008, 0.1, 0.0)$. The combination of parameters utilised is displayed in Table 5.2. The meridional velocity v is not included in the tuning process because of its peculiar role as both a forced and observed variable, and also for its less interesting dynamical behaviour.

Table 5.3: Protocol to assess the performance and the relevance of an idealised forecast-assimilation system, applied to the experiments conducted with the ismodRSW model. The modifications w.r.t. Table 2.2 are in red.

	Operational system	Our idealised system	Relevant?
Prescribed parameters			
Forecast resolution	$\mathcal{O}(1\text{km})$	2.5km	✓
Update frequency	$\mathcal{O}(1\text{hr})$	1hr	✓
Ensemble size, N	$\mathcal{O}(10-100)$	20	✓
Number of observations, p	$\mathcal{O}(10^7)$	38	N/A
State dimension, n	$\mathcal{O}(10^9)$	600	N/A
Rank-deficiency	$N \ll p \ll n$	$N < p < n$	✓
Observation operator	Nonlinear	Nonlinear	✓
Tuning parameters			
Observing system:			
(i) ground obs. density	~ 40km	50km	✓
(ii) satellite to ground obs. ratio	1900/6830 (0.27)	8/30 (0.27)	✓
(iii) obs. error	Correlated and uncorrelated	Uncorrelated	–
Filter configuration:			
(i) Localization (horiz.)	Lengthscale: $\mathcal{O}(100\text{ km})$ (obs. space)	~ 500km (model-space)	✓
(ii) Inflation	Adaptive (RTPP; RTPS) and additive	$\alpha_{RTPP} = 0.5$, $\alpha_{RTPS} = 0.6$, $\gamma_a = 0.5$	✓
Tuning criteria			
(i) $SPR/RMSE$	~ 1	~ 1	✓
(ii) RMEE	Minimum for given lead times	Minimum for 3hrs forecast	–
(iii) CRPS	Minimum for given lead times	Minimum for 3hrs forecast	–
Validation criteria			
Observational influence	Global: $\sim 20\%$; high-res.: $> 20\%$	$\sim 25\%$	✓
Error-doubling time T_d	Global: $\mathcal{O}(1\text{ day})$; high-res.: $\mathcal{O}(1\text{ hr})$	$\sim 5-17$ hrs	✓

In the initial stage, we seek the experiments that satisfy the following tuning criteria (all diagnostics have been defined in full in section §2.4.2):

(i) a SPR/RMSE ratio around 1 (i.e. $0.8 \leq SPR/RMSE \leq 1.2$);

(ii) minimum RMSE;

(iii) minimum CRPS.

Figure 5.3 summarises the values of these three diagnostics across the experiments. Each box represents a 96 hour simulation corresponding to a specific combination of parameters $\{L_{loc}, \alpha_{RTPS}, \gamma_a\}$, which has been averaged over time (excluding an initial spin-up period of 12 hours), space and the model variables σ, u and r.[1] The target values for each criterion are highlighted in white or brighter colours.

The top left panel shows the values of $SPR/RMSE$, in which the experiments satisfying the condition (i) are outlined in red. For all localisation scales, these are located in the top half of the parameter space. In fact, larger ratios are found to correspond to higher inflation values (both additive and RTPS), which is to be expected as inflation acts to increase the ensemble spread. The bottom panels display the corresponding RMSE (bottom left) and the CRPS (bottom right), respectively. The experiments with the lowest values of RMSE and/or CRPS (highlighted in lighter colours) among those already selected in the previous step are contoured in black. This subset of experiments satisfies the tuning criteria listed above and is therefore subject to further examination. In particular, we now turn to select a value of L_{loc}.

To achieve this aim, Figure 5.4 shows a comparison of the correlation matrix obtained from $\mathbf{P}_e^f$ before and after localisation is applied; the matrices of four of the experiments remaining after the previous tuning based on SPR, RMSE and CRPS are plotted, one for each value of L_{loc} (the values of α_{RTPS} and γ_a are reported in the caption). Similarly to what we observed in Chapter 2, a localisation scale of $L_{loc} = 1.0$ seems to represent a good compromise between the excessive suppression of off-diagonal signal for larger L_{loc} values (i.e. $L_{loc} = \{1.5, 2.0\}$, cf. bottom panels in Fig. 5.4) and a value of $L_{loc} = 0.5$ which may not go far enough in dampening the spurious correlations emerging from a small ensemble. Interestingly, the experiments with a value of $L_{loc} = 0.5$ resulted also in larger RMSE and CRPS values, as highlighted by a

---

[1] Note that, differently from Chapter 2, here there is no need to multiply r by a factor 100, as the error across the variables are of similar magnitude.

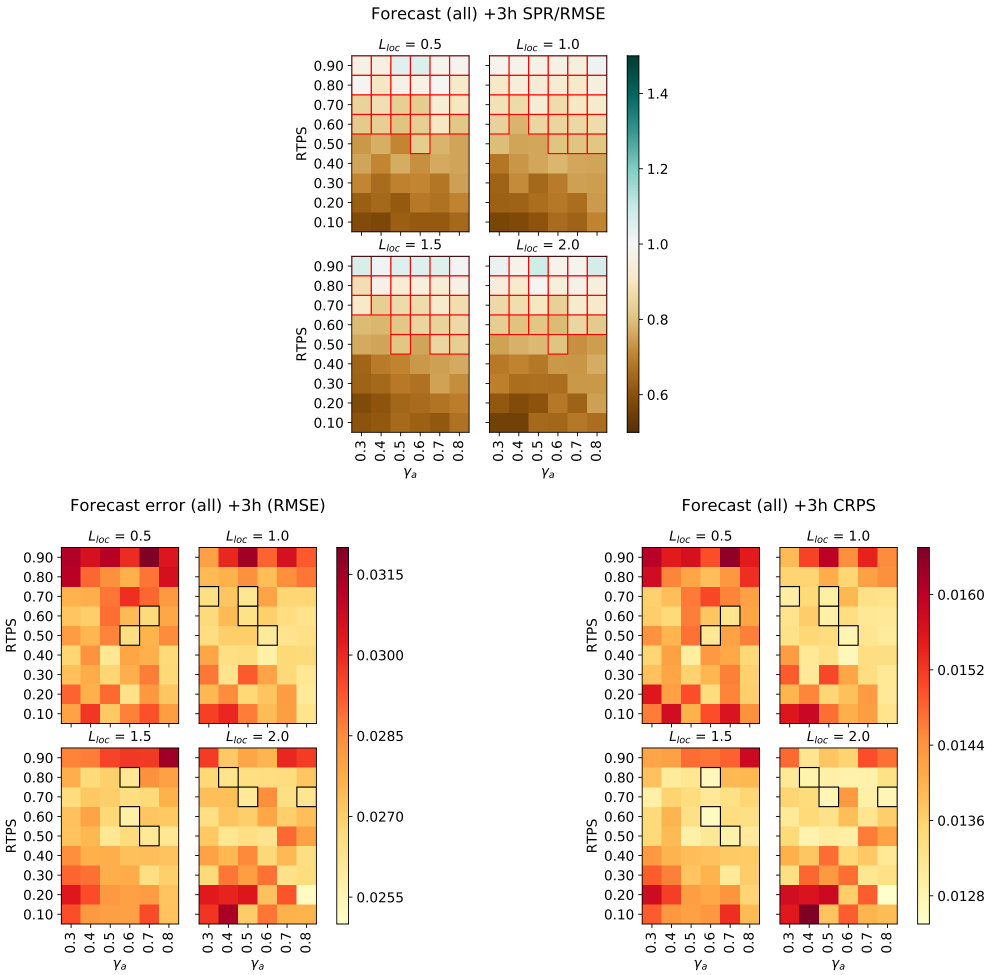


Figure 5.3: Results summary of the 216 experiments obtained with the parameters listed in Table 5.2. Top: $SPR/RMSE$; bottom left: RMSE; bottom right: CRPS. The experiments satisfying the condition $0.8 \leq SPR/RMSE \leq 1.2$ are contoured in red in the top figure; among these, the ones with the lowest values of RMSE and/or CRPS for each value of L_{loc} are contoured in black in the panels below.

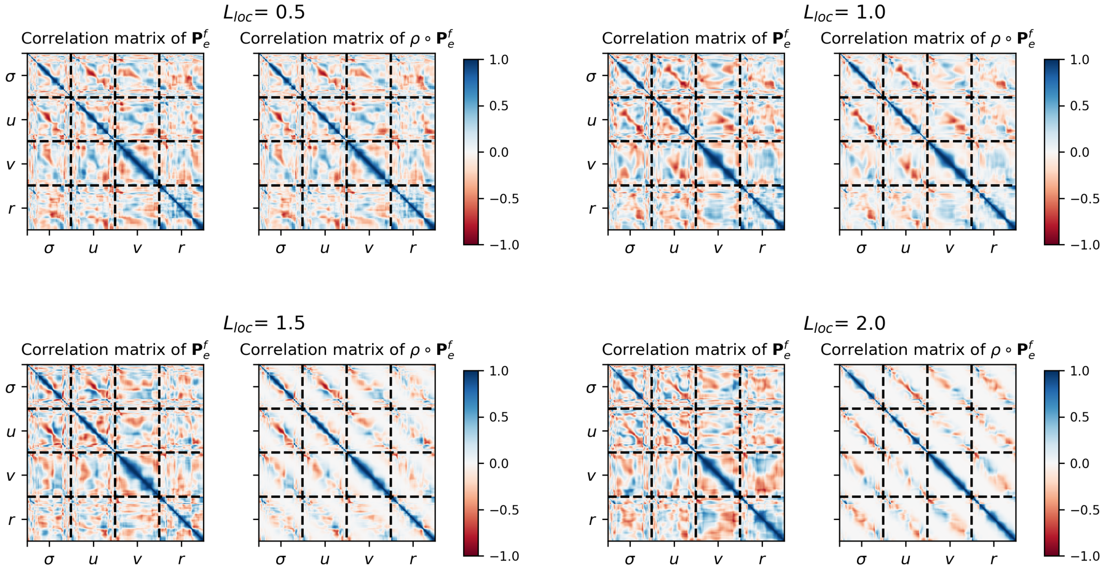
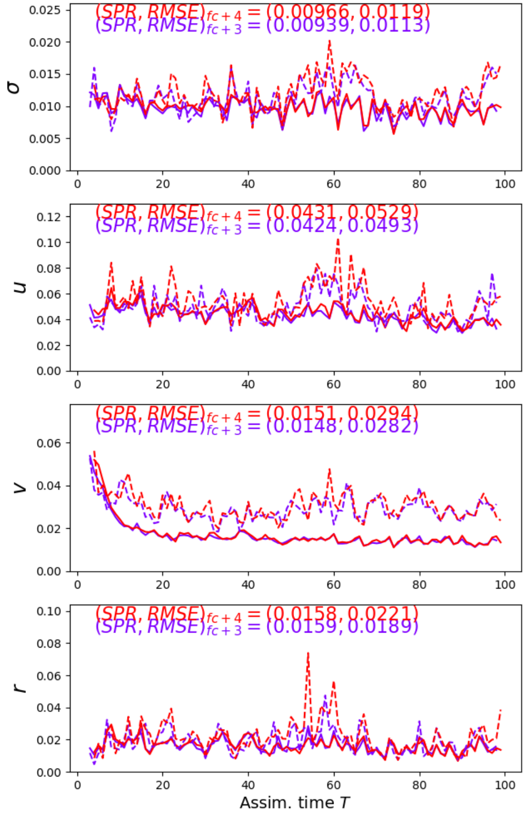


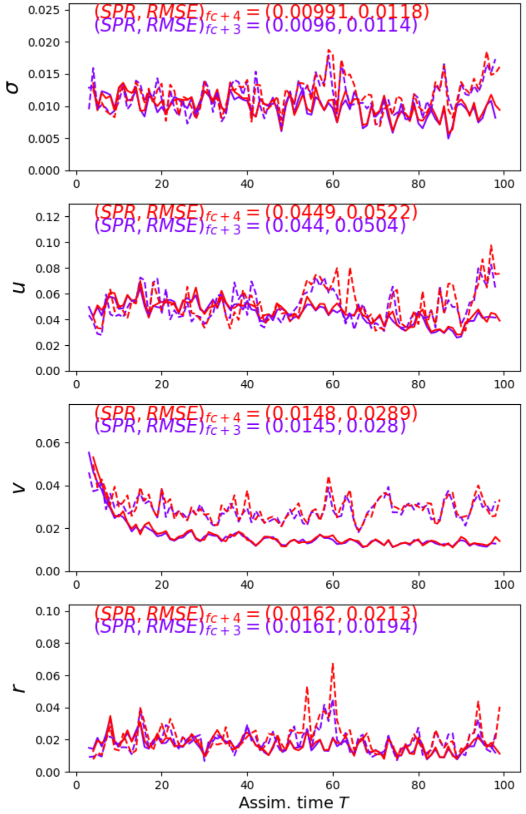
Figure 5.4: Effects of localisation (for different values of the localisation scale L_{loc}) on the forecast-error correlation matrices (derived from the covariance matrices $\mathbf{P}_e^f$ used in the DEnKF). Top left: $\gamma_a = 0.6$, $\alpha_{RTPS} = 0.5$; top right: $\gamma_a = 0.5$, $\alpha_{RTPS} = 0.6$; bottom left: $\gamma_a = 0.6$, $\alpha_{RTPS} = 0.6$; bottom right: $\gamma_a = 0.5$, $\alpha_{RTPS} = 0.7$. For each value of L_{loc}, two correlation matrices are shown: before (left) and after (right) the localisation. Note that since the self-exclusion prescribes the computation of a different covariance matrix $\mathbf{P}_{e,j}^f$ for each j-th ensemble member (cf. §2.2.3), the average matrix is shown instead.

larger number of orange and dark red boxes in the $L_{loc} = 0.5$ panels at the bottom of Fig. 5.3. Selecting $L_{loc} = 1.0$ narrows the search for a well-tuned experiment further, leaving only 4 simulations as candidates (cf. Fig. 5.3). At this point, we focus on evaluating the impact of data assimilation on the quality of the forecasts by looking at the difference in RMSE between two forecasts with different lead times (and valid at the same analysis time). To this end, we repeat the choice made in Chapter 2 and use 3 and 4 hours forecasts (in short: 3hrs and 4hrs). Again, the expectation is that the 3hrs forecasts have a smaller error than 4hrs forecasts, as the former have been initialised more recently benefiting from more recent observations. In Figure 5.5 we show the domain averaged time series of both 3hrs (blue) and 4hrs (red) SPR (continuous lines) and RMSE (dashed lines) of all variables for each experiment. The average values of all scores are also displayed in each panel (these are again computed after excluding the first 12 values to avoid any spin-up effects). As expected, both the spread and the error of the 3hrs forecasts are smaller than for the corresponding 4hrs forecasts, indicating an increase in accuracy due to a more recent assimilation of new observations. Other than being qualitatively visible in all experiments reported in Fig. 5.5, this feature is also confirmed quantitatively in Table 5.4, where both the RMSE values of 3hrs and 4hrs forecasts and the relative percentage reductions are showed. In particular, decreases in a range of 3% to 14% are found for the single variables, with an average reduction of 5%-8% for the overall experiments. It is worth noting that the two variables most prone to non-linear effects, σ and r, display very different behaviour. On the one hand, the pseudo-density σ, which is also the variable most affected by the satellite observations, shows a rather small percentage reduction (3%-6%). On the other hand, the rain mass fraction r displays the largest RMSE improvements across the experiments (9%-14%).

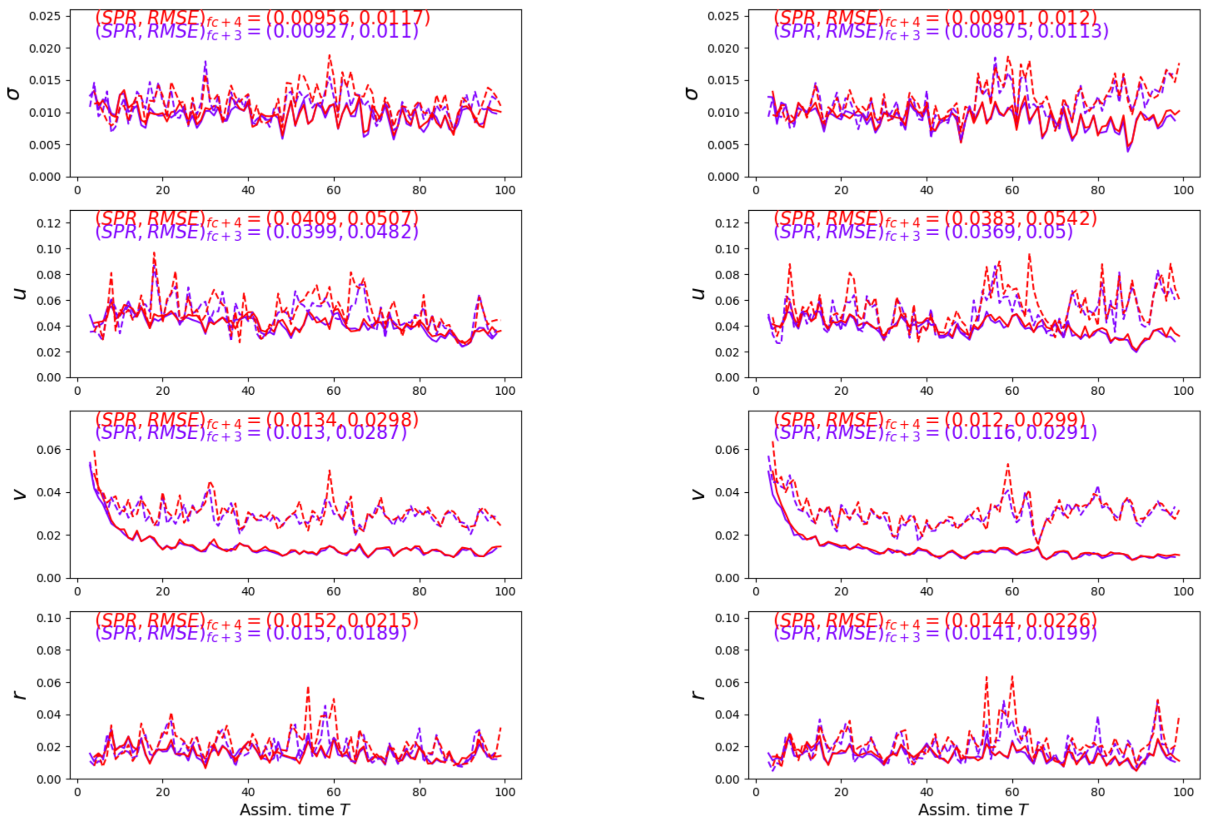
Figure 5.5 can also be used as a further confirmation of the ratio between the ensemble spread and the error being close to one in all experiments, highlighted by a general overlap between the dashed and the continuous lines. Nonetheless, we note that the meridional velocity v, which we previously excluded from the computation of the tuning diagnostics, is considerably underspread in these experiments (that is, it has a $SPR/RMSE$ ratio below 1). Moreover, the oscillatory behaviour displayed by all variables (especially in the RMSE) can be attributed to the cyclical passage of gravity waves moving around and re-entering the (periodic) domain. We also note an increase in RMSE for both lead times at the very end of the time series, affecting σ and u in particular in all experiments shown in Fig. 5.5, with the exception of the simulation



(a) $L_{loc} = 1.0$, $\gamma_a = 0.3$, $\alpha_{RTPS} = 0.7$.



(b) $L_{loc} = 1.0$, $\gamma_a = 0.5$, $\alpha_{RTPS} = 0.7$.



