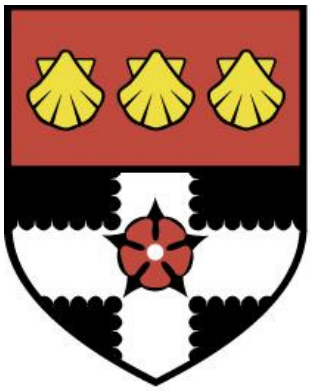
UNIVERSITY OF READING



DOCTORAL THESIS

Information gain that convective-scale models bring to probabilistic weather forecasts

---

Carlo CAFARO

A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Philosophy

in the

School of Mathematical, Physical and Computational Sciences

July 2019

Declaration of Authorship

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Carlo Cafaro

Date: 2 July 2019

"Apri la mente a quel ch'io ti paleso
e fermalvi entro; ché non fa scienza,
sanza lo ritenere, avere inteso. "

Open thy mind to what I now reveal,
and fix it therewithin; for having heard
without retaining doth not knowledge make.

Dante Alighieri, Canto V, Paradiso.

"The great mathematician and explorer of nature Andrei Nikolaevich Kolmogorov used to say to his disciples: before making a mathematical picture of a certain geophysical phenomenon try always to find a person who knows how it occurs in reality. "

Ellis Horwood

UNIVERSITY OF READING

Abstract

School of Mathematical, Physical and Computational Sciences

Doctor of Philosophy

Information gain that convective-scale models bring to probabilistic weather forecasts

by Carlo CAFARO

Ensemble prediction systems, run nowadays at convective-scale by several operational and research forecasting centres, undoubtedly provide a large amount of data. Whether and how the utilisation of these data can lead to additional valuable information for the probabilistic prediction of specific weather phenomena is an open challenge.
Here, the information gain of the Met Office convective-scale ensemble relative to its lower resolution global model counterpart (33 km) is quantified for the sea breeze phenomenon. It occurs on small spatio-temporal scale but influenced by the large-scale conditions, well represented by the global ensembles. Sea breeze is also an atmospheric counterpart of the so called density currents and could lead to high impact weather, especially when colliding with other sea breezes or mesoscale flows.
In the first part of the thesis, a new set of numerical simulations of colliding density currents is presented, with the aim of understanding the dynamics of collision.
A novel parametric formula for predicting the collision angle is shown to agree well with numerical data. This can be useful for future parametrizations of these processes.

Probabilistic forecasts of the occurrence of sea breezes have been then produced, compared and verified over an extended period.
This motivated the development of a novel method for the automatic identification of the sea breeze from convective-scale ensembles. This method can be in principle applied to every coastline. A Bayesian approach is instead used to extract information from the coarser resolution ensemble. It is trained on paired high/low resolution ensemble member, but it can be trained also on observations. This method creates a statistical forecast of the high-resolution ensemble member, given the knowledge of the global ensemble predictors alone.
In the last part of the thesis, the same methodology has been applied to the prediction of wind gusts. Comparison of the two forecasting methods, using a variety of well established verification metrics all lead to the same conclusion: although the Bayesian forecasts have potential skill for the prediction of event occurrence, the convective-scale ensemble is shown to be more skillful.

Acknowledgements

First of all I would like to thank my supervisors: Tom, John, Nigel and Jochen for being always available to help, especially in the difficult moments I experienced in these three years and half.
Thanks to Roberto B., an outstanding scientist and friend. I will never forget the first time I went to the European Centre for Medium Range Weather Forecasts (ECMWF) here in Reading [1] to visit him. His moral and scientific support were undoubtedly beneficial for my career and well-being.
I am grateful to all the people I have met during this journey. Every person I spoke with gave me some insights, feedback and provided me with a bit of inspiration. But only few people really know how I felt while working on this project: Andrea, whose friendship has been truly helpful for me, I hope we can play more tunes in the future; Susanna for sharing with me the pain and the joy of getting a PhD; Grazia for being interested and a very good listener to what I was talking about; Tobias who was always supportive; Francesco, whom I really liked to discuss with; Vincenzo, for making me laugh with his jokes. I would like to mention all the people of the MPE CDT for making this journey really enjoyable.
Furthermore, I want to thank the people I met at UK Met Office during the Training and Research Summer School. In particular my supervisor, Gabriel who was really helpful and supportive during the project and very responsive after, while working on the paper.
I am grateful also to the friends I had the chance to meet in Exeter during my stay there.
I would like to thank my family for supporting my choice. It was not easy to live away from them but I know they are proud of me. Thanks to my grandparents for being always in contact with me.
To conclude I would like to thank whoever/whatever led me to become passionate about this wonderful subject. Wherever I will be, whatever I will do, I will always look at the sky.

This work has been accomplished with the financial support of the Engineering and Physical Sciences Research Council Mathematics of Planet Earth Centre for Doctoral Training (grant reference EP/L016613/1) and the NERC National Centre for Earth Observation.

Exeter, 1 July 2019

[1] 26 November 2014

Contents

List of Abbreviations

BAMS	**B**ulletin of the **A**merican **M**eteorological **Society**
COSMO	**CO**nsortium for **S**mall-scale **MO**delling
CP-DYN	**C**onvection **P**ermitting **DYN**amical
CP-EPS	**C**onvection-**P**ermitting **E**nsemble **P**rediction **Systems**
JAS	**J**ournal of the **A**tmospheric **S**ciences
LAM	**L**imited **A**rea **M**odel
LR-BAY	**L**ow **R**esolution **BAY**esian
LR-EPS	**L**ower-**R**esolution **E**nsemble **P**rediction **Systems**
MOGREPS	**M**et **O**ffice **G**lobal and **R**egional **E**nsemble **P**rediction **S**ystems
MWR	**M**onthly **W**eather **R**eview
NWP	**N**umerical **W**eather **P**rediction
QJRMS	**Q**uarterly **J**ournal of the **R**oyal **M**eteorological Society
TIGGE	**T**he **I**nternational **G**rand **G**lobal **E**nsemble
WAF	**W**eather **A**nd **Forecasting**

To Giulio Regeni, whose willingness to seek the truth was truly inspirational

To everyone else who devoted or is devoting its life fighting for the truth. In particular let me mention Paolo Borsellino, Giovanni Falcone and Ilaria Cucchi, amongst the others.

To my family, for supporting my choice.

To my grandparents, for being close.

Chapter 1

Introduction

1.1 Convection-permitting ensemble prediction systems (CP-EPSs)

In the last decade, increased computer power (e.g. https://www.metoffice.gov.uk/research/technology/supercomputer[1]) has allowed several operational and research forecasting centres to introduce ensemble prediction systems at convection-permitting resolutions. Weisman et al. (1997) found that a grid spacing of $\sim$ 4 km are sufficient to reproduce much of the mesoscale structure of the squall-line type convective systems.
The first experiments with convective-scale ensembles date back to 2007 when the Center for Analysis and Prediction of Storms (CAPS) at the University of Oklahoma generated storm-scale ensemble forecasts (SSEFs) over a near-CONUS (continental United States) domain during the National Oceanic and Atmospheric Administration Hazardous Weather Testbed (NOAA HWT) Spring Experiments (Xue et al., 2007).
At a later time, on 22 May 2012, at the German Weather Service (DWD), a convection-permitting ensemble (COSMO-DE EPS) became operational (Gebhardt et al., 2011). Some months later, during the 2012 London Olympic Games, the UK Met Office introduced the convective-scale version of the Met Office Global and Regional Ensemble Prediction System (MOGREPS-UK) (Golding et al., 2014; Tennant, 2015; Hagelin et al., 2017). Meteo-France has been running first experimentally (during the HyMex campaign) and then operationally since mid-October 2016 its CP-EPS (AROME-EPS: Raynaud and Bouttier, 2016; Bouttier et al., 2016). Other examples include ALADIN-LAEF (Wang et al., 2011; Horányi et al., 2011) and COSMO-LEPS (Montani et al., 2011). More recently, other experiments using CP-EPSs have been conducted by Schwartz et al. (2015) at National Centre for Atmospheric Research (NCAR) in a domain covering the whole United States.
In table 1.1 a more detailed list of CP-EPSs is presented. Many of them are part of the TIGGE-LAM project (https://confluence.ecmwf.int/display/TIGL/Project[2]), which is an extension of the TIGGE project (Swinbank et al., 2016), with the aim of coordinating and supporting the generation of limited area high resolution ensemble models among European countries.
The implementation of CP-EPSs was motivated by the fact that many of high impact hazardous weather are associated with small meso- and convective-scale weather phenomena, the prediction of which can be uncertain even at few hours ahead (Lorenz, 1969;

[1] Accessed online on 28 November 2018
[2] Accessed online on 28 November 2016

Hohenegger and Schar, 2007). Therefore, a probabilistic approach would seem more suitable for quantifying uncertainty associated with short-range small-scale weather events. However, despite being an exciting new forecasting technology, CP-EPSs place a heavy burden on the computational resources of forecasting centres.[3] They also produce large amounts of data which needs to be rapidly digested and utilized by operational forecasters. Knowing when and how the high resolution ensemble is likely to provide useful additional information is key to successful real-time utilisation of this data. Similarly, knowing where equivalent information can be gained (even if partially) from the global ensemble using statistical/dynamical post-processing both extends lead time (due to faster production time) and also potentially provides information in regions where no high resolution ensemble is available. There is therefore a good reason to assess these systems, against the "truth" (usually represented by observations), comparing them with lower-resolution (and thus cheaper) ensemble prediction systems (LR-EPS) to assess any gain in the skill.

The extent to which CP-EPSs are able to provide better probabilistic forecasts is still not clear for some parameters. Therefore, CP-EPS should be assessed in a way that is relevant to the purpose for which they have been constructed, which is the prediction of small-scale weather phenomena (Clark et al., 2016; Bowler et al., 2008).

The meaning of "better" forecast should be defined in advance for the particular small-scale phenomenon being examined. In order to have a statistically significant answer to the question "are CP-EPSs better than LR-EPSs ?", a long enough period to obtain robust results is needed. Obviously these datasets cannot be longer than 10 years (unless they are comprised of re-forecasts). Also, since the real driving motivation for running CP-EPS, is to detect the possibility of small-scale high impact (therefore locally rare) weather events, the dataset becomes even more sparse.

Since their introduction there have been many studies on CP-EPSs. Figure 1.1 shows the percentage of publications per year, in four meteorological journals, containing the phrase "convection-permitting (allowing) ensemble", somewhere in the article. It can be seen the topic of convection-permitting ensembles has received more and more attention over the past years. However, to the best of our knowledge, there is no study focussing on quantification of the added value of CP-EPS: for specific weather phenomena, represented across multiple grid squares by multiple variables; compared to their coarser resolution version and for an extended evaluation period.

Among these studies, some of them have dealt with the verification of CP-EPSs (Gilleland, 2017; Ebert et al., 2013), among the others, but these have not focussed on information gain when compared to statistically enhanced coarser resolution ensembles. Also, most studies have focussed on conventional variables that are directly represented on model grid squares (e.g. surface temperature) and most of the emphasis has been on precipitation accumulation. This is because it can be evaluated against high resolution radar data, but more importantly because it can be used as a proxy for convective storms and their associated hazards for people.

[3] For instance, at the UK Met Office one ensemble member, for one day of forecast for a convective-scale limited area model at grid size of 2.2 km would use 7 node-hours, whereas one global model at grid size of 33 km would use 3 node-hours (Warren Tennant, personal communication, date: 28 November 2018).

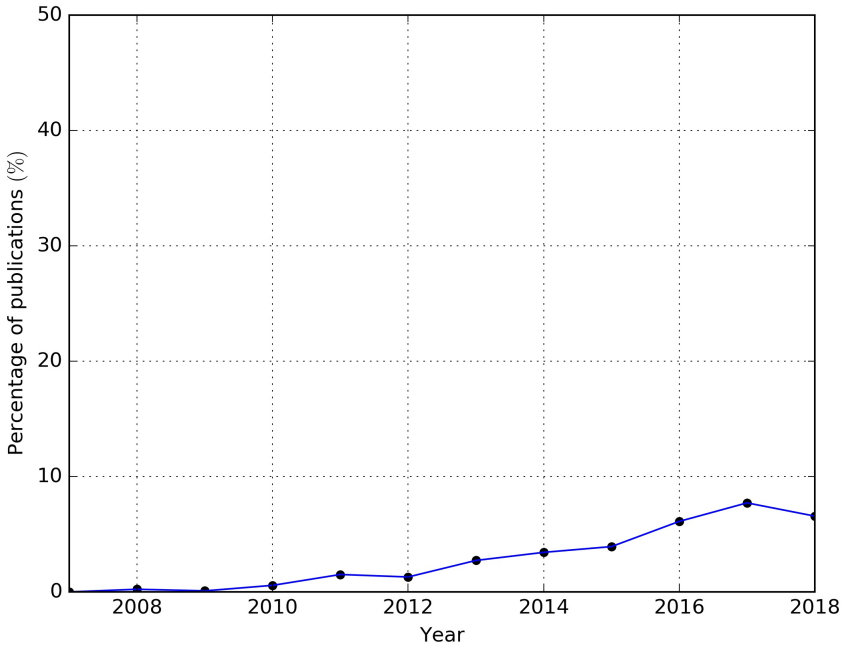


FIGURE 1.1: Percentage of publications per year since 2007 (up to 18 December 2018) including the phrase "convection-permitting ensemble" or "convection-allowing ensemble" for four journals (JAS, MWR, QJRMS, WAF) , according to Google Scholar. The idea of the graph was taken by Guarino M. V. (2018), PhD thesis, accessible at this link: http://centaur.reading.ac.uk/75850/1/22835919_Guarino_thesis.pdf.

Clark et al. (2016) reviewed the state-of-art convection-permitting models for forecasting rainfall and demonstrated their benefit by presenting a few cases studies with a qualitative comparison against radar images.

Among the studies focussing on precipitation, Schellander-Gorgas et al. (2017) showed the benefit of using CP-EPSs for predicting convective precipitation over only 3-month summer period with respect to a coarser resolution model. Dey et al. (2016) looked at the spatial agreement of the precipitation field over the UK. Also, Schwartz et al. (2010) showed the benefit of CP-EPSs for quantitative precipitation forecasts over 35 days of Spring 2007, stating that "there is much to be learnt about what type of information can be extracted from convection-allowing model output and how this information might best be used by operational weather forecasters."

In regard to studies comparing CP-EPSs with global coarser resolution model, Marsigli et al. (2008) compared the skill of COSMO-LEPS to Global EPS for precipitation, concluding that the extent to which COSMO-LEPS is more skillful is still not known and in Marsigli et al. (2005) the added value of COSMO-LEPS with respect to ECMWF-EPS for 24 h precipitation accumulation is quantified. This has been quantified both in terms of Brier Score and ROC (relative operating characteristic) area. It was found that COSMO-LEPS performs better than ECMWF-EPS when comparing maximum precipitation values over boxes of $1.5° \times 1.5°$, while the reverse is true when considering the average precipitation values.

Beck et al. (2016) showed the benefit of combining CP-EPSs from multiple forecast centres for conventional variables (wind, temperature, relative humidity, and precipitation). Gowan et al. (2018) showed the benefit of the NCAR-EPS system ((Schwartz et al., 2015)) for orographic precipitation over West USA for one season (October 2016 - March 2017). **?** focussed on convective precipitation on a sub-daily timescale (rainfall accumulation and rain rate), but on the tropical regions and for an extended period (2 years). They showed the added value provided by a convection-permitting model with respect to a coarser resolution global model, although not in a probabilistic framework.
Finally, Klasa et al. (2018) show the superiority of a CP-EPS run by MeteoSwiss for three precipitating events with respect to ECMWF-EPS, in terms of spatial distribution of the precipitation and spread-skill relationship. Among the few studies that attempted to investigate specific weather phenomena using CP-EPSs, Barrett et al. (2016) investigated four cases of quasi-stationary convective events, but still used precipitation accumulation as model variable to verify against radar observations. Other interesting studies using CP-EPSs to investigate specific weather phenomena are by Gallo et al. (2016) (tornadoes), Munsell et al. (2015) (hurricanes) and Trier et al. (2015) (convection initiation near the dry line). In particular, Gallo et al. (2016) use a CP-EPS to derive tornado probabilities based on thresholding updraft helicity and other environmental factors, finding that the higher-resolution environmental information benefits the probabilities, Munsell et al. (2015) examine the performance of a 60 members CP-EPS for a 5-day forecast of the hurricane Nadine. The track forecast was found to be satisfying, with 50 out of 60 members correctly predicting the trajectory, whereas the intensity forecast was not as skillful and Trier et al. (2015) use a 10-member ensemble of convection-permitting simulations with 3-km horizontal grid spacing to examine afternoon convection initiation (CI) along a central Oklahoma dryline in a case that produced tornadic supercells. They found that a subset of the ensemble members accurately reproduced both the timing and approximate location of CI, while, in others, it was both delayed and less widespread than observed.
Also Sobash et al. (2016) used updraft helicity variable to produce skillful "surrogate" severe weather probabilistic forecasts for day 1 and day 2 (e.g. strong wind gusts, large hail and tornadoes, as this is a surrogate for rotating thunderstorms). These were produced using 30-member convection-permitting ensemble forecasts initialized by an ensemble Kalman filter data assimilation. Ensembles have been shown to be more skillful than two deterministic forecasts, especially on the mesoscale.

1.2 Aims and motivations

It is useful to know what can be learnt from running convective-scale ensembles that would not otherwise be known from running cheaper global lower resolution ensembles (LR-EPSs), especially given the cost.
To what extent increasing the resolution of ensemble forecasts to convective scales allow one to make better quantitative probabilistic predictions about future weather events than the coarse resolution ensemble forecasts in which they are embedded ?
An interesting concept introduced to quantify any additional information provided by the convective-scale ensemble is that of *information gain*. Smith (2006) pointed out that

	Centre	Start (year)	Grid box size (km)	Ensemble size	Forecast length (h)
AROME-EPS	Meteo-France	2016	2.5	12	45
NCAR-ENS*	NCAR	2015	3	10	48
COSMO-DE-EPS	DWD	2012	2.8	20	27
COSMO-E	MeteoSwiss	2016	2.1	21	120
COSMO-LEPS	Arpae-SIMC	2005	7	20	132
MOGREPS-UK**	Met Office	2013	2.2	12	36
HarmonEPS***	Consortium		2.5	11	

TABLE 1.1: Convection-permitting ensemble versions from several research and forecasting centres.
*NCAR is no longer producing daily convection-permitting ensembles since 1 January 2018. A new convection-permitting ensemble is currently being developed at NOAA (for further information see https://rapidrefresh.noaa.gov/internal/pdfs/2018_Spring_Experiment_HRRRE_Documentation.pdf)
**MOGREPS-UK is running on an upgraded configuration since March 2016 (Hagelin et al., 2017)
***HarmonEPS is a relatively new system, used so far mainly for research. It is based on Hirlam Aladin. Regional/Meso-scale Operational NWP In Europe (HARMONIE) forecast system, which is developed within the HIgh Resolution Limited Area Model (HIRLAM), which is a project in collaboration between several European Countries (Ireland, Norway, Sweden, Denmark, Spain, Estonia, Iceland).

.

"There is no doubt that the current operational ensemble systems have value beyond that recognised in industry; this is an opportunity. The question should be seen as one of how to exploit this information content, not as to whether or not it exists". This fits in well with the aim of this thesis, the quantification of the information gain that convective-scale ensembles bring to probabilistic weather forecasts.
There are mainly three reasons why high resolution forecasts could contain more information about the future state of the atmosphere than a low resolution forecast:

1. Because a greater amount of information is extracted from observations at initial time due to the use of high resolution data assimilation.

2. Because the degree of approximation of the physical equations reduces as resolution increases, so less information about the physical processes of the atmosphere is neglected.

3. Because finer scale structure of the orography and topography is resolved, meaning that several local wind phenomena can be resolved (e.g. downslope wind, sea breezes).

When initial conditions for CP-EPSs are generated with a dynamical downscaling method, there is no additional assimilation of high resolution observation. Therefore, in this case, only the second and third type of information gain can exist.
Quantifying this type of information gain is a subtle issue and this is related to the extent to which the synoptic scale flow "controls" sub-synoptic scale dynamical phenomena. It is therefore sensible to compare a CP-EPS with its lower resolution counterpart, on the prediction of that type of sub-synoptic scale phenomena, which may benefit from the better representation of fine-scale structures (e.g. topography, coastline), but still slaved to the advection of large-scale flow. Some past studies (e.g.

In this thesis two weather phenomena have been chosen for the quantification of any information gain by the convective-scale ensembles: sea breezes and convective wind gusts. This choice has been mainly motivated by the following reasons:

- As mentioned earlier, there are only few studies concerning the quantification of specific weather phenomena, rather than simply grid-point variables.

- Sea breezes and wind gusts both lead to high-impact weather. Giving timely and accurate warning of severe weather was the fundamental reason to introduce ensemble models at convective-scale.

- Both the phenomena occur at small spatio-temporal scales. Nonetheless their occurrence is influenced by large-scale conditions. This allows to assess whether the information gain can stem from the dynamical downscaling of the coarser resolution ensemble or the latter is sufficient to provide valuable information.

The research questions of this thesis are:

1. How can the information about different weather events be extracted for an automated forecast production from both the convection-permitting model and its lower resolution counterpart ?

2. How do forecast probabilities calculated from both the ensembles differ ?

3. What is the forecast skill of the convection-permitting ensemble relative to a coarser resolution ensemble ? How does it depend on the predicted variable ? How does the skill evolves with forecast lead time ? Is any gain in the skill significant ?

4. Can be the potential information gain be predicted from the synoptic-scale flow ?

Furthermore, since sea breezes are atmospheric forms of the so called *density currents* and could lead to high-impact weather especially when interacting with other sea breezes or other mesoscale flows, this thesis also includes a study on the theory and simulations of colliding density currents, in an idealized framework. The general aim of this study was to understand the dynamics of collision and in particular the role exerted by the vorticity. The following questions are thus related to this study:

(a) Is a two-dimensional vorticity model able to capture the essential features of the collision dynamics, in an idealized framework similar to the laboratory experiments ?

(b) What is the role of baroclinically generated vorticity in determining and explaining some characteristics of the collision dynamics ?

(c) Can these characteristics be predicted using only the initial conditions, prior to the running of the numerical model ?

In the following chapters each of these questions will be addressed and discussed and possible explanations and insights about the results obtained will be given. Details about the structure of the thesis are described in the next section.

1.3 Outline of the thesis

This thesis is structured as follows into 7 chapters.
Chapter 2 deals with the research project on numerical simulations and theoretical considerations of colliding density currents. This is based on Cafaro and Rooney (2018).
In Chapter 3 the methodology and the ensemble models used in this thesis are described. Chapter 4 and 5 are based on Cafaro et al. (2019). Chapter 4 investigates how the ensemble data (both at convective-scale and lower resolution) have been post-processed in order to extract useful information about the sea breeze prediction and then to produce probabilistic forecasts. A brief description of the sea breeze phenomenology is also presented, focussing on the relevant characteristics to be predicted.
In chapter 5 the verification of the sea breeze probabilistic forecasts is performed against surface station observations. Other sea breeze characteristics (propagation speed and onset timing) are investigated and verified against the same set of observations.
In chapter 6 the same research questions and methodology are applied to a severe weather case, wind gusts in convective and non convective situations.
Finally, in chapter 7 the general conclusions, limitations of this study and suggestions for how to extend it and directions for future research are discussed.

Chapter 2

Density currents in an idealized framework

This chapter presents a new set of numerical simulations of two colliding density currents in a idealized framework, integrating the Boussinesq vorticity equation in a rectangular bounded domain, using a suite of contour-advection Lagrangian numerical models.
These simulations are used to examine the dynamical features of the collision, in the light of recent laboratory experiments. The collision dynamics present various interesting features. Here the interface slope at the front of the two unequal density currents and the maximum height reached by the fluid after the collision have been investigated. For the secondary triggering of atmospheric convection by colliding cold pools from previous convective events, these may affect the positioning and the momentum of the collision uplift, respectively. The aim is to verify the effect of the buoyancy ratio of the fluids on the two collision characteristics. Also, there is need of a theoretical model that could give us more insight about the collision dynamics and also predict the interface slope prior to the running of the numerical model.

The work presented in this chapter was published in the *Quarterly Journal of the Royal Meteorological Society*, with the reference:
Cafaro C. and Rooney G. G., (2018). Characteristics of colliding density currents: a numerical and theoretical study. Q. J. R. Meteorol. Soc.,**144**, 1761-1771.
The research was carried out by the lead author during the Met Office Training and Research Summer School in Summer 2017, under the supervision of G. G. Rooney. The author wrote the first draft of the paper and then discussions took place via e-mail and also in person three times, when the paper was under review. After the first draft, G. G. Rooney commented on the whole draft and re-wrote 30% of the article.

2.1 Introduction

Density currents are primarily horizontal flows generated by a horizontal density gradient. Since they are of practical application to several atmospheric and industrial problems there has been extensive experimental, observational, theoretical and numerical research on density currents, following the early works of von Kármán (1940) and Benjamin (1968). Much of this has been summarised and reviewed by Simpson (1997) and Linden (2012). Cold pool outflows from thunderstorms and sea breezes have been shown

to resemble density currents, by laboratory experiments (Simpson, 1969), field observations (Mueller and Carbone, 1987) and numerical models (Droegemeier and Wilhelmson, 1985).

Many theoretical models, based on energy conservation theory, have considered only irrotational flows. However it is well-known that horizontal density gradients lead to baroclinic vorticity generation (Simpson and Linden, 1989), and density currents may also be modelled using a vorticity formulation (e.g. Kay, 1992; Liu et al., 2003). Rotunno et al. (1988) used the vorticity formulation to model the phenomenon of cold pools in shear flows. Xu and Moncrieff (1994) also recognized the "importance of vorticity dynamics for the local flow in the vicinity of a density current". Furthermore Nasr-Azadani and Meiburg (2016) have developed a vorticity-based model of quasi-steady and supercritical (Fr $>$ 1) gravity currents. Finally Rooney (2018) discussed the role of the vorticity in defining the flow direction.

The interaction between density currents has also received some interest because, for instance, in the atmosphere the collision between mesoscale fronts can be an important trigger for atmospheric convection (Findlater, 1964). Intrieri et al. (1990) used lidar field observations to examine aspects of three different colliding thunderstorm outflows. These revealed large vortex rolls and smaller instabilities in the leading edge of the outflows. In addition, in all cases the colder, more dense outflow passed under the warmer outflow, making the warmer air mass the primary source of the updraft. All these collisions initiated new convection. The collision of sea breezes has been observed by Lapworth (2005) among others, and convective initiation due to colliding land breezes has been modelled by Wapler and Lane (2012). More recently, van der Wiel et al. (2017) have performed laboratory experiments on colliding density currents with different densities and depths. These parameters have an important influence on the characteristics of the collision (see also Shin et al., 2004; Warren, 2014). Laboratory experiments of equal colliding density currents have been performed more recently by Zhong et al. (2018) with the aim of understanding the turbulent mixing, especially in the collision stage.

As discussed by van der Wiel et al. (2017), even the simple case of (quasi) two-dimensional collision in a uniform environment has proved difficult to parametrise with an analytical model. Hence, this basic case is examined here. In the lower atmosphere, wind shear (Xu, 1992; Parker, 1996; Liu and Moncrieff, 1996; Xue et al., 1997; Lee and Wilhelmson, 1997; Bryan and Rotunno, 2014), ambient stratification (Liu and Moncrieff, 2000; Seigel and van den Heever, 2012) and three-dimensional front structure (Droegemeier and Wilhelmson, 1985; Orf et al., 1996; Vermeire et al., 2011) may all modify the cold-pool collision process and effects. These additional processes may begin to be included in the simulations once the basic collision dynamics are described.

Here it will be shown that the dynamics of two colliding density currents can be reproduced with an idealized two-dimensional, vorticity-based model. In section 2.2 numerical simulations in similar situations to those of the laboratory experiments of van der Wiel et al. (2017) are performed. In particular the pre-collision and collision stages are investigated. In section 2.3 some theoretical approaches to predict and explain the shape and orientation of the frontal interface are discussed, and predictions are compared with

the numerical simulations. In section 2.4 some preliminary simulations in a stratified environment are presented. Finally, section 2.5 contains conclusions from this study and ideas to extend it.

2.2 Numerical simulations

The laboratory experiments have been reproduced numerically using the numerical software package Hydra (http://www-vortex.mcs.st-and.ac.uk/software.html[1]) Hydra solves the Boussinesq-Euler equations for the flow of an incompressible fluid in two dimensions, vertical and horizontal. Buoyancy (also known as reduced gravity in a Boussinesq fluid) is approximately conserved following the flow, up to the action of numerical diffusion. The distribution of buoyancy governs the creation of baroclinic vorticity, and the vorticity redistributes the buoyancy. Numerical solution is by the combined Lagrangian advection method, (CLAM, see Dritschel and Fontane, 2010, for more details), which preserves the small-scale flow structures at high accuracy. This allows the problem of density-current collision to be recreated in an idealized framework, with minimal diffusion and no possibility of along-front variation. Hence, the basic processes may be isolated and quantified.

The dynamics produce entrainment and mixing as would be expected, so that fluid elements of different densities can mix quite thoroughly. There is no parametrised diffusion to homogenize the densities at the grid scale following this mixing, although numerical diffusion will act in this manner to some extent. Despite these simplifications, over the relatively short timescales of these simulations, the resultant dynamics are expected to still represent the significant features of the flow, and the numerical results closely resemble the flows in the laboratory experiments of van der Wiel et al. (2017), as shown in the following. Results from the numerical model have also been compared previously with experiments on billows in stratified flow by Carr et al. (2017).