(c) $L_{loc} = 1.0$, $\gamma_a = 0.5$, $\alpha_{RTPS} = 0.6$.

(d) $L_{loc} = 1.0$, $\gamma_a = 0.6$, $\alpha_{RTPS} = 0.5$.

Figure 5.5: Domain-averaged time series of SPR (solid lines) and RMSE (dashed lines) for 3hrs forecasts (blue lines) and 4hrs forecasts (red lines) of the four experiments with $L_{loc} = 1.0$ (parameters values reported in the subcaptions). Variables (from top to bottom, in each subplot): σ, u, v, r. Each panel reports at the top the time averages (which exclude the first 12 assimilation times).

shown in the subplot (c).

In the end, we select the experiment with parameters $L_{loc} = 1.0$, $\alpha_{RTPS} = 0.6$ and $\gamma_a = 0.5$ as our well-tuned simulation to be used as the control in the data-denial experiments which will follow in section §5.6. This choice is motivated by our interest in the impact of satellite observations: indeed, this experiment produces the highest RMSE reduction in the pseudo-density σ (cf. Table 5.4), which we expect to be the most impacted by the assimilation of satellite observations, and therefore the one that can act as the best reference in the data-denial experiments. In addition, this experiment is also the only one in Fig. 5.5 not showing a increase in RMSE at the very end of the time series.

5.3 Relevance for NWP data assimilation

In order to assess the relevance for NWP DA of the experiment selected in the previous section, we turn to two further diagnostics already utilised in Chapter 2: the observation influence (OID) and the error doubling time. Both these quantities have already been defined in detail in section §2.4.2.

Figure 5.6 shows the time series of the OID, including a break-down by observation type. Interestingly, the impact of both u and r observations (blue and purple lines, respectively) fluctuates around 10% for the entire duration of the experiment. Conversely, the impact of satellite observations (red line) is more irregular and displays several oscillations between 0% and 10%, with some occasional peaks above this value. This behaviour depends on the spatial distribution of the satellite observations, which are at times concentrated in areas of the domain where there is neither convection nor rain and the forecast uncertainty is very low. The impact of v observations (cyan line) is initially very high, but it rapidly decreases and subsequently stays low for the rest of the experiment. The reason for its initial large impact is due to an artificially high ensemble spread at the start of the experiment, which is meant to counterbalance the tendency of v to be underspread, as we have observed in the previous section. Overall, the total OID starts at around 40% and gradually decreases until it settles around 20%; overall, these values are in line with the observation influence to be expected in a high-resolution NWP model.

Lastly, Figure 5.7 shows the distribution of the error doubling times for each variable. These are obtained from a series of 36-hour long forecasts for each of the 20 ensemble

Experiment	Lead time	σ	u	r	Average
$\gamma_a = 0.3$, $\alpha_{RTPS} =$ 0.7	3hr	0.0113	0.0493	0.0189	
	4hr	0.0119	0.0529	0.0221	
	% diff	5.0%	6.8%	14.5%	8.8 %
$\gamma_a = 0.5$, $\alpha_{RTPS} =$ 0.6	3hr	0.0110	0.0482	0.0189	
	4hr	0.0117	0.0507	0.0215	
	% diff	6.0%	4.9%	12.1%	7.7%
$\gamma_a = 0.5$, $\alpha_{RTPS} =$ 0.7	3hr	0.0114	0.0504	0.0194	
	4hr	0.0118	0.0522	0.0213	
	% diff	3.4%	3.4%	8.9%	5.2%
$\gamma_a = 0.6$, $\alpha_{RTPS} =$ 0.5	3hr	0.0113	0.0500	0.0199	
	4hr	0.0120	0.0542	0.0226	
	% diff	5.8%	7.7%	11.9%	8.5%

Table 5.4: Summary of the RMSE values (and percentage reduction between the 4hr and the 3hr forecasts) of the four experiments shown in Fig. 5.5.

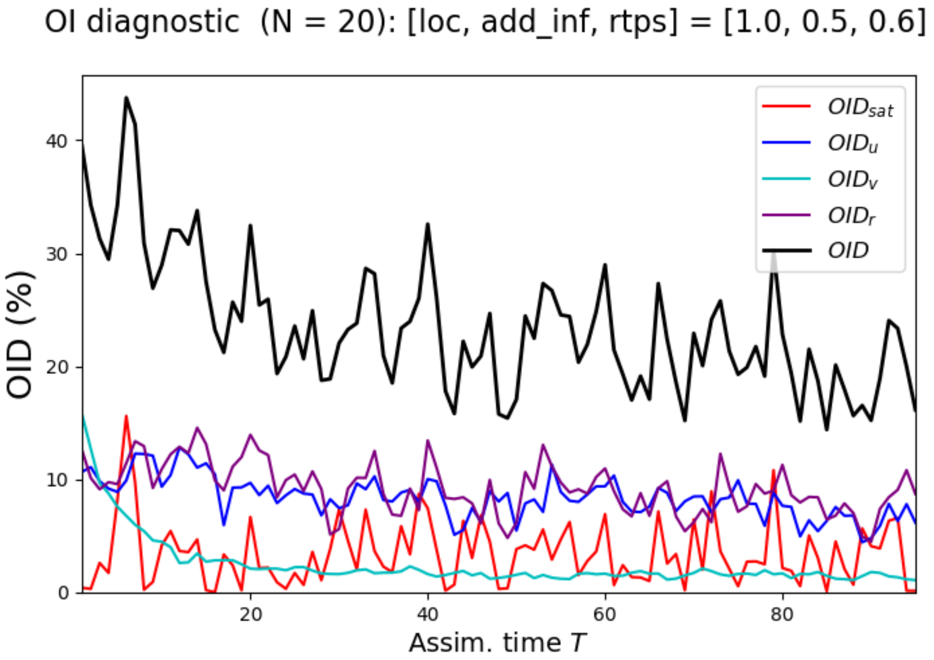


Figure 5.6: Time series of the observational influence diagnostic (OID), as a percentage (%). Overall OID and the OID computed for the single variables are reported in different colours (see legend).

members initialised with the analysis ensemble of the first 60 analysis steps, for a total of 1200 forecasts. The doubling times of u and r are similar to those found in Chapter 2. In particular, by comparing Figures 2.10 and 5.7 we note that the doubling times of u are slightly larger (median from 9 to 10 hours, mean from 9.9 to 11.6 hours), whereas those of r are slightly smaller (median from 6 to 5 hours, mean from 8.1 to 7.1). Similarly to the findings in Chapter 2, r continues to be the variable of the system that is the quickest to double. Conversely, σ and v display considerably larger doubling times, with values around 16-17 hours for both. In the case of the meridional velocity v, the larger doubling times can be explained by the relaxation towards v_{rel} (cf. (3.45)), which forces both the forecast and the nature run towards the same solution, and therefore slows down the error growth. The large doubling time of σ has a less clear explanation. One hypothesis is that the assimilation scheme is still not optimised for σ (which the limited improvement in RMSE highlighted in Table 5.4 might also be a sign of) and therefore the initial analysis error may be too large or close to its saturation value, hampering its ability to grow further in time.

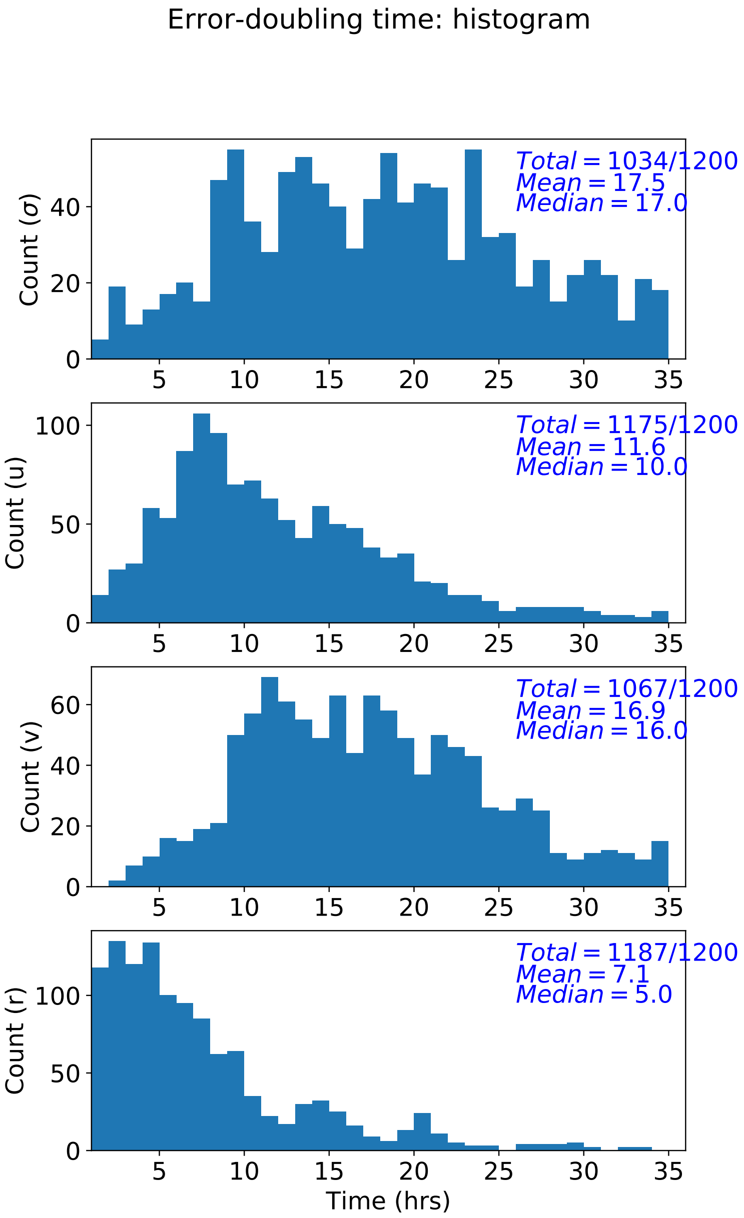


Figure 5.7: Error-doubling time distributions based on 1200 forecasts. Panels from top to bottom: σ, u, v and r. Each panel reports the number of forecasts in which the initial error has doubled within 36 hours, the mean and the median of the distribution.

5.4 Subjective verification

In order to visualise the effects of the DA scheme, we show in Figure 5.8 a snapshot of the model variables (plus I_{sat}) at time $T = 54$ hours, before (left panels) and after (right panels) the assimilation of the observations. The figure displays the nature run trajectory (solid green line), the forecast ensemble mean (solid red line), the observations with their error bars (green dots), the analysis ensemble mean (solid cyan line) and the ensemble members (blue lines).

The top two panels show the action of the observation generator function $\mathcal{G}_s$ in (4.30) in mapping the pseudo-density σ into a set of $p_{sat} = 8$ idealised satellite observations of radiance I_{sat}. In particular, the comparison between σ and I_{sat} reveals the effect of the non-linearity in $\mathcal{G}_s$, which shrinks the values of σ below the convection threshold into quasi-constant values of I_{sat}, contributing to the reduction of the ensemble spread. This distortion therefore has the effect of reducing the impact of the four observations in the first half of the domain ($x < 0.5$) and suggests that the dips in the observation influence diagnostic of Fig. 5.6 are caused by the concentration of satellite observations in areas with no convection nor rain.

Furthermore, Figure 5.8 shows in practice the action of the DA scheme in adjusting the forecast towards the observations (and the nature run). This is clearly noticeable in r, where the observations at $x = 0.6$ and $x = 0.7$ improve both the shape and the peak of the large area of rain in the second half of the domain. A small adjustment is also made in σ around $x = 0.6$, where the forecast mean is struggling to resolve a single convection updraft located at the same location; in this case, the presence of a few satellite observations plays a crucial role in correcting the trajectory. Minor adjustments are also visible in u, whereas the DA seems to have a marginal impact on v, which is expected given the relaxation process being applied to this variable.

5.5 Observations impact and spatial scales

The purpose of this thesis is to investigate the impact of assimilating satellite observations at various spatial scales. More generally, the relationship between the impact of the observations on the analysis and their spatial structure – including their density and error properties – has been already investigated in several studies covering

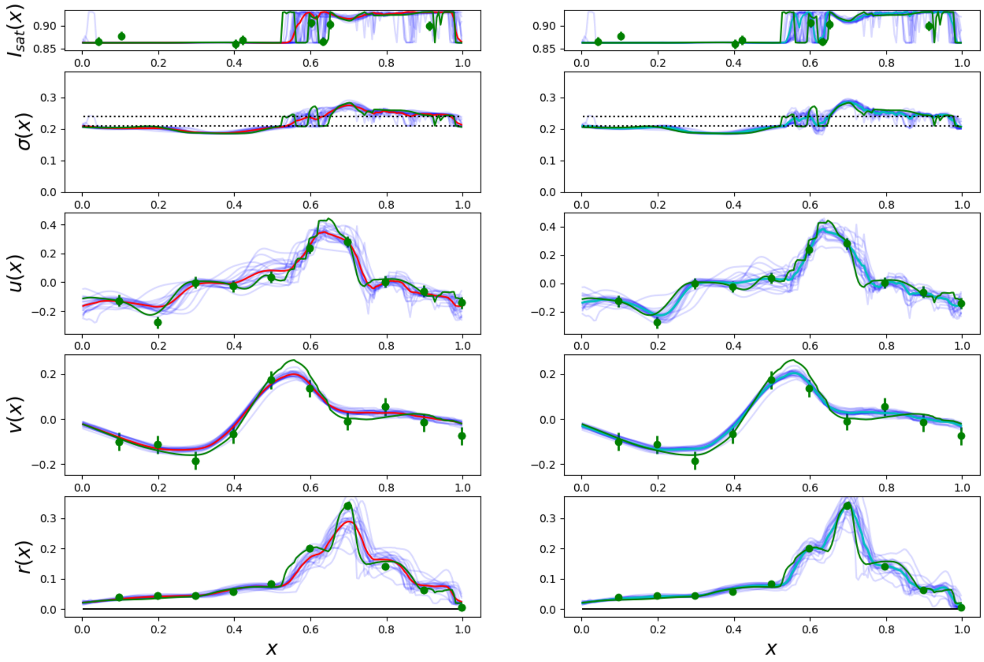


Figure 5.8: Model snapshot valid at $T = 54$. Left: 1hr forecast. Right: analysis. The nature run trajectory is in green, the ensemble trajectories are in blue, the forecast mean is in red and the analysis mean is in cyan. The observations are indicated by green dots with error bars.

various data assimilation topics, such as: the role of both the background and the observation covariance matrices in shaping the analysis, the use of observations thinning techniques, the adoption of ensemble covariance localisation, and the role of representation errors. We present a brief literature review for each of these subjects below.

The interplay between the background error covariance matrix and the impact of the observations on the analysis is nicely summarised in a review article by Bannister (2008). Here it is shown with both theoretical and practical considerations how one of the role of the background error matrix is to spread the information carried by the observations both horizontally and vertically, depending on the strength of the spatial correlations between grid points and model levels.

The impact of the spatial scale, density and geographical distribution of the observations on the analysis is also linked to how the observation error covariance matrix is modelled, and in particular to whether the observation errors are treated as correlated or uncorrelated. For example, a number of studies have shown that while dense observation networks of truly uncorrelated observations lead to a more accurate analysis at the small scales, this ceases to be true when dense correlated observations are treated as uncorrelated (Dando et al., 2007; Forsythe, 2007; Liu & Rabier, 2002, 2003). Treating the observations as explicitly correlated is only a recent development in data assimilation, which can benefit the representation of the small scales in the analysis, as shown by Rainwater et al. (2015) and in section §4.8 of Daley (1993).

In practice, the background and observation error covariance matrices interact with each other and both affect the impact of the observations on the analysis, as noted in Fowler et al. (2018). In this paper, for example, it is shown how a denser network is most useful when the observations provide a better description of the small scales than the background.

The thinning and superobbing of observations are common techniques that allow to reduce the absolute number of observations to be assimilated, as well as optimising the assumptions of uncorrelated observations made in many operational DA systems (Berger et al., 2004; Fowler et al., 2018). As an example, Dando et al. (2007) estimated the optimal thinning distance for the ATOVS satellite observations (assumed as uncorrelated) within the Met Office DA system (with a grid resolution of 60 km), finding a value in the range 110-150 km, and observing a degradation of the forecast error for shorter distances. They also noticed that in the areas with strong horizontal gradients, a shorter thinning distance of 40 km produced the smallest forecast error: a finding in line with the conclusions of an earlier paper by Seaman (1977).

Experiment	# of sat. obs.	FOV [km]
`control`	8	{20,20,40,40,60,60,80,80}
`small scales`	4	{20,20,40,40}
`large scales`	4	{60,60,80,80}
`no sat`	0	—

Table 5.5: Summary of the data denial experiments carried out.

The property of an observation network can also influence the choice of an optimal localisation scale. As we observed in section §2.2.3, the localisation damps the long-distance background error correlations, implicitly reducing the scale over which the information carried by an observation can be spread. In this regard, Perianez et al. (2014) have shown how the optimal localisalisation scale for an EnKF depends on both the observation error and the observation density.

Finally, the representation error constitutes the preeminent source of observation error in data assimilation, is clearly related to the scale of the observations and can impact negatively on the quality of the analysis when is not properly identified and dealt with. Janjić et al. (2018) offer an overview on the subject, providing many examples. As highlighted in the paper, one of the source of representation error lies in the mismatch in spatial scale between the model and the observations, with the latter often reflecting much more local and small-scale atmospheric conditions than NWP models are able to resolve. The paper discusses various methods to identify and quantify the representation errors properly, and highlights the importance of a better understanding of observation-error statistics, including the role of correlated observations.

5.6 Data denial and single-observation experiments

In this section we analyse the results of a series of data denial and single-observation experiments conducted to investigate the role of satellite observations on the idealised forecast-assimilation system described so far. In particular, we are interested in studying the impact of satellite observations of different spatial scales on our DA system.

First, we investigate the impact of a single satellite observation of varying spatial resolution (or FOV), located at the center of the domain, in the presence of a convection updraft. In order to study the impact of the spatial scale of the observation, we compute the analysis increments (analysis minus background) for each FOV value $FOV = \{20, 40, 60, 80\}$ km and for all localisation scales $L_{loc} = \{0.5, 1.0, 1.5, 2.0\}$.
Furthermore, we carry out a series of data-denial experiments, which are summarised in Table 5.5. We use the well-tuned simulation selected in section §5.2 (with $L_{loc} = 1.0$, $\gamma_a = 0.5$, $\alpha_{RTPS} = 0.6$ and $p_{sat} = 8$) as a control simulation, and subsequently compare it against three experiments which differ only in the number (and the FOV) of the satellite observations assimilated. To achieve this aim we run: one experiment in which only four 'small scale' (SS) satellite observations (with FOVs of 20 and 40 km) are assimilated; a second one in which only four 'large scale' (LS) satellite observations (with FOVs of 60 and 80 km) have been used; and a third one in which only ground observations have been included (no sat).
In order to assess the performance of each configuration, we will consider a number of diagnostics. In particular, we will assess the accuracy of the ensemble mean (RMSE), the quality of the ensemble (using both the CRPS and rank histograms) and the impact that the observing system has on the analysis (OID) (more details on their definitions are reported in section §2.4.2). We will compute these diagnostics for both the analysis and 3hrs forecasts in order to understand the impact of satellite observations both at the time when they are assimilated and in a subsequent forecast.
The results presented in this section should be treated as preliminary, and further work should follow.