2.2.1 Model description and initial configuration

The governing equations for an incompressible, inviscid, adiabatic fluid with velocity $\mathbf{u}$ are:

$$\rho \frac{D\mathbf{u}}{Dt} = -\nabla p + \rho \mathbf{g} \tag{2.1}$$

$$\frac{D\rho}{Dt} = 0 \tag{2.2}$$

$$\nabla \cdot \mathbf{u} = 0, \tag{2.3}$$

where $\mathbf{g}$=-g$\mathbf{k}$ is the acceleration due to gravity, ρ and p re the background density and pressure respectively. Then assuming small variations of density (*Boussinesq* approximation), ρ is replaced by ρ_0 a constant, except for the term dealing with vertical motions (the term multiplied by $\mathbf{g}$). Then the buoyancy is defined as:

$$b = g(\rho_0 - \rho)/\rho_0, \tag{2.4}$$

[1] Accessed online on 24 February 2017.

and also a reference pressure, $p_0 = -g\rho_0 z$ in hydrostatic balance with the reference density and then the pressure perturbation is introduced $p' = p - p_0$. Then substituting into equations (2.1-2.3):

$$\frac{D\mathbf{u}}{Dt} = -\frac{1}{\rho_0}\nabla p' + b\mathbf{k} \tag{2.5}$$

$$\frac{Db}{Dt} = 0 \tag{2.6}$$

$$\nabla \cdot \mathbf{u} = 0. \tag{2.7}$$

These are the Boussinesq-Euler equations for an incompressibile fluid. The corresponding equation for vorticity $\omega = \nabla \times \mathbf{u}$ is:

$$\frac{D\omega}{Dt} = -\nabla \times (\frac{1}{\rho_0}\nabla p') + \nabla \times (b\mathbf{k}). \tag{2.8}$$

Provided that the density is uniform, the first term on the r.h.s. vanishes. In a two dimensional plane in the $x - z$ plane, the vorticity has only the component of the normal to the plane of the flow, all other components being zero. Therefore the corresponding vorticity equations are:

$$\frac{D\bar{\eta}}{D\tau} = \frac{\partial \bar{b}}{\partial \bar{x}} \tag{2.9}$$

$$\frac{D\bar{b}}{D\tau} = 0 \tag{2.10}$$

$$\bar{\nabla} \cdot \bar{\mathbf{u}} = 0, \tag{2.11}$$

where overbars denote dimensionless variables, $\mathbf{u} = (u, 0, w)$ is the velocity field in two dimensions. η is the vorticity component normal to the plane of the flow. The dimensionless time, horizontal coordinate and vertical coordinate are denoted τ, $\bar{x}$ and $\bar{z}$ respectively. Their relationships to the equivalent dimensional variables t, x and z, and the non-dimensionalization of the flow variables, will be set out below. There is a free-slip boundary condition at all the boundaries of the rectangular domain. The domain is of height H and length L. Initially, the two patches of fluid denser than the background occupy regions in the lower corners at either side of the domain, of length l and depth D (i.e. the same for each current), see for example figure 2.1.

The current on the l.h.s. is denoted current 1 and the other is denoted current 2. The initial density scales are in proportion $\rho_0 \leq \rho_1 \leq \rho_2$, ρ_0 is the ambient fluid density. That is, current 2 is the more dense.

A uniform ambient density reduces the complexity of the system, since there is no influence of ambient stratification to consider. Its counterpart in the real atmosphere is the deep, well-mixed atmospheric boundary layer, such as occurs in the afternoon on days prone to convective activity. It is therefore also a suitable background state for idealisations of convective cold-pool collisions.

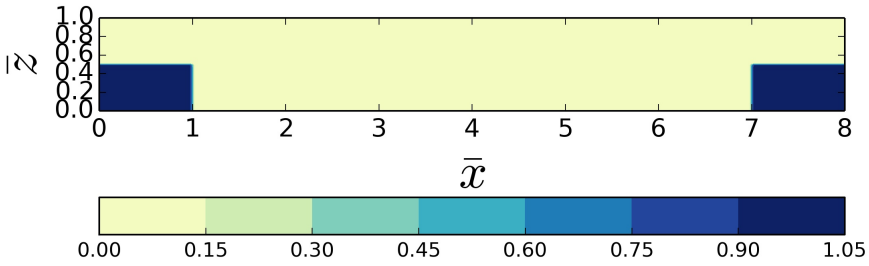


FIGURE 2.1: Initial ($\tau = 0$) distribution of the absolute value of the buoyancy, $|\bar{b}(\bar{x}, \bar{z})|$, for simulation 6 as in table 2.1.

The precise initial distribution of the buoyancy (see figure 2.1) is specified as follows:

$$b(\bar{x}, \bar{z}) = \left[-\frac{1}{2} b_1 \cdot \mathrm{erfc}\{(\bar{x} - l)/\varepsilon\} - \frac{1}{2} \cdot b_2 \, \mathrm{erf}\{(\bar{x} - L + l)/\varepsilon\} - \frac{1}{2} b_2 \right] \cdot \phi(\bar{z}), \tag{2.12}$$

where

$$\phi(\bar{z}) = \frac{1}{2} \mathrm{erfc}\{(\bar{z} - D)/\varepsilon\}, \tag{2.13}$$

describes the variation in height of the buoyancy (to perform partial-depth simulations) and $\varepsilon = 0.02\, l$ describes the thickness of a transition zone between the current and the background flow. b_1 and b_2 are the buoyancies for current 1 and 2 respectively and l the width of each box occupied by the two currents at the two sides of the rectangular domain. Moreover

$$\mathrm{erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^{x} e^{-s^2} ds, \qquad \mathrm{erfc}(x) = 1 - \mathrm{erf}(x), \tag{2.14}$$

are the error function and its complementary, respectively.

D	H	r_g							
		0.99	0.75	0.49	0.33	0.25	0.20	0.15	0.05
1	1	1		2	3	4	5		
1/2	1	6*	7*	8*	9	10*	11*	12*	
1	2	13	14	15	16	17	18	19	20
1	4	21	22	23			24		
1	8	25		26	27				

TABLE 2.1: The numerical simulations and their dimensionless parameters. The domain is of height H and length L, and each current has initial depth D and length l. The buoyancy ratio of the currents is given by r_g. The simulations are numbered as shown in the table body. Simulations marked with an asterisk (*) were repeated with the l.h.s. current only to generate the single-current data.

The variables are non-dimensionalized relative to the denser current,

$$\bar{x} = x/l, \qquad \bar{z} = z/l \tag{2.15}$$

$$\tau = t/\sqrt{l/|b_2|} \tag{2.16}$$

$$\bar{\eta} = \eta/\sqrt{|b_2|/l} \tag{2.17}$$

$$\bar{b} = b/|b_2| \tag{2.18}$$

$$\bar{\mathbf{u}} = \mathbf{u}/\sqrt{|b_2|l} \tag{2.19}$$

The numerical solution will depend on the dimensionless parameters D/H, l/L, H/L and $r_g = b_1/b_2$. Here, simulations for different D/H, H/L and r_g are conducted, varying D and H while $l = 1$ and $L = 8$ are fixed, see table 2.1. As will be seen, variation in r_g was achieved by keeping b_2 fixed and varying b_1.

Simulations 25–27 on the square domain had a horizontal resolution of 256 points, and all others had a horizontal resolution of 512 points. The vertical resolution was in proportion, according to the domain aspect ratio H/L, so as to obtain square grid cells; thus 256 points for simulations 25–27, 64 points for simulations 1–5, etc. The numerical Reynolds number may be estimated from the grid resolution to be $Re_{num} \gtrsim 10^6$ (Carr et al., 2017, § 4.2.2), bearing in mind that these simulations are two-dimensional.

2.2.2 Results

In this section the numerical simulation results are presented for the pre-collision (see figure 2.2) and collision phases. Animations of some of the simulations are available as supplementary material of (Cafaro and Rooney, 2018). For quantitative analysis, only the partial-depth simulations are examined, since these will be more representative of cold-pool type flows in the lower atmospheric boundary layer.

Pre-collision propagation speed

The pre-collision phase is investigated by considering the horizontal propagation speed of current 1 for different r_g. One way to describe this is via the dimensionless horizontal

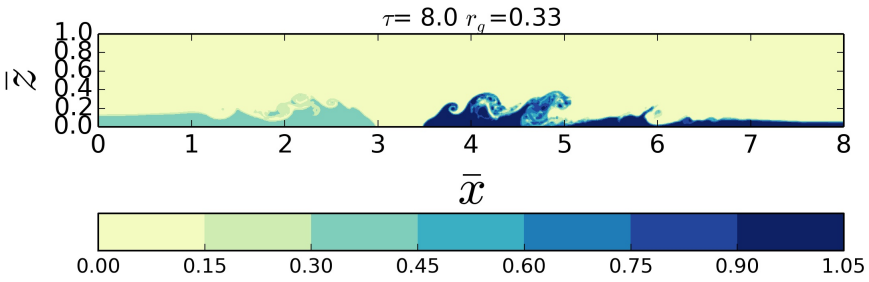


FIGURE 2.2: Snapshot of buoyancy distribution before the collision for simulation 9 at time $\tau = 8$.

propagation speed given by the Froude number (Fr):

$$\mathrm{Fr} = \frac{U}{\sqrt{bD}}, \tag{2.20}$$

where U is the current propagation speed.
Several theoretical predictions of Fr based on the initial fractional depth of the current have been developed (Benjamin (1968), Rottman and Simpson (1983), Shin et al. (2004)). In particular in Shin et al. (2004), for Boussinesq currents, Fr is expressed as follows:

$$\mathrm{Fr} = F_S = \frac{\sqrt{2 - D/H}}{2}, \tag{2.21}$$

where D and H represent the fluid depth and the tank height respectively. Equation 2.21 has been compared by van der Wiel et al. (2017) with values given by laboratory experiments.

Fr was calculated from the numerical simulations using (2.20). For this, b and D are given from the initial state. The speed U was calculated by tracking the leading edge of the density-current front for several timesteps [2], using the buoyancy field as a tracer, and then taking the mean value. The leading edge is tracked at each time step by calculating the buoyancy difference between adjacent grid boxes at the lowest vertical level. The location of the leading edge is where this difference exceeds a threshold, which is $0.1(b_1 -$

[2]The model timestep was not chosen by the user. As written in Carr et al. (2017) "Temporal integration was performed using an adaptive fourth-order Runge–Kutta integration scheme." Further details of the method can be found in Dritschel and Fontane (2010).

b_0), where b_1 and b_0 are the buoyancies of the denser current and the background fluid respectively.

The results of the comparison between (2.20) and (2.21) are shown in figure 2.3. Equation 2.21 is shown to be a good predictor for the magnitude of the propagation speed in the constant-velocity phase[3], especially for $r_g \geq 0.49$ although the agreement for $D/H = 1/2$ is better than that for $D/H = 1/4$.

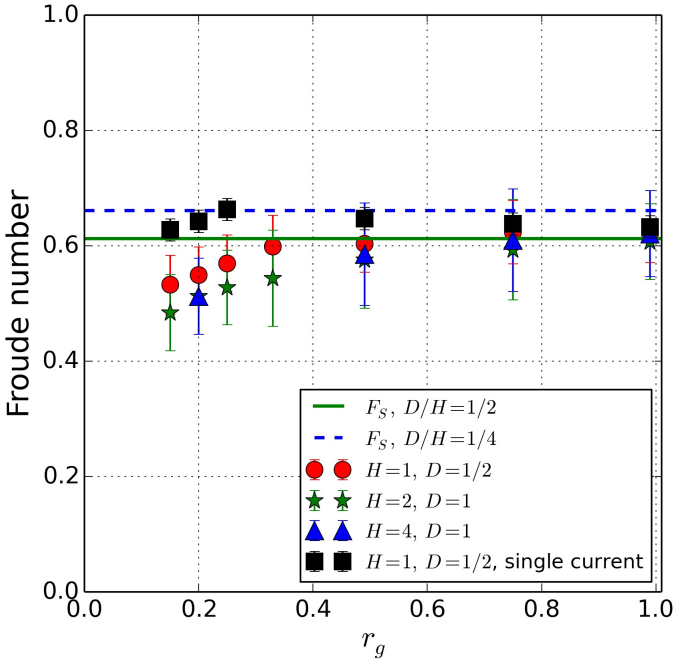


FIGURE 2.3: Froude numbers versus r_g. Error bars show one standard deviation with respect to the mean of density current speeds for different time steps. Continuous lines are theoretical values based on (2.21). The simulations are a subset of those listed in table 2.1.

(2.20) has been calculated also for the single density current case, for $H = 1$ and $D = 1/2$, as also shown in figure 2.3. The single-current simulations are initialised as the two-current simulations but omitting the dense fluid on the r.h.s. In the single-current case the current propagates faster than if there is a current on the other side, as is shown by the Froude number being greater by up to approx. 16% in these experiments. This means that in the two-current case, the presence of the second current exerts a small influence on the propagation of the first current, even before collision. It is expected that this effect is greater here, in a purely two-dimensional domain of limited extent, than it would be in the atmospheric boundary layer.

Given that the single-current speeds appear to be more uniform with r_g, the slight fall-off of the speed of current 1 at lower r_g values seems to be an effect of having an oncoming density current of significantly greater momentum, therefore presumably driving a relatively stronger opposing flow in the ambient.

[3]The constant-velocity phase is the initial phase of the density current propagation as opposed to the subsequent so-called *similarity phase*

Collision characteristics

In this section the numerical results for the collision of two density currents are described. The numerical simulations presented here are able to reproduce the main features of the collisions observed in the laboratory experiments of van der Wiel et al. (2017). For direct comparison between the numerical simulations and laboratory experiments, see the supplementary videos S1 and S3 of (Cafaro and Rooney, 2018), to be compared with with videos S1 and S2 of van der Wiel et al. (2017).

Figure 2.4 shows Hovmöller diagrams based on the lowest level of the gridded model output for simulations 13 and 16. Before the collision the two currents propagate at certain speed (which was found different from the single current case, as shown in the previous section). After the collision the denser current propagates at slightly slower rate than before the collision, for the non symmetric case ($r_g = 0.33$). Also, there is mixing occurring in the collision zone. This is in agreement with van der Wiel et al. (2017), fig. 8.

For comparison with the Hovmöller plots, and with the experiments of van der Wiel et al. (2017), examples of the buoyancy and vorticity fields during symmetric and asymmetric collisions are shown in figure 2.5. The vorticity distributions in these cases show a significant difference at the interface. Namely, in the symmetric case (figure 2.5b) the frontal vortex sheets combine into a vortex pair that propagates vertically upward, whereas in the asymmetric case (figure 2.5d) the stronger vortex sheet at the front of the more dense (r.h.s.) current remains spanning the depth of the interface while the weaker vortex sheet is deflected upward along it.

Interface shape

The interface shape may be characterized to some extent by its angle γ to the horizontal. Figure 2.5 indicates that the angle is not constant with respect to time, and figure 2.6 shows that the interface shape becomes more complex at later times. Nonetheless, the measured angle at a particular dimensionless time post collision indicates how the interface varies from one simulation to another.
The angle between the two fluids, with respect to the horizontal line is given by:

$$\gamma = \arctan\left(\frac{\Delta z}{\Delta x}\right) \tag{2.22}$$

The algorithm to calculate the angle is based on finding the coordinates of discontinuities at two different vertical levels in the buoyancy profiles (i.e. where the two fluids meet).The vertical levels were taken from the interface, identified by the buoyancy discontinuity, in the height range between $0 < \bar{z} < D/2$. Three points in the upper part of this range were considered along with two points in the lower part, to generate six values of the interface angle. From these, a mean and standard deviation were calculated. The angle has been calculated at time $\tau = 0.5$ *after the collision*. The collision time is defined as the time at which the lowest levels of the two fluid meet, i.e. when there is not any grid point occupied by the ambient fluid between the two fluids (see, for instance, the

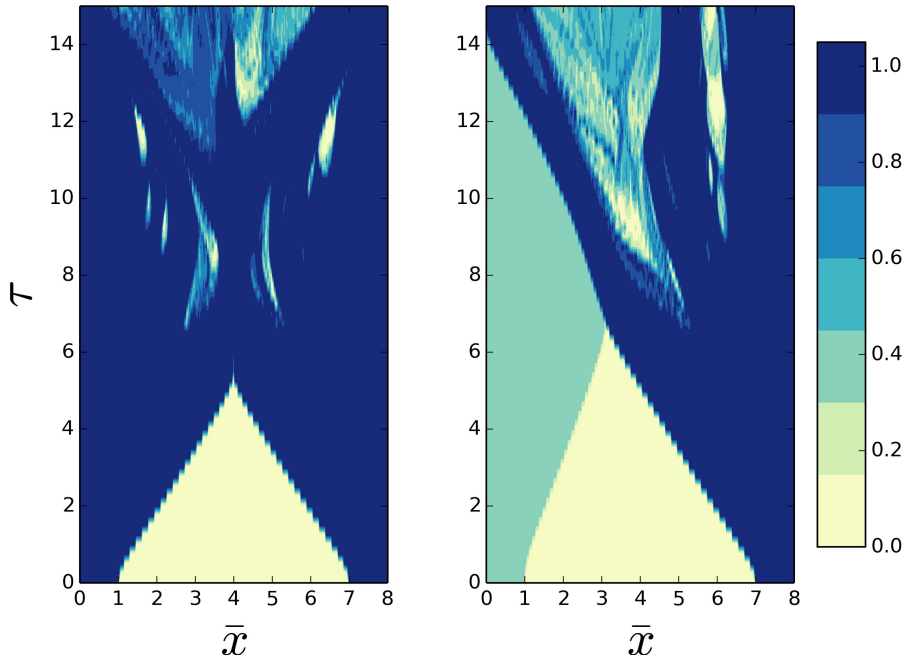


FIGURE 2.4: Hovmöller diagrams of the buoyancy field for simulation 13 (left) and simulation 16 (right) at height $\bar{z} = 0$

first panel of figure 2.5a). The results are shown in figure 2.7, for different r_g and for different H and D. It can be seen that the angle does not have a clear dependence on D/H, although generally increases with increasing r_g.

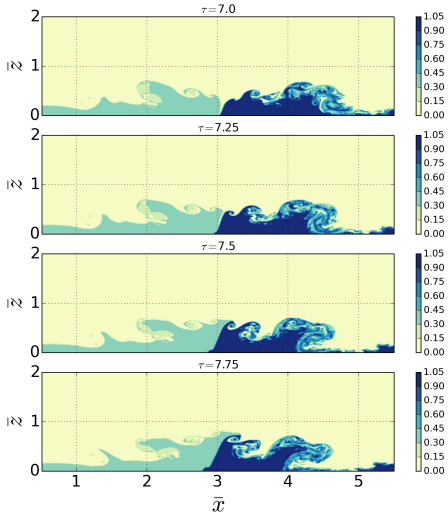
Collision maximum height

Another important feature of the collision is the maximum height reached by the two fluids. In an atmospheric context, this may be a significant factor in whether colliding cold-pool fronts are able to initiate new convective cells.

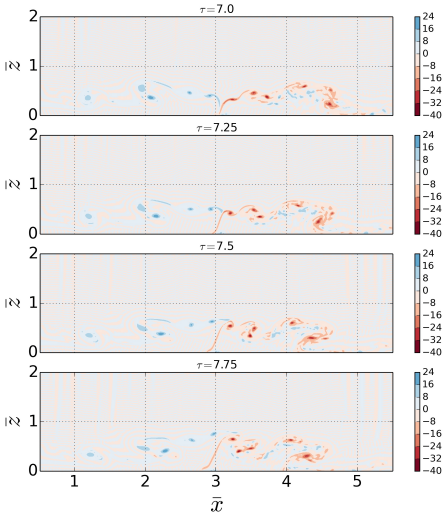
In the symmetric case $r_g = 0.99$, the front is almost vertical (figures 2.5c, 2.7), while in the asymmetric cases ($r_g < 0.99$) the front is tilted. For later time steps in the non-symmetric case the denser fluid propagates underneath the less dense fluid, compare figures 2.5a and 2.5c. Recall that this type overlapping was also observed by Intrieri et al. (1990), and while it means that the perturbation height may be insensitive to r_g, it has a possible bearing on the properties at the source of the updraft. To illustrate this interaction, figure 2.6 shows the simulation evolution after the collision. It can be seen that the maximum height reached by the fluid in the symmetric case (figure 2.6a) is comparable to the maximum height in the asymmetric case (figure 2.6b).

This figure also shows the spatio-temporal domain over which the maximum height has been calculated.

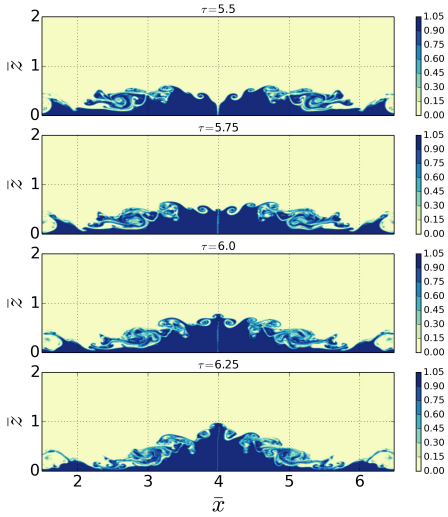
This is done by counting the number of gridboxes G at each level where the the absolute value of the buoyancy is above a threshold of $0.8r_g$, which takes into account that different buoyancies may be present. In general G decreases with height. Then the maximum height is defined as the height at which the last non-zero value of G occurs.



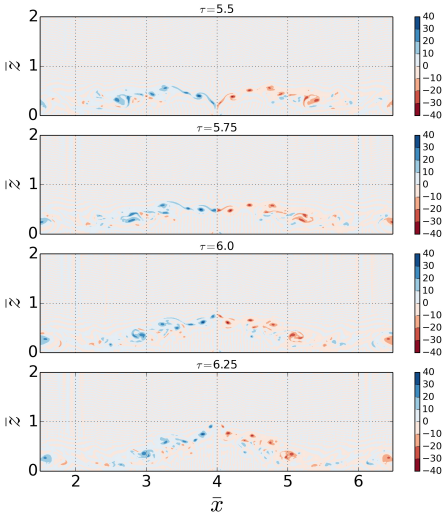
(A) Buoyancy $r_g = 0.33$



(B) Vorticity $r_g = 0.33$

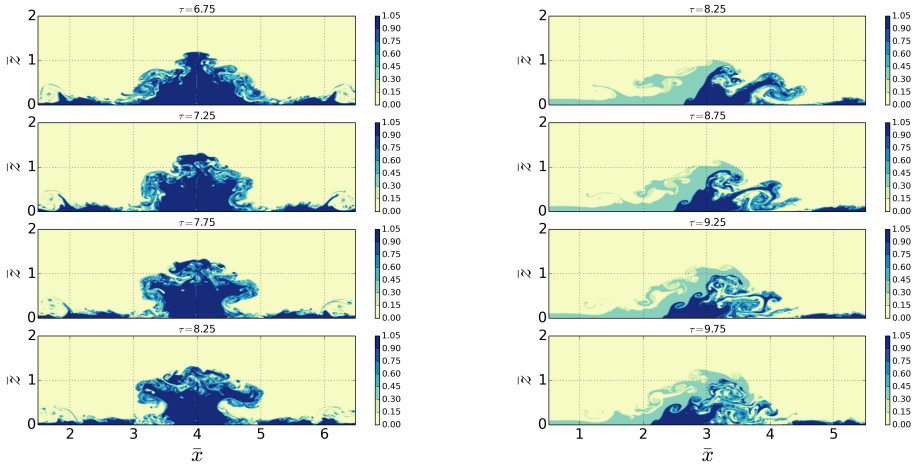


(C) Buoyancy $r_g = 0.99$.



(D) Vorticity $r_g = 0.99$.

FIGURE 2.5: Time sequence of buoyancy and vorticity profiles for two different r_g near the moment of collision for simulations 13 and 16.



(A) Buoyancy $r_g = 0.99$. (B) Buoyancy $r_g = 0.33$.

FIGURE 2.6: Evolution of buoyancy profile after the collision for simulations 13 and 16.

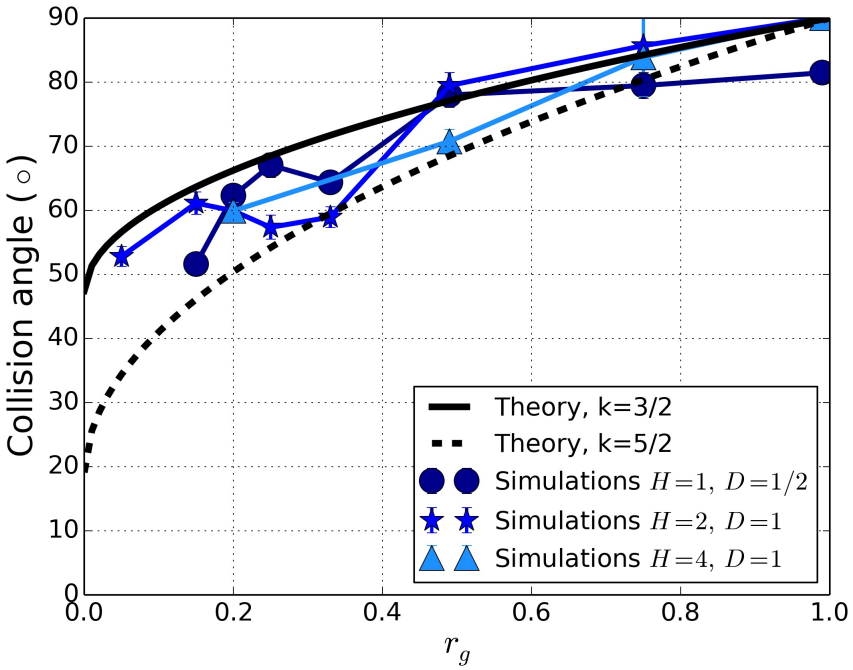


FIGURE 2.7: The collision angle γ calculated from numerical simulations for different values of H and D. Error bars represent one standard deviation with respect to the mean of the different angles measured at different heights of the buoyancy profile. The figure also shows the collision angle values calculated from theory. The dashed and solid lines represent (2.26) for different values of the coefficient k.

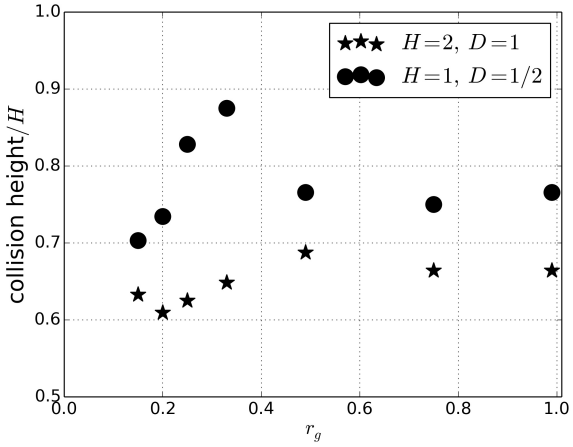
The results of this calculation are shown in figure 2.8. Since both D/H and H/L vary in these simulations, the results are plotted in two ways. Figure 2.8a shows the height normalized by the domain height H when $D/H < 1$ is fixed but H/L varies. Figure 2.8b shows the height normalized by the initial current height D when $D/H < 1$ varies. Figure 2.8c shows the height normalized by the domain height when $D/H = 1$. This last case represents the only direct correspondence in parameter space between the numerical simulations and the laboratory experiments of van der Wiel et al. (2017), whose results are also plotted for comparison.

As previously found by van der Wiel et al. (2017), it is not possible to infer any strong relationship between the maximum collision height and r_g. Figures 2.8a and 2.8b show that, for these data, the collision height seems to scale primarily with D. Figure 2.8b shows that, for $D/H \leq 1/2$, increasing the relative height of the domain does not greatly affect the collision height, relative to the initial current depth. Any effect due to the finite vertical extent of the domain would therefore seem to be quite small, and these simulations all produce maximum heights approximately in the range $1.2D$–$1.4D$, i.e. higher than the initial height of the dense fluid. This contrasts with the $D/H = 1$ simulations and experiments, in which the maximum height is of necessity restricted to be less than or equal to D, and the results shown in figure 2.8c indicate typical values of $0.87D$–$0.98D$. The domain would therefore seem to be restricting the collision effects in this configuration. Zhong et al. (2018) investigated the maximum height after the collision in a full-depth ($H = D$) laboratory experiment of two equal counterflowing density currents. The maximum height was first predicted to be 0.75 H, based on a formula involving frontal vertical velocities. However the observations of the experiments suggested a maximum height of 0.9 H, which is comparable with the range of values in our numerical experiments. Nevertheless figure 2.8c shows that the full-depth simulations generally produce a slightly higher collision height than the laboratory experiments. This could be due to dissipation in laboratory experiments, different boundary conditions (on 3 sides, and interface movement on the 4th), and also 3-dimensional effects in the experiments, which might assist in lowering the height if there is a possibility of sideways movement. Therefore, results of Zhong et al. (2018) are very similar to van der Wiel et al. (2017)'s, despite of the different measurement techniques used.