5.6.1 Single-observation experiments: results

Starting from the same initial condition and using the same filter configuration obtained at the end of section §5.2, we conduct a series of 3 cycles experiments ($t = [0, 0.267]$), assimilating only one, stationary satellite observations located at $x = 0.5$ for each FOV value. Here, we analyse the impact of such observation on the analysis. Figure 5.9 shows the nature run trajectory (green line), the +1hr forecast ensemble members (blue lines) and their ensemble mean (red line) for each variable (including of the satellite radiance $I(x)$) valid at the second analysis step ($t = 0.178$), when a convective updraft appears at the centre of the domain. Because of its lower spatial

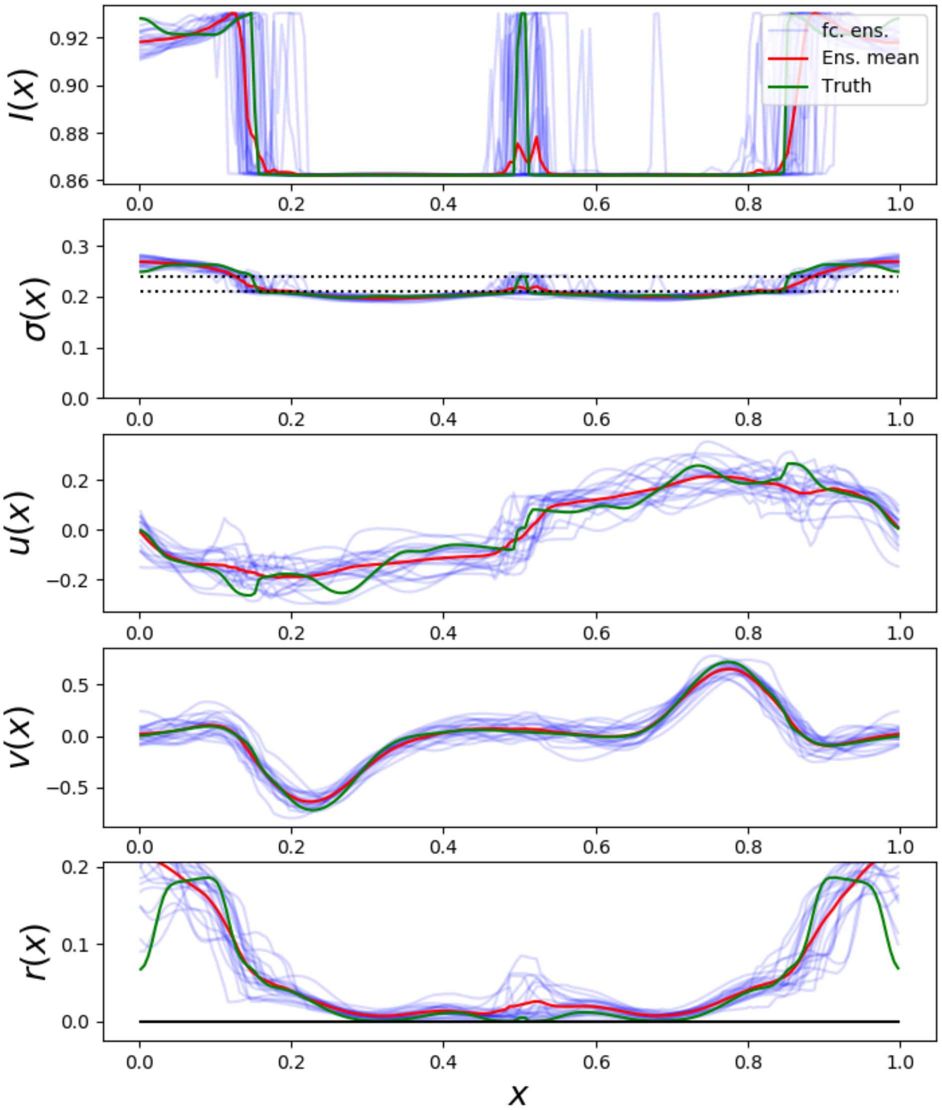


Figure 5.9: Snapshot of model dynamics at $t = 0.178$, displaying the +1hr forecast ensemble members (blue lines), the ensemble mean (red line) and the truth trajectory (green line). Top panel: satellite radiance $I(x)$; second panel: pseudo-density σ; third panel: horizontal velocity u; fourth panel: meridional velocity v; bottom panel: rain r.

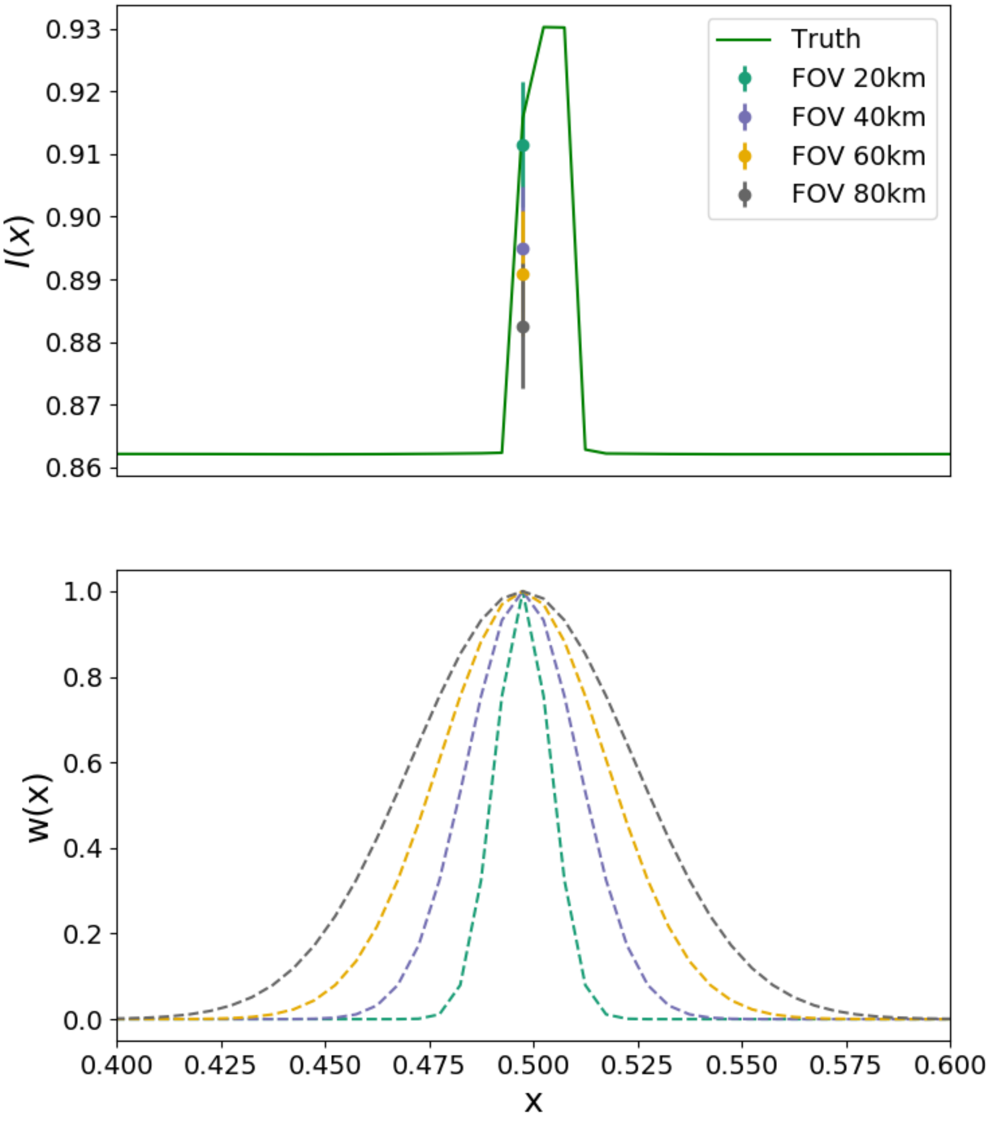


Figure 5.10: Top panel: detail of the nature run trajectory (or truth) of the satellite radiance $I(x)$ at $t = 0.178$ (green line) and pseudo satellite observations I_{sat} (with error bars) valid at the same time, for each FOV value. Bottom panel: weight function $w(x)$ defining the horizontal resolution of the pseudo-satellite observations for each FOV value, as per (4.36).

resolutions and the perturbed initial conditions, the ensemble mean is unable to resolve the peak in $I(x)$ (first panel) and $\sigma(x)$ (second panel) as accurately as the nature run; in fact, a number of ensemble members even predict a value of pseudo-density below the convection threshold, while others show a displaced updraft. As a result, the ensemble mean presents a much broader and shallower area of convection than the truth trajectory and this becomes an excellent test bench for our data assimilation system, as the presence of an observation in the same location as the updraft will help adjust the model towards the nature run.

Figure 5.10 zooms in on the portion of domain where the observations (one for each FOV value) are located, in correspondence of the convective area (top panel). The value of each I_{sat} observation depends on its FOV, that is, its horizontal averaging functions $w(x)$ (cf. (4.36)), which is reported in the bottom panel. Clearly, larger FOVs correspond to smaller values of I_{sat}, as the contribution of the smaller radiance away from the peak increases. As a result, the observation with the highest spatial resolution (a FOV of 20km, green dot) is also the closest to the truth trajectory, while the one with the lowest resolution (a FOV of 80km, grey dot) largely underestimates the value of the nature run at the same location. It is worth noting that all observations are an underestimation of the actual 'true' value of the radiance $I(x)$.

The left panels of Figure 5.11 show the analysis increments (i.e. the difference between the analysis and the forecast) for each variable and each experiment. As expected, the experiment with the smallest FOV (green line) produces the largest adjustment in σ near the observation location (top panel). The experiments with intermediate values of FOV (blue and yellow lines) still produce an adjustment in the right direction, although of smaller magnitude. Finally, the experiment with the largest FOV (grey line) produces an adjustment in the wrong direction. Much smaller increments are generated away from the centre of the domain, but are largest in the experiment with the smallest FOV. A similar pattern is repeated in the rain variable r (bottom panel), whereas wider and more complex adjustments are visible in u and v (middle panels); also in this case, the experiments with the smallest FOV produces the biggest increments.

Finally, the right panels of Figure 5.11 show the analysis increments of each variable for the experiment with the smallest FOV (20 km) when the localisation scale L_{loc} is varied. As expected, the smallest value of L_{loc} (green line, corresponding to a localisation length equal to the domain size) produces significant increments away from the observation location (at $x = 0.5$), as the forecast covariance matrix retains most of

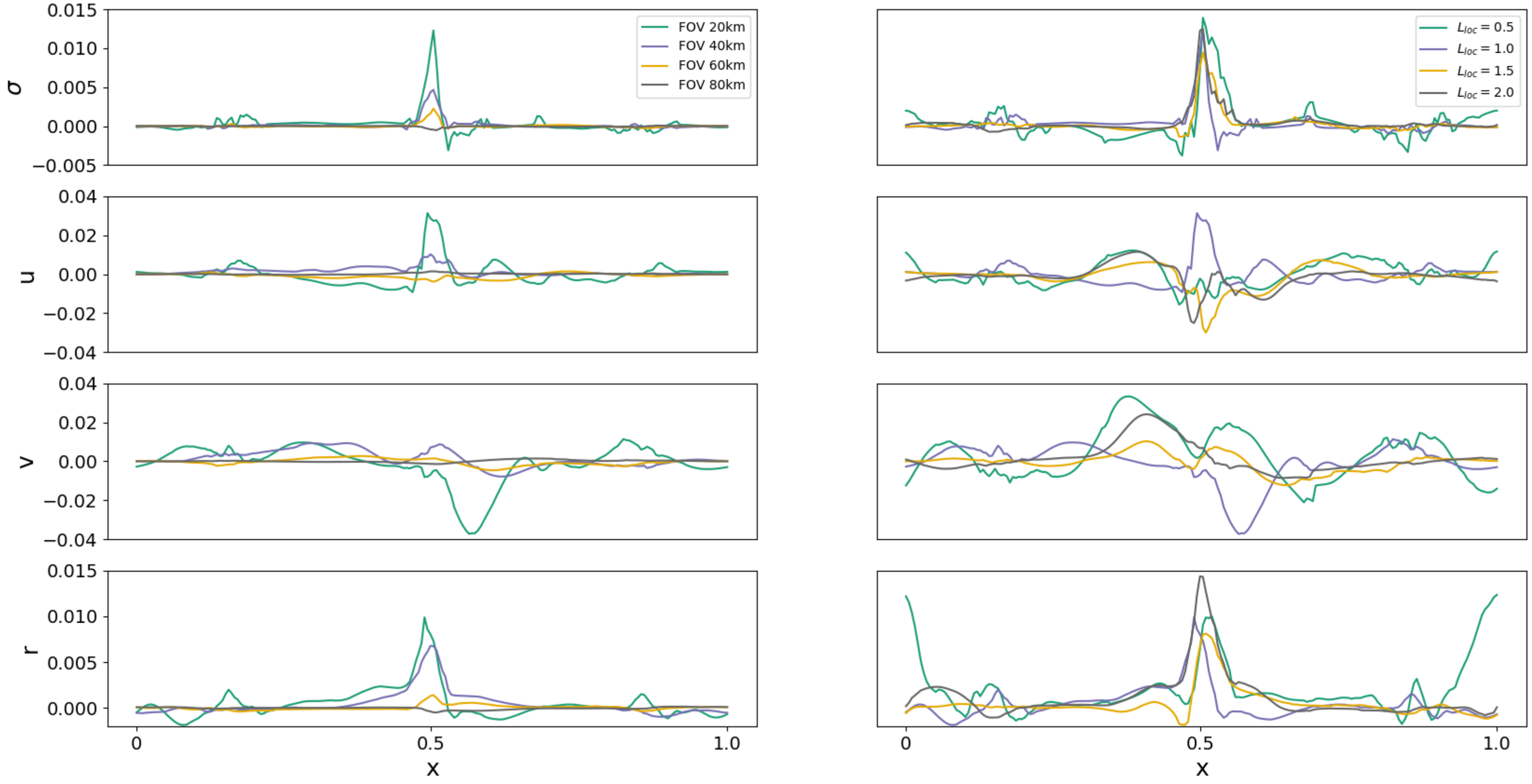


Figure 5.11: Analysis increments at $t = 0.178$ for each model variable as a function of the FOV (left panel) and of the localisation scale L_{loc} (right panel).

the off-diagonal signal which leads to long-range adjustments in the analysis; this is especially visible in r, where a large adjustment is made around the end of the domain. Conversely, as the value of L_{loc} increases and the localisation length shortens, the analysis increments become smaller away from the centre of the domain.

5.6.2 Data-denial experiments: results

RMSE

Figure 5.12 shows the time series of the analysis (domain-averaged) RMSE for each model variable and for all the experiments listed in Table 5.5. To facilitate the comparison between them, each panel is complemented with its time-averaged values in the plot legends.

Overall, the differences in RMSE between the experiments are very small and more work needs to be done to ascertain the statistical significance of these results. However, the large scale only simulation (LS, blue line) consistently underperforms the control (green line) across all variables, displaying an increase in RMSE of 1%-5%. Instead, it is more difficult to interpret the performance of the small scale (SM, yellow line) and the no sat (gray line) simulations: the former slightly overperforms the control in σ and u (with a reduction in RMSE between 0% and 2%), underperforms the control in v and produces the same results for r; the latter displays larger RMSE values in all variables (that is, an increase of 1%-4%) but in v, where it shows the lowest error.

Figure 5.13 reports equivalent RMSE time series for 3hrs forecasts. Even at this time from the initialisation, the LS experiment continues to underperform the control in all variables, with an increase in RMSE of 1%-8%. Interestingly, the 'no sat' experiment, which relies only on ground observations, appears to be more accurate than the control in all variables, with a small reduction in RMSE (i.e. 0%-2%) in σ, u and r, and a larger decrease (around 3%) in v, suggesting a sub-optimal tuning configuration. Lastly, the small scale experiment appears to be marginally better (RMSE reduction of 0%-1%) than the control in all variables but σ.

CRPS

Figure 5.14 displays the time series of the (domain-averaged) analysis CRPS. Again, the differences between the experiments appear small, and the interpretation of the

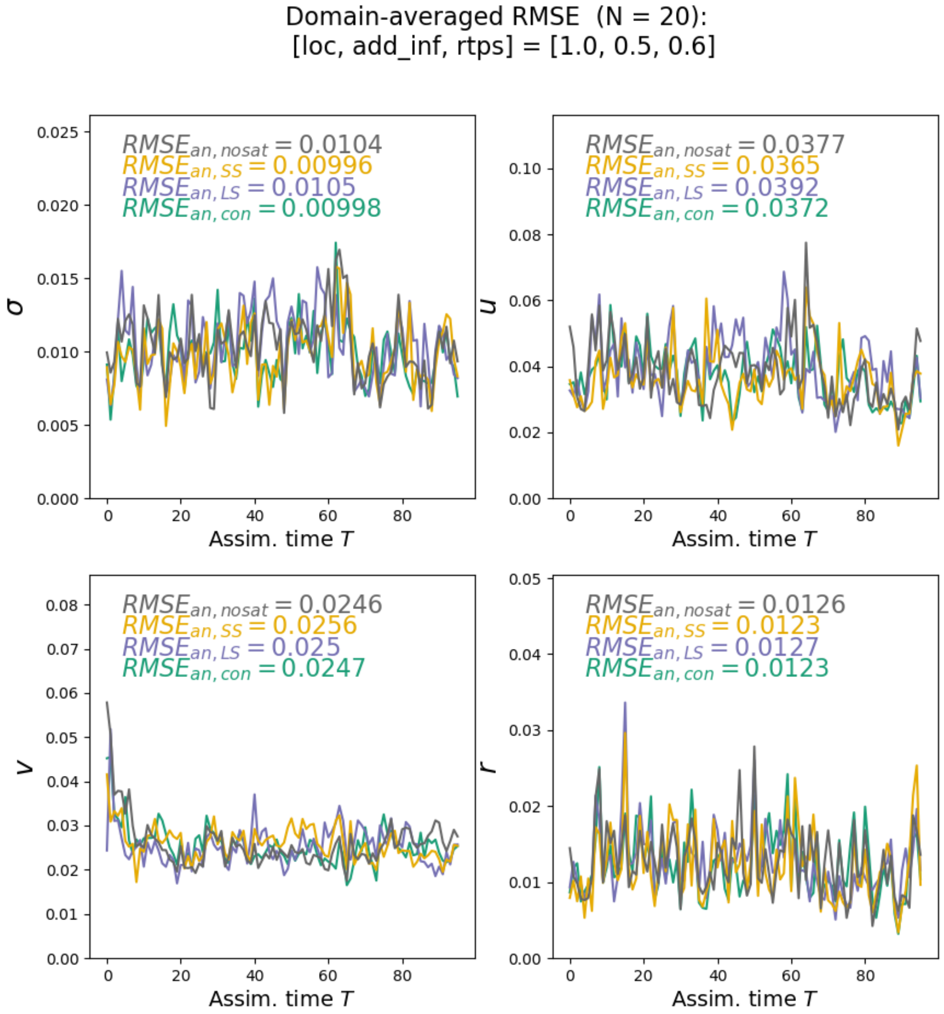


Figure 5.12: Time series of (domain-averaged) analysis RMSE for each model variable: σ (top left), u (top right), v (bottom left) and r (bottom right). All experiments listed in Table 5.5 are shown, together with a time average to facilitate their comparison. Note that there are no direct observations of σ, as I_{sat} is observed instead.

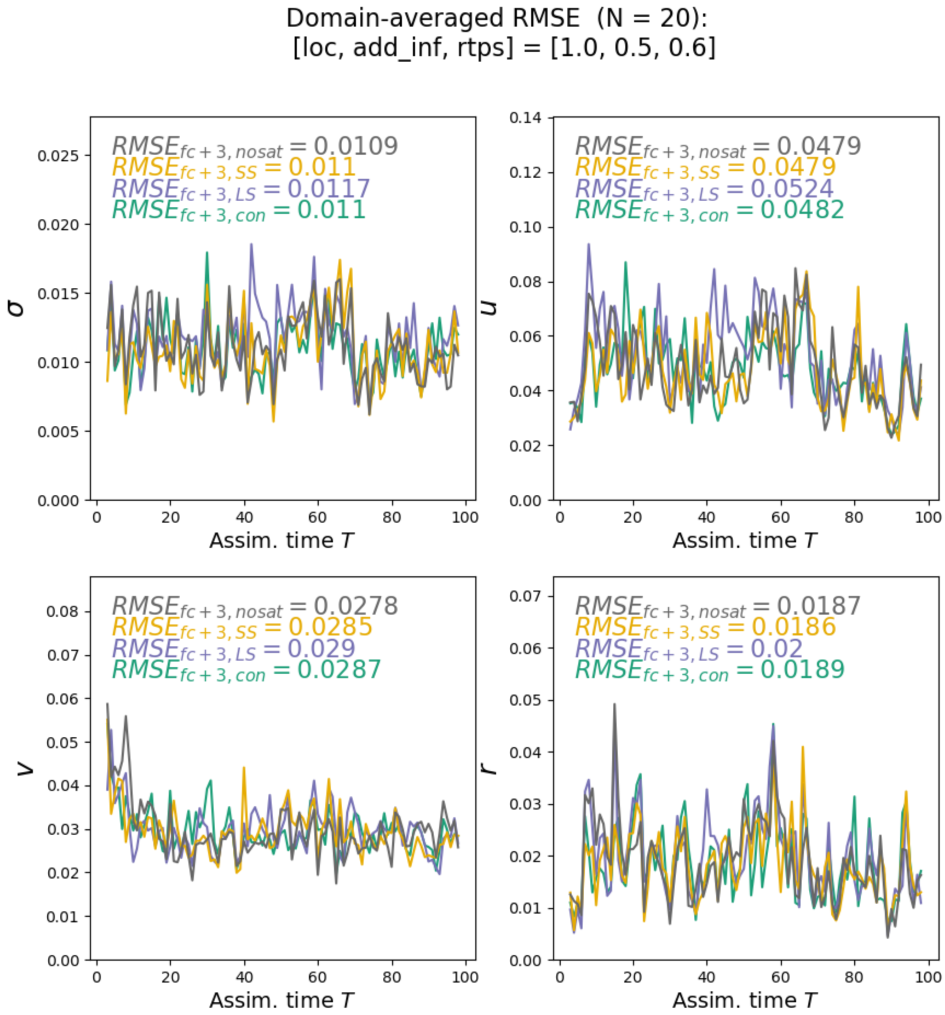


Figure 5.13: As in Figure 5.12, but for 3hrs forecasts.

results is not straightforward. However, similarly to what was observed in Fig. 5.12, the LS experiment underperforms the control and produces higher CRPS values by 3%-8% in all variables except v. Conversely, the SS experiment produces better results than the control (except in v), with a CRPS reduction in σ, u and r of 1%-5%; this experiments has also a lower CRPS than the other two experiments (both LS and 'no sat') in all variables but v.

The CRPS values for 3hrs forecasts are shown in Figure 5.15. Once again, the LS simulation displays the largest CRPS values (with the exception of v) and underperforms the control by 4%-11%, indicating that assimilating large scale satellite observations only may have a negative effect on the ensemble quality for this observational configuration. The results of the other experiments are less clear, although it is worth noting that the 'no sat' simulation produces the smallest CRPS values across all variables at this lead time, hinting that some adjustments to the observing system might be needed.

Rank histogram

Figure 5.16 shows the rank histograms of each model variable and each data denial experiment computed with the analysis ensemble and the value of the nature run sub-sampled at each observation location.

A rank histogram is a graphical tool used to assess the quality of an ensemble forecast. In particular, it is used to verify to what extent the verifying reference, usually observations or an independent analysis, fall within the range predicted by the ensemble members. An extensive discussion on the interpretation of rank histograms can be found in Hamill (2001). A brief description of how the rank histograms of figures 5.16-5.17 have been obtained follows.

In an operational system, comparing the output of an ensemble forecast against independent observations $\mathbf{y}^o$ or an independent analysis is a natural choice. In our idealised system, however, the 'true' state of the atmosphere is represented by the nature run simulation $\mathbf{x}^t$, and the observations are obtained by just adding a random perturbation to it. For this reason, in this work it seemed more appropriate to use the truth $\mathbf{x}^t$ (conveniently sub-sampled at the observation location) to generate a rank histogram. Therefore, for each observation, we have sub-sampled the true trajectory $\mathbf{x}^t$ at the same location and subsequently create a set of values by merging it with the $N = 20$ ensemble members valid at the same time and location. By sorting these values from

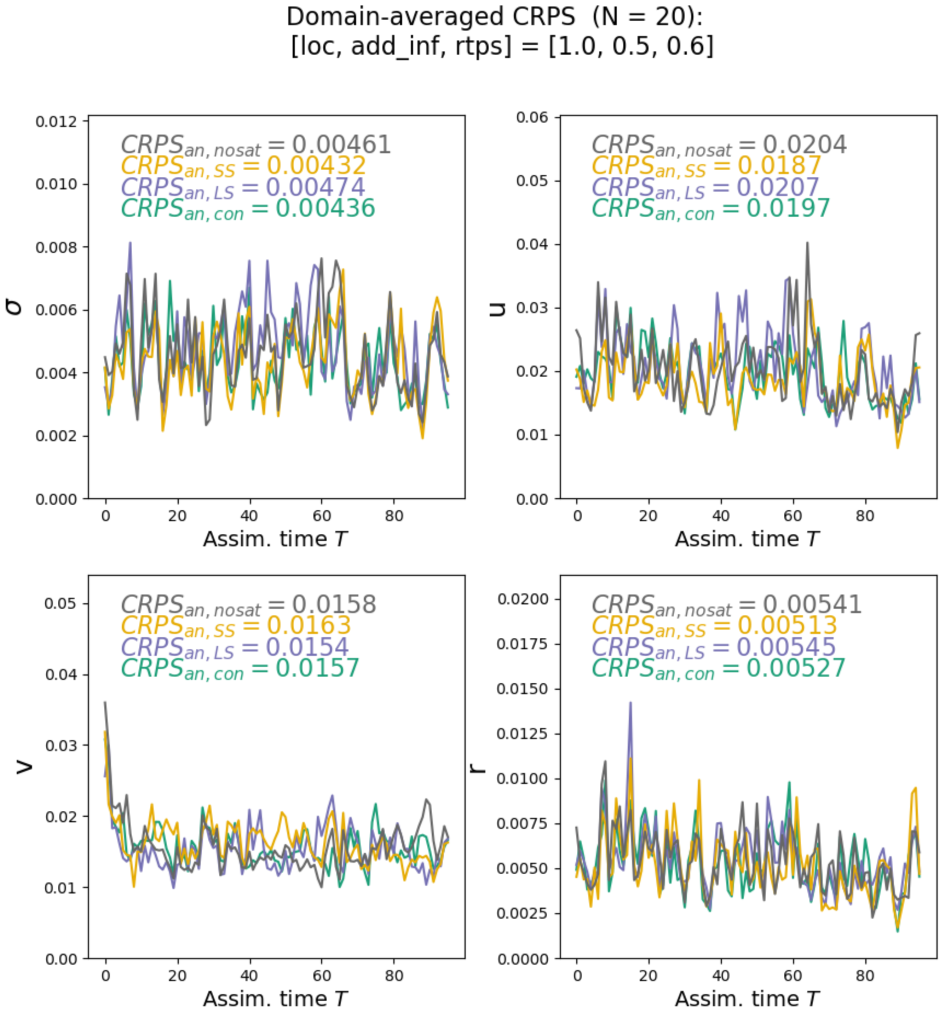


Figure 5.14: Time series of analysis CRPS for each model variable: σ (top left), u (top right), v (bottom left) and r (bottom right). All experiments listed in Table 5.5 are shown, together with time-averaged values included in the legends to facilitate their comparison.

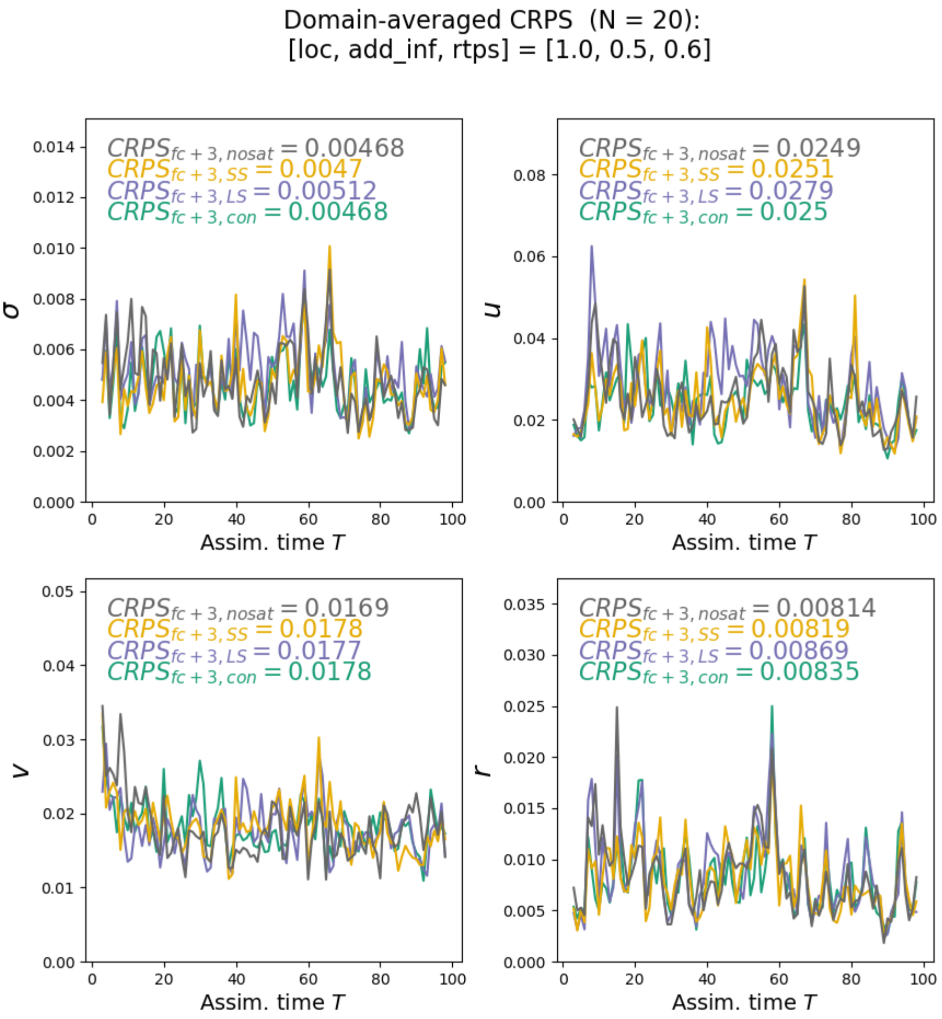


Figure 5.15: As in Figure 5.14, but for 3hrs forecasts.

the smallest to the largest, we could assign an integer (or rank) between 1 and $N+1$ to the sub-sampled truth, depending on the its position within the set. By repeating this procedure for all observations, it is possible to build a histogram with the distribution of the rank values obtained (cf. figures 5.16-5.17).

The principle behind a rank histogram is that if the observations and the ensemble members are drawn from the same distribution, then all the rank values between 1 and $N + 1$ are equally likely, and therefore the expectation for a good ensemble forecast is that the histogram should be flat (rank uniformity). In general, the shape of the histogram is descriptive of the performance (or problems) of an ensemble: for example, a U-shaped histogram is a sign that the ensemble has insufficient spread, as many observations fall outside its range of values; on the other hand, a bell-shaped histogram is associated with an ensemble with too much spread, in which most observations fall in the central region. Lastly, rank histograms that show an overpopulation of large (small) rank values are associated with an underforecasting (overforecasting) bias, as most observations fall in the upper (lower) region of the predicted values.

The pseudo-density σ is close to rank uniformity in all experiments, although a peak of low rank values in both the control and the LS simulation indicates a small positive bias (i.e. the forecast ensemble values are typically too large compared to the nature run); interestingly, this effect seems to be attenuated in the SS experiment. The zonal velocity u is close to rank uniformity in all experiments, whereas the meridional velocity v is visibly overconfident in all simulations (a result of the forecast ensemble being underspread, as already noted in Fig. 5.5). Finally, the rain variable r appears to suffer from some overdispersion effects in all experiments, with histograms peaking around the central rank values. A different behaviour can be detected in the 'no sat' simulation, which displays a positive bias, with the number of occurrences decreasing at high rank values.

Figure 5.17 displays the same diagnostics computed with 3hrs forecast ensembles. Overall, we observe a shift towards underdispersion in all variables, especially in σ and u. A tendency towards negative bias (i.e. forecasting values that are typically too small compared to the nature run) appears in r across all data denial experiments (LS, SS, no sat): a sign that the 3hrs ensemble systematically underestimates the nature run.

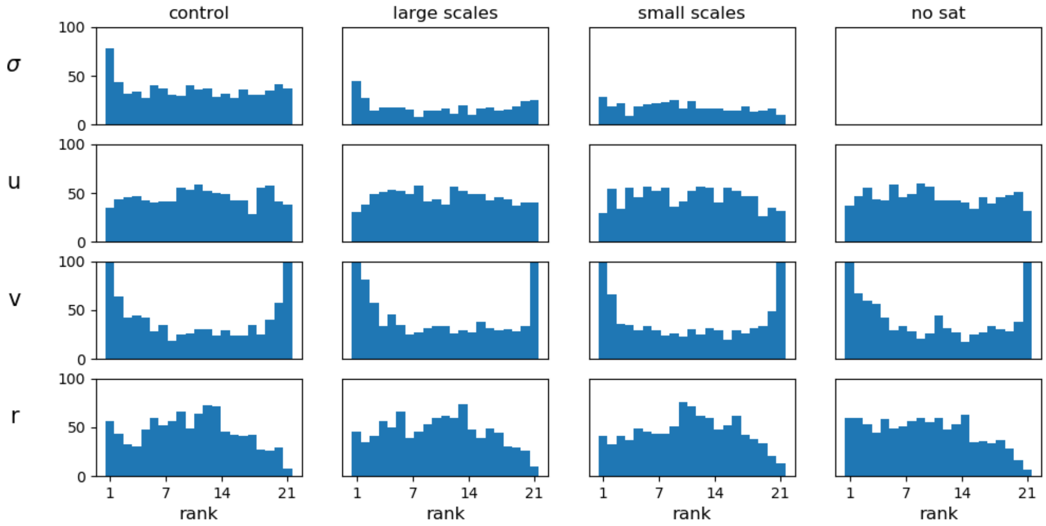


Figure 5.16: Rank histograms computed with the analysis ensemble for each data-denial experiment and for each model variable. The rank histogram for σ of the no sat experiments is blank as no satellite observations are assimilated. The interpretation of a rank histrogram is covered in Hamill (2001), while a description of how it is computed can be found in the text.

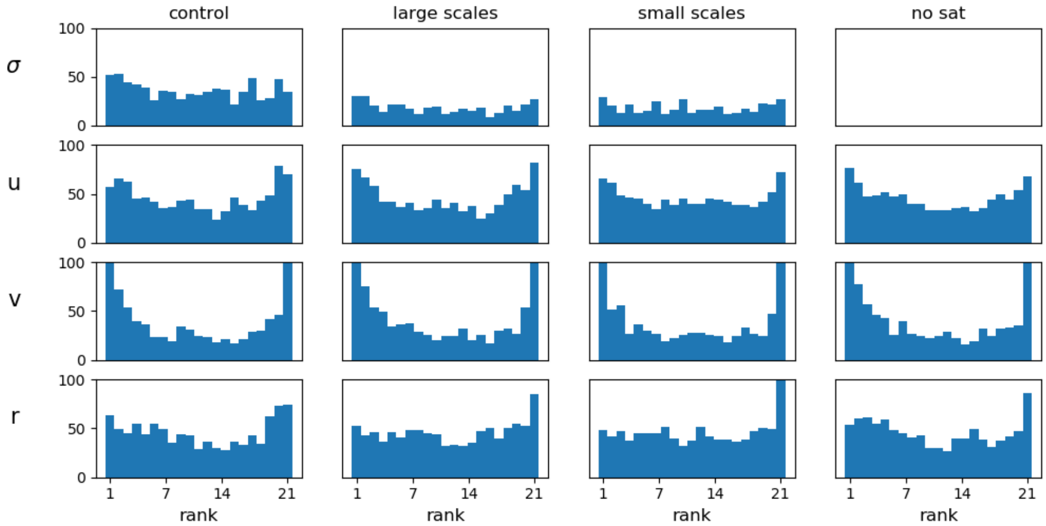


Figure 5.17: As in Fig. 5.16, but computed for 3hrs forecasts.

OID

Figure 5.18 shows the time series of the observation influence diagnostics (OID) of each type of observation in each data denial experiment. The impact of satellite observations is highest in the control simulation and is clearly zero in the 'no sat' experiment; moreover, the satellite observations appear to have a slightly larger impact in the small scale experiment than in the large scale one. All experiments show a succession of peaks and dips which we already observed in Fig. 5.6.

In addition, the 'no sat' experiment compensates the lack of satellite observations by using the other observations more, as proven by its larger OID values in u, v and r. The overall OID value of the 'no sat' experiments is approximately the same as the LS experiment (i.e. $OID_{nosat} = 25.8\%$, $OID_{LS} = 25.7\%$) whereas the other two experiments produce smaller impacts ($OID_{con} = 23.6\%$, $OID_{SS} = 23.8\%$). Finally, we note that the LS simulation has larger OID values than the SS one.

5.6.3 Discussion

The main research question of this thesis concerned whether there exists a relative benefit in assimilating satellite observations at small spatial scales rather than large scales, or vice versa. The results of the experiments presented above highlight only small differences between the various configurations and therefore provide only a preliminary answer. Nevertheless, both the single-observation and data denial experiments considered in this thesis represent an interesting starting point for further work and a more in-depth investigation.

Some evidence that limiting the assimilation to just large scale satellite observations may have a detrimental effect on the idealised forecast-assimilation system based on the ismodRSW model has emerged in the data denial experiments conducted so far. In particular, the accuracy of the ensemble mean and the quality of the ensemble both deteriorate with respect to the control simulation, at both the analysis time and after 3 hours from the initialisation, as the results for the RMSE and the CRPS reported in section §5.6.2 demonstrate. Moreover, the LS experiment tends to perform worse than the 'no sat' simulation, suggesting that, in the current setup, the assimilation of satellite observations of large spatial scales seem to have a net negative impact on the overall system.