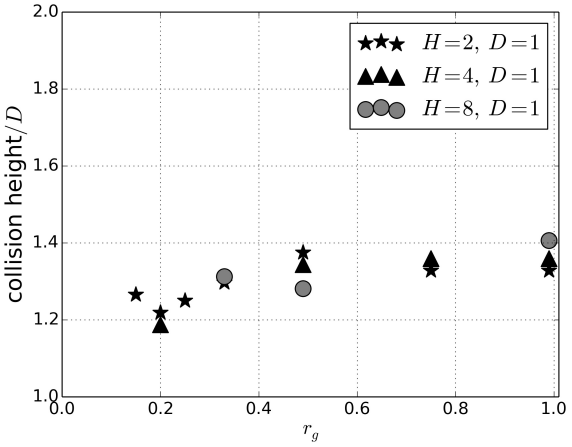
2.3 Theoretical considerations

In the previous sections the output of the numerical model solving (2.9–2.11) has been presented. In particular in figure (2.7) the variation of the collision angle with the density ratio r_g of the two density currents. In this section two different models of the collision angle are presented, based on the properties of the two incoming density currents. This will aid the future parametrization of these processes in large-scale models which are unable to resolve the details of such flows.

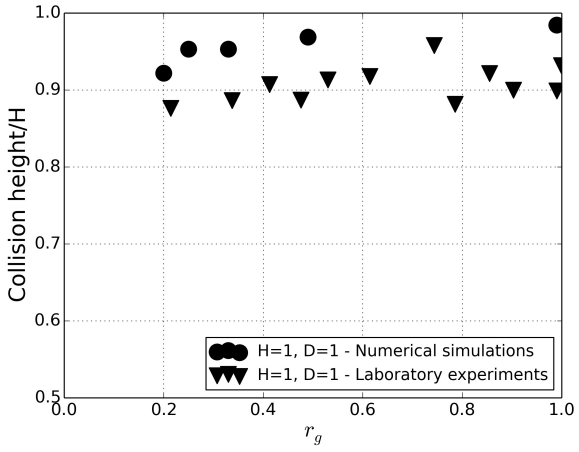
The similarity between the numerical results presented here and the experiments of van der Wiel et al. (2017) indicates that the idealized equations capture the essential features of the flow. Since the numerical model has buoyancy and vorticity as its dynamic prognostic variables, this indicates that the baroclinic vorticity plays a crucial role in the



(A) Numerical simulations $D/H = 1/2$.



(B) Numerical simulations $D = 1, H = \{2, 4, 8\}$.



(C) Full-depth numerical simulations compared with the full-depth laboratory experiments of van der Wiel et al. (2017).

FIGURE 2.8: Comparison of maximum collision heights: (a) numerical simulations with $D/H = 1/2$ for two values of H, (b) numerical simulations with the same D but different H, and (c) comparison of full-depth ($H = D = 1$) maximum collision heights between numerical simulations and laboratory experiments. The maximum collision height is scaled with H in (a) and (c) and with D in (b).

front and collision dynamics. Thus the flow has been modelled in terms of the vortex dynamics.

For a steadily translating density current, the vorticity generation is balanced to first order by the horizontal current motion. It is conjectured that during a density-current collision, this balance is lost, and takes some time to be recovered. This implies that the dynamics during the collision are controlled by the motion of the current-head vortices which are already established.

Note that in the rest of this section the main concern is with the collision and its aftermath, hence time is taken as relative to the moment of collision. That is, $\tau = 0$ is re-defined as the collision moment, identified as the first contact of the density currents at the lowest level.

2.3.1 Vortex-pair model

It is well-known (e.g. Lamb, 1945, Article 155) that two isolated line vortices rotate around the centre of their system with angular velocity

$$\Omega = \frac{\Gamma}{2\pi\delta^2} \tag{2.23}$$

where Γ is the sum of the vortex circulations and δ is the distance between them. For vortices with different-signed circulations, the centre of the system is on the far side of the stronger vortex compared to the position of the weaker vortex. Here, if the heads of the density currents are likened to regions of two-dimensional vorticity, considering the shear across the current indicates that the magnitude of the circulation is proportional to the product of the current speed and head height, $\Gamma_i \propto U_i h$. When equal-height currents interact in a collision, therefore, the net angular velocity of fluid in the collision zone may be approximated by

$$\Omega \propto \frac{U_2 - U_1}{h} \tag{2.24}$$

assuming that the separation of the centres of circulation is also proportional to the head height, i.e. $\delta \sim h$. For two vortices, the net circulation sets the direction of rotation of the pair. Likewise, the direction of the circulation here will be in the same sense as the circulation in the faster i.e. denser current. This implies that the relatively less-dense current will move over the top of the denser current, as is observed.

The currents have very similar front shapes and heights prior to collision so it is assumed that the collision interface starts from the vertical, and that the angle will evolve simply as the product of Ω and time. This gives the front angle β to the vertical at time t after collision as

$$\begin{aligned}
\beta &= k_0 \frac{U_2 - U_1}{h} t \\
&= k_1 \mathrm{Fr} \sqrt{b_2} \left(1 - \sqrt{b_1/b_2}\right) \frac{t}{\sqrt{h}} \\
&= k_1 \mathrm{Fr} \left(1 - r_g^{1/2}\right) \tau \left(\frac{l}{h}\right)^{1/2}
\end{aligned} \tag{2.25}$$

where Fr is the Froude number defined in (2.20) and k_0 is an unknown proportionality constant and $k_1 = \sqrt{2}k_0$ which absorbs the ratio $h/D \approx 0.5$ in the Fr definition. It has been assumed that, with the Boussinesq approximation, the Froude numbers of the two currents are approximately equal, which is a common assumption, see for example (2.21). (Here they have small differences in the range 0–0.1.)

This model is approximate, and is only expected to apply for a short interval during the collision. Nonetheless, it has two implications. The first, from (2.23), is that, at a given dimensionless time after collision, the angle is proportional to the net circulation in the collision zone. This follows fairly immediately from the application of this type of model. The second, from (2.25), is a parabolic dependence of the measured angle on the density ratio, with a vertical interface in the case $r_g = 1$.

The first of these relationships may be tested by considering the total, or net, vorticity in a region of the simulation domain (always of the same area) at a particular time after the collision, and covering the point of collision. An example of the vorticity evolution in such a region is shown in figure 2.9.

Figure 2.10 shows how this net vorticity, taken to be equivalent to the circulation per unit (constant) area of the box over which the vorticity is summed, varies with the mean interface angle measured at the same time. In this plot the net vorticity is multiplied by the time interval since the collision to correspond more closely to (2.25). It can be seen that the data appear to follow a general linear trend with increasing net vorticity, which is in agreement with the model prediction.

For the second, the angle to the vertical at any particular time τ is given by (2.25). Taking $\tau = 0.5$ the angle to the horizontal is then

$$\gamma = \frac{\pi}{2} - \beta = \left(\frac{\pi}{2} - k\right) + k\, r_g^{1/2}. \tag{2.26}$$

The coefficient

$$k = \frac{k_0}{2}\mathrm{Fr}\left(\frac{l}{h}\right)^{1/2}. \tag{2.27}$$

may be determined from the numerical results at $\tau = 0.5$.

Figure 2.7 shows the comparison of the theory with the numerical values. The relatively small range of $3/2 \leq k \leq 5/2$ covers most of the spread of angles calculated numerically, hence a value of $k = 2$ may be taken as representative. Hence

$$k_0 \approx 4\,\mathrm{Fr}^{-1}\left(\frac{l}{h}\right)^{-1/2}. \tag{2.28}$$

This simple model of the angle dependence obtained from approximating the vorticity dynamics would not seem to be derivable from considerations of conservation of energy or momentum alone, but despite its simplicity and approximate nature it fits the data reasonably well. Also, as has been shown, the present results from a numerical model in which the main variables are vorticity and buoyancy reproduce many of the features of the experiments of van der Wiel et al. (2017) (see also the supplementary flow animations). This demonstrates that the consideration of vorticity is crucial in fully understanding the dynamics of colliding density currents.

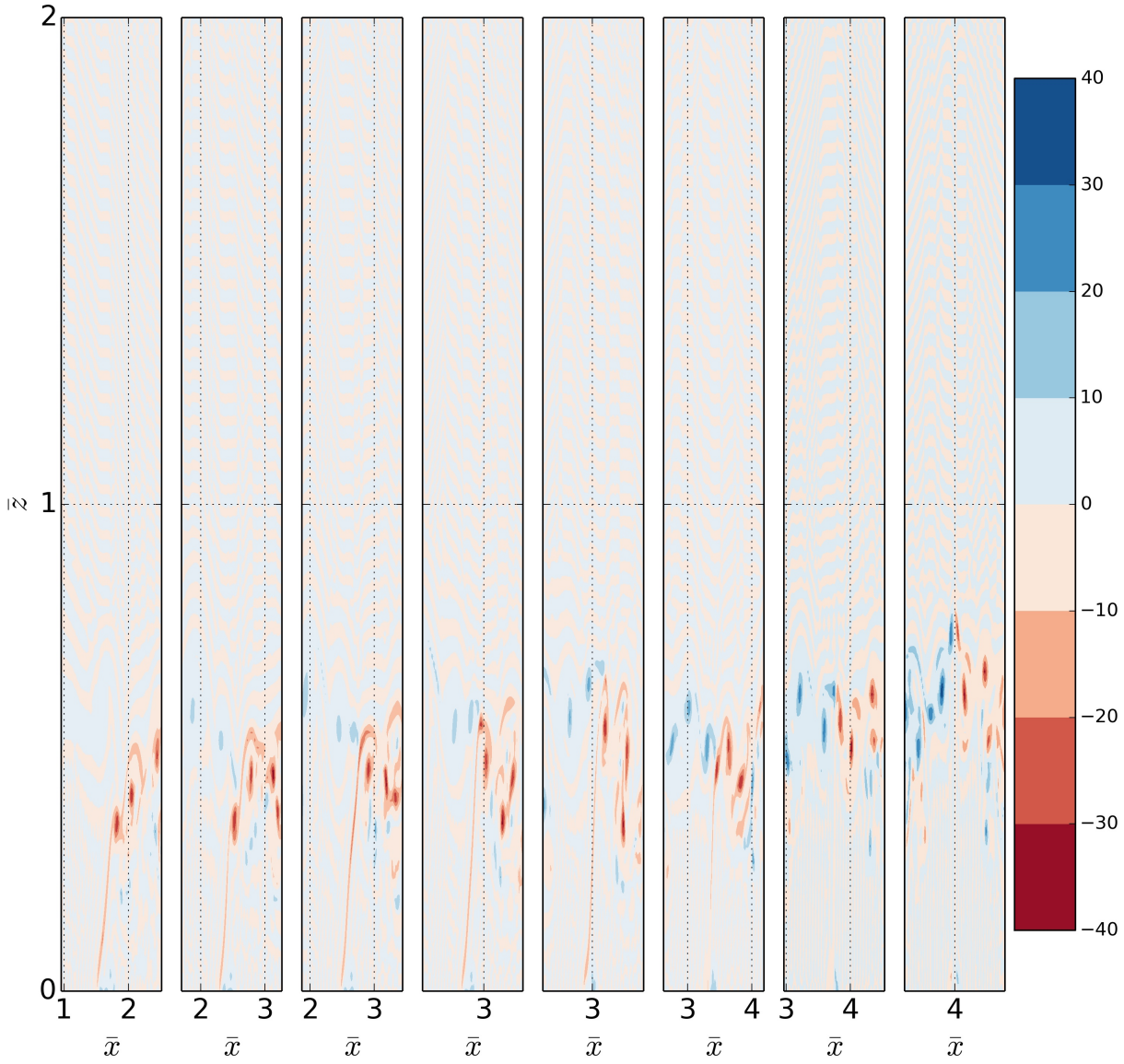


FIGURE 2.9: Snapshot of the vorticity field 2τ after the collision for different r_g (simulations 13-20 in order from right to left), in the two dimensional domain where η has been integrated as in figure 2.10. The width of the domain is the same for all r_g, and equal to $0.2L$.

2.3.2 An alternative approach for the interface shape

Consider (2.9). First note that:

$$\frac{D\bar{\eta}}{D\tau} \approx \frac{\bar{\eta}}{\tau} \tag{2.29}$$

Remembering that $\bar{\eta} = \omega \cdot \mathbf{j} = \dfrac{\partial \bar{u}}{\partial \bar{z}} - \dfrac{\partial \bar{w}}{\partial \bar{x}}$ and discretizing the (2.9):

$$\frac{\Delta \bar{u}}{\Delta \bar{z}} - \frac{\Delta \bar{w}}{\Delta \bar{x}} = \tau \frac{\Delta \bar{b}}{\Delta \bar{x}} \tag{2.30}$$

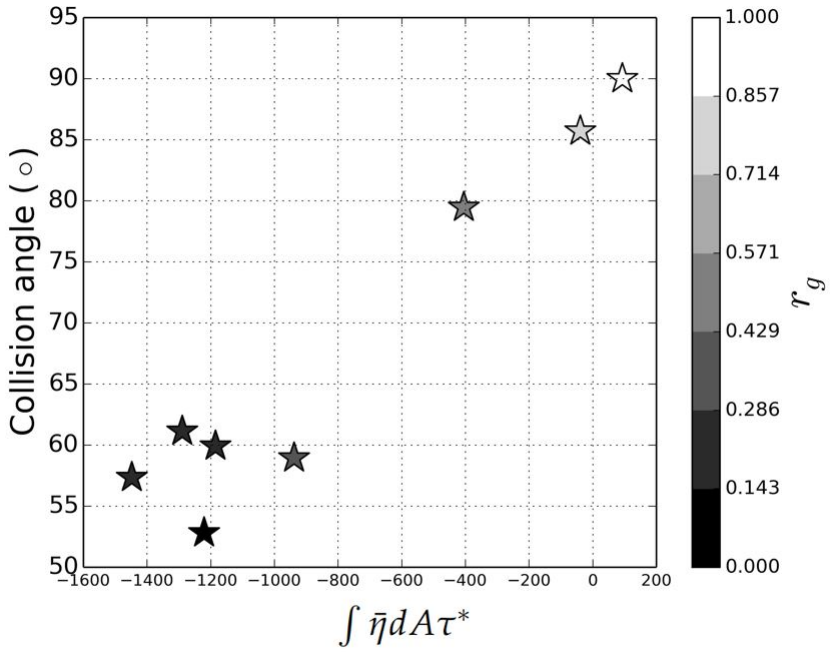


FIGURE 2.10: Relationship between the integral of the vorticity in the 2D box as described in figure 2.9 at $\tau^* = 0.5$ post-collision and the collision angle γ measured for simulations 13-20 as in table 2.1, with a colorbar representing r_g.

In order to calculate γ the angle formed by the two fluids with the horizontal line, it can be seen that

$$\tan\gamma = \frac{\Delta\bar{z}}{\Delta\bar{x}}.$$

Rearranging (2.30):

$$\tan\gamma = \frac{\Delta\bar{z}}{\Delta\bar{x}} = \frac{\Delta\bar{u}}{\tau\Delta\bar{b} + \Delta\bar{w}}, \tag{2.31}$$

where $\Delta\bar{b} = \bar{b_2} - \bar{b_1}$ is the buoyancy difference across the interface, $\Delta\bar{u} = \bar{u_C} - \bar{u_B}$ the difference in the horizontal velocity along the vertical line and $\Delta\bar{w} = \bar{w_B} - \bar{w_A}$, the difference of the vertical velocity along the horizontal line (see figure 2.3.2 for a schematic representation of the terms involved) and τ the moment at which the angle is calculated.

Since the main focus of this study is in the collision moment, as initial condition the situation at time $\tau = 0.5$ after the collision is considered.

In regard to $\Delta\bar{u} = \bar{u_C} - \bar{u_B}$, it can be simply noticed that

$$\bar{u_C} = U_1 \tag{2.32}$$

$$\bar{u_B} = -U_2, \tag{2.33}$$

where U_1 and U_2 are the propagation speed of the light and dense current respectively with the sign indicating the direction of propagation.

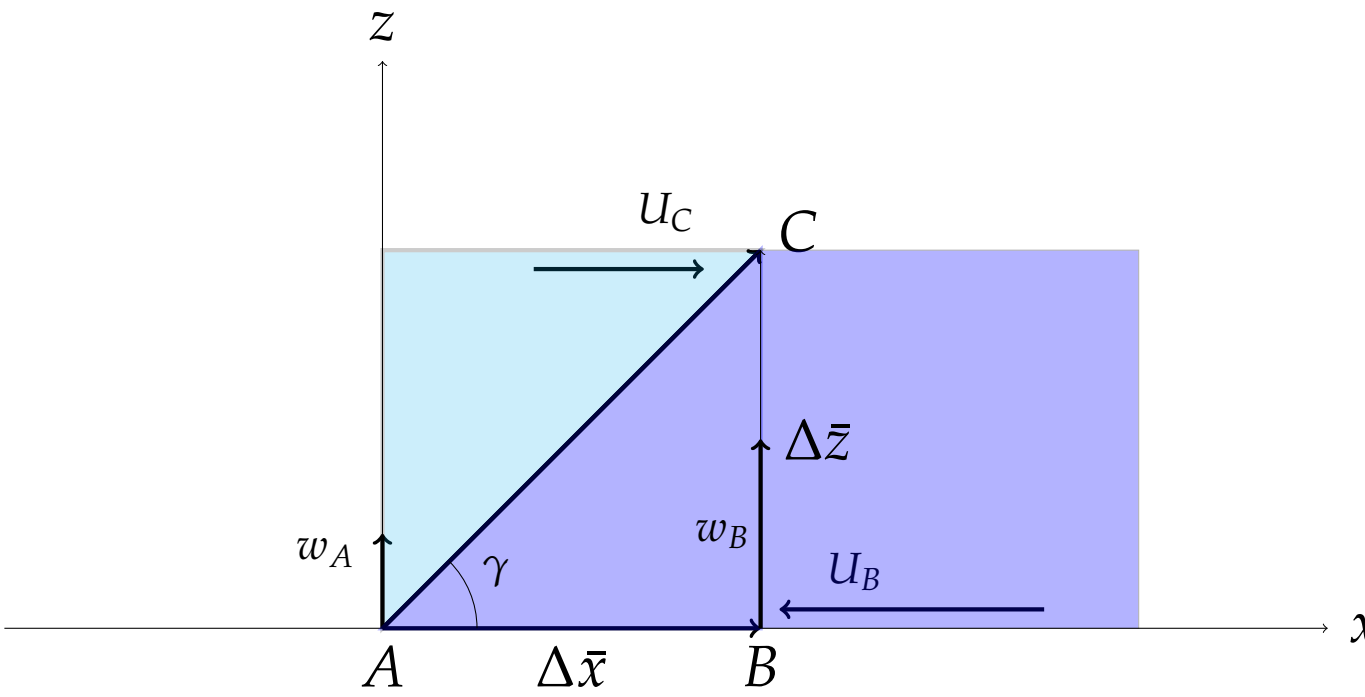


FIGURE 2.11: Schematic diagram at the collision moment of the two currents (in blue the less dense current on the left side and in purple the denser current on the rigth side). The line AC separates the two currents.

For Δw the following assumptions are made:

$$w_B = k_w|U_2|, \tag{2.34}$$

$$w_A = k_w|U_1| \tag{2.35}$$

which is a direct consequence of the continuity equation 2.11, where k_w is a factor to be determined. [4]

Finally $\tau = 0.5$ will be used, since this time step has been used to calculate the angle from numerical simulations with which the predictive formula (2.31) will be compared.

Therefore substituting (2.35), (2.32), (2.33) into (2.31):

$$\gamma = \arctan\left(\frac{U_1 + U_2}{\tau\Delta b + k_w(|U_2| - |U_1|)}\right) \tag{2.36}$$

where the propagation speed u_i is predicted using the Froude number Fr as follows:

$$U_i = \mathrm{Fr}\sqrt{b_i D} \tag{2.37}$$

k_w has been varied in the interval $[0, 1]$, calculated γ for each value of k_w and then the mean of the different values of γ calculated for different k_w has been considered. The maximum standard deviation with respect to the mean value, calculated for different values of r_g, D, H is found to be $\sim 2.70°$. The comparison between (2.36) and the experiments are shown in figure 2.12. It is clear that also this formula is able to predict the

[4] Please note that if the denser current is put on the left then the sign of propagation would change accordingly and this will change also the angle formed with the horizontal.

relation the collision angle and r_g.

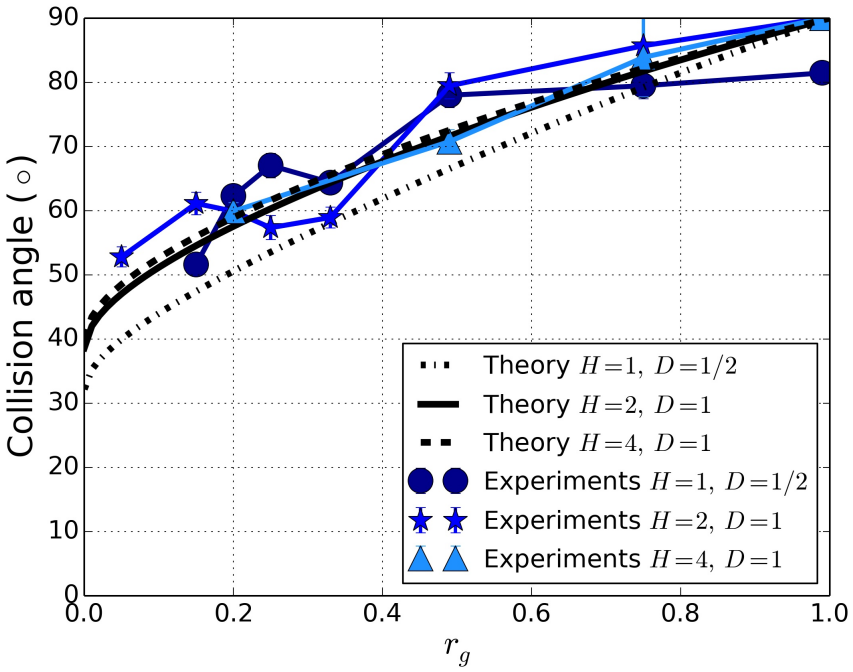


FIGURE 2.12: Collision angle γ measured to the horizontal as calculated in the experiments. The error bar describes one standard deviation with respect to the mean of the different angles measured at different heights of the buoyancy profile. The dashed lines show (2.36) the mean value of the angle for $0 \leq k_w \leq 1$.

2.4 Simulation in a stratified environment

In the numerical simulations presented in the previous sections, a uniform ambient fluid with density ρ_0 has been considered. Here a non uniform ambient fluid is considered. Only a visualization of the fluid features is presented, without performing any quantification of the fluid characteristics (e.g. propagation speed, maximum height, interface angle).

This is in order to provide evidence that, for the future work, stratification effects can be included also in the Hydra numerical simulations.

The initial configuration is shown in figure 2.13. In this case the ambient fluid has two layers: one in between the two symmetric density currents ($H < 1$ with density $\rho_s = \frac{1}{2}(\rho_1 + \rho_2)$ and another layer for $H > 1$ with the same density ρ_0 as in the previous simulations.

The evolution of the buoyancy profile before and after collision is shown in figure 2.14.

The most interesting feature to note is the evolution of the heights reached by the ambient fluid and by the colliding currents. It can be noticed that the ambient stratified fluid reaches its maximum height before or at the moment of collision, whereas when the colliding fluids reach their maximum height, the height of the ambient fluid is lower.

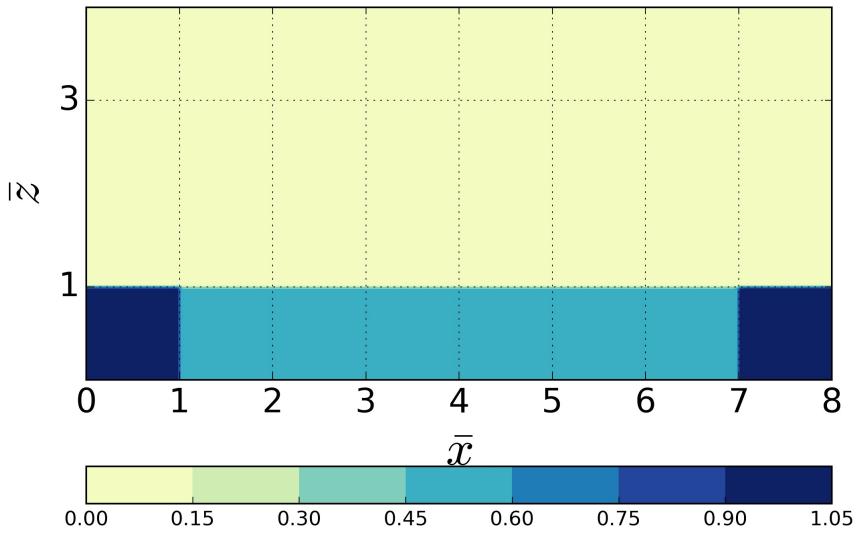


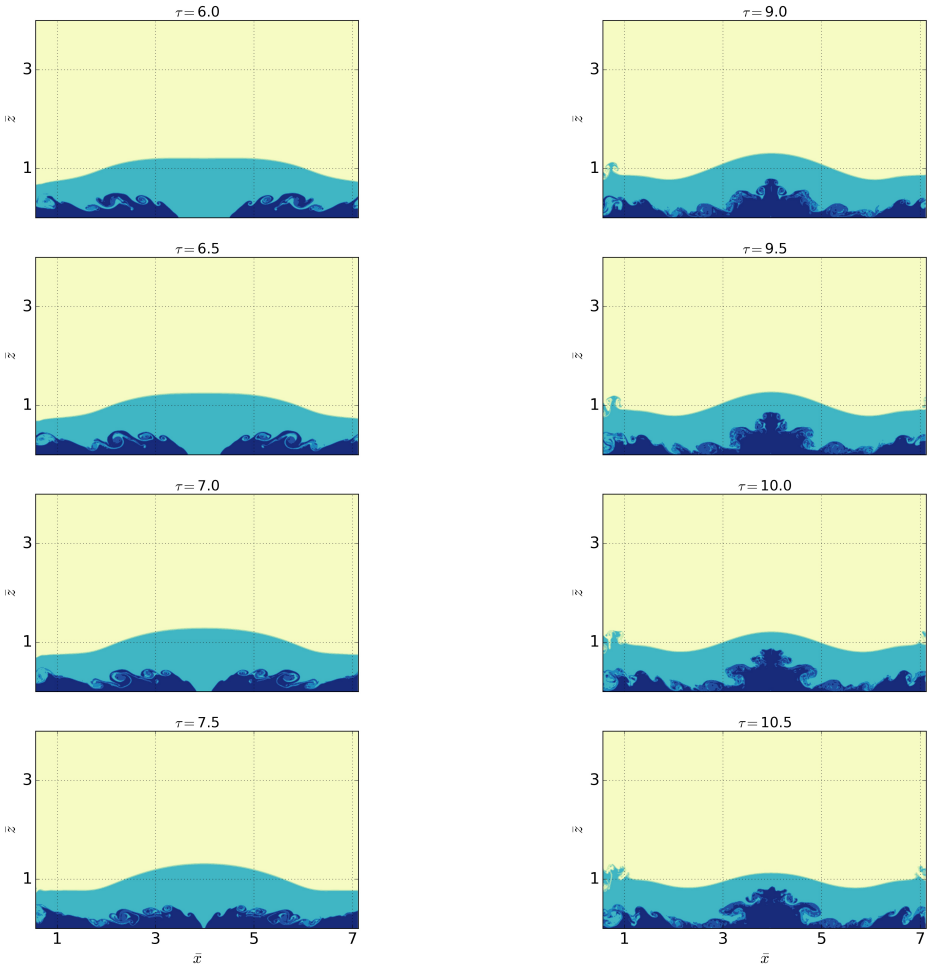
FIGURE 2.13: Initial ($\tau = 0$) distribution of the absolute value of the buoyancy, $|\bar{b}(\bar{x}, \bar{z})|$, with $r_g = 0.99$, $H = 4$, $D = 1$.

This could offer an insight when considering onset time of collision-induced convection triggering.

2.5 Conclusions and future work

In this study, numerical simulations of two colliding density currents in a limited rectangular domain for different buoyancy ratios r_g were performed. These were based around the recent laboratory experiments performed by van der Wiel et al. (2017). Beside the buoyancy ratios, also the height of the domain and the relative current depth have been varied, to investigate the effect of the rigid top lid and also to make the simulation closer to a realistic atmospheric situation. Only currents of equal initial depth have been considered, and of one initial length relative to the domain length, to keep the parameter space to manageable proportions, and to simplify the modelling requirement. Investigating the other parameters could be the subject of future work.

Two main characteristics have been analysed: the slope of the collision front that forms since the denser current deflects the other fluid upward, and the maximum height reached by the fluid raised after the collision. While a large variation in the initial angle of collision with $r_g$has been found, this is not the case for the maximum collision height. This is consistent with an interpretation of the front collision being governed by the interaction of the vortex sheets at the current heads, which leads to unequal currents tending to circulate around one another rather than to "splash" upward. It is worth to remark that this has possible relevance in determining which current interacts with the source of any resulting convective updraft.



(A) Before collision (B) After collision

FIGURE 2.14: Time sequence of buoyancy profile for $r_g = 0.99$ in a stratified environment. Colour scale as in figure 2.1.

As noted in the introduction however, while the collision height is insensitive to r_g in the basic case, other factors may produce secondary effects. For instance, the position or maximum of collision height may vary if the fronts have appreciable curvature (Droegemeier and Wilhelmson, 1985), and the height perturbation may be transmitted vertically and horizontally by gravity waves in a stratified environment (Liu and Moncrieff, 2000). Further work in this area could involve the investigation of stratification effects (also possible with Hydra), or wind-shear and front-curvature effects (perhaps with a large-eddy model). More realistic simulations, including the effect of topography or Earth's rotation, are encouraged. A theoretical model, based on a vortex-pair argument, has been proposed to predict the interface slope at the front, taking as input either the buoyancy ratio or the relative propagation speeds of the currents. This has been shown to agree well with the results of the simulations.

In the wider context of convection parametrization, increasing model resolution (in time and space) produces a requirement for better understanding of the dynamics, rather than merely the thermodynamics, of atmospheric convection. Hence, triggering by convective cold pools has become an area of current interest (e.g. Grandpeix and Lafore, 2010; Rio et al., 2013), since it represents one means by which a dynamical "memory" prolongs the convective state. One proposed mechanism for cold pools overcoming convective inhibition is kinematic uplift, either at the front of an advancing cold pool or where cold pools collide. It has been suggested recently (Torri et al., 2015, see also references therein) that this is the most important process for initiating parcel ascent from the surface layer, with thermodynamic processes reinforcing the convective motion at higher levels. The insights into collision dynamics obtained here will be useful in designing the future parametrization of the distribution, in position and strength, of collision-induced convective triggering.

Chapter 3

Ensemble forecasting methodology

In this chapter the ensemble prediction systems and the methodologies to extract information from them are described. Furthermore, the method to verify this information against observations is presented.

3.1 Met Office Global and Regional Ensemble Prediction System (MOGREPS)

The forecast data used in this thesis come from the Met Office Global and Regional Ensemble Prediction System (MOGREPS), which consists of a global ensemble prediction system (MOGREPS-G) and a convection-permitting limited area version, nested into it (MOGREPS-UK). It is one of the CP-EPSs introduced in the recent years, as mentioned in chapter 1.

One of the key reasons for implementing CP-EPSs is the quantification of uncertainty at small spatio-temporal scales. Three sources of uncertainty have to be take into account when constructing a CP-EPS: lateral boundary conditions (LBCs); model formulation (i.e. physics perturbations); initial condition uncertainty (ICs).

LBCs are necessary because, CP-EPSs are usually run on limited area domains, due to computation constraints. Therefore they require LBCs to close the dynamical equations and it is important to take into account uncertainty stemming from the boundaries. LBCs are provided by the driving coarser resolution ensemble. The topic of the influence of LBCs on forecast errors has been addressed in many studies. Davies (2014) conclude that errors arising from LBCs are a tiny part of the overall error (due to ICs and model error). However, particularly for longer lead times than 12 hours, LBC uncertainty can dominate over ICs uncertainty, whereas for the first six forecast hours the impact of ICs is largest and dominate the physics perturbations and LBCs. (e.g. Hohenegger et al., 2008; Kühnlein et al., 2014).

How to best represent the uncertainty in the ICs at convective-scale is an ongoing research question. For coarser resolution global ensembles many techniques have been proposed (e.g. singular vectors by Buizza et al., 2005; Leutbecher and Palmer, 2008). However, the assumptions these techniques are based on are generally no longer valid at convective-scale, due to reduced predictability and non-linear error growth (Hohenegger and Schar, 2007).

One widely used simple technique to initialise CP-EPSs is to dynamically downscale the information from the coarser resolution ensembles (e.g. Hohenegger et al. (2008); Peralta

et al. (2012)). In this case, small-scale perturbations are initially missing and are created by the high resolution model during the forecast integration. Kühnlein et al. (2014) demonstrated the use of downscaled approach for IC perturbations leads to improved results for convective-scale precipitation ensemble forecast with respect to a deterministic model at same convection-permitting resolutions.

As described in Hagelin et al. (2017), MOGREPS became operational in 2005 and comprised a global model (MOGREPS-G) and a 24 km 24 member regional ensemble (MOGREPS-R: Bowler, 2008; Bowler et al., 2009). Both were run every 12 hours, MOGREPS-G at 00UTC and 12UTC and MOGREPS-R 6 hours later.

When MOGREPS-UK was introduced in July 2012, it was decided to run it every 6 hours (since focussed on short-range time scale) and to halve the number of members, to make it affordable. It was originally nested into MOGREPS-R and was introduced with the main aim to improve the prediction of the uncertainty associated with high-impact weather events, especially convective precipitation and the resultant flooding at river basin level. Since January 2013 MOGREPS-UK has been nested into MOGREPS-G. Both MOGREPS-G and MOGREPS-UK models are configurations of the Met Office Unified Model (UM; Wood et al. 2014; Davies et al. 2005), a non-hydrostatic grid-point model.

In the configuration used in this thesis, MOGREPS-G used a horizontal grid spacing of 33 km, 70 vertical levels, and was run 4 times a day, every 6 hours starting from 00 UTC, with 12 members (1 control + 11 perturbed members) on each cycle run to T+174 (7 days, 6 hours). ICs perturbations for MOGREPS-G are generated using an Ensemble Transform Kalman Filter (ETKF: Wang and Bishop, 2003; Bowler, 2008), whereas model uncertainty is taken into account by using two stochastic physics schemes: random parameter (RP) and stochastic convective vorticity (SCV) (Bowler, 2008).

MOGREPS-UK had a horizontal grid spacing of 2.2 km. The model domain consists of 532×654 grid points east-west and north-south respectively, with Arakawa C staggering. The grid is set on a rotated latitude-longitude projection such that grid boxes are roughly square in physical space. The area covered by MOGREPS-UK includes British Isles and north-west France and extends up to North Sea up to the coast of Norway. The relatively large jump between the two ensembles has been made possible by using a variable resolution that adjusts smoothly from the inner domain with 2.2 km grid size to 4 km at the edges (see figure 3.1 for the MOGREPS-UK inner domain.) This reduces the effects of discontinuities at the boundaries and the need of an intermediate model between the global and the convection-permitting one (Davies, 2014).

MOGREPS-UK was run every 6 h up to 36 h , by dynamically downscaling starting from the 3 h forecast of the matching MOGREPS-G ensemble members ICs. The initial forecast times are 0300, 0900, 1500 and 2100 UTC. This means that MOGREPS-UK was initialized using the ICs and BCs from an interpolated 3 h forecast of the matching MOGREPS-G member run. This also means that ICs and BCs are "taken from the most recent global model run (3 hours after the nominal time)" (Tennant, 2015).

MOGREPS-UK used a set of 70 vertical levels with variable grid size, with a lid at 40 km. These levels are terrain following near the surface. Vertical levels use Charney-Phillips staggering. Primary field advection uses a semi-implicit Lagrangian scheme which allows a longer time step to be used, so reducing the computational cost.

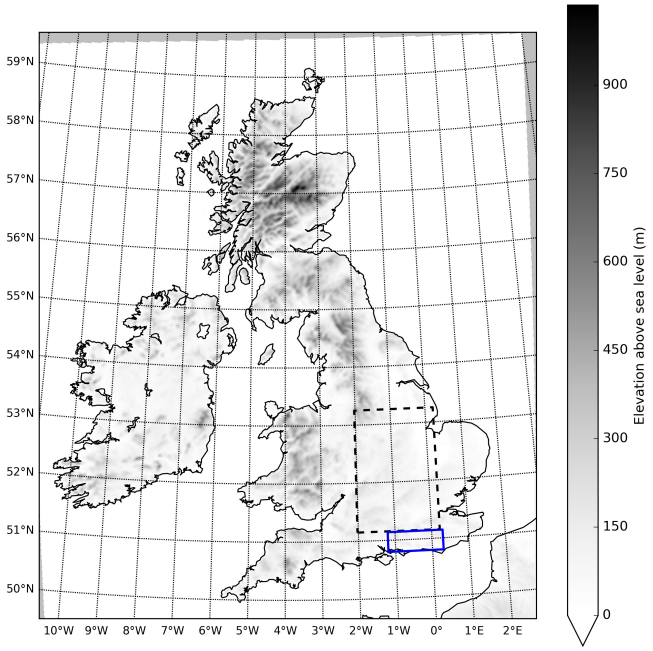


FIGURE 3.1: MOGREPS-UK inner domain (2.2 km × 2.2 km), with elevation shaded. The dashed box indicate the sub-domain used to verify wind gust forecasts, whereas the blue solid box is used for the sea-breeze forecasts.

Parametrized physical processes include long- and short-wave radiation, mixed-phase cloud microphysics, a boundary-layer turbulence scheme (Hagelin et al., 2017).
MOGREPS-UK has stopped from being a pure dynamical downscaler of MOGREPS-G in March 2016, when large-scale perturbations from MOGREPS-G were started being centered on the UKV analysis. This provided high resolution details in the assimilation cycle and Tennant (2015) showed the benefit of the new configuration over the downscaling approach, especially in the first few hours of the integration cycle.
Another upgrade was the introduction of the RP stochastic physics scheme, with the aim to perturb selected parameters in order to represent uncertainties in key physical processes (e.g. fog, low cloud). Then, in November 2016, with the increased computer capacity, the run length was increased from 36h to 54 hours and also the domain covered, for a better spin-up of convective events originating outside the inner domain (Hagelin et al., 2017).
In this thesis the data used for producing probabilistic forecast cover the period 2013-2015, when MOGREPS-UK was purely a simple downscaler from MOGREPS-G. Also, in the configuration used here there is no stochastic component in the MOGREPS-UK model physics and it exactly follows the UKV physics.
Since in this configuration (i.e. using a dynamical downscaling approach), there is no additional high resolution data assimilation, the information gain from CP-EPS could come from a better representation of the dynamics and ancillary fields like coastline, topography and land-use dataset. It is therefore constrained by observations only at scales

greater than 33 km, through the global initial and boundary conditions.

3.2 How to identify weather phenomena in CP-EPSs ?

The post-processing of CP-EPSs is a crucial step in the chain of the weather forecasting process to provide reliable probabilistic forecasts of the local weather event of interest. An example of probabilistic products for an intense rainfall event is reported by Golding et al. (2016).
Probabilistic forecasts from MOGREPS-UK are routinely generated for the Operational Meteorologists via the Gridded Post-Processing system. Products include: probabilities of exceeding, or being below, pre-defined thresholds for rainfall rates and accumulations, visibility, temperature and wind gusts.
Since, as mentioned in chapter 1, CP-EPSs provide large amounts of data and represent a heavy burden on forecasting centres' computational resources, there is need to fully exploit them in order to extract valuable information as fully as possible. There is also need to assess whether these higher-resolution ensembles can provide better quantitative probabilistic forecasts of key weather. The final aim is to provide early reliable indication of severe weather for the emergency response in hazardous weather situations.
In this thesis the prediction of small-scale weather phenomena is investigated, in particular those whose occurrence is often strongly tied to large-scale environmental factors, yet also locally generated.
Very fine gridded data ($\sim 1-3$ km) allow, on one hand, to have a more realistic representation of small-scale structures (e.g. convection, sea-breezes, local fog and wind gusts), therefore to more easily identify these phenomena in a more objective way.
Despite being localized both in space and in time, these phenomena reveal a certain spatio-temporal coherence, which is key to separating them from the plethora of small-scale structures, which are also represented by CP-EPSs.
Another factor is that knowing in advance the synoptic weather regime allows a more successful identification of the weather event to predict.
It enables to reduce the number of false alarms, by excluding weather events that should be not be expected with particular regimes (e.g. sea breezes in a cyclonic regime).
Although knowing a physical relationship with the larger scale environment helps, detection of weather events such as fog, sea breezes, mesoscale-convective systems, banded precipitation, is not straightforward. Therefore a combination of weather variables associated with the event occurrence is suggested as a means of defining the weather event, since usually there is a physical relationship between the local phenomenon and environmental conditions. However, uncertainty in definition of a particular phenomenon is still an issue that needs to be addressed.
This uncertainty stems from the fact that different events can resemble the same distinctive features (e.g. sea breezes, cold fronts). For instance, sea breezes and cold fronts both may have a wind speed increase, shift in the direction and a drop in temperature.
Therefore, the information coming from the synoptic weather regime is vital to be able to help to discriminate one event from another. Not less important is the geographical

information (e.g. coasts, high orography region, valley).
As these "ancillary" fields (e.g. land-sea mask, topography) are better represented by CP-EPSs, then small-scale weather features, associated with local topography, are more realistic than the ones represented on larger grid box sizes.

In this thesis sea breezes and wind gusts probabilistic predictions are performed and analysed, to assess whether the convective-scale ensemble forecast can provide more information than a cheaper coarser resolution ensemble.
Also, in order to determine the statistical significance of information gain from the CP-EPS, the approach used in this thesis is not based on individual case studies. A more systematic and longer period of evaluation is considered. Thus, there is need for automated methods to identify the phenomena of interest and these algorithms can be used in operational forecast, since it would be too time consuming to do it visually.
Therefore the analysis conducted in this thesis consist of four main steps:

1. Identification of weather phenomenon of interest, using an automatic algorithm, in a convection-permitting ensemble forecast.

2. Extraction of the information about the event occurrence from a coarser resolution model using a Bayesian approach, due to model's inability to properly resolve the small-scale weather event.

3. Generation of probabilistic forecasts of the event occurrence using both methodologies.

4. Comparison and verification of probabilistic forecasts against each other and observations, calculating proper scores and assessing any significance in the scores differences.

The algorithms used to detect sea breezes and wind gusts will be presented in chapter 4 and 6 respectively.

3.3 From CP-EPSs to probabilistic forecasts

Once weather events have been defined for each ensemble member, the next step is to translate the information contained in each member into a probability distribution function. For instance, Bröcker and Smith (2008) discuss different kernel methods to build this distribution function, considering ensembles as a source of information and not as the possible scenarios of the reality.
In this study the focus is on binary predictions, i.e. the full distribution function is not estimated. Instead, if the occurrence of the event is considered as a success, the probability p of the success given N trials (i.e. the different ensemble members) has to be estimated. This can be estimated by the Binomial distribution with parameters p and N[1]. p can be estimated from the ensembles, calculating the maximum likelihood estimator (MLE) of

[1]It is worth to notice that the assumption of the Binomial distribution is that the different trials are independent. This can be not true with ensemble members.

the Binomial distribution, which is just the fraction of successes, i.e. the fraction of ensemble members that satisfy the predetermined condition. In the Appendix A a modified MLE to estimate p will be presented.

3.4 Bayesian model

In order to produce probabilistic forecasts from the coarser resolution model a Bayesian approach based on environmental predictors is used, mainly for two reasons:

1. the grid box of the coarse resolution model is too large to properly directly represent the small scale weather phenomenon to be predicted.

2. To assess to extent to which the small-scale phenomenon is predictable by using only salient information from the large-scale conditions, which still exert an important influence on the occurrence of the phenomenon. A Bayesian model based on a few predictors is thus implemented and compared with the dynamical convection-permitting model that represents directly the physical process.[2]

This Bayesian model is trained on CP-EPSs data for each forecast lead time. In particular the relationship between events as defined in CP-EPSs models and the parameters, which are given by the coarser resolution models, is found. In this Bayesian framework the prior probability distribution is represented by the high resolution model climatology. This is then updated by the information given by the coarse resolution parameters from the forecast.
The aim of this approach is then to model the conditional distribution, given a set of predictors.
The two most used statistical techniques to estimate the predictive probability density function are the Bayesian model averaging (BMA; Raftery et al. (2005)) and non-homogeneous regression, also referred as model output statistics (EMOS; Gneiting et al. (2005)).
For both the methods a parametric form of the functional relationship between the predictors and the predictand needs to be estimated. An advantage of this "data-driven" method presented in this thesis is that such parametric equations of the predictive distribution are not necessary.
Other "data-driven" methods are becoming popular in atmospheric science, such as neural networks. In this case functional relationship between the predictors and the predictand are "learned in data-driven way rather than requiring prespecified link functions" (Rasp and Lerch, 2018).

3.5 Observational data

Observations are taken from the UK land surface observing network, with information available at this link (`https://www.metoffice.gov.uk/guide/weather/observations-guide/uk-observations-network`). The average distance between the surface synoptic stations

[2]For certain weather phenomena the convection-permitting model may not fully resolve the physical process either.

is 40 km, which is approximately the size of the coarse resolution grid box (33 km). Provided that the distribution of the stations is reasonably uniform, this means that there should be at least one station for each grid box.
Temperature and wind measurements are available at every hour. More precisely temperature is the instantaneous value at every hour, whereas wind speed and direction are a 10 minutes average between 20 and 10 minutes to the hour and the max gust speed is the max wind speed in the same 10 minutes period.
In order to contruct verification datasets for the two phenomena, automatic algorithms for detecting sea breezes and wind gusts are proposed and will be presented in the following chapters.
For the sea breeze case the synoptic stations data are aggregated, in order to verify sea breeze occurrence for any point of the domain, whereas for the wind gust case the verification at each coarse resolution grid box is performed. This will be explained in the following chapters.

3.6 Verification

After extracting the signal of the weather phenomenon from the observations in an automatic way, it is possible to build the verification dataset, against which to compare the probabilistic forecasts calculated with two different methodologies, as explained in the previous sections.
The general aim of our analysis is to calculate two scores from the probabilistic verification, then calculate the differences and assess whether any difference in the scores is significant. A null hypothesis of no difference between the two probabilistic forecasts is used to test significance.
In regard to the scores Ignorance score, the Brier Score and the Area under the ROC curve (AUC) are calculated. This provides information about reliability and resolution (discussed in chapter 5) and focusses on additional useful information provided by the convection-permitting ensemble forecast.
In order to calculate any significance in the verification scores difference (Information gain, Brier score difference and AUC difference) two different techniques are considered. For the probabilistic scores, the empirical bootstrap technique is used. N_{boot} sequences of the forecast-observation pair are considered, each one obtained from the original by sampling with replacement. Then the Information gain and Brier Score difference are computed for each sequence and finally the 95% confidence interval.
In regard to the comparison of the area under the ROC curve, since in this case the areas are correlated, the method described in (DeLong et al., 1988) to compute the variance is used. Then, following (Zhou et al., 2008) a Z-test is used to compare the two areas under the ROC curves and to test the null hypothesis the two probabilistic forecasts having the same AUC.

Chapter 4

Probabilistic forecasts of sea breezes

In this chapter the information about sea-breeze occurrence is first extracted from a dynamically downscaled convection-permitting ensemble forecast (CP-EPS) using a novel automatic tracking algorithm and then compared with a Bayesian model, taking as input coarser resolution data and trained on convective-scale data.
In essence the Bayesian method forecasts the high resolution member based on large-scale variables from low resolution model (LR-EPS). The aim of this is twofold: -firstly to develop a method to extract information from the low resolution prior to the running of the convective-scale forecast for real time forecasting ; secondly to provide an estimate of the information gained by running the convective-scale forecast beyond that which is contained in the large-scale flow conditions.
The work presented in this chapter is part of a paper which is published in the *Quarterly Journal of the Royal Meteorological Society*, with the reference:
Cafaro C, Frame THA, Methven J, Roberts N, Bröcker J, (2019). The added value of convection-permitting ensemble forecasts compared to a Bayesian forecast driven by the global ensemble. Q. J. R. Met. Soc., doi:10.1002/qj.3531
The research included in this paper was conducted by the main author under the supervision of T. H. A. Frame, J. Methven, N. Roberts and J. Bröcker via weekly meetings and discussions. The lead author wrote the first draft of the paper, prepared all the figures and had overall control of the submitted paper. The other authors commented on draft version of the paper and the lead author updated the manuscript accordingly.

4.1 Introduction

The sea breeze is a phenomenon that has been known about long time, already documented by Aristotle (see Neumann (1973) and references therein). Yet it still raises interesting open questions, in particular about how it is identified and predicted.
A sea breeze is a mesoscale circulation caused by the temperature contrast between the land and the sea during the day, i.e. the differential heating due to different heat capacity of the water and land. Land heats more quickly during the day and cools more quickly at night. Sea breezes are an appealing choice of phenomena for model comparison because they are geographically constrained to initiate at the coast and occur quite frequently in summer months. Also, sea breeze forecasting is important for several reasons: its impacts on air quality, , since it affects transports of pollutants e.g. (Loughner et al., 2014; Kambezidis et al., 1998; Clappier et al., 2000). It is also important for health, being a relief

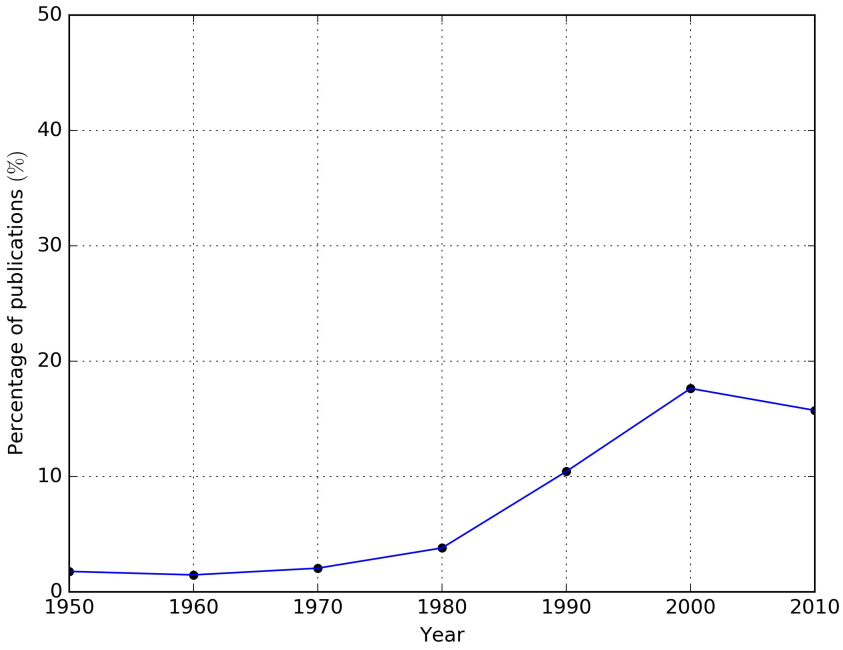


FIGURE 4.1: Percentage of publications every 10 years years since 1950 (up to 18 December 2018) including the phrase "sea breeze" according to Google Scholar for four journals (BAMS, JAS , MWR, QJRMS). Each dot represents the percentage in the following 10 years (except for year 2010).

from oppressive hot weather (e.g. Papanastasiou et al. (2010); Meir et al. (2013)) and as a possible trigger for convective storms or enhancing rainfall totals from existing storms, especially when interacting with other mesoscale flows (Warren, 2014; Birch et al., 2015). Miller et al. (2003) stated that forecasting sea breezes consists of three main aspects: occurrence, propagation speed and direction and distance of penetration. In this study the main focus is on the occurrence. Other sea breeze characteristics are examined in the section 4.6.

Furthermore, this analysis can benefit from the fact there have been lots of studies on the sea breeze, especially in the past 40 years (see figure 4.1), as more and more people are living coastal urban areas. Among these studies, Azorin-Molina and Tijm (2011, Table 1) include a list of studies on sea breeze forecasting and Crosman and Horel (2010, Table 1) on sea breeze numerical studies). This has permitted to understand the phenomena, how best to detect it and what drives it.

For instance, Miller et al. (2003, Table 3) presents a list of factors controlling the sea-breeze occurrence and inland penetration. The correct representation of these factors in NWP models depend on the model grid size. In particular, topography, shape of the coastline require fine grid scale to be properly represented. This is crucial to reduce forecast errors (e.g. Persson and Grazzini (2007) reports an "over-sized"[1] sea breeze during the forecast integration, due to the grid box size of 25 km of the ECMWF model).

Since sea breezes usually occur on a time scale of about 1 day and spatial scale between

[1]This is directly quoted from the authors' document. Over-sized in this context could mean a sea-breeze which extends too far inland.

10 to 100 km (Lin, 2007), it is not expected that coarse resolution model (with grid spacing larger than 20 km) are able to properly resolve the sea breeze circulation. The key distinction being that the CP-EPSs can explicitly develop a sea breeze, whereas the global ensemble can only predict the large-scale conditions that give rise to sea breeze.
In this study the degree to which CP-EPSs provide additional information about the occurrence of sea-breeze beyond that which could be determined from the larger-scale environment is examined. This means to post-process direct model output variables to extract information and create a "sea breeze occurrence" variable based on them.
The information extracted from a CP-EPS is compared with a Bayesian model based on LR-EPS model variables. The Bayesian model should perform better than the LR-EPS alone and is still cheap compared to running a CP-EPS, and therefore provides a much more stringent test of the benefit of the CP-EPSs, compared with a more intelligent use of the LR-EPSs than simply extracting raw model output.
Although the sea breeze is locally forced by the land-sea temperature contrast, the synoptic scale flow plays a fundamental role in controlling the initiation and the evolution of the sea breeze itself. There are of course other geographical factors involved, like coastline topography, surface roughness, latitude, season. However Azorin-Molina and Chen (2008) deem the impact of the large-flow to be greater than these, even than the thermal gradient. This is related to question the source of sea-breeze predictability. More precisely it is assessed whether sea-breeze occurrence is due to better representation of local factors and sea-breeze dynamics in CP-EPSs or if it can be instead derived solely from the knowledge of the most influential large-scale conditions (synoptic wind and land-sea temperature contrast) which are sufficiently well represented even on coarse grid boxes. These are used as predictors in the Bayesian model.
In this context ensembles are used mainly for two reasons: they provide inherent uncertainty in initiation and subsequent development because of dependence on environmental factors that are themselves open to uncertainty; they provide a larger sample to get a more robust comparison between the two forecasting approaches. Therefore a more appropriate question to answer is "how much more information is the CP-EPS bringing compared to the LR-EPS and is this significant ?". The null hypothesis is that the two EPSs provide the same amount of information. Therefore sea breeze occurrence becomes a test, by which this hypothesis can be either rejected or accepted.
The rest of the chapter is organized into five sections. In section 4.2 the forecast and observational data used are described, section 4.3 presents the method for identifying sea breezes in observations and CP-EPSs and the Bayesian model, section 4.4 contains a first qualitative comparison between the two probabilistic forecasts and subjective verification against station observations. A summary of the results and conclusions is given in section 4.7.

4.2 Data set

In this section forecasts and observational data sets are presented.

4.2.1 Model data

The forecast data used in this paper come from the Met Office Global and Regional Ensemble Prediction System (MOGREPS), which consists of a global ensemble prediction system (MOGREPS-G) and a nested convection-permitting limited-area version (MOGREPS-UK). Both ensembles are constructed using the Met Office Unified Model (Wood et al., 2014; Davies et al., 2005). Descriptions of both configurations, the initial condition perturbations and stochastic physics can be found in (Hagelin et al., 2017),(Tennant, 2015) and (Bowler et al., 2008). These data also form part of the THORPEX Interactive Grand Global Ensemble (TIGGE) database (Swinbank et al., 2016). This database of operational ensemble forecasts from leading NWP centers was set up to enhance the development of probabilistic forecasting of high-impact weather. In 2014 a TIGGE-Limited Area Model (LAM) panel was also established to apply TIGGE concepts to limited-area model ensembles. Both the datasets are available for research and education purposes at the following link (`https://www.ecmwf.int/en/research/projects/tigge`[2]). The MOGREPS-G forecasts were produced 4 times a day, every 6 hours starting from 00 UTC, with 12 members (1 control + 11 perturbed members) on each cycle run to T+174 (7 days, 6 hours), using a horizontal grid spacing of 33 km and 70 vertical levels. MOGREPS-UK has been running routinely since 2012 and has a horizontal grid spacing of 2.2 km. It was run every 6 h up to 36 h , by dynamically downscaling starting from the 3 h forecast of the matching MOGREPS-G ensemble members. This means that the initial conditions for each MOGREPS-UK member were simply interpolated to the high-resolution grid using the 3 h MOGREPS-G forecast at 0300/0900/1500/2100 UTC. Note that this configuration of MOGREPS-UK was operational from 16 January 2013 until 15 March 2016, after which the MOGREPS-UK ensemble was generated by perturbing around the control forecast intialised using a high resolution analysis (Hagelin et al., 2017). Since there was no additional high resolution data assimilation in the period studied here (2013-2015), the information gain from CP-EPS can only come from better representation of the dynamics and/or ancillary fields such as the coastline, topography and the land-use dataset. The CP-EPS is therefore only constrained by observations at scales greater than 33 km through the global initial and boundary conditions. Since we are focussing on sea breeze, we use only summer (JJA) data for the years 2013-2015. Data from some months during this time period were either partially or completely missing from the archive so are not included in this analysis. These are June 2013, June 2014 and 1st-16th of July 2014. Probabilistic forecasts of sea breeze are made once a day, for each forecast lead time, at any time between 0900 and 2100 UTC, in which the sea breeze is expected to occur. In total 201 forecasts are made across the three summer seasons. Figure 4.2 shows a schematic diagram of the different lead times we have considered. For instance, the forecast labelled with "T+12" is the MOGREPS-G forecast issued at 00 UTC (black dots) or the MOGREPS-UK forecast initialised three hours later (red dots). The parameters for the Bayesian forecast are computed at 12 UTC valid time (black squares). From MOGREPS-G we have data every 3 hours from the initialization time. So the possible times are 0900UTC, 1200UTC,

[2] Accessed online on 22 February 2016

1500UTC, 1800UTC and 2100UTC within the 0900-2100 period. Since the aim is to predict the occurrence based solely on the daily atmospheric conditions, it has been chosen a time which is believed to be the most representative of these conditions. Times after 12UTC were excluded since by that time sea-breeze will have already occurred. Therefore 12UTC was preferred over 09UTC since in a typical summer day (which is the season being investigated) land temperature is higher at that time, leading to a higher temperature contrast.
Note that the T+36h forecast lead time is excluded from our analysis since the valid "time window" extends beyond it.