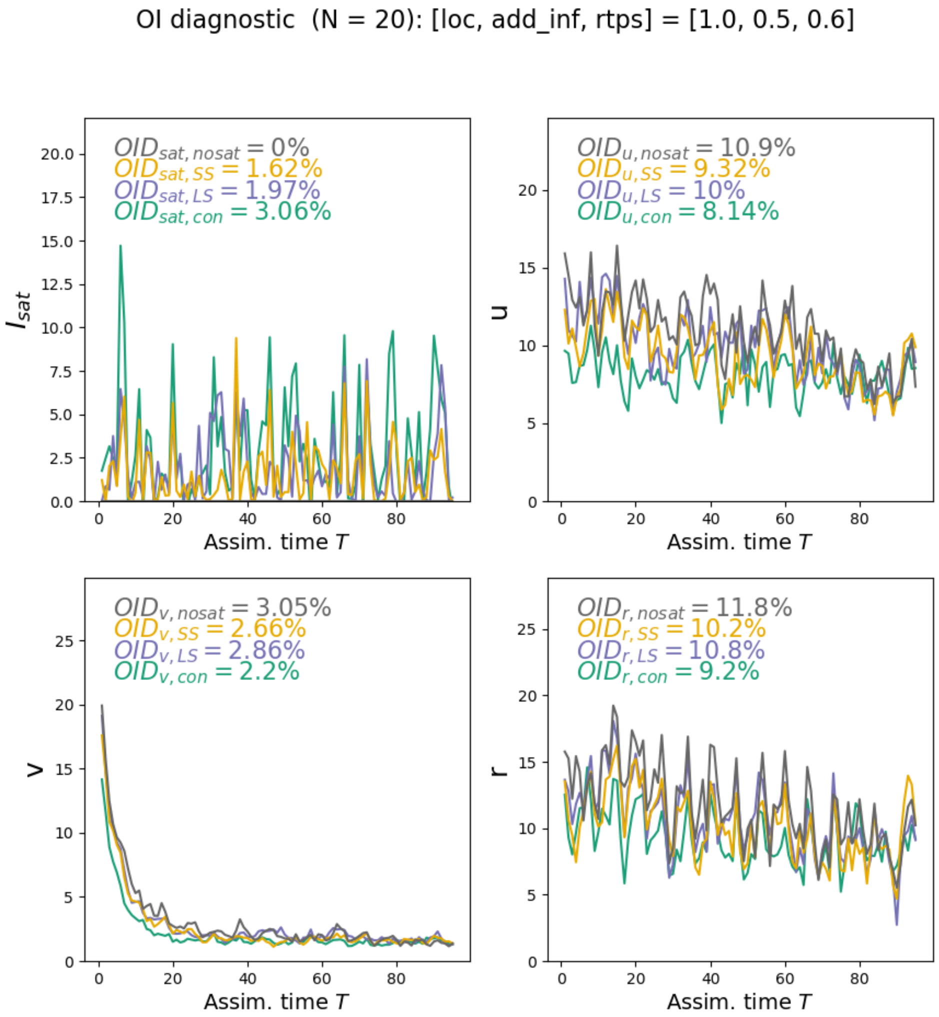


Figure 5.18: Time series of the observation influence diagnostics (OID) for each type of observation: I_{sat} (top left), u (top right), v (bottom left) and r (bottom right). All experiments listed in Table 5.5 are showed, together with time-averaged values included in the legends to facilitate their comparison.

At the same time, it is less clear whether assimilating only small scale satellite observations bring a benefit to the DA system, although there is some evidence supporting this hypothesis too. For example, the small scale-only simulation slightly overperforms the control in terms of both RMSE and CRPS at the analysis time in all variables except v, and displays a more uniform rank histogram in σ with respect to the control, indicating a better match between the forecast ensemble and the nature run. However, the results for the 3hrs forecasts are more neutral, with the SS experiment performing similarly to the control simulation. The potential advantage of the small scall satellite observations over the large scale ones seem to be related to their better ability in resolving narrow convection updrafts, as highlighted in the single-observations experiments presented section §5.6.1.

In all experiments, one interesting aspect is the role of the meridional velocity v, which behaves differently from the other variables: this is not surprising given the peculiar role of v as both observed and forced variable.

Overall, in order for these experiments to produce clearer results, some features of the observing system (such as the observation error) might need further adjustments and more tuning. Indeed, running additional data denial experiments for a wider range of configurations might help to determine the sensitivity and the robustness of our preliminary findings.

Finally, the OID results indicate a larger impact from all observation types when large scale satellite observations only are used, with respect to small scales only; this is a sign of the LS experiment somehow increasing the forecast uncertainty, driving a larger contribution of the observations to the final analysis.

Chapter highlights and summary

- At the start of this chapter, the data assimilation algorithm already used in Chapter 2 has been revised in order for it to include the assimilation of the psuedo satellite observations. The modifications concern the observing system, the observation operator and the way the ensemble-space localisation is performed.

- A new series of forecast-assimilation experiments has been carried out using the ismodRSW model, in which a combination of ground and satellite observations are assimilated. The protocol and the tuning process introduced in Chapter 2 has been re-applied here to the new configuration, and a well-tuned simulation has been selected at the end of the process.

- The initial research question of this thesis asked whether there is a relative benefit in focussing on the assimilation of satellite observations at small or large spatial scales. To address this question, a series of single-observation and data-denial experiments are used to assess the impact that the scale of the satellite observations have on the system. The results show some evidence of degradation of both the analysis and short-range forecasts when only large scale satellite observations are assimilated.

Chapter 6

Conclusions and future work

Satellites constitute one of the most important sources of observations in current NWP DA systems and their use is considered crucial in the day-to-day production of accurate weather forecasts. As operational centres move to extend their use beyond clear sky scenes, including in precipitating conditions, and the number of usable satellite observations increases, it is important to understand more about how to make best use of this growing amount of data.

In this thesis we set out to investigate the impact of satellite observations at different spatial scales on a DA system. In particular, we were interested in finding benefits in focussing on the assimilation of small scale rather than large scale observations (or vice versa), in view of improving future strategies for satellite DA research.

The work conducted in this thesis was based on an idealised model of the atmosphere using modified shallow water equations. On the one hand, the use of an idealised model has allowed us to avoid the logistical challenges posed by the complexity and the computational costs of state-of-the-art NWP models and their DA schemes. On the other hand, working with an idealised configuration has required us to put a considerable amount of effort in identifying – and subsequently imitating – the indispensable characteristics of real-world systems (including the properties of operational systems and those of satellite observations), so that our study would be relevant for NWP satellite DA research.

This final chapter has three purposes: summarising the content of the thesis, revisiting the objectives that were presented in the introduction (cf. §1.2) and concluding with an overview of future work which may stem from this study.

6.1 Summary

This thesis has originated from the idealised model of the atmosphere for convective-scale DA research based on modified shallow water equations described in Kent (2016) and Kent et al. (2017). Chapter 2 summarised the dynamics and the numerics of such a model (called modRSW, for modified Rotating Shallow Water), including the thresholds-based mechanisms used to mimic convection and precipitation. The model description was followed by an illustration of the so-called 'twin-setting' experiments, which made it possible to conduct DA experiments in an idealised setup. In this configuration, a single high-resolution model integration (referred to as nature run simulation) was used to generate a set of observations, which were subsequently combined in a Deterministic Ensemble Kalman filter (Sakov & Oke, 2008) with en ensemble of forecasts obtained by running the same model at a lower resolution.
In the final part of Chapter 2, we presented a protocol to assess the performance and the relevance of DA experiments conducted with an idealised model, which we subsequently applied to a series of forecast-assimilation simulations performed with the modRSW model. Hence, we showed that the setup described in Chapter 2 is both well-tuned and relevant for convective-scale NWP DA research, with a summary of the results reported in Table 2.2.

The modRSW model was initially considered as our test-bed for investigating satellite DA. Unfortunately, it displayed some limitations (such as the lack of a built-in fluid temperature definition) which we have outlined at the beginning of Chapter 3. A revised model was presented thereafter, still based on shallow water equations, but altered in two respects: the replacement of an isopycnal fluid (i.e. constant density) with an isentropic one (i.e. constant potential temperature), and the addition of a second motionless layer capped by a rigid lid. Next, this new, isentropic $1\frac{1}{2}$-layer model (in the absence of convection and precipitation) was manipulated to be put in conservative form in order to facilitate its numerical implementation (which is described in Appendix C), and the observations-based scaling at the basis of its mathematical derivation (not included in this work) is discussed. In addition, both the analytical and the numerical consistency of the model were checked: first, the isopycnal model

was recovered by imposing $\kappa = R/c_p = 1$; second, the numerical model was verified against an independent analytical solution based on the stationary wave solutions discussed in Shrira (1981, 1986).
The final part of Chapter 3 presented the fully modified ismodRSW (isentropic modified Rotating Shallow Water) model (which includes convection and precipitation) and illustrated a prototype nature run simulation based on the scaling discussed earlier in the chapter, which included rotation and was forced by a relaxation solution in the meridional velocity v_{rel}. As a final check, the covergence of the full ismodRSW model with $\kappa = 1$ towards the modRSW model was verified.

Chapter 4 started with a brief summary of basic principles of radiative transfer and atmospheric radiation, followed by a discussion on how they are exploited by meteorological satellites. In addition, a short historical overview of satellite DA was presented, from the first missions in the 1960s to the most recent developments regarding the assimilation of clouds and precipitation.
Later in the chapter, the actual generation of pseudo satellite observations was discussed, with the aim to mimicking the measurements of passive microwave radiation carried out by polar-orbiting satellites. This aim was achieved by: developing a simple radiative transfer model based on the Rayleigh-Jeans law, generating spatially varying observations that travel along the domain and re-enter it periodically, and defining the measured radiance as a horizontally weighted average in order to imitate the FOV of satellite observations. In addition, clouds were implicitly modelled by exploiting the built-in system of thresholds and by taking into account their impact (as well as the effect of precipitation) on the measured microwave radiation.

Chapter 5 started with a description of the modifications required for the DA scheme outlined in Chapter 2 to deal with the assimilation of the pseudo satellite observations described in Chapter 4. These adjustments included a revision of the observing system, the new (nonlinear) observation operator, the implementation of the the modulated ensemble technique presented in Bishop et al. (2017) to preserve the background-space localisation, and the new model-error covariance matrix $\mathbf{Q}$.
Next, the chapter revisited the protocol presented in Chapter 2 (slightly modified to take into account the revised observing system, such as the ratio between conventional and satellite observations) and discussed the tuning of a series of new forecast-assimilation experiments from which a well-tuned simulation emerged. The new setup

was also validated to verify its relevance in the context of NWP DA systems and is subsequently used as a control simulation in a series of data denial experiments which try to address the initial research question of this thesis. To this end, the control experiment was compared against three other experiments which differed only in the number and the resolution (i.e. the FOV) of the satellite observations assimilated: small scale only, large scale only and no satellite observations. In addition, a series of short-range single-observation experiments are conducted, in which only one pseudo satellite observations of various FOVs is assimilated, and the analysis increments produced by each of them subsequently compared.

The single-observations experiments showed the importance of small scale observations in resolving correctly the narrowest convection updrafts. The data denial experiments produced some evidence of a degradation caused by the assimilation of large scale satellite observations in both the analysis and a 3 hour forecast, whereas the assimilation at the small scales produced a small positive benefit to the analysis. In general, the differences between the different configurations were small and we considered these findings as preliminary. Further tuning of the control simulation and additional experiments might be needed in order to confirm the robustness of our results.

6.2 Objectives revisited

In Chapter 1, we set out a list of objectives to orientate the research conducted in this thesis. Here, we revisit each of them, discussing if (and how) they have been met.

1. *Show that it is possible to obtain a well-tuned idealised experiment with the modRSW model which is also relevant for operational NWP DA research and formulate a protocol to assess other idealised systems in the same way.*
   The forecast-assimilation experiments performed with the modRSW model were discussed in sections §2.4-2.5 and the results were summarised in Table 2.2, where the criteria to assess the tuning and the relevance of an experiment for convective-scale NWP DA were established. After the systematic comparison of a large number of simulations, we selected a well-tuned experiment (with localisation and inflation parameters: $L_{loc} = 1.0$, $\alpha_{RTPS} = 0.7$ and $\gamma_a = 0.15$) that satisfied all the requirements: a $SPR/RMSE$ ratio near 1, minimum CRPS and RMSE (of the ensemble mean), an observation influence around 30% and an error doubling time of about 6-9 hours. The process described in Chapter 2 can

be re-applied to a different idealised systems, as we did in Chapter 5, where we adopted the same procedure for the experiments performed with the ismodRSW model.

2. *Show that the modRSW model needs to be modified for satellite DA research, and then develop a new, revised version of the model that is fit for purpose.*
   We achieved both these aims in Chapter 3. First (section §3.1), we showed that the modRSW model was not suitable for satellite DA since the lack of an in-built definition of fluid temperature (which is essential for the formulation of any radiative transfer model) led to an unrealistic scaling when a simpler temperature definition was adopted. Afterwards, the ismodRSW model was developed, on the basis of two modifications which made it a much more suitable option for satellite DA: the replacement of the isopycnal fluid with an isentropic one (which ensured a physically consistent temperature definition, cf. (3.3)), and the addition of a passive second layer, which increased the vertical complexity of the model. The ismodRSW model adopted the same thresholds mechanism used in the modRSW model to imitate convection and precipitation.

3. *Construct a complex (and idealised) observing system, comprising both simulated ground and satellite observations, in which the characteristics of real-world observing systems and satellite observations are imitated closely, and include the effect of clouds.*
   Sections §5.1.1 described the new idealised observing system, which included both ground observations at fixed locations of fluid velocities and rain (u, v and r) and spatially varying satellite observations ($I_{sat}(\sigma)$) recreating passive microwave radiation measured by polar-orbiting satellites (widely assimilated at most operational centres). The way these satellite observations were generated was described thoroughly in Chapter 4, where we also modelled the impact that clouds and precipitation have on these observations, exploiting the in-built system of threshold for convection and rain (cf. §4.4.4). In order to make the observing system more realistic, the ratio of satellite to conventional observations assimilated in the experiments of Chapter 5 reflected the ratio of 0.27 found in the UKV model run at the Met Office (Milan et al., 2020).

4. *Conduct new forecast-assimilation experiments in which satellite observations are assimilated that are relevant for operational NWP systems.*

A series of new forecast-assimilation experiments which include the use of pseudo satellite observations is outlined in section §5.2. Table 5.3 summarises the criteria utilised to assess the performance and the relevance of the experiment selected at the end of the tuning process (with localisation and inflation parameters: $L_{loc} = 1.0$, $\alpha_{RTPS} = 0.6$ and $\gamma_a = 0.5$). The results show a $SPR/RMSE$ ratio near 1, minimum CRPS and RMSE, an observation influence around 25% and an error doubling time of about 5-17 hours.

5. *Perform a series of data denial experiments – i.e. simulations in which a portion of the observations are intentionally excluded from the assimilation – to investigate the impact of satellite observations at different spatial scales.*
   The data denial experiments are discussed in section §5.6. The experiment selected at the end of the tuning process in objective 4 is used as the control and three more configurations are tested: one with no satellite observations, one that assimilates only small scale satellite observations, and one that assimilates only large scale satellite observations. The results showed a degradation of the experiment which assimilated only large scale satellite observations, with an increase in both RMSE and CRPS with respect to the control, both at the analysis time and at a lead time of 3 hours. Conversely, the experiment in which only small scale satellite observations were used produced marginal improvements against the control, and mostly at the analysis time. Overall, further tuning of the control simulation and additional experiments might be needed to establish the robustness of the current findings.

6.3 Future work

As outlined at the end of Chapter 5, in order to confirm the validity of the results of the data denial experiments reported therein, additional simulations should be conducted in the future, following further tuning of the DA scheme combined with adjustments to the observing system. A list of possible modifications to the current configuration comprises:

- varying the number of ensemble members;
- varying the observation error of any of the observed quantities;

- varying the number and spacing of ground and/or satellite observations;
- revisiting the role of the meridional velocity v as both an observed and forced variable;
- modifying the observation generator $\mathcal{G}_s$ to allow for hourly cycles with no satellite observations assimilated, akin to real systems (cf. Fig. 11 in Milan et al. (2020));
- redefining the functions $\alpha_k(\sigma)$ in the measured radiance expression I_{sat} – eq. (4.37) – in terms of simpler analytical functions, resulting in an expression that is ideally linear in the coefficients.

Moreover, in order to obtain more robust results in the future, it might be necessary to repeat the same set of simulations more than once (for example by generating a new set of observations, or with different random perturbations), to avoid the risk of drawing conclusions from possible random effects or fluctuations.

The flexibility of an idealised model also opens up a series of possibilities which go beyond the scope of this thesis. Among them, it is worth mentioning:

- **Explore alternative types of satellite observations**. In this thesis, we have chosen to focus on microwave radiation measured by polar-orbiting satellites, but as we have seen in section §4.1.4, many other types of satellite observations are currently assimilated at operational centres. In this regard, for example, geostationary satellites (with their constant view over the same portion of the Earth) constitute a more reliable (and continuous) source of observations for limited area models such as the UKV. Therefore, by redefining the observation generator function $\mathcal{G}_s$, it is possible to imitate geostationary observations and subsequently investigate their impact on the DA system. Incidentally, as most instruments mounted on geostationary satellites measure infrared radiation, a new (and nonlinear) radiative transfer model would also be needed, as the Rayleigh-Jeans law is not valid for infrared radiation.
- **Explore more complex observation error**. Since radiances started to be directly assimilated into operational DA systems, a lot of research has been done on how to model their observation errors. In particular, it has become common practice to inflate the errors of the radiance observations that are believed to be

contaminated by clouds and precipitations. This approach was not included in this thesis, but can represent a possible way to increase the level of complexity and explore the impact of a widely adopted solution on our idealised system.

- **Explore bias correction algorithms**. Bias correction has been an important area of research in satellite DA (Dee, 2004; Derber & Wu, 1998), as both instruments and radiative transfer models can produce systematic errors. Since most DA algorithms are based on the assumption that the observations are unbiased, correcting for this type of error is crucial in order to obtain reliable analysis estimates. Biases can therefore artificially added to the idealised satellite observations and the DA scheme conveniently adapted in order to investigate their impact on the system or to explore other solutions to correct for biased observations.

- **Adding topography to the ismodRSW model**. The ismodRSW model presented in Chapter 3 and used in the experiments of Chapter 5 did not include topography. Adding topography to the system would expand the number of dynamical configurations that can be used as test-bed in future DA experiments. In order to include topography, the numerical scheme needs to be revisited, as an additional term appears in the momentum equation (3.15b), stemming from the full definition of M_2 (cf. 3.5d); moreover, the well-balancedness of the model (i.e. the preservation of the rest state) needs to be investigated, as this issue was present in the zero order Discontinuous Galerkin scheme used in Kent (2016) and Kent et al. (2017), as further discussed in Kent & Bokhove (2020), and led to the adoption of the method proposed by Audusse et al. (2004).

Appendix A

Equivalence of DEnKF with ‘no-perturbation’ EnKF with RTPP $\alpha_{RTPP} = 0.5$

Note that this appendix is present in a similar form in Kent et al. (2020) as Appendix B.