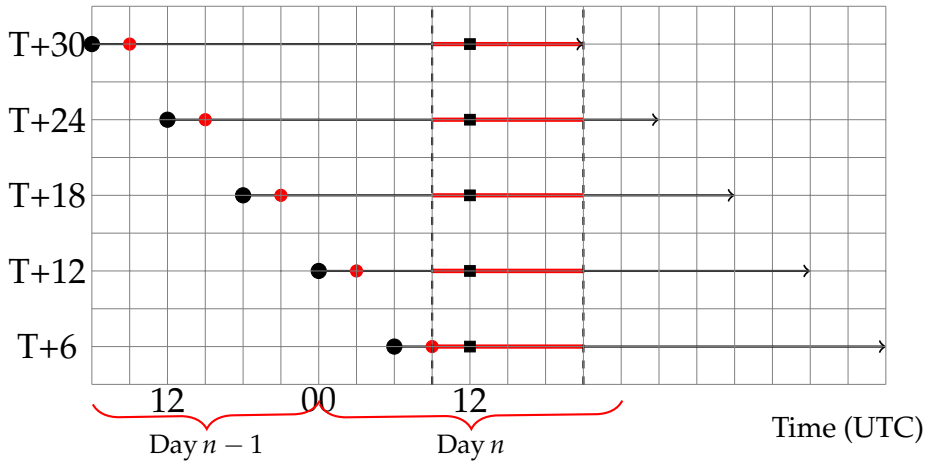


FIGURE 4.2: Schematic diagram of the forecast lead times. The black (red) dot indicates the MOGREPS-G (MOGREPS-UK) initialization time and the red bar represents the sea-breeze detection window. The black square indicates the valid time at which the MOGREPS-G parameters are computed, for all the different initialization times. The solid black line with arrow covers the total MOGREPS-UK forecast length (36 hours).

4.2.2 Observational data

The probabilistic forecasts for sea-breeze occurrence have been verified using the time series from four synoptic stations (see table 4.1 for details), with station data from Met Office Integrated Data Archive System (MIDAS) and available at this link: http://archive.ceda.ac.uk/ [3]. Weather variables are recorded every hour. Table 4.1 shows the details of the stations considered in this study for verification and figure 4.3 their location on the map, with the rectangular box indicating the spatial domain where sea breeze has been detected and verified and elevation shaded. This domain has been chosen because of the near straight line shape of the coast and also for having been studied by Simpson et al. (1977), finding that sea breeze can travel northwards inland up to 100 km.

4.3 Sea breeze identification and prediction

Before looking at the sea breeze occurrence prediction problem over a large sample it is necessary to have a method to identify the sea breeze automatically.

[3] Accessed online on 25 September 2017

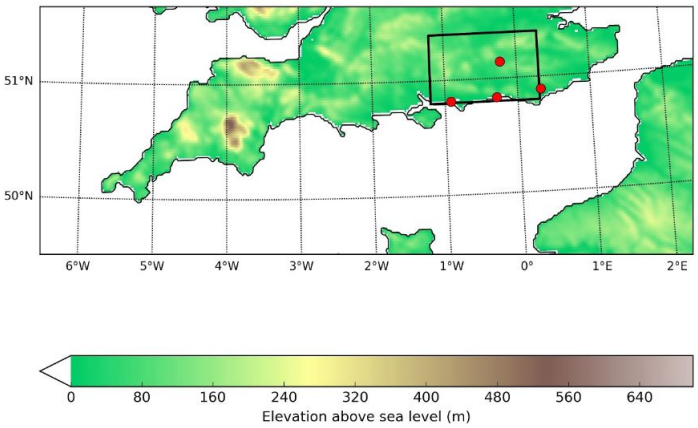


FIGURE 4.3: A map showing the orography over the south UK domain. Orography data are from MOGREPS-UK. The solid box encloses the sub-domain used in this study with red dots indicating weather station observations.

WMO number	Name	Location	Distance from the coast (km)
03876	Shoreham Airport	50.8356 N, 0.29194 W	1.3
03872	Thorney Island	50.8142 N, 0.92098 W	1.2
03882	Herstmonceux: West End	50.8904 N, 0.31818 E	9
03769	Charlwood	51.1435 N, 0.22786 W	35

TABLE 4.1: Location and beginning of the data record of the surface weather stations used for verification. The shortest distance from the coast has been considered. See figure 4.3 for their location on a map.

Many criteria have been developed for the sea breeze identification in observational data, however these are not tailored to automatic detection in high resolution gridded forecast. These criteria involve meteorological variables which are associated with the sea breeze front passage. In (Azorin-Molina and Tijm, 2011, Table 1) there is a list of the past studies, each with the objective criteria described. Some are directly involved with the sea breeze passage (air humidity, air temperature, wind speed and direction), others are used as predictors for the inhibition of sea breeze occurrence (total cloudiness, precipitation). However, Azorin-Molina and Tijm (2011) pointed out that these techniques are clearly dependent on the nature of test criteria adopted by each researcher and on the region of interest. Also, they rely significantly on specifying thresholds for the meteorological variables used. More recently, Coceal et al. (2018) proposed a new method for sea breeze detection in stations observations using fuzzy-logic[4], but still relying on thresholds of the meteorological variables affected by sea breeze passage.

For the purpose of this study and to make effective general use of CP-EPSs, there is need for a more general identification method that makes use of a few key variables that can in principle be applied to any coastal region of the world, as long as a cross-coastal direction can be found. Steele et al. (2014) implemented a novel identification filter method, based on a combination of both model simulations and analysis. They claim that the method, in principle, can be applied on any coastline of the world. However, again, the use of

[4]fuzzy-logic is a type of logic where the variable can take any real value between 0 and 1 and not just 0 or 1, as in the classical Boolean logic.

the thresholds, like for the land-sea temperature contrast depend on the region being considered.
In this study the focus is on the south coast of UK (see figure 4.3) and a new sea breeze identification algorithm that does not rely significantly on the thresholds of the variables involved is presented. It could be useful for the current state of the art CP-EPS data and observations.

4.3.1 Sea breeze identification in observations

Sea breezes were identified using the weather station data described above. Three stations have been chosen as they are located within a few kilometers of the coast, where the sea breeze phenomenon is found. The fourth station is inland by 35 km, approximately matching the scale of the MOGREPS-G forecast grid box size. Hourly 10 m wind speed and direction, station air pressure, 2 m dry bulb temperature, and 2 m dew point were used to calculate hourly changes of wind speed and direction, specific humidity and dew point depression. As an example of two cases of sea-breeze days detected in the station observations, Figure 4.4 shows a clear signal of an early and a relatively late onset of sea breeze front, in that there is a simultaneous shift in wind direction towards being more normal to the coast, an increase in wind speed and specific humidity and a reduction in dew-point depression. In order to construct the verification data set we developed an automatic algorithm for detecting sea breeze in stations observations. The algorithm can be summarized as follows:

1. We select wind direction, wind speed, specific humidity and dew point depression. We then check that the wind changes from having an offshore component to have an onshore component and at the same time the wind speed and specific humidity increase and dew point depression decrease. This is done for each station independently, for the three stations on the coast and the station inland.

2. If sea breeze is detected at least at one coastal station, then we check if it has also been detected inland. Then we require that time of the arrival of the sea breeze front inland is at least two hours after the onset time at the coast. Given that the station inland is approximately 35 km from the coast (see table 4.1), a 1-hour interval would allow propagation speeds that are too fast and therefore perhaps pick up spurious signals; a 3-hour interval would make identification more difficult because the sea breeze may have propagated well beyond the inland station at the time of measurement, and also gives less guarantee the same feature is being identified. This would imply an approximate propagation speed of $35/2$ km/h ~ 17 km/h $\sim 4.7 m/s$

3. Finally, in order to avoid spurious detections (e.g. cold fronts), we also require that the sea breeze day is also a dry day at the coast, meaning that no rainfall accumulation is recorded in the time window relevant to the sea-breeze passage at the stations near the coast only. Thus the possibility of having convective rain further inland, maybe due to sea-breeze front interacting with other mesoscale and synoptic-scale flow, is not excluded in this analysis. (e.g. the case reported by Pedgley (2003).

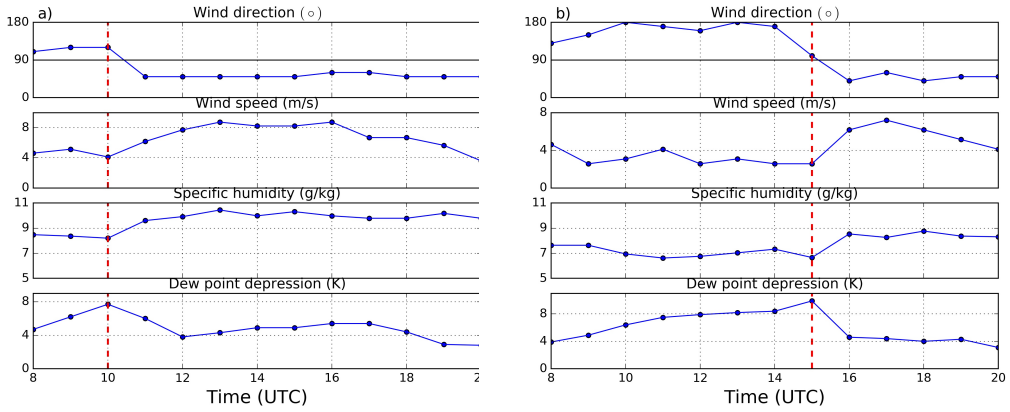


FIGURE 4.4: Shoreham airport station observation time series on a) 27 June 2015 and b) 23 June 2015. Wind direction is defined such that values of $0, 90$ and 180° correspond to onshore, along the shore and offshore flow, respectively. The horizontal solid line indicates the separation between the northerly and southerly direction. Sea breeze onset time is indicated by the vertical dotted red line.

4.3.2 Sea breeze identification in MOGREPS-UK

In this subsection we will describe the sea breeze identification algorithm using CP-EPS data. The method is designed to identify, at each time step, the position of the leading edge of sea breeze front. Due to the different grid spacing of the observational network ($\sim$ 40 km, see again Figure 4.3), and the CP-EPS (2.2 km), we have developed a different algorithm, with the same aim to detect the sea-breeze initiation and its propagation inland.

The passage of the sea breeze front is typically marked by a drop in the air temperature, an increase in wind speed, a shift in the wind direction and increase in specific humidity. In Figure 4.5 the hourly rate of change of air temperature, wind direction and wind speed are shown for all ensemble members from an example T+12 forecast. A sea breeze front is clear in several of the panels (see members 0, 4, 7, 9 in particular) where one can see the extrema of the hourly rate of change of all three variables in an elongated "strip" approximately parallel to the coast. Note also that the changes with respect to the different variables occur in the same location. The sea-breeze signal is less clear in other members, for instance members 1 and 11. Figure 4.6 shows the same variables at two different time steps for member 0 (the control member). It is clear that two hours later the sea breeze front has progressed further inland. The algorithm for sea-breeze detection in the MOGREPS-UK data is based on an analysis of the spatial patterns of the hourly rate of change in the 2 metre temperature and 10 metre windspeed and direction and then on the tracking of their propagation. A formal mathematical description is contained in supplementary material, but the method can be summarized as follows. We first define an along-coast and cross-coast coordinate. Since we are looking at the south-coast of England, these are taken to be zonal and meridional coordinates of the model's rotated equatorial grid, respectively. At each along-coast point in turn the sea breeze detection is performed as follows:

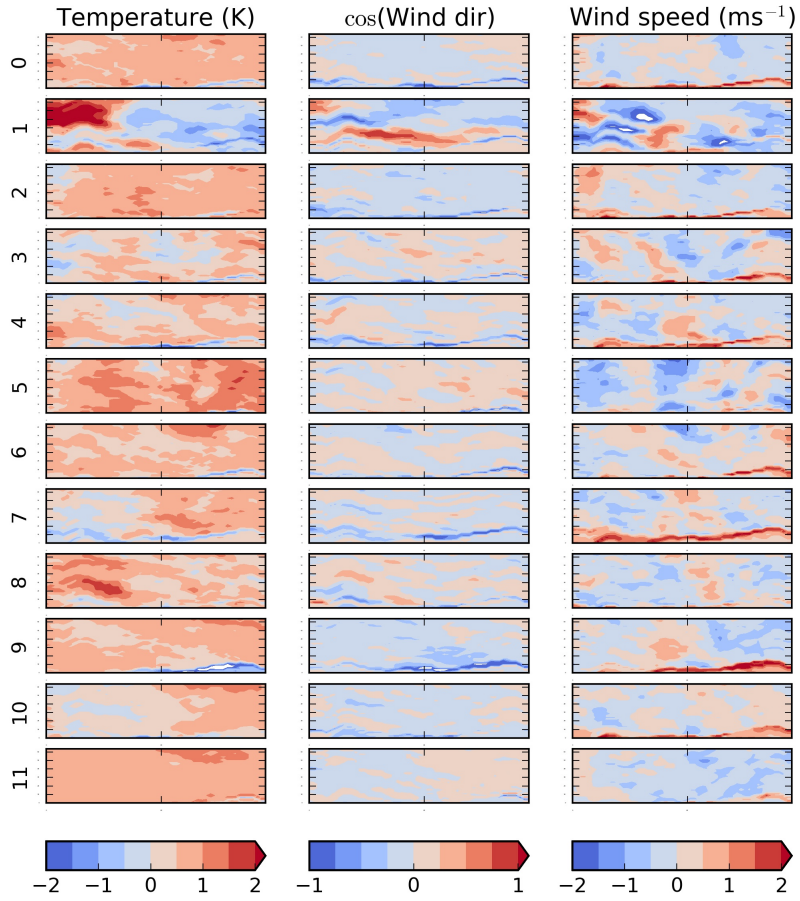


FIGURE 4.5: Hourly differences of the three variables (2 m temperature, 10 m wind speed and direction) of the 12 ensemble members on 27/06/2015 for the T+12 forecast, calculated between 1100UTC and 1200UTC. More precisely the values at 1100UTC are subtracted from the ones at 1200UTC. The rectangles correspond to the black box in Figure 4.3. The ticks on the y-axis are every 10 km (from 10 up to 60 km), whereas tick on the x-axis is at 60 km from the left edge of the domain.

- We find the location of the minimum in the hourly change of the cosine of wind speed occurring within 30 grid points (66 km) from the coast in the cross-coast direction to identify the wind shift. If this occurs within four grid points (8.8km), and the minimum hourly change in temperature and maximum hourly change in wind speed are also found to be within four grid points of the coast at that time or one hour later, the time is flagged as a potential sea-breeze onset at that along-coast point.

- The next step is to test for propagation of the sea breeze. This done by checking that one hour later the minimum in cosine wind direction, the minimum in temperature change and maximum in wind speed change have moved between 1 and 10 grid points (up to 22 km) further inland. This is equivalent to express that sea breeze does not propagate faster than 22 km/h $\sim 6.1 m/s$. This threshold is different from the one based on observations (17 km/h $\sim 4.7 m/s$). This is because the model grid spacing is Δx= 2.2 km and therefore the location of a sharp front at any snapshot

is uncertain to at least Δx. The distance travelled between any two time frames is therefore uncertain to $2\Delta x$ and this motivates a model threshold that is larger by $2\Delta x$. Also, we check that the minimum of cosine of wind direction and temperature and maximum in wind speed move coherently in space inland, by requiring that their position is not further than 6 grid points apart. If this is satisfied for at least the subsequent two time steps from the onset time then it is confirmed that the potential sea breeze front propagates inland in a manner consistent with a sea breeze. If it is not satisfied then no sea breeze is detected. In Figure 4.7 the positions of the extrema hourly rate of change with respect to time are shown for a particular along-coast grid point.

- The final step is to test that the potential sea breeze is spatio-temporally coherent in the along-coast direction. This is done by requiring that the potential sea breeze occurs at least four adjacent along coast points, meaning that it has an extent of at least 8.8km and that the range of the onset times of these coastal points is small (by requiring the standard deviation has to be less than 2 hours). If these criteria are satisfied then the detection of a sea breeze is confirmed. If it is not then no sea breeze is detected.

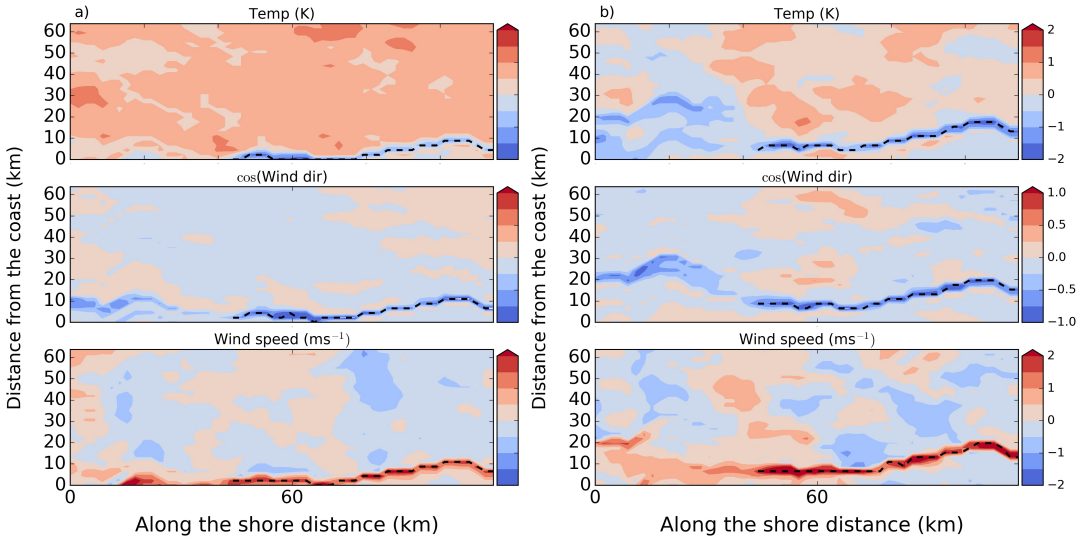


FIGURE 4.6: Snapshots map of hourly changes of temperature (temp), cosine of wind direction (wind dir) and wind speed for member 0 on 27/06/2015 for a) 11-12 UTC and b) 13-14 UTC. The dashed lines indicate approximately the position of the sea-breeze front for the along the coast grid points where the conditions of the sea-breeze detection algorithm are satisfied. The rectangles, as in Figure 4.5, correspond to the black box in Figure 4.3.

Sea breeze tracking algorithm: mathematical description Herewith the mathematical description of the sea breeze algorithm using convection-permitting model data is presented:

1. Three variables output from MOGREPS-UK have been selected: 2 metres dry bulb temperature ($T(x_j, \widetilde{y_k}, t_i)$), 10 metres wind speed ($|\mathbf{v}|(x_j, \widetilde{y_k}, t_i)$) and direction ($\phi(x_j, \widetilde{y_k}, t_i)$),

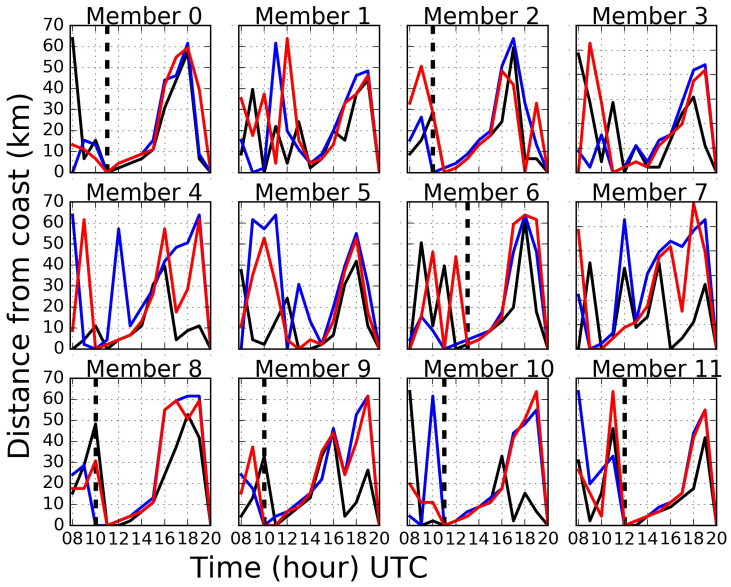


FIGURE 4.7: Postage stamp plot of the maxima hourly rate of change location against time for temperature (black), wind direction (blue) and wind speed(red) for T+12 MOGREPS-UK forecast on 27/06/2015 and for a particular coastal grid point. The vertical dotted line identifies the members satisfying the sea-breeze conditions and represents the onset time identified from the model .

where t_i is the forecast leadtime, x_j the along coast direction and $\widetilde{y_k} = y_k - y_{0k}$ the latitude on the model grid relative to y_{0k}, the latitude of the coast, for $1 \leq i \leq N$, $0 \leq j \leq L, 0 \leq k \leq P$, where N is the time dimension and L, P are spatial dimension of our domain of interest. Then the temporal change is calculated:

$$\Delta T(x_j, \widetilde{y_k}, t_{i+1}) = T(x_j, \widetilde{y_k}, t_{i+1}) - T(x_j, \widetilde{y_k}, t_i), \tag{4.1}$$

$$\Delta \phi(x_j, \widetilde{y_k}, t_{i+1}) = \cos(\phi(x_j, \widetilde{y_k}, t_{i+1})) - \cos(\phi(x_j, \widetilde{y_k}, t_i)), \tag{4.2}$$

$$\Delta |\mathbf{v}|(x_j, \widetilde{y_k}, t_{i+1}) = |\mathbf{v}|(x_j, \widetilde{y_k}, t_{i+1}) - |\mathbf{v}|(x_j, \widetilde{y_k}, t_i) \tag{4.3}$$

$\forall 1 \leq i \leq N, 1 \leq j \leq L, 1 \leq k \leq P$, and $t_{i+1} = t_i + \Delta t$, with $\Delta t = 1$ hour for MOGREPS-UK.

Then calculate [5]:

$$Y_{|\mathbf{v}|_{max}}(x,t) = \underset{y}{\operatorname{argmax}}\, \Delta |\mathbf{v}|(x,y,t) \tag{4.4}$$

$$Y_{\phi(x,t)_{min}} = \underset{y}{\operatorname{argmin}}\, \Delta \phi(x,y,t), \tag{4.5}$$

$$Y_{\Delta T(x,t)_{min}} = \underset{y}{\operatorname{argmin}}\, \Delta T(x,y,t). \tag{4.6}$$

[5]This is where the assumption of the orientation of the coast is used. In the case of north-south oriented coast argmin and argmax are calculated with respect to x. In the more general case of slantwise oriented coast argmin and argmax should be calculated both with respect to x and y. Alternatively, new along-coast and cross-coast directions need to be defined and the algorithm can be applied as in the case described here.

a. Let $A := \{t \in [t_0, t_n] : Y_{\phi(x,t)_{min}} \leq G\}$, where $[t_0, t_n]$ is the time interval between two different forecast leadtimes and G a selected threshold on the number of grid points the feature has to be found within. $G = 4$ in this case.
If $A = \emptyset$, then it is stated that sea breeze signal is not detected and the algorithm is stopped. Otherwise, since $|A| < \infty$, $t^* = \min A$ can be defined.
Similarly $B := \{t \in [t_0, t_n] : 0 \leq Y_{T(t,x)_{min}} \leq G\}$ and C are defined for the wind speed. Then t_1^* and t_2^* are defined for B and C respectively. Since the sea-breeze signal is anticipated by the change in wind direction, it is required that $t_1^* - t^* \leq 1$ and $t_2^* - t^* \leq 1$.

b. Once selected t^* as described above, the following condition is checked:

$$Y_{T(x,t^*)_{min}} < Y_{T(x,t)_{min}} < M \qquad \text{for} \quad t = t^* + h\Delta t,\ h \in \{1, 2\} \tag{4.7}$$

and similarly for the other variables. $M = 10$ is the upper threshold on the maximum propagation speed, mentioned in the main text.
Also, it is required that:

$$|Y_{T(t,x)_{min}} - Y_{\phi(t,x)_{min}}| < R \tag{4.8}$$

$$|Y_{\phi(t,x)_{min}} - Y_{\mathbf{v}|_{max}(t,x)}| < R \tag{4.9}$$

$$|Y_{|\mathbf{v}|_{max}(t,x)} - Y_{\phi(t,x)|_{max}}| < R, \tag{4.10}$$

where $R = 6$ is a threshold on the maximum number of grid boxes separating wind and temperature hourly changes. If the conditions 4.7 and 4.8-4.10 are not satisfied, no sea breeze is detected.

2. Let $X := \{(x_j, \widetilde{y_k}) : \widetilde{y_k} = 0, \forall k, j\}$, i.e. the set of coastal grid boxes. Then the previous steps are repeated $\forall x \in X$.

3. Then the sea breeze diagnostic variable based on (4.7) is defined.

- Let $D := \{x_l \in X : (4.7)\text{is satisfied for some } l\}$. Let also $E := \{x_j \in D : x_{j+1} = x_j + 1, \forall 1 \leq j \leq |X|\}$. Clearly $E \subseteq D$. E_j is the set of the adjacent coastal grid boxes. Then the following variables are defined:

$$S2 := \max_j |E_j| + 1 \tag{4.11}$$

It is required that $S2 > 4$, i.e. the number of the adjacent coastal grid boxes where the sea breeze is detected is at least four. If this condition is not satisfied then no sea breeze is detected.
Let $\mathcal{O} = \{t^*(x_j), x_j \in E\}$ the set of onset times of sea breeze along the adjacent grid points for each coastal grid box x_j. Finally, it is required that

$$\sqrt{\frac{\sum_{j=1}^{N} (\overline{t^*(x_j)} - t^*(x_j))^2}{N-1}} < \beta, \tag{4.12}$$

where $N = |\mathcal{O}|$ and β a variable threshold. In this case $\beta = 2$.

These tuning parameters G, M, R and β have been chosen based on the following rationale:

- G represents the maximum number of grid boxes far from the coast that sea breeze can be detected at its onset time. The choice was motivated by the fact only hourly data are available. Therefore at each time step model output sea breeze at certain speed. Looking at figure 4.7 it can be seen that near the coast sea breeze is slower than further from the coast, with a propagation speed in the range of $3 - 5\ km/h \sim 0.83 - 1.4\ m/s$. Therefore a value of $G = 4\Delta x = 8.8\ km$ is chosen, also to account for hourly data leading to underestimation of the sea-breeze propagation speed.

- M is motivated by the maximum propagation speed observed for sea breezes. This corresponds to a propagation speed of 22 $km/h \sim 6m/s$. This value is chosen to be above the maximum propagation speed observed in the south of England (Simpson et al., 1977) to take into account the uncertainty derived from having hourly data as for the choice of G.

- R measures how many grid boxes apart can be the kinematic and thermodynamic fronts during the penetration inland. Ideally R should not be very large, since it is not physically expected that changes in temperature and wind are too far apart. R was chosen in the interval $[5, 11]$.
  β measures how different are the onset times between adjacent coastal grid boxes where sea breeze is detected. This difference should be, since the front is a coherent phenomenon, so it is expected that adjacent grid boxes have a similar time when sea breeze initiates. β was chosen in the interval $[0.5, 2]$.
  Then a score for the probabilistic forecasts generated [6] for all the combination of R and β have been computed. The difference between the minimum and the maximum value of this score is $\sim 10^{-2}$. Also, the score is found to be more sensitive to β than to R. A value of $R = 6$ was chosen since, given hourly data are available, it corresponds to a propagation speed of $6\Delta x/h = 13.6km/h \sim 3.6m/s$, which is in the range of typical sea breeze propagation speed (Simpson et al., 1977). Therefore, since sea breeze arrival is anticipated by waves (Miller et al., 2003), it means that wind changes are usually ahead of the temperature change. It is therefore reasonable to assume that the distance between them corresponds to the distance travelled by the sea breeze at each model time step output (e.g. one hour).

Sea breeze probabilistic forecast The sea breeze detection method described in the subsection § 4.3.2 provides a definition of sea breeze occurrence which is used to create a daily forecast of sea breeze. The detection algorithm is applied to each CP-EPS member over the twelve hour period starting at 9UTC valid time. This defines a binary predictand

$$X(m,n) := \begin{cases} 1 & \text{if sea breeze occurs,} \\ 0 & \text{otherwise} \end{cases},$$

[6] The area under the ROC curve to be explained in the next chapter.

where m the ensemble member and n the valid day of the forecast. Using the standard formula for calculating probabilities from ensembles, the forecast probability of a sea breeze occurring is

$$\textbf{HR-DYN}(n) := \frac{1}{M}\sum_{m=1}^{M} X(m,n), \tag{4.13}$$

where M is the ensemble size (12 in the case of MOGREPS-UK). Due to the small ensemble size, events with very small or very large probability of occurrence are likely to result in forecast probabilities of zero or one. This can be problematic for information-based skill measures. Therefore, as in (Bröcker and Smith, 2008), the vanishing value of probability is substituted with a value of 1/(3M). An alternative method will be proposed in the appendix A. This alternative approach has not been used in this thesis.

4.3.3 Sea breeze prediction using MOGREPS-G

In this subsection the methodology used to produce the Bayesian forecast of sea breeze occurrence using MOGREPS-G variables as input is described. Due to the coarser grid box size and inability to properly represent sea breezes explicitly, the method applied to MOGREPS-UK, described in the previous section, is not appropriate.
Therefore large-scale predictors of sea breeze are sought. These have to be related directly to the underlying dynamics and resolvable in the global model.
Two candidates, as mentioned in section 4.1, are the synoptic wind and the temperature contrast between the land and the sea. The temperature contrast is the driving mechanism of the sea breeze, whereas the synoptic wind exerts an important influence on its occurrence and evolution.
The temperature contrast between the land and the sea is needed to create the density and pressure gradient in the direction perpendicular to the shoreline. This is the same mechanism driving the so called *density* or *gravity currents*. There have been many studies showing sea breezes behave like atmospheric density currents (Simpson, 1969; Simpson and Britter, 1980). The comparison between sea breezes and density currents is based on calculating the ratio between the sea breeze propagation speed and the density current speed, which is referred to in the literature as the *Froude number*. In (Simpson and Britter, 1980) laboratory experiments have been carried out considering the effect of head/tail ambient wind. The effect of the head wind is to reduce the rate of advance of the sea breeze front by about three fifths of the value of the opposing flow.
Arritt (1993) examined the effect of the ambient wind on the development of characteristic features of the sea breeze using a two dimensional numerical model. He found that both the direction with respect to the coast (offshore or onshore) and the magnitude of the ambient wind exert a strong influence on the sea breeze behaviour. The synoptic regime associated with the inland penetration of the sea breeze consists of either calm conditions or weak (i.e. up to 6 ms^{-1}) opposing (offshore) synoptic flow . With strong ($6-11$ ms^{-1}) or very strong ($> 11ms^{-1}$) opposing flow the sea breeze does not penetrate inland (remaining on the shoreline) or does not occur at all. More subtle is the case with onshore synoptic flow. In (Crosman and Horel, 2010, Table 4), a review of the studies investigating the effect of onshore synoptic flow on sea-breeze circulation is offered. Generally, an

onshore flow tends to weaken the sea-breeze circulation, since it leads to frontolysis. For onshore geostrophic wind of $2-4\ ms^{-1}$ sea breezes becomes almost indistinguishable from the large-scale flow.

More recently, Porson et al. (2007) investigated the effect of opposing wind on the sea-breeze inland propagation speed and occurrence. They developed an index for the sea-breeze occurrence, using time-integrated surface heat flux instead of land-sea temperature contrast. There are two different threshold values of this index for the sea-breeze occurrence, depending on the existence or not of an opposing geostrophic wind.

Other indices had been developed by Biggs and Graves (1962), Walsh (1974) and Frysinger et al. (2003), amongst the others. Biggs and Graves (1962) derived, using the Buckinghams' π theorem, different dimensionless ratios useful for understanding the sea-breeze behaviour. They claim the ratio $U^2/c_p\Delta T$ to be the dominant one and they called it "lake-breeze index". U^2 is the square of the the large-scale wind, c_p is the specific heat capacity of the air and ΔT the land-water temperature difference. They also claim the ratio to be the ratio of the inertial force and the buoyancy force, which is actually not. [7] Later, Frysinger et al. (2003), derived the ratio $U^2/\Delta T$, from the Bernoulli's equation. They included also a sign to discriminate the synoptic wind direction (positive for offshore winds, negative otherwise). Here, $U|U|$ and $c^2 = \alpha\frac{gH\Delta T}{T}$ are used as predictors. $|U|$ is the cross-coast offshore magnitude of the synoptic wind and $\frac{U}{|U|}$ its sign (positive for offshore direction and negative for onshore). c[8] represents the density current propagation speed, $\frac{g\Delta T}{T}$ is the reduced gravity (von Kármán, 1940; Keulegan, 1957; Benjamin, 1968), $\Delta T = T_{land} - T_{sea}$, H the fluid depth, in this case the height of sea breeze front head and α a density current coefficient to be calculated experimentally. For instance, Atkins and Wakimoto (1997) calculated α for offshore ambient winds cases, finding an average value of 0.7. The results of this chapter do not depend on the specific value chosen because the predictors are rescaled by the covariance matrix as part of the calculation used to create the forecast, as explained later. [9] Therefore, for simplicity, here $\alpha = 1$ is assumed.

In static ambient conditions, c gives directly the speed of inland propagation of the sea breeze (once it has initiated). However, in the presence of offshore winds sea breeze progress is slowed. The effect of the ambient wind is taken into account by the other predictor $U|U|$. Another factor leading to sea breezes propagating slower than the predicted c has been investigated by Robinson et al. (2013). They found, by means of 2D numerical simulations, that a gradual input of heat from a continental surface, rather than an initially specified density contrast as typical in laboratory analogs, can explain real sea breezes propagating at slower rate than the one predicted by the formula. However, their simulations did not consider other factors that could help to simulate real sea breezes, namely topography or surface roughness.

The two predictors are computed from MOGREPS-G at 12 UTC valid time for each forecast initialised at different times. To find ΔT, the 2 m temperature from the model is used,

[7] Both the numerator and denominator have units of m^2s^{-2}, which are not the units of the force. Furthermore, an expression for the specific buoyancy is given in the Equation 2.4.

[8] It should be noted that c is defined only if $\Delta T \geq 0$. If $\Delta T < 0$, c is not defined. In this case c^2 should not be interpreted mathematically as the square of c, but as another variable $f := -c^2$. For the sake of notation, the same symbol has been used when $\Delta T < 0$.

[9] Only the 1:1 line shown for reference in Figure 4.8 is affected by the choice of scaling.

above the land and the sea at a fixed distance from the coast, averaged along the coast direction. For H and T standard values ($H = 1000$ m and $T = 300$ K) are used, i.e. they are not output of the model. Note, as with the choice of α the results do not depend on the specific values chosen for H and T, however values are chosen to be physically realistic to aid physical interpretation of Figure 4.8. $H = 1000$ has been used for instance used by Miller et al. (2003), whereas Simpson and Britter (1980) used H=700, an empirical mean value over 54 cases. Also, Zhong and Takle (1993) show the sea breeze depth starting at 500-600 m, increasing to 1000 m and then decreasing to 800 m later. For U, the model wind output at 850 hPa pressure level (above the boundary layer height) is used, with the magnitude being averaged over the area of interest and taken with positive or negative sign if the mean direction is offshore or onshore respectively.

These two predictors define a two dimensional parameter space. The aim is to calculate the conditional probability of sea breeze occurrence given these predictors. It is therefore natural to apply a Bayesian framework.

Firstly the distributions of the occurrences and non occurrences of sea breeze in this space have to be estimated. This has been done in the past studies using mainly observational data (e.g. Frysinger et al. (2003)).

In this study, in order to train the Bayesian model, MOGREPS-UK data are used. This is because the main aim of this study is not to produce the most skillful Bayesian forecast, but to examine whether, with only few predictors, it is possible to retrieve the information content of a much higher resolution model. Also training on model data permits to have a larger sample size. In particular, given $X(m,n)$ from MOGREPS-UK, let $f(x|\theta_{sb})$ the distribution of $\{X(m,n) = 1\}$ (occurrences) and $f(x|\theta_{nsb})$ the distribution of $\{X(m,n) = 0\}$ (non occurrences), where $x = (c^2, U|U|)$ the vector of the two MOGREPS-G predictors defining the parameter space.

To estimate the two distributions, a *Gaussian kernel density estimation is performed*. It is a statistical method to estimate the probability density functions based based on the data (Parzen, 1962; Rosenblatt, 1956). The kernel is a smoothing non-negative function which is evaluated on the data points. A parameter called bandwidth regulates the level of smoothness of the function , with bandwidth calculated using Scott's rule (Scott, 1992), for both the occurrences and non occurrences categories, applied to the *training* dataset. The Scott's bandwidth is adaptive in the sense that it is dependent on the data density in the parameter space. In the next chapter it will be shown that the factor chosen corresponds to the highest value of a metric used to verify probabilistic forecasts. The Gaussian kernel density estimation has been implemented using the Python SciPy library (Jones et al., 2001), where the user can specify the rule to estimate the bandwidth (e.g. Scott in this case). As part of the standard kernel density estimation procedure the data are rescaled by their covariance matrix which eliminates any dependence of the smoothed PDF on the choice of scaling applied to the physical parameters (i.e. the values chosen for α, H and T). [10]

[10]The whole dataset is partitioned into the training dataset, which consist of the data used to build up model parameters (in this case of the Gaussian kernel). The rest of the data is used only to assess the performance of the model.

In order to test the differences between the two distributions a two dimensional Kolmogorov-Smirnov test (Justel et al., 1997) has been implemented. The null hypothesis is that two dimensional samples (occurrences and non-occurrences) are drawn from the same distribution. Also, to validate the Bayesian model the *leave-p-out* cross validation technique is adopted, with p =1 year (Shao, 1993; Celisse and Robin, 2008). More precisely the whole dataset comprises the years 2013, 2014 and 2015. A three-fold cross-validation is performed with training on two years and validation on the third year, in each cross-validation run. For instance if the year 2013 is validated, the training dataset consists of the years 2014 and 2015. In figure 4.8 the distributions of occurrences and non occurrences are shown with their respective density estimations.
Figure 4.8f shows the density estimations of all the data together, occurrences and non occurrences. The peak of the this PDF is located in the subregion corresponding to positive ΔT, which is expected as the summer seasons are considered, when land is warmer than the sea during the day.
Then, looking at distributions of occurrences (figure 4.8d) and non occurrences (figure 4.8e), it is clear that these two PDFs are different. This means that these parameters can provide some information about the sea breeze occurrence. The Kolmogorov-Smirnov test gives a p-value of $2.72 \cdot 10^{-72}$, which implies that the null hypothesis can be rejected at 1% level of significance. This means the two samples (occurrences and non-occurrences) are drawn from different distributions. The diagonal line in figure 4.8 is the 1:1 line, dividing the parameter space in subregions. The peak of the distribution of the occurrences is in the subregion of the space ($0 < U|U| < c^2$) (hereafter called the *wedge*) where the sea breeze is expected to occur, consistent with dynamical mechanisms, where the temperature contrast is positive and the opposing ambient wind not strong enough to prevent its initiation. On the other hand, the peak of the non occurrences distribution is shifted towards higher magnitudes of the opposing wind and lower temperature contrast values.
In order to establish whether there is a relationship between the two distributions (occurrences and non-occurrences) and their position in the parameter space, a χ^2-test of independence is performed based on the contingency table 4.2. The null hypothesis is that the rows and columns of the table are independent, implying that being inside or outside the *wedge* is not related to the sea-breeze condition. The test gives a p-value of $2.098 \cdot 10^{-73}$, implying there is enough evidence to reject the null hypothesis of independence. It can be therefore concluded, from the two statistical tests performed, that the two samples of occurrences and non-occurrences are drawn from two different distributions and this difference is not due to the different sample size, but to the different meteorological conditions leading (preventing) to the occurrence (non-occurence) of sea breezes. The 1:1 slope is not calculated empirically from the data, was added as a visualization

	Inside	Outside
Occurrences	260	93
Non-occurrences	509	1544

TABLE 4.2: Number of data points used in the figure 4.8 according to the sea-breeze definition and the position in the parameter space. Inside and outside are relative to the *wedge*.

tool to highlight the region where the peak of the distribution of occurrences is and to illustrate the relationship between the parameters and sea-breeze occurence. Other slopes, or equivalently other values of the ratio $\frac{U|U|}{c^2}$ which can discriminate better between the occurrences and non-occurrences, can be tested to see how they affect the entries of the table 4.2 and the alternative forecast which will be presented later in the chapter. Anyway such sensitivity tests are beyond the aim of this thesis and can be considered for future work.

These two distributions can be used to calculate the probabilistic forecast of sea breeze conditioned on the information given by $x = (c^2, U|U|)$. Let θ_{sb} be a discrete binary variable, representing the event $X(m,n) = 1$, i.e. the sea breeze has occurred in the MOGREPS-UK forecast. As prior knowledge also $\pi(\theta_{sb})$, the climatological probability of sea breeze occurring in the MOGREPS-UK forecasts is also available. The aim is then to update this information, after observing x.

Therefore it is natural to calculate the posterior distribution using Bayes' formula:

$$\pi(\theta_{sb}|x) = \frac{f(x|\theta_{sb})\pi(\theta_{sb})}{f(x)}, \tag{4.14}$$

where

$$f(x) = f(x|\theta_{sb})\pi(\theta_{sb}) + f(x|\theta_{nsb})\pi(\theta_{nsb}), \tag{4.15}$$

where $\pi(\theta_{nsb}) = 1 - \pi(\theta_{sb})$.

$f(x|\theta_{sb})$ and $f(x|\theta_{nsb})$ are the probability distribution functions (PDFs) of the occurrences and non occurrence respectively calculated using a Gaussian kernel density estimation, mentioned beforehand. In particular $f(x|\theta_{sb})$ is the so called likelihood function.

The distribution of $\pi(\theta_{sb}|x)$ in the parameter space is shown in figure 4.8c. It can be seen the higher values of probability are in correspondence of high values of c^2 (and so, high values of ΔT) and low values of $U|U|$.

The use of Bayesian model and/or kernel density estimation to calculate conditional probabilities is not new. For instance Hamill and Church (2000) used the Bayesian framework to calculate conditional probabilities of tornadic storms without using a kernel approach, whereas, more recently, Taszarek et al. (2017) used the kernel approach but not Bayes' rule. Also, the scatter plots in figure 4.8a,b are essentially similar to Frysinger et al. (2003, Fig. 3). In that case the aim was to find a proper threshold of the ratio $\frac{U|U|}{\Delta T}$, such that for any value below this threshold most of the non occurrences were excluded. By using this kernel approach it is not necessary to find such a threshold, so again it can be *in principle* generalized to any part of the world.

In operational contexts where these CP-EPS data are not available for some reason, coastal station observational data could be used instead. An example of the same plot, using observations rather than model data is shown in figure 4.9. It is worth nothing that it shows very similar result to when the distributions are trained with MOGREPS-UK data. This means that this method has potential applicability even in regions where CP-EPS data are not available.

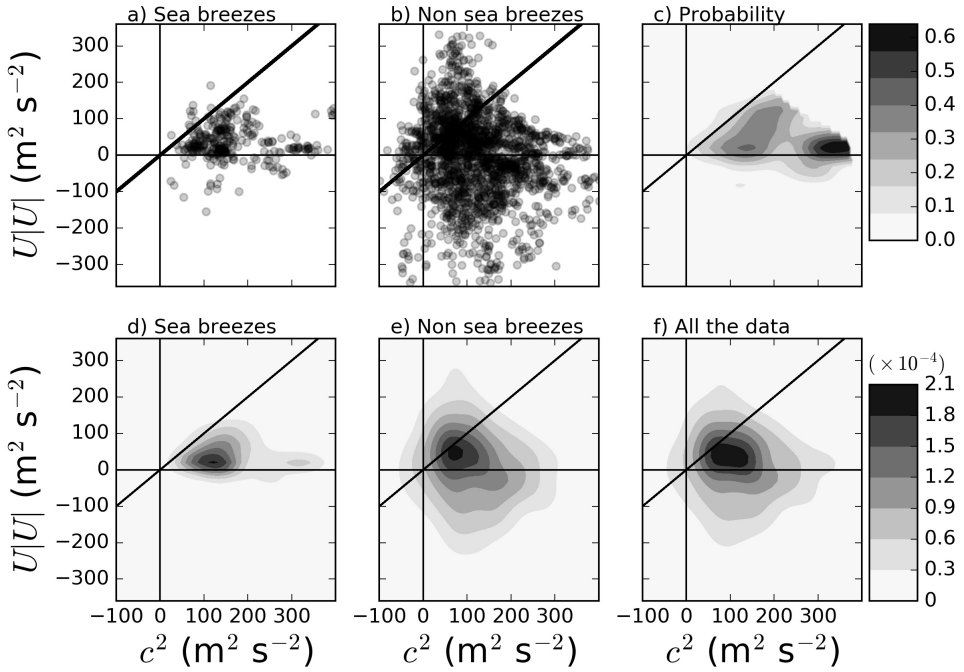


FIGURE 4.8: Distribution of a) sea breeze occurrences and b) non occurrences in the parameter space defined by c^2 and $U|U|$ for all the years of our study at 12 hours ahead. d), e) and f) show the kernel density estimation for a) and b) and the total distribution respectively. c) represents the conditional probability values $\pi(\theta_{sb}|x)$ in this parameter space. As kernel bandwidth the one proposed by (Scott, 1979) was used. All ensemble members and all the dates of the period studied here have been considered in this plot. Contour fields are limited to values of the kernel density estimation higher than 10^{-6}. The solid line represents the 1:1 line.

Equation (4.14) was used to calculate the probabilistic forecast of sea breeze using MOGREPS-G predictors (hereafter called **LR-BAY**):

$$\textbf{LR-BAY}(n) := \frac{1}{M}\sum_{m=1}^{M}\pi(\theta_{sb}|x(m,n)), \tag{4.16}$$

where M is the number of ensemble members, 12 in case of MOGREPS-G.

4.4 Comparison

In this section a qualitative comparison is shown between **LR-BAY** and **HR-DYN** forecasts for the sea breeze occurrence for different lead times against station observations. In Murphy (1993, Table 2) a list of the properties related to the quality of probabilistic forecasts is presented. Some of these properties depend solely on the forecast distribution, others also on the observation. A property that it is interesting to compare for **HR-DYN** and **LR-BAY** is *sharpness*. Sharpness is the tendency of a forecast to predict values close to 0 or 1 and it is a property that depends on the forecast distribution alone.
This can be shown by binning the probabilistic forecast into categories and counting the instances for each category. The result is shown in figure 4.10 for four different lead times.

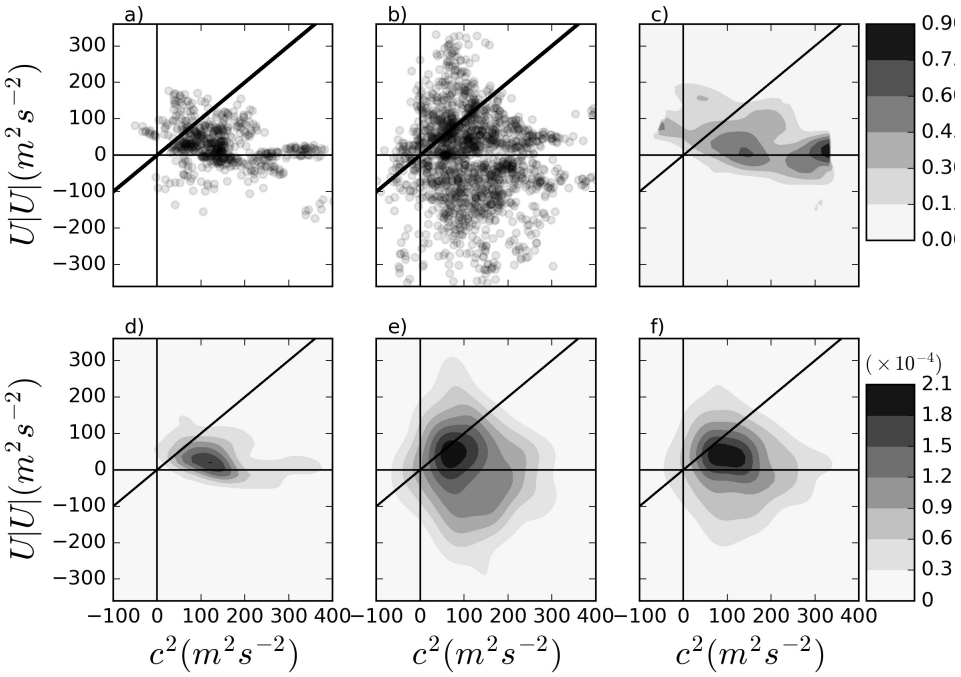


FIGURE 4.9: Same as in figure 4.8, with the Bayesian model trained on observations. Sea-breeze identification using observations is explained in the section 4.3.1.

It can be notice that **HR-DYN** is sharper than **LR-BAY** for all lead times, i.e. it tends to forecast values near 0 or 1 more often than **LR-BAY**. Thus there is more uncertainty in **LR-BAY** about sea-breeze occurrence. By looking again at figure 4.8d, e, this can be explained by the overlapping of the two distributions (occurrences and non occurrences). A more definite separation between the two distribution would have resulted into a sharper Bayesian forecast.

Another comparison that does not depend on observation is the Spearman correlation coefficient r to assess whether there is a monotonic relationship between **HR-DYN** and **LR-BAY**. The results are shown in figure 4.11. A perfect correlation would have a value of $r = 1$. It can be seen that the highest correlation ($r \approx 0.3$) corresponds to $T + 12$h and $T + 24$h forecast, whereas at $T + 18$h and $T + 30$h, the forecasts are not correlated. This implies that the forecast performance will be more similar at $T + 12$ and $T + 24$ than for other lead times. It will be seen in the next chapter that is actually the case and a possible reason for this behaviour will be offered.

For a first subjective comparison against observations, in figure 4.12 two visible satellite pictures for two different days are shown. These correspond to days with the highest probability of occurrence (according to **HR-DYN**). In both cases it can be noticed the typical cloud free strip along the south England and French coasts associated with a sea breeze, due to penetration of the cooler air mass from the sea. This distinctive feature has been used also to detect sea breeze fronts in many studies (e.g. Bigot and Planchon (2003); Planchon et al. (2006); Corpetti and Planchon (2011)).

For a more quantitative verification, figure 4.13 shows the T+12h time series forecasts for the summer season 2015. This plot also shows the instances when sea breezes were

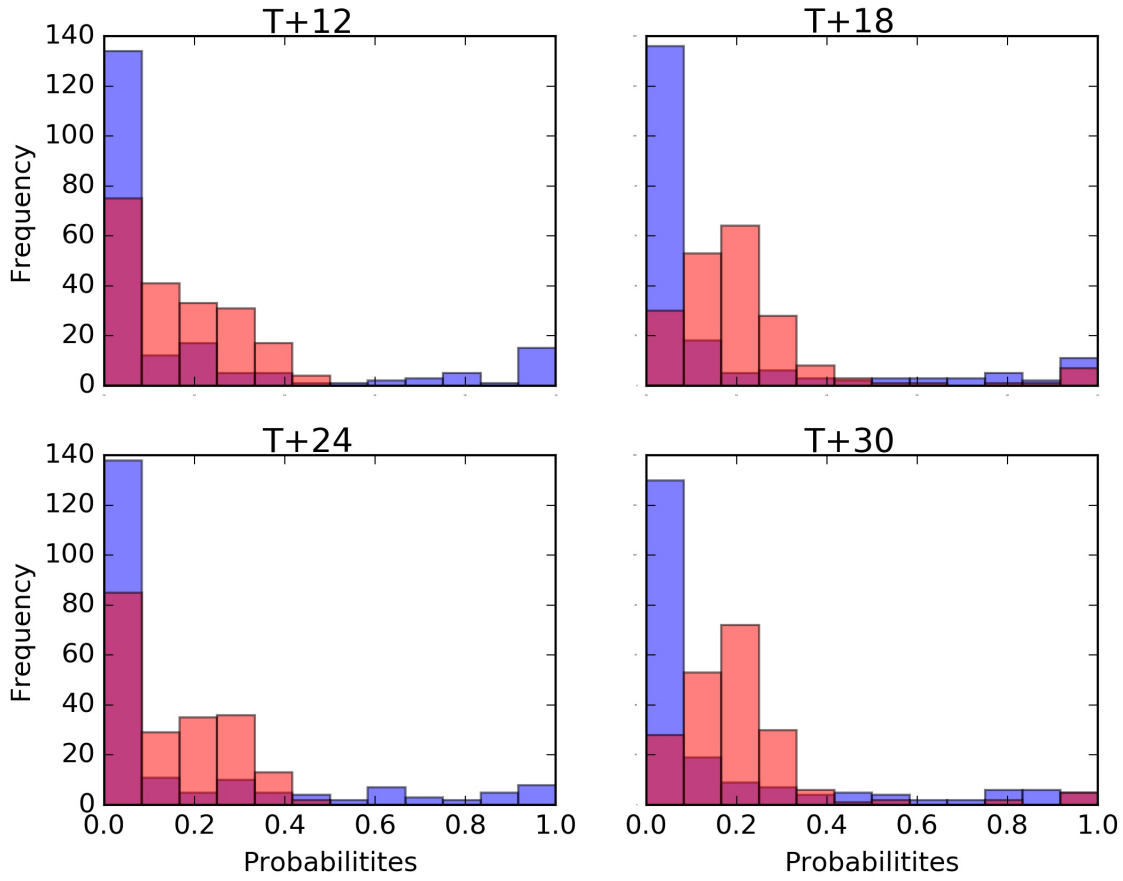


FIGURE 4.10: Sharpness diagram for **HR-DYN** (blue) and **LR-BAY** (red) for different lead times over all the period considered. Purple bars correspond to the overlapping of the blue and red bars.

observed or not, according to the algorithm just described. The probability values on occasions when a sea breeze is observed in surface stations are plotted above the horizontal line, whereas if the sea breeze is not observed the values are plotted below. It can be also seen that **HR-DYN** is a better detector of sea-breeze occurrence, with a greater proportion of blue bars above the 0.5 line with compared to LR-BAY.

However, it should be noticed that, unlike for **HR-DYN**, **LR-BAY** never exceeds 0.5. Therefore in order to make a first comparison the number of occasions when sea breeze is observed (or not) and probability values exceed (or not) a threshold is counted. This threshold has chosen to be the half of the maximum respective probability values ($t_1 = 0.5$ and $t_2 = 0.21$) for the two models. The results are shown in table 4.3. This table can be interpreted as an usual contingency table. Firstly, it can be noted that **HR-DYN** has much fewer false alarms than **LR-BAY** (3 and 41 respectively). On the other hand, the misses are more comparable (29 and 23). More precisely, when sea breeze is observed **HR-DYN** has proportion of probabilities exceeding the threshold of $24/53 \approx 0.45$, whereas **LR-BAY** of $30/53 \approx 0.57$. When sea breeze is not observed the proportions are more different. **HR-DYN** has a proportion of $144/147 \approx 0.98$ of low probabilities, where **LR-BAY** a proportion of $106/147 \approx 0.72$. This already suggests **HR-DYN** and **LR-BAY** are similarly able to discriminate events occuring, but **HR-DYN** is more able to discriminate

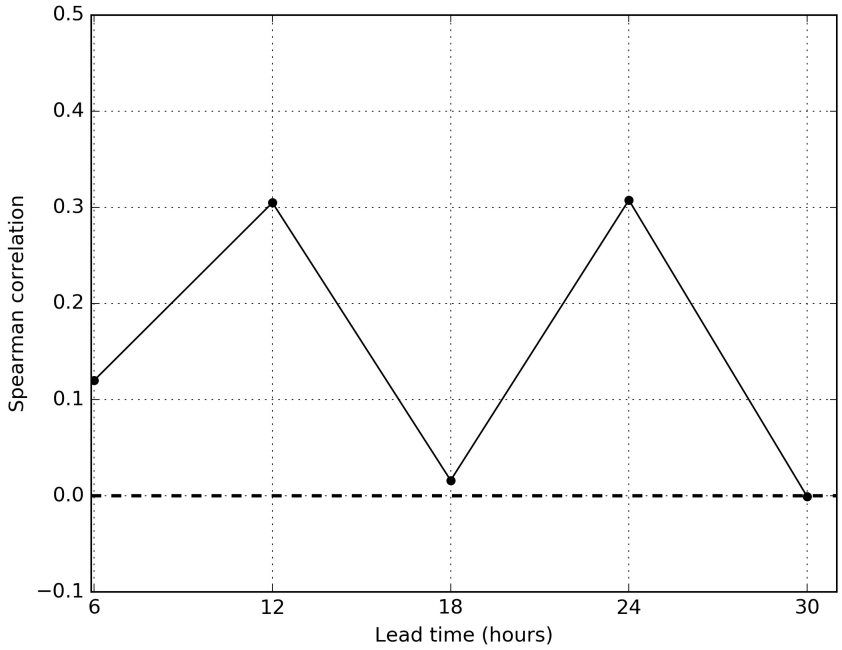


FIGURE 4.11: Spearman correlation coefficient between **HR-DYN** and **LR-BAY** for different lead times.

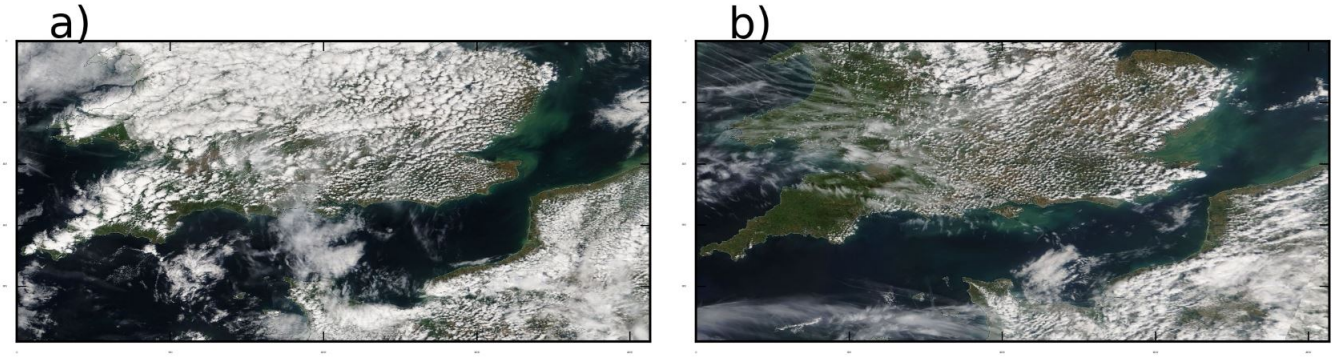


FIGURE 4.12: MODIS visible image of south UK and English Channel on a) 6 August 2013 at 1133 UTC and b) 7 August 2014 at 1145 UTC. Source: https://worldview.earthdata.nasa.gov/

no-events than LR-BAY. This property will be more fully described and quantified in the next chapter.

	Event observed	Event non observed
HR-DYN $> t_1$	24	3
HR-DYN $\leq t_1$	29	144
LR-BAY$> > t_2$	30	41
LR-BAY $\leq t_2$	23	106

TABLE 4.3: Number of occurrences and non-occurrences of sea breeze in relation to probability values exceeding half of the maximum probability value. This refers to the whole dataset for the T+12 forecast.