In this appendix we show that the EnKF without perturbed observations (i.e. ‘no-perturbation’ EnKF) together with an adaptive RTPP inflation with $\alpha_{RTPP} = 0.5$ (cf. (2.26)) is formally equivalent to the Deterministic Ensemble Kalman Filter (DEnKF) developed by Sakov & Oke (2008), even when using self-exclusion (cf. section §2.2.3). This configuration is currently in use at the Met Office within the global ensemble forecast system (i.e. MOGREPS-G, cf. Bowler et al. (2017)).
In order to prove this, we note that the analysis step of the ‘no-perturbation’ EnKF reads (by taking the self-exclusion into account):

$$\mathbf{x}_j^a = \left(\mathbf{I} - \mathbf{K}_{e,\widehat{j}}\mathbf{H}\right)\mathbf{x}_j^f + \mathbf{K}_{e,\widehat{j}}\mathbf{y}^o, \tag{A.1}$$

with the ensemble mean $\overline{\mathbf{x}}^a$ being:

$$\overline{\mathbf{x}}^a = \left(\mathbf{I} - \mathbf{K}_{e,\widehat{j}}\mathbf{H}\right)\overline{\mathbf{x}}^f + \mathbf{K}_{e,\widehat{j}}\mathbf{y}^o. \tag{A.2}$$

The j^{th} column of the analysis perturbation matrix $(\mathbf{X}^a)_{\widehat{j}}$ can therefore be expressed as:

$$(\mathbf{X}^a)_{\widehat{j}} = \mathbf{x}^a_j - \overline{\mathbf{x}}^a = \left(\mathbf{I} - \mathbf{K}_{e,\widehat{j}}\mathbf{H}\right)(\mathbf{X}^f)_{\widehat{j}}, \tag{A.3}$$

in which $\mathbf{X}^f$ is the forecast perturbation matrix. The RTPP equation (2.26) together with the above result yields:

$$(\mathbf{X}^a)_{\widehat{j}} = (1 - \alpha_{RTPP})\left(\mathbf{I} - \mathbf{K}_{e,\widehat{j}}\mathbf{H}\right)(\mathbf{X}^f_{\widehat{j}}) + \alpha_{RTPP}(\mathbf{X}^f_{\widehat{j}}). \tag{A.4}$$

For $\alpha_{RTPP} = \frac{1}{2}$, we obtain:

$$(\mathbf{X}^a)_{\widehat{j}} = (\mathbf{X}^f)_{\widehat{j}} - \frac{1}{2}\mathbf{K}_{e,\widehat{j}}\mathbf{H}(\mathbf{X}^f)_{\widehat{j}}, \tag{A.5}$$

which is Eq. (15) in Sakov & Oke (2008), generalized to include self-exclusion.

Appendix B

The Gaspari-Cohn taper function

The localisation matrix $\boldsymbol{\rho}$ described in section §2.2.3 is a symmetric matrix with entries $\rho_{i,j}$ based on the Gaspari-Cohn taper function (Gaspari & Cohn, 1999), that is:

$$\rho(z,c) = \begin{cases} f_1(z/c) & \text{for } 0 \leq z \leq c; \\ f_2(z/c) & \text{for } c \leq z \leq 2c; \\ 0 & \text{for } 2c \leq z; \end{cases} \tag{B.1.a}$$

where:

$$\begin{aligned} f_1(z/c) = &-\frac{1}{4}\left(\frac{z}{c}\right)^5 + \frac{1}{2}\left(\frac{z}{c}\right)^4 + \frac{5}{8}\left(\frac{z}{c}\right)^3 \\ &- \frac{5}{3}\left(\frac{z}{c}\right)^2 + 1, \end{aligned} \tag{B.1.b}$$

$$\begin{aligned} f_2(z/c) = &\frac{1}{12}\left(\frac{z}{c}\right)^5 - \frac{1}{2}\left(\frac{z}{c}\right)^4 + \frac{5}{8}\left(\frac{z}{c}\right)^3 \\ &+ \frac{5}{3}\left(\frac{z}{c}\right)^2 - 5\left(\frac{z}{c}\right) + 4 - \frac{2}{3}\left(\frac{z}{c}\right)^{-1}, \end{aligned} \tag{B.1.c}$$

in which z represents the distance between 2 grid points, and c is a length-scale that determines the amount of localisation. The function $\rho(z,c)$ is similar to a half-Gaussian function (cf. Fig. 2.4), which has a maximum at $\rho(0) = 1$ and subsequently decreases as z increases, with values that go to zero beyond a 'cut-off' distance of twice the characteristic length-scale c.

In the case of the experiments described in Chapter 2, c is a function of the (dimensionless) localisation parameter L_{loc}, that is: $c = N_{el}\Delta x/(2L_{loc})$, with Δx being the grid spacing. As a result, and according to (B.1), the value of $\rho(z,c)$ drops to zero beyond $z = 2c = N_{el}\Delta x / L_{loc}$. For example, a localisation scale of $L_{loc} = 1.0$ translates into a cut-off distance $2c$ equal to the scale of the horizontal domain (i.e. 500 km), beyond which $\rho = 0$ and thus every correlation is suppressed.

Appendix C

Numerics of the ismodRSW

Note that this appendix appears also in Cantarello et al. (2020) as Appendix B.

In this appendix we summarize the numerical methods utilised to integrate the *ismodRSW* model. Despite some modifications, the scheme has remained the same used in Kent (2016); Kent et al. (2017), which we refer the reader to for a more comprehensive and satisfactory description.

C.1 Classic shallow water

To integrate numerically the model in absence of convection and rain, a 0-degree discretization of the Discontinuous Galerkin Finite Element Method (DGFEM) developed by Rhebergen et al. (2008) is used (equivalent to a finite volume method), in combination with HLL fluxes (Harten et al., 1983). As we saw in section §3.3.3, the shallow water model is hyperbolic and can be put in conservative form.
We split the horizontal domain $[0, L]$ into N_{el} open elements $K_k = (x_k, x_{k+1})$ of constant length $|K_k| = x_{k+1} - x_k$ with $k = 1, 2, ..., N_{el}$, delimited by $N_{el} + 1$ nodes where $0 = x_1, x_2, ..., x_N, x_{N_{el}+1} = L$. Therefore, we derive the weak formulation of equation (3.14) (see §3.1.2 of Kent (2016) and more extensively Zienkiewicz et al. (2014)). The zero-degree discretization (henceforth DG-0) implies that the function $\mathbf{U}$ in (3.14) is

approximated with a piece-wise constant function within each element K_k such as:

$$\mathbf{U}_h(x,t) = \overline{U}_k = \frac{1}{|K_k|}\int_{K_k} \mathbf{U}(x,t)dx. \tag{C.1}$$

In the end, the DG-0 discretization for each element $|K_k|$ reads:

$$\frac{\mathrm{d}\overline{U}_k}{\mathrm{d}t} + \frac{\mathcal{F}_{k+1} - \mathcal{F}_k}{|K_k|} + T(\overline{U}_k) = 0, \tag{C.2}$$

where $\mathcal{F}_k$ represents the numerical flux computed at each element's node, which in the case of the HLL fluxes is defined as:

$$\mathcal{F}_k = \begin{cases} F^L & \text{if } S^L > 0, \\ F^R & \text{if } S^R < 0, \\ F^{HLL} & \text{if } S^L < 0 < S^R, \end{cases} \tag{C.3}$$

with F^{HLL} defined as:

$$F^{HLL} = \frac{F^L\,S^R - F^R\,S^L + S^L\,S^R\,(U^R - U^L)}{S^R - S^L}, \tag{C.4}$$

in which S^L and S^R are the numerical velocities arising from the eigenvalues in Eq. (3.22):

$$S^L = \min\left(u_k - \sqrt{\partial_\sigma \mathcal{E}_k},\; u_{k+1} - \sqrt{\partial_\sigma \mathcal{E}_{k+1}}\right); \tag{C.5.a}$$

$$S^R = \max\left(u_k + \sqrt{\partial_\sigma \mathcal{E}_k},\; u_{k+1} + \sqrt{\partial_\sigma \mathcal{E}_{k+1}}\right). \tag{C.5.b}$$

It is worth noticing that in order to compute the fluxes $\mathcal{F}$ and the numerical velocities S^L, S^R, the non-dimensional pressure η needs to be calculated at each time step from the pseudo-density σ. However, since inverting analytically eq. (3.11) is not possible, an alternative which is efficient enough needs to be found. Here we chose to precompute the corresponding values of σ and η (with a resolution of $\delta\sigma = 0.0001$) and to perform a linear interpolation during the model integration. A polynomial interpolation is also possible.

We refer to Kent (2016) for the (adaptive) time step implementation.

C.2 NCPs for the full model

We noted in section §3.4 that the model in its full form – comprising convection and rain – cannot be put in conservative form, and that non-conservative products arise

and need to be handled numerically (see (3.42)). To achieve this aim, notwithstanding the DGFEM method outlined above, the DLM theory introduced by Dal Maso et al. (1995) is used. Again, more details about the mathematical formulation of this scheme can be found in Kent (2016) and Kent et al. (2017), although in this case (because of the missing topography term) there is no need to apply the theory about state reconstruction used therein and the scheme developed by Audusse et al. (2004).
The following semi-discrete space-DGFEM scheme for a single element K_k is found:

$$\frac{\mathrm{d}\overline{U}_k}{\mathrm{d}t} + \frac{\mathcal{P}^p_{k+1} - \mathcal{P}^m_k}{|K_k|} + S(\overline{U}_k) = 0, \tag{C.6}$$

in which the numerical fluxes $\mathcal{P}^p_i$ and $\mathcal{P}^m_i$ read as:

$$\mathcal{P}^p_i = \mathcal{P}^{NC}_i + \frac{1}{2}\int_0^1 G_{ij}(\boldsymbol{\phi})\frac{\partial \phi_j}{\partial \tau}d\tau, \tag{C.7}$$

$$\mathcal{P}^m_i = \mathcal{P}^{NC}_i - \frac{1}{2}\int_0^1 G_{ij}(\boldsymbol{\phi})\frac{\partial \phi_j}{\partial \tau}d\tau. \tag{C.8}$$

In the expressions above, $G_{ij}(\boldsymbol{\phi})$ indicates the $\{i,j\}$ matrix element of the NCP $\mathbf{G}$ matrix of (3.43), with $\boldsymbol{\phi}$ being a Lipschitz continuous path connecting the left and the right state: $\boldsymbol{\phi}(\tau;\mathbf{U}^L,\mathbf{U}^R) = \mathbf{U}^L + \tau\left(\mathbf{U}^R - \mathbf{U}^L\right)$, as per the DLM theory. Moreover, the NCP flux contributions $\mathcal{P}^{NC}_i$ read as:

$$\mathcal{P}^{NC}_i(\overline{U}^L_i,\overline{U}^R_i) = \begin{cases} F^L_i - \frac{1}{2}V^{NC}_i, & \text{if } S^L > 0; \\ F^{HLL}_i - \frac{1}{2}\frac{S^L+S^R}{S^R-S^L}V^{NC}_i, & \text{if } S^L < 0 < S^R; \\ F^R_i + \frac{1}{2}V^{NC}_i, & \text{if } S^R < 0. \end{cases} \tag{C.9}$$

with the numerical velocities S^L and S^R (cf. eq. (3.44)) being:

$$S^L = \min\left(u^L - \sqrt{\partial_\sigma \mathcal{E}|^L + c_0^2\widetilde{\beta}|^L}, u^R - \sqrt{\partial_\sigma \mathcal{E}|^R + c_0^2\widetilde{\beta}|^R}\right), \tag{C.10}$$

$$S^R = \max\left(u^L + \sqrt{\partial_\sigma \mathcal{E}|^L + c_0^2\widetilde{\beta}|^L}, u^R + \sqrt{\partial_\sigma \mathcal{E}|^R + c_0^2\widetilde{\beta}|^R}\right). \tag{C.11}$$

$\mathbf{V}^{NC}$ is a vector containing the worked out contribution of the NCP integral expressions $\int_0^1 G_{ij}(\boldsymbol{\phi})\frac{\partial \phi_j}{\partial \tau}d\tau$:

$$\mathbf{V}^{NC} = \begin{bmatrix} 0 \\ -c_0^2[\![r]\!]\{\!\{\sigma\}\!\} \\ 0 \\ -\beta[\![u]\!]\Theta([\![u]\!])\left(\sigma^R I_\beta + [\![\sigma]\!]I_{\tau\beta}\right) \end{bmatrix}, \tag{C.12}$$

where: $\Theta(\cdot)$ indicates the Heaviside function, $[\![\cdot]\!] = (\cdot)^L - (\cdot)^R$ the jump across the node and $\{\!\{\cdot\}\!\} = \frac{1}{2}\left((\cdot)^L + (\cdot)^R\right)$ the average quantity. I_β and $I_{\tau\beta}$ are still expressed by

(C18) and (C22) as per appendix C of Kent et al. (2017), with analogous derivation upon the redefinition of X and Y as $X = \sigma^R - \sigma^L$ and $Y = \sigma^L - \sigma_r$. The derivation of the elements in $\mathbf{V}^{NC}$ is also analogous to the one performed in the same appendix, by replacing all references to z and h with σ.

Appendix D

Square root of localisation matrix

This appendix summarises the method used to compute the square root of the localisation matrix $\boldsymbol{\rho}$, that is, the matrix $\mathbf{W}$ such that:

$$\boldsymbol{\rho} = \mathbf{W}\mathbf{W}^T. \tag{D.1}$$

The procedure described here is a formalisation of the coded algorithm found in Jeff Whitaker's Github repository[1] and is based on a low-rank approximation of the spectral theorem.

First, the eigenvalues and the eigenvectors of $\boldsymbol{\rho}$ are computed:

$$\lambda_1, \lambda_2, \lambda_3, ..., \lambda_N \qquad \text{eigenvalues of } \boldsymbol{\rho}; \tag{D.2}$$

$$v_1, v_2, v_3, ..., v_N \qquad \text{eigenvectors of } \boldsymbol{\rho}. \tag{D.3}$$

Subsequently, the negative eigenvalues of $\boldsymbol{\rho}$ are set to zero and all eigenvalues are summed together, giving the matrix trace:

$$\Lambda = \sum_{i=1}^{N} \lambda_i. \tag{D.4}$$

In order to exclude very small eigenvalues for numerical reasons, we order the eigenvalues in descending order and consider the first K, such that their sum approximates Λ, that is:

$$S = \sum_{i=1}^{K} \lambda_i \geq a \cdot \Lambda, \tag{D.5}$$

---

[1] URL: https://github.com/jswhit/L96/blob/master/L96ensrf.py

with a a factor close to 1 (for example, one can take $a = 0.99$). Finally, the square root matrix $\mathbf{W}$ is built by defining the column vectors w_i, that is:

$$w_i = v_i \cdot \sqrt{\frac{\lambda_i}{S}} \quad i = 1, 2, ..., K. \tag{D.6}$$

References

Akvilonova, A.B., Basharinov, A.Y., Gorodetskiy, A., Gurvich, A.S., Krylova, M.S., Kutuza, B.G., Matveyev, D.T. & Orlov, A. (1973). Cloud parameters measured from the 'Cosmos-384' satellite. Izvestiya, Atmospheric and Oceanic Physics, **9**, 187–189. 136

Allison, J.L. & Neil, A.E. (1962). Final report on the TIROS I meteorological satellite system. Tech. rep., National Aeronautics and Space Administration, available at `https://ntrs.nasa.gov/citations/19640007992`. 122

Andersson, E., Pailleux, J., Thépaut, J.N., Eyre, J.R., McNally, A.P., Kelly, G.A. & Courtier, P. (1994). Use of cloud-cleared radiances in three/four-dimensional variational data assimilation. Quarterly Journal of the Royal Meteorological Society, **120**, 627–653. 125

Andrews, P.L. & Bell, R.S. (1998). Optimizing the united kingdom meteorological office data assimilation for ers-1 scatterometer winds. Monthly Weather Review, **126**, 736–746. 127

Audusse, E., Bouchut, F., Bristeau, M.O., Klein, R. & Perthame, B. (2004). A fast and stable well-balanced scheme with hydrostatic reconstruction for shallow water flows. SIAM Journal on Scientific Computing, **25**, 2050–2065. 35, 196, 203

Ballard, S.P., Macpherson, B., Li, Z., Simonin, D., Caron, J.F., Buttery, H., Charlton-Perez, C., Gaussiat, N., Hawkness-Smith, L., Piccolo, C., Kelly, G., Tubbs, R., Dow, G. & Renshaw, R. (2012). Convective scale data assimilation and nowcasting. In Proceedings, Seminar on Data Assimilation for Atmosphere and Ocean, 265–300. 53

Bannister, R.N. (2008). A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances. Quarterly Journal of the Royal Meteorological Society, **134**, 1951–1970. 169

Bannister, R.N. (2017). A review of operational methods of variational and ensemble-variational data assimilation. Quarterly Journal of the Royal Meteorological Society, **143**, 607–633. 25

Bannister, R.N. (2020). The ABC-DA system (v1. 4): a variational data assimilation system for convective scale assimilation research with a study of the impact of a balance constraint. Geoscientific Model Development Discussions, 1–42. 27

Basharinov, A.E., Gurvich, A.S. & Egorov, S.T. (1969). Determination of geophysical parameters from thermal radio emission measurements on the artificial earth satellite 'Cosmos-243'. Doklady Akademii Nauk, **188**, 1273–1276. 136

Bauer, P., Geer, A.J., Lopez, P. & Salmond, D. (2010). Direct 4D-Var assimilation of all-sky radiances. Part I: Implementation. Quarterly Journal of the Royal Meteorological Society, **136**, 1868–1885. 128, 130

Bauer, P., Auligné, T., Bell, W., Geer, A., Guidard, V., Heilliette, S., Kazumori, M., Kim, M.J., Liu, E.H.C., McNally, A.P., Macpherson, B., Okamoto, K., Renshaw, R. & Riishøjgaard, L.P. (2011). Satellite cloud and precipitation assimilation at operational NWP centres. Quarterly Journal of the Royal Meteorological Society, **137**, 1934–1951. 129, 130

Bauer, P., Thorpe, A. & Brunet, G. (2015). The quiet revolution of numerical weather prediction. Nature, **525**, 47–55. 22

Beck, A. & Ehrendorfer, M. (2005). Singular-vector-based covariance propagation in a quasigeostrophic assimilation system. Monthly Weather Review, **133**, 1295–1310. 27

Berger, H., Forsythe, M. & Healy, S. (2004). A superobbing scheme for atmospheric motion vectors. In Proc. of the Seventh Int. Winds Workshop, Citeseer. 169

Bick, T., Simmer, C., Trömel, S., Wapler, K., Hendricks Franssen, H.J., Stephan, K., Blahak, U., Schraff, C., Reich, H., Zeng, Y. & Potthast, R. (2016). Assimilation of 3D radar reflectivities with an ensemble Kalman filter on the convective scale. Quarterly Journal of the Royal Meteorological Society, **142**, 1490–1504. 53, 54

Bishop, C.H., Whitaker, J.S. & Lei, L. (2017). Gain form of the ensemble transform Kalman filter and its relevance to satellite data assimilation with model space ensemble covariance localization. Monthly Weather Review, **145**, 4575–4592. 152, 153, 191

Bloom, S.C., Takacs, L.L., Da Silva, A.M. & Ledvina, D. (1996). Data assimilation using incremental analysis updates. Monthly Weather Review, **124**, 1256–1271. 47

Bokhove, O. (2002). Eulerian variational principles for stratified hydrostatic equations. Journal of the Atmospheric Sciences, **59**, 1619–1628. 78

Bokhove, O. & Oliver, M. (2009). A parcel formulation of Hamiltonian layer models. Geophysical and Astrophysical Fluid Dynamics, **103**, 423–442. 78, 80

Bokhove, O., Cantarello, L. & Tobias, S. (2021). An idealized $1\frac{1}{2}$-layer isentropic model with convection and precipitation for satellite data assimilation research. Part II: model derivation, available at `https://eartharxiv.org/repository/view/2007/`. 77, 78, 81, 84, 85, 96