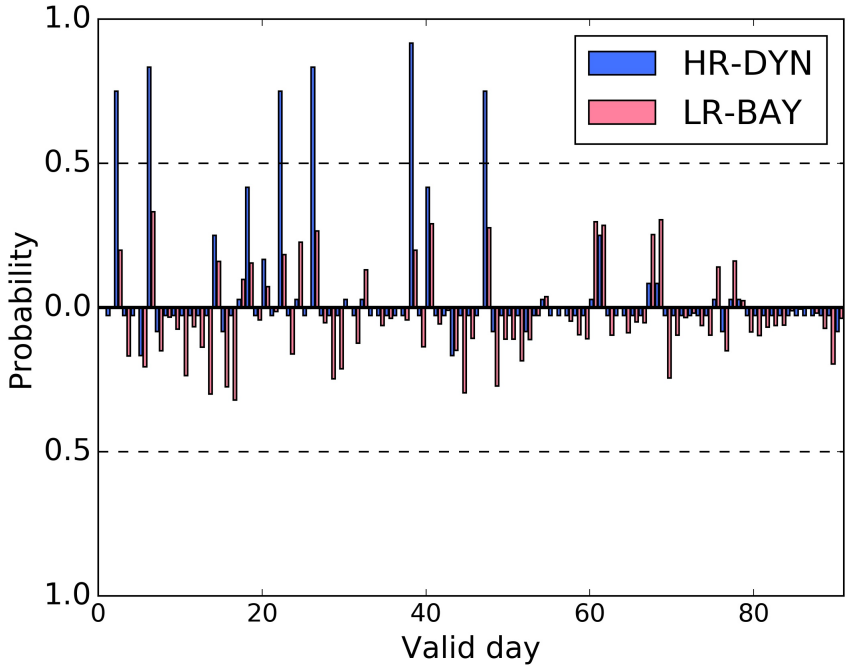


FIGURE 4.13: Time series of HR-DYN and LR-BAY for T+12 (hours) forecast lead time. The values above the black solid line refer to the *observed* sea breeze occurrence, whereas the values below correspond to days when sea breeze has not been observed.

4.5 An alternative method for sea-breeze occurrence forecast

Here an alternative sea-breeze occurrence forecasting method is proposed. The rationale for this method lies in the fact explained in section 4.3.3 about the relationship between sea-breeze occurrence and the physical parameters c^2 and $U|U|$. In particular it has been shown there is dependence between the sea-breeze occurrence and the region of the parameter space defined *wedge* earlier. Therefore the following variable is defined:

$$Y(m,n) := \begin{cases} 1 & \text{data point inside the } \textit{wedge}, \\ 0 & \text{otherwise} \end{cases},$$

where m the ensemble member and n the valid day of the forecast. Using the standard formula for calculating probabilities from ensembles the probability forecast of a sea breeze occurring is

$$\mathbf{WF}(n) := \frac{1}{M} \sum_{m=1}^{M} Y(m,n). \tag{4.17}$$

4.6 Other sea breeze characteristics

In this section other sea breeze characteristics beyond the occurrence have also been investigated. In particular, propagation speed and onset times have been calculated empirically from MOGREPS-UK. The inland propagation speed and onset time are other two important characteristics of the sea-breeze phenomenon that are useful to predict. These

two combined can give information about the time of the arrival of the sea-breeze front at locations inland.

Propagation speed and onset times The sea breeze propagation speed (u_{sb}) is calculated for sea breezes that have penetrated inland for at least 33 km (one MOGREPS-G grid box size). Following the description of the sea-breeze algorithm, the position of the sea breeze front is defined as the position of the leading edge based only on the wind direction criterion.
More precisely, as explained in the section 4.3.2, sea breeze is first detected at time t^* within G gridboxes from the coast. Then the propagation speed, from MOGREPS-UK, is simply the inverse of the time taken to travel between the coast and 33 km inland. This distance has been chosen, since one station inland is located that distance (see again table 4.1), making the verification easier.
The speed is then calculated for each coastal grid point on the coast where the sea breeze occurs (i.e. the set E defined in the section 4.3.2). Therefore, each coastal grid box will, in principle, have its own propagation speed value. Then, the mean value of the different speeds of these adjacent grid boxes is considered. Thus, the propagation speed in the early stage of the inland penetration is calculated. However, it can be noticed in figure 4.7 the fact that sea breeze tends to accelerate further inland (e.g. by noticing the steeper wind and temperature profiles). This acceleration has been documented by Simpson et al. (1977, Fig. 15) in observed sea breezes and Robinson et al. (2013) in numerical simulations in the continuous heating case.
The onset time is defined as the time at which the sea breeze front is located within G grid boxes from the coast, as explained in the section 4.3.2. For example, the onset time is represented by the vertical dashed line in figure 4.7. Similarly as for the propagation speed, the mean value of the onset times of adjacent grid boxes, where the sea breeze has occurred, is considered.
It is worth to note, however, that for the quantification of speed and timing only hourly data were available, both from the model and observations. This is a source of error in the estimation which should be considered in the future work for a more careful assessment.

Froude number and inland penetration These other two characteristics have been calculated in order to compare with values found in the literature and not for the probabilistic prediction itself.

Froude number *Froude number* has been already introduced in chapter 2 and following the notation of this chapter can be rewritten as:

$$\mathrm{Fr} = \frac{u_{sb}}{\sqrt{\frac{gH\Delta T}{T}}} \tag{4.18}$$

u_{sb} has been estimated from MOGREPS-UK as explained in the previous section, whereas ΔT from MOGREPS-G, as describer earlier in this chapter, with $H = 1000m$ and $T = 300K$.

Inland penetration The estimation of maximum inland penetration is based on the data already shown in figure 4.7. For each member that satisfies the sea breeze condition, the furthest grid point inland reached by sea breeze front (tracked by the wind direction profile) is flagged as the maximum inland penetration.
The frequency distributions of the these four characteristics are shown in figure 4.14. In

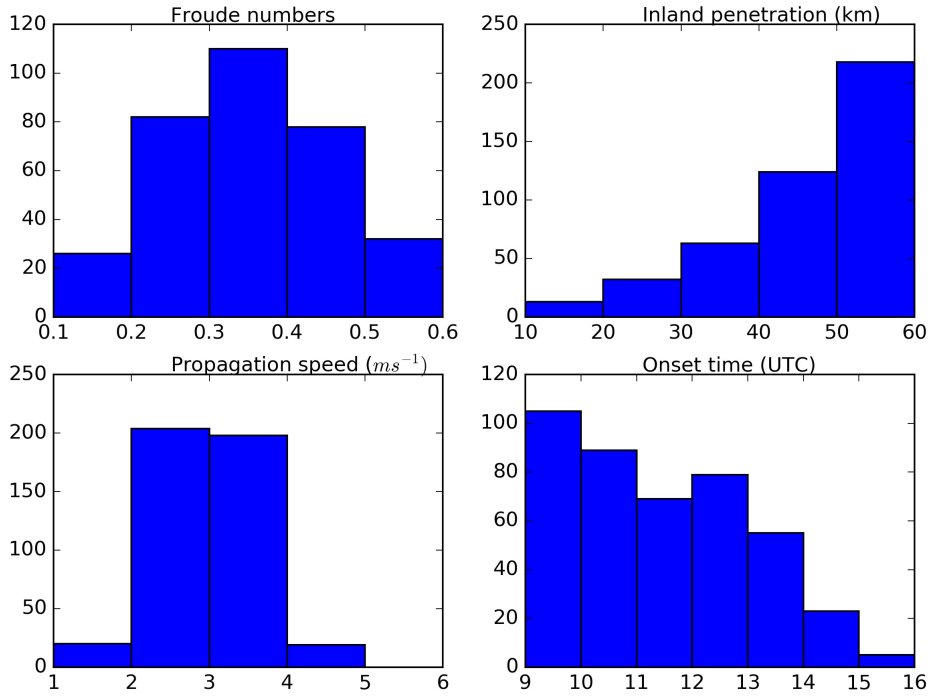


FIGURE 4.14: Frequency distribution for different sea breeze characteristics: Froude numbers, inland penetration, propagation speed and onset time calculated from T+6 and T+12 MOGREPS-UK forecasts.

regard to the propagation speed, most of the values are in the range $2-4\ ms^{-1}$, which are consistent with values found by Simpson et al. (1977) in the South of England ($3ms^{-1}$), whereas Finkele (1998) reported values of $1-2\ ms^{-1}$ from aircraft data on the long, straight coastline of the Coorong region near Woods Well in South Australia.
The Froude numbers are also in agreement with values found by (Robinson et al., 2013) for the continuous heating case in the numerical simulations, which is dynamically closer to the real sea breeze than the lock exchange experiment (see figure 2.3 for Froude number values of the lock exchange numerical simulations.)
Then sea breezes are predicted to penetrate inland up to 60 km most of the time, reaching the edge of the sub-domain considered in this study, which again is not uncommon for Southern England, as shown by (Simpson et al., 1977). It is worth to note that the sea-breeze tracking algorithm stops at the upper edge of the domain considered in this study. This is because the focus is on the early stage of the sea-breeze penetration. However it is possible that sea-breezes can propagate further inland.

In regard to the onset times, most of the sea breezes initiate in the morning (before 12 UTC), and very rarely after 14 UTC. This is not surprising, considering the very early sunrise time at this latitudes in Summer. However, most of the sea breezes occur between 9 and 10 UTC. This might suggest that an earlier time window should be considered.
To conclude, the consistent values found here for the propagation speeds and inland penetration are another clue of the validity of the sea-breeze tracking algorithm.

4.6.1 Probabilistic forecasts of onset times and propagation speeds

Probabilistic forecasts of propagation speeds and onset times exceeding a defined threshold have been calculated from both MOGREPS-UK and MOGREPS-G have been computed.

MOGREPS-UK Let **OT** and $\mathbf{SB}_{speed}$ the variables describing the onset time and the propagation speed respectively and $\bar{t}$, $\bar{s}$ the respective thresholds.
Therefore the following probabilities can be calculated:

$$\mathbf{P}(\mathbf{OT} > \bar{t}) = \frac{1}{M}\sum_{m=1}^{M} X(m,n) \tag{4.19}$$

$$\mathbf{P}(\mathbf{SB}_{speed} > \bar{s}) = \frac{1}{M}\sum_{m=1}^{M} Y(m,n), \tag{4.20}$$

where

$$X(m,n) := \begin{cases} 1 & \text{if} \quad \mathbf{OT}(m,n) > \bar{t}, \\ 0 & \text{otherwise} \end{cases},$$

and similarly for $Y(m,n)$ with $\mathbf{SB}_{speed}$.

MOGREPS-G In regard to MOGREPS-G, in analogy with the occurrence case, the ability of the parameters $U|U|$ and c^2 to provide some information about the propagation speed and the onset times is tested. Figure 4.15 shows the distribution of the propagation speed and onset times values as a function of the two parameters. The values of the propagation speed and onset times, calculated for each MOGREPS-UK ensemble member as explained in the previous section, are put in relation with the matching MOGREPS-G member. Then an elliptical shaped neighborhood [11], with the centre in each MOGREPS-UK data point, is defined. Then the mean value of all the points within each neighborhood is calculated and plotted as a function of $U|U|$ and c^2.

The relationship between the sea breeze propagation and the MOGREPS-G parameters is not very clear. In regard to the onset times, it can be immediately seen that the the earliest onset times correspond to the lowest positive values of $U|U|$. Again this is physically consistent with the fact that strong offshore winds could prevent or delay the onset time of the sea breeze.

[11] The two half width of the major and minor axes of the ellipse have chosen to be $2gH/T$ m^2/s^2 and 10 m^2/s^2.

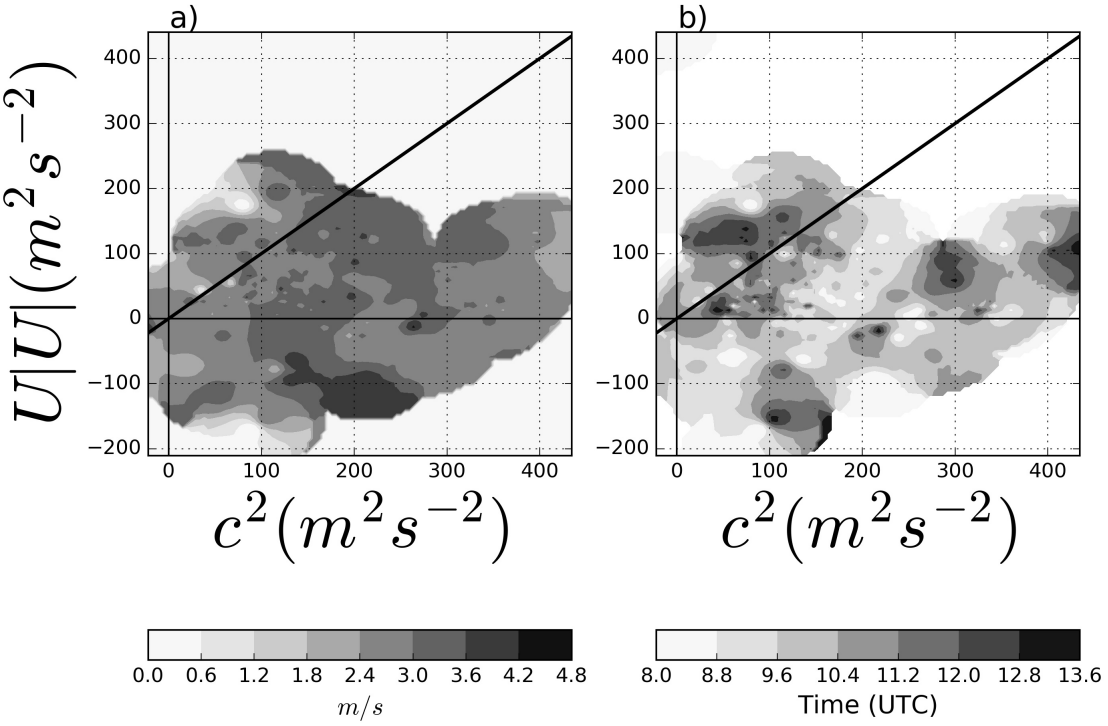


FIGURE 4.15: Distribution of a) propagation speed and b) onset time of sea breeze in the parameter space defined by $U|U|$ and c^2 at 12 hours ahead.

With this method, the functions $\widehat{OT} = f(c^2, U|U|)$ and $\widehat{SB}_{speed} = g(c^2, U|U|)$ can be estimated. The two events to be predicted are conditioned on the event that the sea breeze has occurred. Also in the observations, non-occurrences have been excluded from the dataset. This is simply because onset times and propagation speed are defined only when the sea breeze occurs.

Therefore, similarly as with MOGREPS-UK, the following probabilities can be calculated:

$$\mathbf{P}(\widehat{OT} > \bar{t}|\theta_{sb} = 1) = \frac{1}{M}\sum_{m=1}^{M}\widehat{X(m,n)} \tag{4.21}$$

$$\mathbf{P}(\widehat{SB}_{speed} > \bar{s}|\theta_{sb} = 1) = \frac{1}{M}\sum_{m=1}^{M}\widehat{Y(m,n)}, \tag{4.22}$$

where

$$\widehat{X(m,n)} := \begin{cases} 1 & \text{if} \quad \widehat{OT}(m,n) > \bar{t}, \\ 0 & \text{otherwise} \end{cases},$$

and similarly with $\widehat{Y(m,n)}$ for $\widehat{SB}_{speed}$.
The verification of these probabilistic forecast will be shown in the next chapter.

In regard to observations, onset time is defined for each coastal station independently. An example for one coastal station, for two different days is shown in figure 4.4, where the onset time is described with a vertical dotted line. Then a mean onset, between the coastal stations detecting sea breezes is calculated. For the propagation speed, it is calculated as the inverse of difference between the times of detection at coastal station and the station inland.

4.7 Summary and conclusions

In this chapter the post-processing of two ensemble prediction systems for the generation and comparison of sea-breeze probabilistic forecasts has been presented.
Two different methodologies have been applied to the two EPSs with different grid sizes. The sea breeze is a mesoscale phenomenon, which is realistically represented only at convection-permitting resolutions. This has allowed an automatic detection of the sea-breeze phenomenon in MOGREPS-UK, using a novel tracking algorithm. The algorithm makes use of only very few thresholds on the model variables, which are mainly required to avoid any spurious detection (false alarms or misses). In principle, this tracking algorithm can be applied to any coastline, provided with very fine resolution gridded model data.[12]
In regard to the coarser resolution model a different approach has been proposed to extract information about sea-breeze occurrence. Two model parameters reflecting the key elements of the meteorological environment (temperature contrast and large-scale wind) have been chosen as predictors. Then the conditional probability of the sea breeze occurrence based on these two parameters has been estimated in a Bayesian framework.
A simple algorithm for detecting automatically sea breezes in surface stations has also been presented, using again few thresholds on the meteorological variables involved. Finally, an alternative method for forecasting sea-breeze occurrence was also presented. It is based on the relationship between the sea-breeze occurrence and its relationship with the large-scale predictors, without using any kernel smoothing.
The main conclusions from this chapter can be summarised as follows:

1. the two selected predictors for the sea-breeze occurrence have shown some ability to discriminate between sea-breeze occurrence and non-occurrences as defined either by MOGREPS-UK data or observations.

[12] This algorithm is now currently being tested at the Bureau of Meteorology (Australia) by Dr Peter Steinle (personal communication), using a version of the local regional model ($\sim$ 4 km grid size), based on the UK Met Office Unified model.

2. The correlation between the convection-permitting probabilistic forecast and the Bayesian forecast varies with lead time, peaking at 12 and 24 hours ahead (≈ 0.3) and vanishing at other lead times.

3. A preliminary comparison with observations, suggests that they are equally able to discriminate event occurrences, with convection-permitting forecast more able to discriminate no-events than the Bayesian model (lower false alarm ratio)

4. Other sea-breeze characteristics (inland penetration, propagation speeds and onset times) are shown to have consistent values with the previous literature.

.

In summary, this confirms that in a particular synoptic regime (positive land-sea temperature contrast and light opposing wind) the sea breeze is much more likely to occur. However, as shown in figure 4.8 sea-breeze could also occur outside the *wedge*, meaning that this condition is not sufficient to indicate the sea-breeze occurrence. However whether the convection-permitting model is providing additional information is still not known at this stage.
In order to quantify this information gain it is thus necessary to verify the two probabilistic forecasts against the station observations.

Chapter 5

Verification of sea breeze probabilistic forecasts

In this chapter the verification of probabilistic forecasts, calculated in chapter 4, is carried out. The inter-comparison between the convective-scale forecast and the Bayesian model is analyzed in terms of *reliability* and *resolution*, two necessary attributes of a skillful probabilistic forecast.
Statistical tests are performed to assess these two properties, firstly for the two forecasts individually. Then the difference in the resolution property is assessed by comparing the areas under the ROC curve.
Finally, Brier Score differences and Information gain are computed with respect to observations for the forecasts pair defined in chapter 4 and for each probabilistic forecast relative to the climatology.
The work presented in this chapter is part of the same peer-reviewed paper as in chapter 4.

5.1 Introduction

Once computed probabilistic forecasts, which quantifies how likely is the occurrence of the event to be predicted, it is necessary to assess how well the probabilistic forecasts are performing. Quantifying the forecast skill in a objective way is the aim of the *forecast verification*. Since the introduction of NWP in 1950s, it has been an active research topic. Joliffe and Stevenson (2012) offer a review of the forecast verification theory, along with different methodologies to perform it practically.
In this chapter, the verification of probabilistic forecasts of a binary predictand (sea breeze occurrence) is performed. In the previous chapter the forecasts and observations dataset have been described.
Two properties that probabilistic forecasts (not only weather/climate related) should have are *reliability* and *resolution*. A qualitative description of these two properties can be found in Murphy (1993). Broadly speaking, *reliability* measures how close the distribution of the observations, conditional on the forecasts, is to forecast distribution. *Resolution* is the ability of a forecast to sort or resolve classes of events with different observed frequencies.[1]
Reliability and resolution together determine the forecast skill. This is measured usually

[1] http://www.cawcr.gov.au/projects/verification/; Accessed online on 20 December 2018.

relative to a reference forecast and by defining a metric, called *score*. A score is a summary measure. Here, two scoring rules are used for the verification: Brier Score (Brier, 1950) and ignorance score (Roulston and Smith, 2002). They have been chosen for two reasons. Firstly they are both *proper* (Bröcker and Smith, 2007b). Being proper (strictly proper) means that the reward associated with the score is maximised if (and only if) the true probability distribution is forecast. This encourages the forecaster to issue his or her true belief.
Secondly, they are related. Benedetti (2010) proved the Brier score is a second order approximation of the ignorance score. However, there are some differences. The main difference is that the ignorance score is *local*, whereas Brier score is not. Local scores depends only on the probability assigned to the event that actually occurs and not on the full probability distribution. However, in case of binary predictands, as in this thesis, all the scores are local. Also, the two functions the scores are based upon are different. Therefore some differences between the two scores are expected.
Roulston and Smith (2002) and Benedetti (2010) showed that the two scores judge forecasts differently, especially for very rare or very frequent forecasts. In particular ignorance score penalises poor sharp forecast more than the Brier score.
Throughout this chapter no serial correlation between forecast and observation pair is assumed. This is a reasonable assumption, since sea breeze occurrence is verified every 24 hours, which is longer than the typical sea breeze time scale. This assumption allows to neglect correlations when calculating the standard error of the differences with the bootstrap method.
Also, the results presented in this chapter have been cross-validated. More precisely, to verify the Bayesian forecast, as explained in the previous chapter, a three-fold cross-validation is performed with training the model on two years and validation on the third year, in each cross-validation run. For instance if the verification is performed for the year 2013, the training dataset consists of the years 2014 and 2015.
The rest of the chapter is structured as follows. In section 5.2 reliability property will be defined and verified for both the probabilistic forecast individually, in section 5.3 resolution is first defined and then tested for both the forecasts individually and then compared for the forecast pair, section 5.4 contains the summary scores results. In section 5.5 the verification of the alternative forecast presented in the previous chapter is performed, whereas section 5.7 contains the verification of the sea breeze onset time and propagation speeds. The sensitivity of the Bayesian model performance to the kernel bandwidth is then presented in section 5.8 and finally in section 5.9 conclusions are discussed.

5.2 Reliability diagram

The reliability diagram is a tool to visualize the reliability property, by plotting the conditional observed frequencies against the predicted probabilities (see (Murphy and Winkler, 1977) for example). For a reliable system points should lie along the diagonal; meaning, for example over all the occasions in which an 80% probability of an event is forecast, the event should occur 80% of the time. In order to plot the reliability diagrams, the probabilistic forecasts are first partitioned into bins (here we use 4 bins). Then observed

frequencies and the mean value of the probabilities forecasts are computed for each bin. Bröcker and Smith (2007a) propose a new method to take into account sampling fluctuations in the reliability diagram, since these can lead to misleading evaluation. Even for a perfectly reliable forecast, due to limited statistics, it is not expected that all the points lie along the diagonal line. Therefore this method helps to assess whether any deviation from the diagonal line is due to small sampling or to an unreliable forecasting system. In other words, being close to the diagonal line does not necessarily imply that the system is reliable.
The method consists into resampling with replacement from the original pair of forecasts and observations. A pair of surrogate observations and forecast is created and the surrogate observed frequencies are recorded. This procedure is repeated N_{boot} times and the range of the possible surrogate observed frequencies is plotted for each bin, represented by the vertical $5\% - 95\%$ error bar. A reliable system is the one for which the observed frequencies lie within the range of possible surrogate observed frequencies, i.e. within the bar.
Reliability diagrams of **HR-DYN** and **LR-BAY** for all lead times are shown in figure 5.1. Firstly, it can be seen that for **HR-DYN** most of the observed frequencies, for all lead times, lie in the 90% confidence interval. However, for all lead times, in the lowest probability bin, the observed frequency is significantly higher than the forecast probability, i.e. the forecast is overconfident, even though it is close to the diagonal it is outside the error bar. It can be seen that for the T+12 forecast the observed frequency in the second bin is further from the diagonal than the one in the first bin. However, it is inside the range of expected frequencies. This is just an example of how this resampling method can reinforce reliability.
This makes the forecast not reliable overall, despite good reliability in the other bins (except for the second lowest bin at T+6 and T+30). **LR-BAY** is not reliable. In addition to the lowest bin, also the bin with the higher probability values is problematic, at 6, 18 and 30 hours ahead. In this case it can be seen that the observed frequency is significantly lower than the forecast probability, thus indicating that the forecast is overconfident in this case.

5.3 Resolution

In this section the *resolution* property of both forecasting systems is tested. The null hypothesis to test is that that they have no resolution. Hereafter the verification dataset will be indicated by Y in the equations.
Thus the mathematical definition of *no resolution* is introduced:

$$\mathbb{P}(Y = k, p = p_d) = \mathbb{P}(Y = 1)\mathbb{P}(p = p_d) \tag{5.1}$$

where $k = \{0, 1\}$ in this case and $\{p_d\}$ is the set of finite values that the forecasting scheme could take, after binning it, for $1 <= d <= D$.
This condition describes *independence* between the probabilistic forecast and the observations. In other words, if equation (5.1) holds, it means that the probabilistic forecast

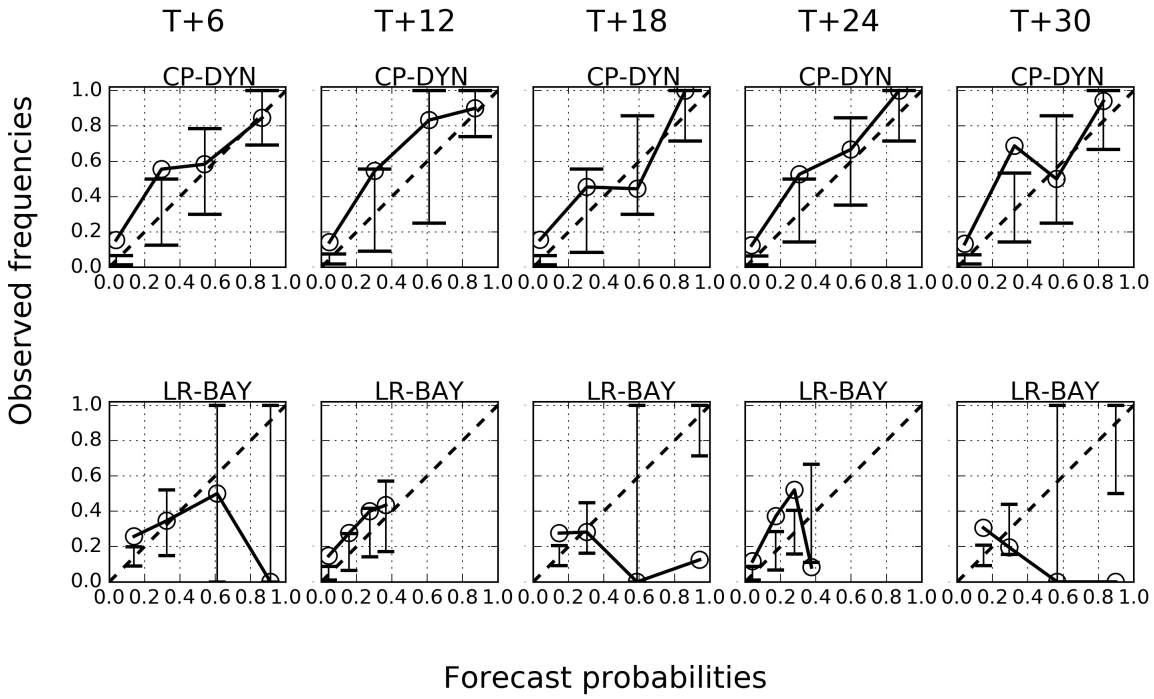


FIGURE 5.1: Reliability diagrams for **HR-DYN** and **LR-BAY** for all lead times. The dots represent the calculated value from our sample, whereas error bars indicates the 5%-95% confidence interval for the consistency resampling, as described in Bröcker and Smith (2007a).

distribution and observations are independent, meaning that the probabilistic forecast has no ability to discriminate between events and no events. The aim is to test the validity of this null hypothesis.

R statistic test A way to test the null hypothesis 5.1 is to create a *statistics* that quantifies the distance between the two probability distributions on the left and right hand side of equation 5.1. This quantity has been introduced by Bröcker (2015) and is defined as follows:

$$\mathcal{R} = \sum_{k,d} -\log\left\{\frac{\mathbb{P}(Y=k)\mathbb{P}(p=p_d)}{\mathbb{P}(Y=k, p=p_d)}\right\}\mathbb{P}(Y=k, p=p_d). \tag{5.2}$$

$\mathcal{R}$ is the *Kullback-Leibler divergence* (Kullback and Leibler, 1951) between the two sides of equation 5.1. It measures how different are the two distributions and the higher the value of $\mathcal{R}$, the less tenable is the equality 5.1. A statistical test is therefore performed to quantify $\mathcal{R}$ and to assess whether the value is significantly high to reject the null hypothesis of no resolution. The test can be summarized as follows:

- The probabilities in the equation 5.1 are replaced by the observed frequencies. These are computed by constructing the so-called *two way contingency table*. Two examples are given in tables 5.1,5.2. Each entry of the table C_{kd} represents the number of occurrences when $Y = 0, 1$ and $p = p_d$ at the same time. Let $C_{\bullet d}$ and $C_{k\bullet}$ be the two marginals of the table. Let N be the total number of occurrences.

- The terms in equation 5.2 can be estimated with the relative observed frequencies. For instance $\mathbb{P}(Y = k, p = p_d) = C_{kd}/N$, $\mathbb{P}(Y = k) = C_{k\bullet}/N$ and $\mathbb{P}(p = p_d) = C_{\bullet d}/N$. Thefore an estimator for $\mathcal{R}$ is given by:

$$\widehat{\mathcal{R}} := \sum_{kd} -\log\left(\frac{C_{\bullet d}C_{k\bullet}}{NC_{kd}}\right) NC_{kd} \tag{5.3}$$

- If forecast and observations are independent then $2N\widehat{\mathcal{R}}$ has an asymptotic χ-square distribution with $(K-1)(D-1)$ degrees of freedom (Mood et al., 1974).