Bowler, N.E., Clayton, A.M., Jardak, M., Lee, E., Lorenc, A.C., Piccolo, C., Pring, S.R., Wlasak, M.A., Barker, D.M., Inverarity, G.W. & Swinbank, R. (2017). Inflation and localization tests in the development of an ensemble of 4D-ensemble variational assimilations. Quarterly Journal of the Royal Meteorological Society, **143**, 1280–1302. 45, 47, 53, 54, 65, 197

Burgers, G., Van Leeuwen, P.J. & Evensen, G. (1998). Analysis scheme in the ensemble Kalman filter. Monthly Weather Review, **126**, 1719–1724. 40

Butterworth, P. & Ingleby, N.B. (2000). Recent developments in the use of satellite winds at the UK Met Office. In Proceedings of the Fifth International Winds Workshop, 151–159. 126

Butterworth, P., English, S., Hilton, F. & Whyte, K. (2002). Improvements in forecasts at the Met Office through reduced weights for satellite winds. In Proceedings of the 6th International Winds Workshop, Madison, available from EUMETSAT, Darmstadt, Germany. 126

Cantarello, L., Bokhove, O. & Tobias, S. (2020). An idealized $1\frac{1}{2}$-layer isentropic model with convection and precipitation for satellite data assimilation research. Part I: model dynamics, available at `https://eartharxiv.org/repository/view/2006/`. 201

Caplan, P., Derber, J., Gemmill, W., Hong, S.Y., Pan, H.L. & Parrish, D. (1997). Changes to the 1995 NCEP operational medium-range forecast model analysis–forecast system. Weather and Forecasting, **12**, 581–594. 127

Cardinali, C., Pezzulli, S. & Andersson, E. (2004). Influence-matrix diagnostic of a data assimilation system. Quarterly Journal of the Royal Meteorological Society, **130**, 2767–2786. 55, 57

Carrassi, A., Bocquet, M., Bertino, L. & Evensen, G. (2018). Data assimilation in the geosciences: An overview of methods, issues, and perspectives. Wiley Interdisciplinary Reviews: Climate Change, **9**, e535. 23

Chang, A.T.C. & Wilheit, T.T. (1979). Remote sensing of atmospheric water vapor, liquid water, and wind speed at the ocean surface by passive microwave techniques from the Nimbus 5 satellite. Radio Science, **14**, 793–802. 136

Chelton, D.B. & Freilich, M.H. (2005). Scatterometer-based assessment of 10-m wind analyses from the operational ECMWF and NCEP numerical weather prediction models. Monthly Weather Review, **133**, 409–429. 127

Chelton, D.B., Freilich, M.H., Sienkiewicz, J.M. & Von Ahn, J.M. (2006). On the use of QuikSCAT scatterometer measurements of surface winds for marine weather prediction. Monthly Weather Review, **134**, 2055–2071. 127

Courtier, P., Thépaut, J.N. & Hollingsworth, A. (1994). A strategy for operational implementation of 4D-Var, using an incremental approach. Quarterly Journal of the Royal Meteorological Society, **120**, 1367–1387. 24

Curry, J.A., Ardeel, C.D. & Tian, L. (1990). Liquid water content and precipitation characteristics of stratiform clouds as inferred from satellite microwave measurements. Journal of Geophysical Research: Atmospheres, **95**, 16659–16671. 137

Dal Maso, G., Lefloch, P.G. & Murat, F. (1995). Definition and weak stability of nonconservative products. Journal de Mathématiques Pures et Appliquées, **74**, 483–548. 35, 203

Daley, R. (1993). Atmospheric data analysis. 2, Cambridge university press. 169

Dando, M.L., Thorpe, A.J. & Eyre, J.R. (2007). The optimal density of atmospheric sounder observations in the Met Office NWP system. Quarterly Journal of the Royal Meteorological Society, **133**, 1933–1943. 169

Dee, D.P. (2004). Variational bias correction of radiance data in the ECMWF system. In Proceedings of the ECMWF workshop on assimilation of high spectral resolution sounders in NWP, Reading, UK, vol. 28, 97–112. 196

Derber, J.C. (2014). Assimilation of satellite data for meteorology. In ECMWF Seminar on the Use of Satellite Observations in NWP, 8–12 September 2014, available at `https://www.ecmwf.int/sites/default/files/elibrary/2015/9004-assimilation-satellite-data-meteorology.pdf`. 122

Derber, J.C. & Wu, W.S. (1998). The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system. Monthly Weather Review, **126**, 2287–2299. 125, 196

Djurić, D. & Damiani Jr, M.S. (1980). On the formation of the low-level jet over texas. Monthly Weather Review, **108**, 1854–1865. 85

Duc, L., Saito, K. & Hotta, D. (2020). Analysis and design of covariance inflation methods using inflation functions. Part 1: Theoretical framework. Quarterly Journal of the Royal Meteorological Society, **146**, 3638–3660. 44

Ehrendorfer, M. & Errico, R.M. (2008). An atmospheric model of intermediate complexity for data assimilation studies. Quarterly Journal of the Royal Meteorological Society, **134**, 1717–1732. 27

Eigenwillig, N. & Fischer, H. (1982). Determination of midtropospheric wind vectors by tracking pure water vapor structures in METEOSAT water vapor image sequences. Bulletin of the American Meteorological Society, 44–58. 126

Errico, R.M. (1997). What is an adjoint model? Bulletin of the American Meteorological Society, **78**, 2577–2592. 25

Errico, R.M., Bauer, P. & Mahfouf, J.F. (2007). Issues regarding the assimilation of cloud and precipitation data. Journal of the Atmospheric Sciences, **64**, 3785–3798. 128

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. Journal of Geophysical Research: Oceans, **99**, 10143–10162. 24

Eyre, J.R. (1994). Assimilation of radio occultation measurements into a numerical weather prediction system. ECMWF, Tech. Memorandum. 128

Eyre, J.R. (1997). Variational Assimilation of Remotely-Sensed Observations of the atmosphere. Journal of the Meteorological Society of Japan, **75**, 331–338. 125

Eyre, J.R. (2007). Progress achieved on assimilation of satellite data in numerical weather prediction over the last 30 years. In ECMWF Seminar Proceedings: Recent developments in the use of satellite observations in numerical weather prediction, 1–27, available at `https://www.ecmwf.int/en/elibrary/15614-progress-achieved-assimilation-satellite-data-nwp-over-last-30-years`. 122

Eyre, J.R., Kelly, G.A., McNally, A.P., Andersson, E. & Persson, A. (1993). Assimilation of TOVS radiance information through one-dimensional variational analysis. Quarterly Journal of the Royal Meteorological Society, **119**, 1427–1463. 124

Eyre, J.R., English, S.J. & Forsythe, M. (2020). Assimilation of satellite data in numerical weather prediction. Part I: The early years. Quarterly Journal of the Royal Meteorological Society, **146**, 49–68. 122, 124, 127

Forsythe, M. (2007). Atmospheric motion vectors: past, present and future. In Proceedings of ECMWF Seminar on Recent Developments in the Use of Satellite Observations in Numerical Weather Prediction, 3–7 September 2007, Reading, UK.

Reading: ECMWF., available at https://www.ecmwf.int/en/elibrary/9445-atmospheric-motion-vectors-past-present-and-future. 169

Fowler, A.M. (2019). Data compression in the presence of observational error correlations. Tellus A: Dynamic Meteorology and Oceanography, **71**, 1634937. 26

Fowler, A.M., Dance, S.L. & Waller, J.A. (2018). On the interaction of observation and prior error correlations in data assimilation. Quarterly Journal of the Royal Meteorological Society, **144**, 48–62. 169

Gaspari, G. & Cohn, S.E. (1999). Construction of correlation functions in two and three dimensions. Quarterly Journal of the Royal Meteorological Society, **125**, 723–757. 45, 199

Geer, A.J. & Bauer, P. (2011). Observation errors in all-sky data assimilation. Quarterly Journal of the Royal Meteorological Society, **137**, 2024–2037. 129

Geer, A.J., Bauer, P. & Lopez, P. (2010). Direct 4D-Var assimilation of all-sky radiances. Part II: Assessment. Quarterly Journal of the Royal Meteorological Society, **136**, 1886–1905. 130

Geer, A.J., Baordo, F., Bormann, N., Chambon, P., English, S., Kazumori, M., Lawrence, H., Lean, P., Lonitz, K. & Lupu, C. (2017). The growing impact of satellite observations sensitive to humidity, cloud and precipitation. Quarterly Journal of the Royal Meteorological Society, **143**, 3189–3206. 26, 130, 131

Geer, A.J., Lonitz, K., Weston, P., Kazumori, M., Okamoto, K., Zhu, Y., Liu, E.H., Collard, A., Bell, W., Migliorini, S., Chambon, P., Fourrié, N., Kim, M.J., Köpken-Watts, C. & Schraff, C. (2018). All-sky satellite data assimilation at operational weather forecasting centres. Quarterly Journal of the Royal Meteorological Society, **144**, 1191–1217. 26, 129, 130, 131, 133

Goodliff, M., Amezcua, J. & Van Leeuwen, P.J. (2015). Comparing hybrid data assimilation methods on the Lorenz 1963 model with increasing non-linearity. Tellus A: Dynamic Meteorology and Oceanography, **67**, 26928. 27

GREENWALD, T.J., STEPHENS, G.L., VONDER HAAR, T.H. & JACKSON, D.L. (1993). A physical retrieval of cloud liquid water over the global oceans using Special Sensor Microwave/Imager (SSM/I) observations. Journal of Geophysical Research: Atmospheres, **98**, 18471–18488. 137

GREENWALD, T.J., STEPHENS, G.L., CHRISTOPHER, S.A. & VONDER HAAR, T.H. (1995). Observations of the global characteristics and regional radiative effects of marine cloud liquid water. Journal of Climate, **8**, 2928–2946. 137

GRODY, N. (1976). Remote sensing of atmospheric water content from satellites using microwave radiometry. IEEE Transactions on Antennas and Propagation, **24**, 155–162. 136

GRODY, N.C., GRUBER, A. & SHEN, W.C. (1980). Atmospheric water content over the tropical Pacific derived from the Nimbus-6 scanning microwave spectrometer. Journal of Applied Meteorology, **19**, 986–996. 136

GUSTAFSSON, N., JANJIĆ, T., SCHRAFF, C., LEUENBERGER, D., WEISSMANN, M., REICH, H., BROUSSEAU, P., MONTMERLE, T., WATTRELOT, E., BUČÁNEK, A., MILE, M., HAMDI, R., LINDSKOG, M., BARKMEIJER, J., DAHLBOM, M., MACPHERSON, B., BALLARD, S., INVERARITY, G.W., CARLEY, J., ALEXANDER, C., DOWELL, D., LIU, S., IKUTA, Y. & FUJITA, T. (2018). Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres. Quarterly Journal of the Royal Meteorological Society, **144**, 1218–1256. 40, 52, 53, 54

HALEM, M., KALNAY, E., BAKER, W.E. & ATLAS, R. (1982). An assessment of the FGGE satellite observing system during SOP-1. Bulletin of the American Meteorological Society, **63**, 407–426. 124

HAMILL, T.M. (2001). Interpretation of rank histograms for verifying ensemble forecasts. Monthly Weather Review, **129**, 550–560. 21, 179, 183

HAMILL, T.M. & SNYDER, C. (2000). A hybrid ensemble Kalman filter–3D variational analysis scheme. Monthly Weather Review, **128**, 2905–2919. 45

HAMILL, T.M., WHITAKER, J.S. & SNYDER, C. (2001). Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. Monthly Weather Review, **129**, 2776–2790. 45

Harper, K., Uccellini, L.W., Kalnay, E., Carey, K. & Morone, L. (2007). 50th anniversary of operational numerical weather prediction. Bulletin of the American Meteorological Society, **88**, 639–650. 22

Harten, A., Lax, P.D. & Leer, B.v. (1983). On upstream differencing and Godunov-type schemes for hyperbolic conservation laws. SIAM Review, **25**, 35–61. 201

Haslehner, M., Janjić, T. & Craig, G.C. (2016). Testing particle filters on simple convective-scale models. Part 2: A modified shallow-water model. Quarterly Journal of the Royal Meteorological Society, **142**, 1628–1646. 51

Healy, S.B. & Thépaut, J.N. (2006). Assimilation experiments with CHAMP GPS radio occultation measurements. Quarterly Journal of the Royal Meteorological Society, **132**, 605–623. 128

Healy, S.B., Eyre, J.R., Hamrud, M. & Thépaut, J.N. (2007). Assimilating GPS radio occultation measurements with two-dimensional bending angle observation operators. Quarterly Journal of the Royal Meteorological Society, **133**, 1213–1227. 128

Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. Weather and Forecasting, **15**, 559–570. 56

Hoffman, R.N. & Leidner, S.M. (2005). An introduction to the near–real–time QuikSCAT data. Weather and Forecasting, **20**, 476–493. 127

Hohenegger, C. & Schar, C. (2007). Atmospheric predictability at synoptic versus cloud-resolving scales. Bulletin of the American Meteorological Society, **88**, 1783–1794. 55, 67

Houtekamer, P.L. & Mitchell, H.L. (1998). Data assimilation using an ensemble Kalman filter technique. Monthly Weather Review, **126**, 796–811. 27, 43, 45, 151

Houtekamer, P.L. & Mitchell, H.L. (2001). A sequential ensemble Kalman filter for atmospheric data assimilation. Monthly Weather Review, **129**, 123–137. 151

Houtekamer, P.L. & Mitchell, H.L. (2005). Ensemble Kalman filtering. Quarterly Journal of the Royal Meteorological Society, **131**, 3269–3289. 151

Huang, R. & Liou, K.N. (1983). Polarized microwave radiation transfer in precipitating cloudy atmospheres: Applications to window frequencies. Journal of Geophysical Research: Oceans, **88**, 3885–3893. 137

Janjić, T., McLaughlin, D., Cohn, S.E. & Verlaan, M. (2014). Conservation of mass and preservation of positivity with ensemble-type Kalman filter algorithms. Monthly Weather Review, **142**, 755–773. 40

Janjić, T., Bormann, N., Bocquet, M., Carton, J.A., Cohn, S.E., Dance, S.L., Losa, S.N., Nichols, N.K., Potthast, R., Waller, J.A. & Weston, P. (2018). On the representation error in data assimilation. Quarterly Journal of the Royal Meteorological Society, **144**, 1257–1278. 170

Jones, A.S. & Vonder Haar, T.H. (1990). Passive microwave remote sensing of cloud liquid water over land regions. Journal of Geophysical Research: Atmospheres, **95**, 16673–16683. 137

Kållberg, P.W., Uppala, S., Gustafsson, N. & Pailleux, J. (1982). The impact of cloud track wind data on global analyses and medium range forecasts. 34, ECMWF, available at `https://www.ecmwf.int/en/elibrary/10594`. 126

Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. Journal of Basic Engineering, **82**, 35–45. 24

Kalman, R.E. & Bucy, R.S. (1961). New results in linear filtering and prediction theory. Journal of Basic Engineering, **83**, 95–108. 24

Kalnay, E. (2003). Atmospheric modeling, data assimilation and predictability. Cambridge University Press. 23, 24, 25, 124

Kelly, G. & Thépaut, J.N. (2007). Evaluation of the impact of the space component of the Global Observing System through Observing System Experiments. ECMWF Newsletter, **113**, 16–28. 25

Kelly, G.A.M., Mills, G.A. & Smith, W.L. (1978). Impact of Nimbus-6 temperature soundings on Australian region forecasts. Bulletin of the American Meteorological Society, **59**, 393–406. 124

KENT, T. (2016). An idealised fluid model for convective-scale NWP: Dynamics and data assimilation. Ph.D. thesis, School of Mathematics, University of Leeds. 2, 28, 31, 33, 34, 35, 49, 67, 72, 92, 190, 196, 201, 202, 203

KENT, T. & BOKHOVE, O. (2020). Ensuring 'well-balanced' shallow water flows via a discontinuous Galerkin finite element method: issues at lowest order, arXiv preprint available at `https://arxiv.org/abs/2006.03370`. 35, 196

KENT, T., BOKHOVE, O. & TOBIAS, S. (2017). Dynamics of an idealized fluid model for investigating convective-scale data assimilation. Tellus A: Dynamic Meteorology and Oceanography, **69**, 1369332. 2, 28, 31, 34, 35, 49, 72, 76, 84, 92, 93, 97, 190, 196, 201, 203, 204

KENT, T., CANTARELLO, L., INVERARITY, G., TOBIAS, S. & BOKHOVE, O. (2020). Idealized forecast-assimilation experiments for convective-scale Numerical Weather Prediction, available at `https://eartharxiv.org/repository/view/1921/`. 32, 47, 197

KIDDER, S.Q., M, R. & VONDER HAAR, T.H. (1995). Satellite meteorology: an introduction. Academic Press. 104, 110, 111, 118, 121

LADWIG, D.S. (1980). Cyclogenesis and the Low-Level Jet over the Southern Great Plains. Tech. rep., Air Force Institute of Technology Wright-Patterson Air Force Base, OH, available at: `https://apps.dtic.mil/sti/pdfs/ADA106159.pdf`. 85

LANGLAND, R.H. & BAKER, N.L. (2004). Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. Tellus A: Dynamic Meteorology and Oceanography, **56**, 189–201. 130

LE DIMET, F.X. & TALAGRAND, O. (1986). Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. Tellus A: Dynamic Meteorology and Oceanography, **38**, 97–110. 24

LE FLOCH, P. (1989). Shock waves for nonlinear hyperbolic systems in nonconservative form. Tech. Rep. 593, Institute for Mathematics and its Applications, Minneapolis, MN. 35

LEUTBECHER, M. & PALMER, T.N. (2008). Ensemble forecasting. Journal of computational physics, **227**, 3515–3539. 53

LeVeque, R.J. (2002). Finite volume methods for hyperbolic problems, vol. 31. Cambridge University Press. 34

Liou, K.N., Yeh, H.Y., Chen, F.M., Hutchison, K. & Astling, E. (1980). Development of Infrared and Microwave Techniques for Cloud Parameter Inference from Satellite Imagery and Sounder Data. Tech. rep., Utah University, Salt Lake City, Department of Meteorology. 137

Liu, Z.Q. & Rabier, F. (2002). The interaction between model resolution, observation resolution and observation density in data assimilation: A one-dimensional study. Quarterly Journal of the Royal Meteorological Society, **128**, 1367–1386. 169

Liu, Z.Q. & Rabier, F. (2003). The potential of high-density observations for numerical weather prediction: A study with simulated observations. Quarterly Journal of the Royal Meteorological Society, **129**, 3013–3035. 169

Lorenc, A.C. (1986). Analysis methods for numerical weather prediction. Quarterly Journal of the Royal Meteorological Society, **112**, 1177–1194. 25

Lorenc, A.C., Ballard, S.P., Bell, R.S., Ingleby, N.B., Andrews, P.L.F., Barker, D.M., Bray, J.R., Clayton, A.M., Dalby, T., Li, D., Payne, T.J. & Saunders, F.W. (2000). The Met Office global three-dimensional variational data assimilation scheme. Quarterly Journal of the Royal Meteorological Society, **126**, 2991–3012. 125

Lorenz, E.N. (1963). Deterministic nonperiodic flow. Journal of Atmospheric Sciences, **20**, 130–141. 23, 27

Lorenz, E.N. (1995). Predictability: a problem partly solved. In Seminar on Predictability, ECMWF, Reading, United Kingdom, available at `https://www.ecmwf.int/en/elibrary/10829-predictability-problem-partly-solved`. 27

Lupu, C. & McNally, A.P. (2012). Assimilation of cloud-affected radiances from Meteosat-9 at ECMWF. Tech. Rep. 25, European Centre for Medium-Range Weather Forecasts. 130

McNally, A.P. (2002). A note on the occurrence of cloud in meteorologically sensitive areas and the implications for advanced infrared sounders. Quarterly Journal of the Royal Meteorological Society, **128**, 2551–2556. 26, 129

Melbourne, W.G., Davis, E.S., Duncan, C.B., Hajj, G.A., Hardy, K.R., Kursinski, E.R., Meehan, T.K., Young, L.E. & Yunck, T.P. (1994). The application of spaceborne GPS to atmospheric limb sounding and global change monitoring. Tech. rep., National Aeronautics and Space Administration, available at `https://ntrs.nasa.gov/api/citations/19960008694/downloads/19960008694.pdf`. 128

Menzel, W.P. (2001). Cloud tracking with satellite imagery: From the pioneering work of Ted Fujita to the present. Bulletin of the American Meteorological Society, **82**, 33–47. 122, 125

Menzel, W.P., Schmit, T.J., Zhang, P. & Li, J. (2018). Satellite-based atmospheric infrared sounder development and applications. Bulletin of the American Meteorological Society, **99**, 583–603. 122

Migliorini, S. & Candy, B. (2019). All-sky satellite data assimilation of microwave temperature sounding channels at the Met Office. Quarterly Journal of the Royal Meteorological Society, **145**, 867–883. 26, 131

Milan, M., Macpherson, B., Tubbs, R., Dow, G., Inverarity, G., Mittermaier, M., Halloran, G., Kelly, G., Li, D., Maycock, A., Payne, T., Piccolo, C., Stewart, L. & Wlasak, M. (2020). Hourly 4D-Var in the Met Office UKV operational forecast model. Quarterly Journal of the Royal Meteorological Society, **146**, 1281–1301. 150, 193, 195

Mugnai, A. & Smith, E.A. (1988). Radiative transfer to space through a precipitating cloud at multiple microwave frequencies. Part I: Model description. Journal of Applied Meteorology, **27**, 1055–1073. 137

Nieman, S.J., Menzei, W.P., Hayden, C.M., Gray, D., Wanzong, S.T., Velden, C.S. & Daniels, J. (1997). Fully automated cloud-drift winds in NESDIS operations. Bulletin of the American Meteorological Society, **78**, 1121–1134. 126

Njoku, E.G. & Swanson, L. (1983). Global measurements of sea surface temperature, wind speed and atmospheric water content from satellite microwave radiometry. Monthly Weather Review, **111**, 1977–1987. 137

Ohring, G. (1979). Impact of satellite temperature sounding data on weather forecasts. Bulletin of the American Meteorological Society, **60**, 1142–1147. 123

Ott, E., Hunt, B.R., Szunyogh, I., Zimin, A.V., Kostelich, E.J., Corazza, M., Kalnay, E., Patil, D. & Yorke, J.A. (2004). A local ensemble Kalman filter for atmospheric data assimilation. Tellus A: Dynamic Meteorology and Oceanography, **56**, 415–428. 27

O'Dell, C.W., Wentz, F.J. & Bennartz, R. (2008). Cloud liquid water path from satellite-based passive microwave observations: A new climatology over the global oceans. Journal of Climate, **21**, 1721–1739. 137

Pan, F. & Smith, R.B. (1999). Gap winds and wakes: SAR observations and numerical simulations. Journal of the Atmospheric Sciences, **56**, 905–923. 76

Pangaud, T., Fourrie, N., Guidard, V., Dahoui, M. & Rabier, F. (2009). Assimilation of AIRS radiances affected by mid-to low-level clouds. Monthly Weather Review, **137**, 4276–4292. 130

Pavelin, E.G., English, S.J. & Eyre, J.R. (2008). The assimilation of cloud-affected infrared satellite radiances for numerical weather prediction. Quarterly Journal of the Royal Meteorological Society, **134**, 737–749. 130

Perianez, A., Reich, H. & Potthast, R. (2014). Optimal localization for ensemble Kalman filter systems. Journal of the Meteorological Society of Japan. Ser. II, **92**, 585–597. 170

Petrie, R.E., Bannister, R.N. & Cullen, M.J.P. (2017). The "ABC model": a non-hydrostatic toy model for use in convective-scale data assimilation investigations. Geoscientific Model Development, **10**, 4419–4441. 27

Petty, G.W. (2006). A first course in atmospheric radiation. Sundog Pub. 104, 106

Petty, G.W. & Katsaros, K.B. (1994). The response of the SSM/I to the marine environment. Part II: A parameterization of the effect of the sea surface slope distribution on emission and reflection. Journal of Atmospheric and Oceanic Technology, **11**, 617–628. 139

Prabhakara, C., Chang, H. & Chang, A.T. (1982). Remote sensing of precipitable water over the oceans from Nimbus 7 microwave measurements. Journal of Applied Meteorology, **21**, 59–68. 137

Prabhakara, C., Wang, I., Chang, A. & Gloersen, P. (1983). A statistical examination of Nimbus-7 SMMR data and remote sensing of sea surface temperature, liquid water content in the atmosphere and surface wind speed. Journal of Climate and Applied Meteorology, **22**, 2023–2037. 137

Rabier, F. (2005). Overview of global data assimilation developments in numerical weather-prediction centres. Quarterly Journal of the Royal Meteorological Society, **131**, 3215–3233. 25

Rainwater, S., Bishop, C.H. & Campbell, W.F. (2015). The benefits of correlated observation errors for small scales. Quarterly Journal of the Royal Meteorological Society, **141**, 3439–3445. 169

Reich, S. & Cotter, C. (2015). Probabilistic forecasting and Bayesian data assimilation. Cambridge University Press. 23

Rennie, M.P. (2010). The impact of GPS radio occultation assimilation at the Met Office. Quarterly Journal of the Royal Meteorological Society, **136**, 116–131. 128

Renshaw, R.J., Barwell, B.R., English, S.J. & Grant, J.R. (1997). Developments in assimilating global TOVS data at the UK Met Office. In Technical Proceedings of the 9th International TOVS Study Conference, 20–26 February 1997, Igls, Austria, 405–416. 125

Rhebergen, S., Bokhove, O. & van der Vegt, J.J.W. (2008). Discontinuous Galerkin finite element methods for hyperbolic nonconservative partial differential equations. Journal of Computational Physics, **227**, 1887–1922. 35, 201

Rife, D.L., Pinto, J.O., Monaghan, A.J., Davis, C.A. & Hannan, J.R. (2010). Global distribution and characteristics of diurnally varying low-level jets. Journal of Climate, **23**, 5041–5064. 85

Ripa, P. (1993). Conservation laws for primitive equations models with inhomogeneous layers. Geophysical & Astrophysical Fluid Dynamics, **70**, 85–111. 78

Rodgers, C.D. (1976). Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation. Reviews of Geophysics, **14**, 609–624. 124

Rohn, M., Kelly, G. & Saunders, R.W. (2001). Impact of a new cloud motion wind product from Meteosat on NWP analyses and forecasts. Monthly Weather Review, **129**, 2392–2403. 126

Rosenkranz, P.W., Staelin, D.H. & Grody, N.C. (1978). Typhoon June (1975) viewed by a scanning microwave spectrometer. Journal of Geophysical Research: Oceans, **83**, 1857–1868. 136

Ruckstuhl, Y.M. & Janjić, T. (2018). Parameter and state estimation with ensemble Kalman filter based algorithms for convective-scale applications. Quarterly Journal of the Royal Meteorological Society, **144**, 826–841. 51

Rudd, A.C., Roulstone, I. & Eyre, J.R. (2012). A simple column model to explore anticipated problems in variational assimilation of satellite observations. Environmental modelling & software, **27**, 23–39. 28

Sakov, P. & Oke, P.R. (2008). A deterministic formulation of the ensemble Kalman filter: an alternative to ensemble square root filters. Tellus A: Dynamic Meteorology and Oceanography, **60**, 361–371. 28, 40, 45, 190, 197, 198

Salby, M.L. (1996). Fundamentals of atmospheric physics. Elsevier. 104, 115, 118

Salman, H., Kuznetsov, L., Jones, C.K.R.T. & Ide, K. (2006). A method for assimilating lagrangian data into a shallow-water-equation ocean model. Monthly Weather Review, **134**, 1081–1101. 27

Saunders, R. (2021). The use of satellite data in numerical weather prediction. Weather, **76**, 95–97. 122

Savage, R.C. & Weinman, J.A. (1975). Preliminary calculations of the upwelling radiance from rainclouds at 37.0 and 19.35 GHz. Bulletin of the American Meteorological Society, **56**, 1272–1274. 137

Schmetz, J., Holmlund, K., Hoffman, J., Strauss, B., Mason, B., Gaertner, V., Koch, A. & Van De Berg, L. (1993). Operational cloud-motion winds from Meteosat infrared images. Journal of Applied Meteorology, **32**, 1206–1225. 126

Schraff, C., Reich, H., Rhodin, A., Schomburg, A., Stephan, K., Periáñez, A. & Potthast, R. (2016). Kilometre-scale ensemble data assimilation for the COSMO model (KENDA). Quarterly Journal of the Royal Meteorological Society, **142**, 1453–1472. 53, 54

Schur, J. (1911). Bemerkungen zur Theorie der beschränkten Bilinearformen mit unendlich vielen Veränderlichen. Journal für die reine und angewandte Mathematik, **1911**, 1–28. 45

Seaman, R.S. (1977). Absolute and differential accuracy of analyses achievable with specified observational network characteristics. Monthly Weather Review, **105**, 1211–1222. 169

Shrira, V. (1981). Propagation of long nonlinear waves in a layer of rotating fluid. Sov. Phys. - Izvestija, **17**, 55–59. 89, 101, 191

Shrira, V. (1986). On the long strongly nonlinear waves in rotating ocean. Sov. Phys. - Izvestija, **22**, 285–305. 89, 101, 191

Simmons, A.J. & Hollingsworth, A. (2002). Some aspects of the improvement in skill of numerical weather prediction. Quarterly Journal of the Royal Meteorological Society, **128**, 647–677. 25

Smith, E.A. & Mugnai, A. (1988). Radiative transfer to space through a precipitating cloud at multiple microwave frequencies. Part II: Results and analysis. Journal of Applied Meteorology, **27**, 1074–1091. 137

Smith, P.J., Fowler, A.M. & Lawless, A.S. (2015). Exploring strategies for coupled 4D-Var data assimilation using an idealised atmosphere–ocean model. Tellus A: Dynamic Meteorology and Oceanography, **67**, 27025. 28

Smith, W., Woolf, H. & Jacob, W. (1970). A regression method for obtaining real-time temperature and geopotentlal height profiles from satellite spectrometer measurements and its application to Nimbus 3 “SIRS” observations. Monthly Weather Review, **98**, 582–603. 123

Spencer, R., Olson, W., Rongzhang, W., Martin, D., Weinman, J. & Santek, D. (1983). Heavy thunderstorms observed over land by the Nimbus 7 Scanning

Multichannel Microwave Radiometer. Journal of Climate and Applied Meteorology, **22**, 1041–1046. 137

Spencer, R.W. (1984). Satellite passive microwave rain rate measurement over croplands during spring, summer and fall. Journal of climate and applied meteorology, **23**, 1553–1562. 137

Spencer, R.W. (1986). A satellite passive 37-GHz scattering-based method for measuring oceanic rain rates. Journal of Climate and Applied Meteorology, **25**, 754–766. 137

Spencer, R.W., Goodman, H.M. & Hood, R.E. (1989). Precipitation retrieval over land and ocean with the SSM/I: Identification and characteristics of the scattering signal. Journal of Atmospheric and Oceanic Technology, **6**, 254–273. 18, 137, 139, 140

Staelin, D.H., Cassel, A.L., Kunzi, K.F., Pettyjohn, R.L., Poon, R.K.L., Rosenkranz, P.W. & Waters, J.W. (1975). Microwave atmospheric temperature sounding: Effects of clouds on the Nimbus 5 satellite data. Journal of the Atmospheric Sciences, **32**, 1970–1976. 136

Staelin, D.H., Kunzi, K.F., Pettyjohn, R.L., Poon, R.K.L., Wilcox, R.W. & Waters, J.W. (1976). Remote sensing of atmospheric water vapor and liquid water with the Nimbus 5 microwave spectrometer. Journal of Applied Meteorology, **15**, 1204–1214. 136

Stamnes, K., Thomas, G.E. & Stamnes, J.J. (2017). Radiative transfer in the atmosphere and ocean. Cambridge University Press. 104, 115

Stewart, L.M., Dance, S.L. & Nichols, N.K. (2013). Data assimilation with correlated observation errors: experiments with a 1- shallow water model. Tellus A: Dynamic Meteorology and Oceanography, **65**, 19546. 27

Stoffelen, A. & Anderson, D. (1997a). Ambiguity removal and assimilation of scatterometer data. Quarterly Journal of the Royal Meteorological Society, **123**, 491–518. 127

STOFFELEN, A. & ANDERSON, D. (1997b). Scatterometer data interpretation: Estimation and validation of the transfer function CMOD4. Journal of Geophysical Research: Oceans, **102**, 5767–5780. 127

STOFFELEN, A. & ANDERSON, D. (1997c). Scatterometer data interpretation: Measurement space and inversion. Journal of Atmospheric and Oceanic Technology, **14**, 1298–1313. 127

SUN, J., XUE, M., WILSON, J.W., ZAWADZKI, I., BALLARD, S.P., ONVLEE-HOOIMEYER, J., JOE, P., BARKER, D.M., LI, P.W., GOLDING, B., XU, M. & PINTO, J. (2014). Use of NWP for nowcasting convective precipitation: Recent progress and challenges. Bulletin of the American Meteorological Society, **95**, 409–426. 53

TAKEDA, T. & LIU, G. (1987). Estimation of Atmospheric Liquid-Water Amount by Nimbus 7 SMMR Data. Journal of the Meteorological Society of Japan. Ser. II, **65**, 931–947. 18, 137, 138, 139

TALAGRAND, O. & COURTIER, P. (1987). Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. Quarterly Journal of the Royal Meteorological Society, **113**, 1311–1328. 24

TANG, Y., LEAN, H.W. & BORNEMANN, J. (2013). The benefits of the Met Office variable resolution NWP model for forecasting convection. Meteorological Applications, **20**, 417–426. 150

TIPPETT, M.K., ANDERSON, J.L., BISHOP, C.H., HAMILL, T.M. & WHITAKER, J.S. (2003). Ensemble square root filters. Monthly Weather Review, **131**, 1485–1490. 40

TJEMKES, S.A., STEPHENS, G.L. & JACKSON, D.L. (1991). Spaceborne observation of columnar water vapor: SSM/I observations and algorithm. Journal of Geophysical Research: Atmospheres, **96**, 10941–10954. 137

TOMASSINI, M., KELLY, G. & SAUNDERS, R. (1999). Use and impact of satellite atmospheric motion winds on ECMWF analyses and forecasts. Monthly Weather Review, **127**, 971–986. 126

Vallis, G.K. (2017). Atmospheric and oceanic fluid dynamics. Cambridge University Press. 77

Van Leeuwen, P.J. (1999). Comment on "Data assimilation using an ensemble Kalman filter technique". Monthly Weather Review, **127**, 1374–1377. 43

Van Leeuwen, P.J. (2009). Particle filtering in geophysical systems. Monthly Weather Review, **137**, 4089–4114. 25

Veersé, F. & Thépaut, J.N. (1998). Multiple-truncation incremental approach for four-dimensional variational data assimilation. Quarterly Journal of the Royal Meteorological Society, **124**, 1889–1908. 24

Vetra-Carvalho, S., Migliorini, S. & Nichols, N. (2011). Ensemble data assimilation in the presence of cloud. Computers & Fluids, **46**, 493–497. 28

Wark, D.Q. (1970). SIRS: An experiment to measure the free air temperature from a satellite. Applied optics, **9**, 1761–1766. 123

Wentz, F.J. (1983). A model function for ocean microwave brightness temperatures. Journal of Geophysical Research: Oceans, **88**, 1892–1908. 136

Wentz, F.J. (1997). A well-calibrated ocean algorithm for special sensor microwave/imager. Journal of Geophysical Research: Oceans, **102**, 8703–8718. 137

Whitaker, J.S. & Hamill, T.M. (2002). Ensemble data assimilation without perturbed observations. Monthly Weather Review, **130**, 1913–1924. 41

Whitaker, J.S. & Hamill, T.M. (2012). Evaluating methods to account for system errors in ensemble data assimilation. Monthly Weather Review, **140**, 3078–3089. 44, 51

Wilheit, T.T. & Chang, A.T.C. (1980). An algorithm for retrieval of ocean surface and atmospheric parameters from the observations of the scanning multichannel microwave radiometer. Radio Science, **15**, 525–544. 136

Wilheit, T.T., Chang, A.T.C., V. Rao, M.S., Rodgers, E.B. & Theon, J.S. (1977). A satellite technique for quantitatively mapping rainfall rates over the oceans. Journal of applied meteorology, **16**, 551–560. 137

WU, R. & WEINMAN, J.A. (1984). Microwave radiances from precipitating clouds containing aspherical ice, combined phase, and liquid hydrometeors. Journal of Geophysical Research: Atmospheres, **89**, 7170–7178. 18, 137, 139, 140

WÜRSCH, M. & CRAIG, G.C. (2014). A simple dynamical model of cumulus convection for data assimilation research. Meteorologische Zeitschrift, 483–490. 27, 31, 51

ŽAGAR, N., GUSTAFSSON, N. & KÄLLÉN, E. (2004). Dynamical response of equatorial waves in four-dimensional variational data assimilation. Tellus A: Dynamic Meteorology and Oceanography, **56**, 29–46. 27

ZAPLOTNIK, Ž., ŽAGAR, N. & GUSTAFSSON, N. (2018). An intermediate-complexity model for four-dimensional variational data assimilation including moist processes. Quarterly Journal of the Royal Meteorological Society, **144**, 1772–1787. 27

ZAPOTOCNY, T.H., JUNG, J.A., LE MARSHALL, J.F. & TREADON, R.E. (2008). A two-season impact study of four satellite data types and rawinsonde data in the NCEP Global Data Assimilation System. Weather and Forecasting, **23**, 80–100. 126

ZENG, Y., JANJIĆ, T., DE LOZAR, A., BLAHAK, U., REICH, H., KEIL, C. & SEIFERT, A. (2018). Representation of model error in convective-scale data assimilation: additive noise, relaxation methods, and combinations. Journal of Advances in Modeling Earth Systems, **10**, 2889–2911. 52

ZENG, Y., JANJIĆ, T., SOMMER, M., DE LOZAR, A., BLAHAK, U. & SEIFERT, A. (2019). Representation of Model Error in Convective-Scale Data Assimilation: Additive Noise Based on Model Truncation Error. Journal of Advances in Modeling Earth Systems, **11**, 752–770. 52

ZHOU, X., ZHU, Y., HOU, D., LUO, Y., PENG, J. & WOBUS, R. (2017). Performance of the new NCEP Global Ensemble Forecast System in a parallel experiment. Weather and Forecasting, **32**, 1989–2004. 22

ZHU, Y., LIU, E., MAHAJAN, R., THOMAS, C., GROFF, D., VAN DELST, P., COLLARD, A., KLEIST, D., TREADON, R. & DERBER, J.C. (2016). All-sky microwave radiance assimilation in NCEP's GSI analysis system. Monthly Weather Review, **144**, 4709–4735. 131

Zienkiewicz, O., Taylor, R. & Nithiarasu, P. (2014). The Finite Element Method for Fluid Dynamics. Butterworth-Heinemann, Oxford, seventh edn. 201