- Even though this is an asymptotic result, the Cochran (1954)'s rule requires that $C_{kd} \geq 5$. It can be seen that this condition is satisfied by the table 5.2, but not by the table 5.1. Alternative solutions on how to deal with small samples in contingency tables are offered in (Bröcker, 2015) and (Kroonenberg and Verbeek, 2018). Therefore the statistical test is applied only to the Bayesian model.

The results of the χ^2-test are given in table 5.3. It can be therefore concluded that the null hypothesis of no resolution of the Bayesian model can be rejected at 1% level of significance for all lead times.

	$0-0.25$	$0.25-0.5$	$0.5-0.75$	$0.75-1$
Y=0	138	6	1	2
Y=1	23	6	5	18

TABLE 5.1: Contingency table for the convection-permitting forecast at 12 hours ahead.

	$0-0.113$	$0.113-0.226$	$0.226-0.339$	$0.339-0.45$
Y=0	74	35	23	13
Y=1	14	10	17	12

TABLE 5.2: Contingency table for the Bayesian forecast at 12 hours ahead.

Forecast lead time (hours)	6	12	18	24	30
$\widehat{\mathcal{R}}$	0.04	0.04	0.04	0.03	0.04
p value	0.0007	0.00289	0.002	0.006	0.001

TABLE 5.3: R statistic test results for **LR-BAY** for all lead times.

Area under the ROC curve Equation (5.1) can be rewritten in terms of H and F, the hit rate and false alarm rate respectively (Bröcker (2015)):

$$H(\omega) = F(\omega), \tag{5.4}$$

where $H(\omega) = \mathbb{P}(p \geq \omega | y = 1)$ and $F(\omega) = \mathbb{P}(p \geq \omega | y = 0)$ and ω is a variable probability threshold. It can be shown by a simple manipulation that equation (5.1) and (5.4) are equivalent.

Starting from equation (5.4) and rewriting it using Bayes' formula:

$$\begin{aligned}
\frac{\mathbb{P}(y=1|p\geq w)P(p\geq w)}{\mathbb{P}(y=1)} &= \frac{\mathbb{P}(y=0|p\geq w)P(p\geq w)}{\mathbb{P}(y=0)} \iff \\
\mathbb{P}(y=1|p\geq w)\mathbb{P}(y=0) &= \mathbb{P}(y=0|p\geq w)\mathbb{P}(y=1) \iff \\
\mathbb{P}(y=1|p\geq w)(1-\mathbb{P}(y=1)) &= \mathbb{P}(y=0|p\geq w)\mathbb{P}(y=1) \iff \\
\mathbb{P}(y=1|p\geq w) &= \mathbb{P}(y=1)(\mathbb{P}(y=0|p\geq w)+\mathbb{P}(y=1|p\geq w)) \iff \\
\mathbb{P}(y=1|p\geq w) &= \mathbb{P}(y=1),
\end{aligned} \tag{5.5}$$

where in the last equality the fact that $(\mathbb{P}(y=0|\Gamma\geq w)+\mathbb{P}(y=1|\Gamma\geq w))=1$ is used. Since this holds for all $w\in[0,1]$, the equivalence is thus proven.
Then $H(\omega)$ is plotted against $F(\omega)$, obtaining the so called Receiver Operating Characteristics (ROC) curve (parametrized by the varying threshold ω). The performance of the two probabilistic forecasts is then evaluated by calculating the area under the ROC curve (AUC, (Egan, 1975), (Mason and Graham, 2002)).
In the previous the χ^2-test suggests that there is evidence to reject the null hypothesis of no resolution for the Bayesian model. The test was not performed for the convection-permitting forecast due to violation of the *Cochran's rule*. Therefore here another way to test the validity of equation 5.4 (equivalent to equation 5.1) is proposed. An example of ROC curves, for both the forecasts for T+24 forecast lead time, is shown in figure 5.2a. The diagonal line represents the condition of no resolution, as described by the equation (5.4). A perfect forecast is the one for which the ROC curve passes through the top left corner (the point (0,1), i.e. when the hit rate is 1 and the false alarm rate is 0.), whereas a non skillful forecast (i.e. with no resolution) is the one for which the hit rates are equal to false alarm rates. In the first case, AUC$=1$, whereas for a non skillful forecast AUC$=0.5$. Figure 5.2b shows the AUC evolution against lead times for both the forecasts. It can be seen that the convective-scale forecast has a greater AUC than the Bayesian forecast for all lead times. The aim is then to test whether this difference in AUC is significant. For this purpose this statistic is used:

$$\frac{\mathrm{AUC}_h-\mathrm{AUC}_l}{\sqrt{\mathrm{Var}(\mathrm{AUC}_h-\mathrm{AUC}_l)}}, \tag{5.6}$$

where the subscripts h and l indicate **HR-DYN** and **LR-BAY** forecasts respectively. $\mathrm{Var}(\mathrm{AUC}_h-\mathrm{AUC}_l)=\mathrm{Var}(\mathrm{AUC}_h)+\mathrm{Var}(\mathrm{AUC}_l)-2\mathrm{Cov}(\mathrm{AUC}_h,\mathrm{AUC}_l)$. For the variance and covariance, expressions by (DeLong et al., 1988) have been used.
The results of this comparison are shown in figure 5.2c. It can be seen that $\mathrm{AUC}_h-\mathrm{AUC}_l>0$ is positive for all lead times. Furthermore the 95% confidence intervals, calculated using equation 5.6, indicate that the inequality $\mathrm{AUC}_h>\mathrm{AUC}_l$ is significant for all lead times, suggesting that **HR-DYN** is more skillful than **LR-BAY** in discriminating between events and non-events, since it has more resolution.
The χ^2 test reveals that **LR-BAY** has resolution and this test based on the AUC differences showed that **HR-DYN** has significantly more resolution than **LR-BAY**. Thus the null hypothesis of no resolution can be rejected also for **HR-DYN**.

A property related the resolution is discussed and tested in the appendix B.

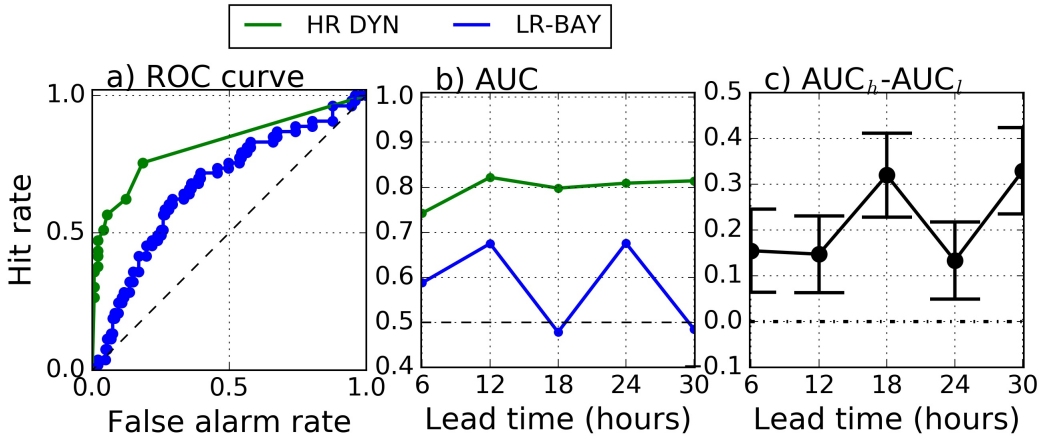


FIGURE 5.2: a) ROC curve for **HR-DYN** (green) and **LR-BAY** (blue) at 12 hours lead time. The dots indicate the different probability thresholds. b) shows area under the ROC curve values against forecast lead time and c) difference between the two areas. Positive (negative) values indicate the area for the **HR-DYN** (**LR-BAY**) forecast is greater. The error bars show the 95% confidence interval based on equation (5.6) .

5.4 Summary scores

In this subsection we compute two probabilistic scores, namely the Brier Score (BS: Brier (1950)) and Ignorance score (IS: Roulston and Smith (2002)). First we compute the classic decomposition of the scores into reliability and resolution terms, to examine whether the conclusions drawn in the previous subsections about reliability and resolution can be confirmed. Then we compute the differences in the scores for each forecast relative to the climatological forecast. The scores and relative differences (BSD and ISD) with respect to a *reference* (ref) system are defined as follows:

$$\mathrm{BS}_p = \frac{1}{N}\sum_{n=1}^{N}(p(n) - Y(n))^2 \tag{5.7}$$

$$\mathrm{IS}_p = \frac{1}{N}\sum_{n=1}^{N} D_{KL}(p(n), Y(n)) \tag{5.8}$$

$$\mathrm{BSD} = \mathrm{BS}_{ref} - \mathrm{BS}_p \tag{5.9}$$

$$\mathrm{ISD} = \mathrm{IS}_{ref} - \mathrm{IS}_p. \tag{5.10}$$

Here $p(n)$ is the probabilistic forecast to be evaluated, $Y(n)$ is the observation on day n, N is total length of the dataset. $\mathrm{D}_{KL}(p, Y) = Y \log\left(\frac{Y}{p}\right) + (1 - Y)\log\left(\frac{1-Y}{1-p}\right)$ is the Kullback-Leibler divergence (Kullback and Leibler, 1951). BS is a mean-squared error and measures the distance between forecasts and observations. IS measures the ignorance possessed (or the information content) by a forecast after the verification against the true outcome. The amount of ignorance decreases monotonically with increasing probability

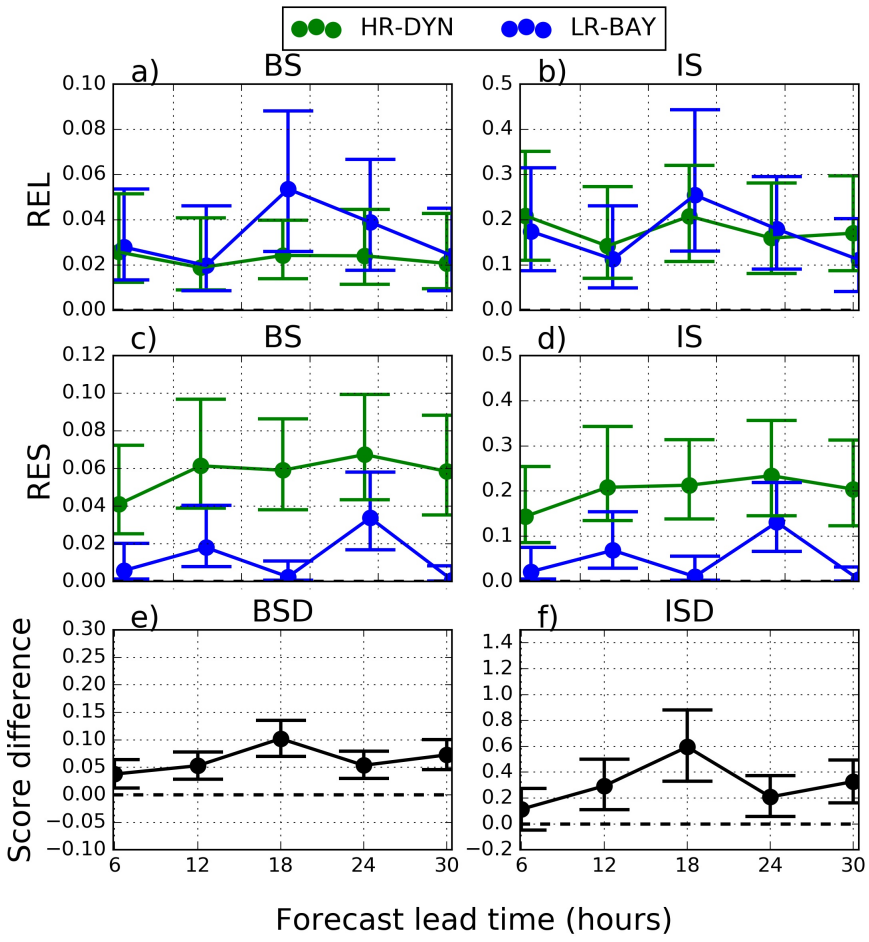


FIGURE 5.3: Reliability (a) and b)) and resolution (c) and d)) terms calculated for both the forecasts calculated using Brier Score (BS) and ignorance score (IS). e) and f) are the Brier and ignorance score differences respectively based on equations 5.9 and 5.10. Error bars represent the 90% bootstrap confidence interval.

value. [2] Both IS and BS are negatively oriented, which means the lower the score the more skillful the forecast. Therefore BSD>0 or ISD>0 implies that the convective-scale forecast is more skillful than the reference system, whereas if BSD <0 or ISD<0 the reference system is more skillful. Both BS and ISDN can be decomposed into reliability (REL), resolution (RES) and uncertainty (UNC) components (Bröcker, 2009; Weijs et al., 2010) as follows:

$$\mathrm{BS} = \mathrm{BS}_{REL} - \mathrm{BS}_{RES} + \mathrm{BS}_{UNC} \tag{5.11}$$

$$= \frac{1}{N}\sum_{k=1}^{K} n_k(p(k) - \overline{Y_k})^2 - \frac{1}{N}\sum_{k=1}^{K} n_k(\overline{Y_k} - \overline{Y})^2 + \overline{Y}(1-\overline{Y}) \tag{5.12}$$

$$\mathrm{IS} = \mathrm{IS}_{REL} - \mathrm{IS}_{RES} + \mathrm{IS}_{UNC} \tag{5.13}$$

$$= \frac{1}{N}\sum_{k=1}^{K} n_k D_{KL}(p(k), \overline{Y_k}) - \frac{1}{N}\sum_{k=1}^{K} n_k D_{KL}(\overline{Y_k}, \overline{Y}) - \left(\overline{Y}\log(\overline{Y}) + (1-\overline{Y})\log(1-\overline{Y})\right). \tag{5.14}$$

[2] For instance if $p(n) \to 1$ and $Y(n) = 1$ (the event actually occurs), then IS $\to 0$.

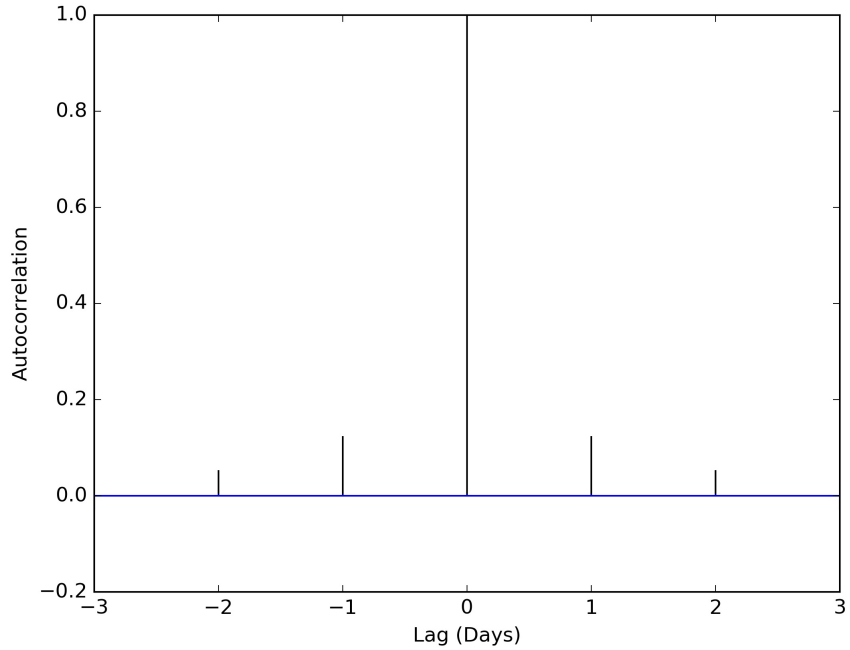


FIGURE 5.4: Lag – 1 and Lag – 2 autocorrelation function of the score differences timeseries at T+12 h.

Here, K is the number of categories (bins) which the probabilistic distribution is divided into and $p(k)$ is the value of each category k. n_k is the number of occurrences and $\overline{Y_k}$ is the observed frequency for each category, respectively. $\overline{Y} = \frac{1}{N}\sum_{n=1}^{N} Y(n)$ is simply the observed frequency over the entire data set. In the calculation of the decomposition and the score differences we have assumed no serial correlation. "The effect of serial correlation is to inflate the variances of the sampling distributions of the two statistics, so that uncorrected confidence intervals are too narrow" (Wilks, 2010). This assumption allows us to neglect correlations when calculating the standard error of the differences with the bootstrap method. This is a reasonable assumption, since we are verifying sea breeze occurrence every 24 hours, which is longer than the typical sea breeze time scale. A lag – 1 autocorrelation of the ignorance score differences (according to equation 5.10) at 12 hours ahead. Results are shown in figure 5.4. Results for other lead times are similar. The autocorrelation at lag 1 day is $\lesssim 0.2$. In Wilks (2010), Fig. 4 it is shown the sensitivity of BS variance to autocorrelation. It can be seen that for lag – 1 autocorrelation ≤ 0.2, the probability coverage (i.e. the width of the confidence intervals) does not change much. Therefore the assumption assumed for the rest of the chapter is reasonable.

The results of the decomposition are shown in Figures 5.5a–5.5d. Recalling that for an ideal forecast REL $= 0$ and RES > 0, it can be seen that both the forecasts have reliability significantly different from 0 but have significant resolution, for all lead times. It is also clear, that the resolution term for the convective-scale forecast is greater than the Bayesian one although not significantly so at all lead times. This implies that the Bayesian forecast is less able than the convective-scale forecast to discriminate between events and non-events but more able to do so than a climatological forecast.

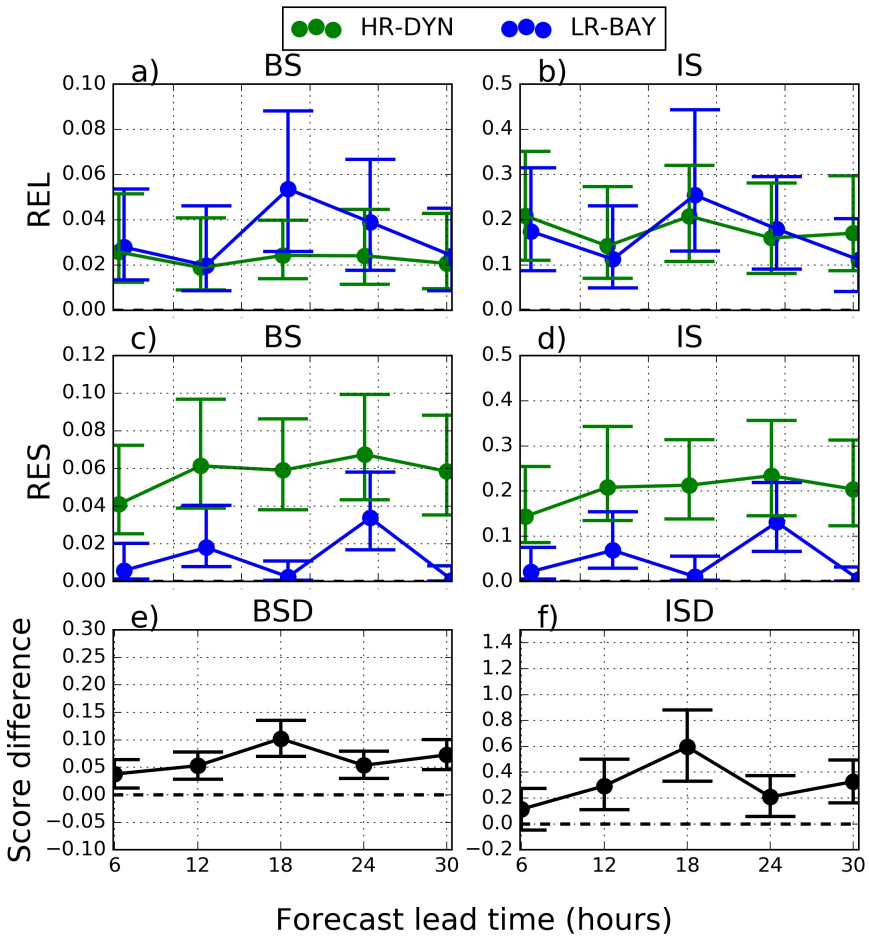


FIGURE 5.5: Reliability (a) and b)) and resolution (c) and d)) terms calculated for both the forecasts calculated using Brier Score (BS) and ignorance score (IS). e) and f) are the Brier and ignorance score differences respectively based on Equations 5.9 and 5.10. Error bars represent the 90% bootstrap confidence interval.

Figures 5.5e and 5.5f show the differences between the BS and IS scores obtained for the two forecasting methods, using Equations 5.9 and 5.10 with **LR-BAY** taken as reference and **HR-DYN** as p. The results are shown in Figure 5.5e, f. It can be seen that for both ISD and BSD the difference in the score is significant (except for ISD at 6 hours ahead), implying that the convective-scale forecast has more value than the Bayesian forecast for sea-breeze occurrence.

After comparing and quantifying the skill of the convection-permitting forecast with respect to the Bayesian forecast, it is interesting to assess how both the systems perform with respect to a third reference, the climatological forecast. By computing the resolution terms in the previous section, we have already shown that both the forecasts have significantly more resolution than the climatological forecast but less reliability. Therefore, this further comparison is useful to assess whether the probabilistic forecasts are performing better than the climatological forecasts, or the lack of reliability is masking their additional resolution. The climatological forecast, i.e. the percentage of times sea breeze occurs over all the data, has been calculated using either observations or convection-permitting model data, leaving one-year-out for the verification itself (for instance, if we

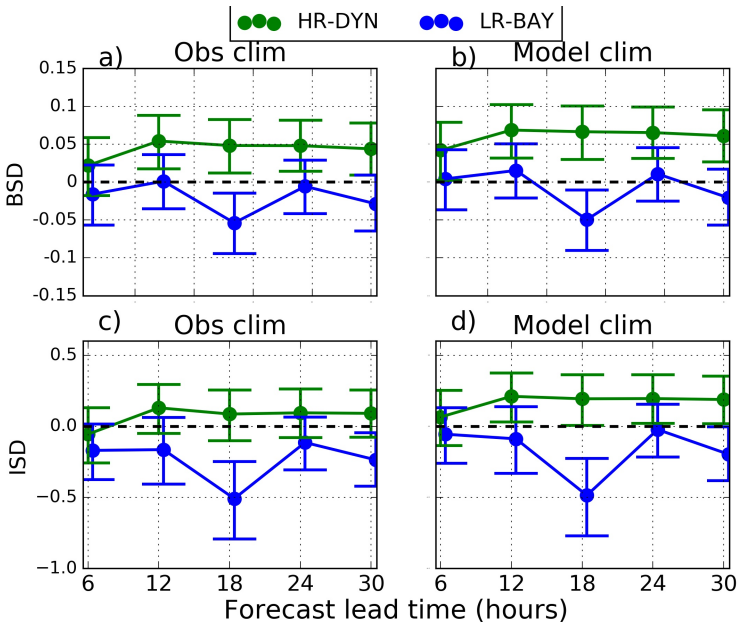


FIGURE 5.6: Score differences between each probabilistic forecast and a climatological forecast (estimated either from the observations or the model). a) and b) show these differences calculated using the Brier score, whereas c) and d) show the score differences calculated with the ignorance score. Positive (negative) differences indicate that the probabilistic forecast is more (less) skillful than the climatological forecast. Error bars represent the 90 % confidence interval calculated using bootstrap technique.

are validating on 2013 then 2013 is not included in the calculation of the climatology). The difference in the skill has been quantified using the Brier score and ignorance score, using Equations 5.9 and 5.10, with climatology as a reference, in place of LR-BAY. The results are shown in Figure 5.6. In terms of Brier score difference, the convective-scale forecast is significantly more skillful than the climatological forecast (based either on model or observations), for all lead times, except 6 hours ahead. On the other hand, the Bayesian forecast is shown to be generally less skillful than climatology, although significantly only at 18 hours ahead. With regard to the ISD, the conclusions are essentially same. The convection-permitting forecast is still more skillful than the climatology (although significantly only compared to model climatology for T+12, T+24 and T+30) and the Bayesian forecast significantly worse than the climatology not only at 18 hours ahead, but also at 30 hours ahead. These slight differences between ISD and BSD could be due to the fact that the two scores assign rewards and penalties to the forecast in a different way, because of the different function they are based on. Roulston and Smith (2002) show that IS penalizes sharp [3] forecasts much more than BS. Since even the Bayesian forecast is sharp compared to a climatological forecast, it is likely that this is the reason for the differences in the ISD and BSD measures. Although the Bayesian forecast has more resolution than a climatological forecast it is not sufficient to compensate for its lack of reliability.

[3]Sharpness is the tendency of a forecast to predict extreme values (0 or 1).

5.5 Verification of the alternative forecast of sea-breeze occurrence

Reliability and resolution have been tested for **WF**, presented in chapter 4. Reliability diagrams, for all lead times, are shown in figure 5.7. First of all it can be noticed, by comparing it with figure 5.1, that **WF** reaches higher probability values than **LR-BAY** for all lead times. This is expected, since, unlike in the Bayesian case, the method to produce forecasts is dichotomous, so a data point could be either in the *wedge* or not. Therefore if on a selected day n all the ensemble members are in the *wedge*, then **WF**(n)=1, but **LR-BAY**(n) < 1, because there are also non-occurences in the *wedge* which are taken into account in the Equation 4.16.

Also **WF** is not reliable, with the last bin containing the highest probability values being

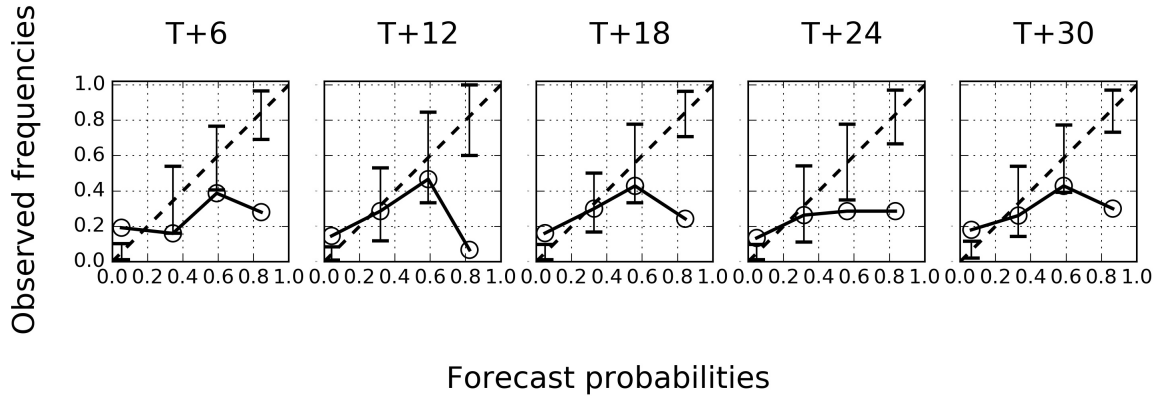


FIGURE 5.7: Same as in figure 5.1 for **WF** forecast.

problematic. **WF** is overconfident and this can be explained simply by looking at figure 4.9b, where it can be noticed that there are many non-occurrences in the observations inside the wedge.

In regard to the resolution, areas under the ROC curve are computed for all lead times and shown in figure 5.8. By comparing with figure 5.2 it can be seen that **WF** has an higher AUC score than **LR-BAY** for all lead times. However, **HR-DYN** is still performing significantly better, expect for 6 hours ahead.

5.6 Forecast verification against the control member

In the previous sections probabilistic forecasts have been performed against observations. The alternative is to consider the high-resolution member as observation. This is to verify whether the added value of the convection-permitting ensembles is due to the verifying system or to its intrinsic ability to resolve the sea-breeze circulation. This is because the Bayesian forecast has been trained on high-resolution data and not on observations. This might seem a disadvantage when verifying it against observations.

Also, when comparing against the high-resolution control member, the frequency bias is removed and also the so called representativity error.

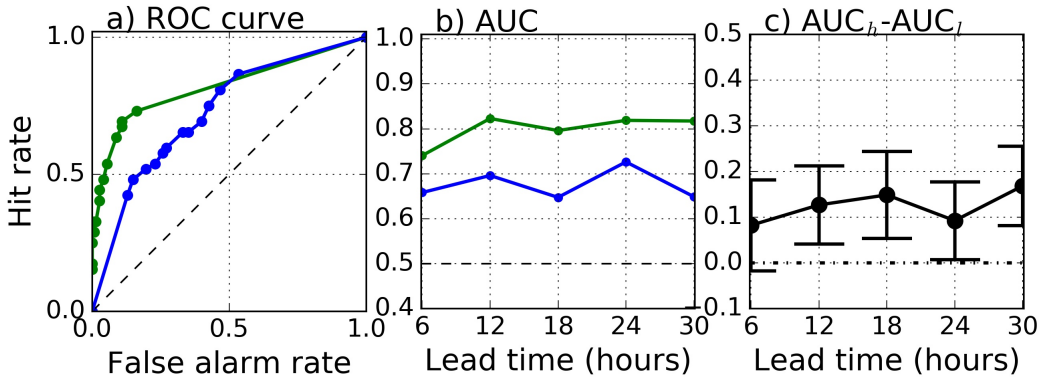


FIGURE 5.8: Same as in figure 5.2 with the alternative forecast **WF** in place of the Bayesian forecast **LR-BAY**.

Therefore, reliability diagram and ROC curves have been computed for both the forecasts evaluated against the high-resolution control member. The results are shown in Figures 5.9 and 5.10. It can be seen that in this case, unlike when comparing against observations, are both reliable. In regard to the ROC curves, it can be seen that now the

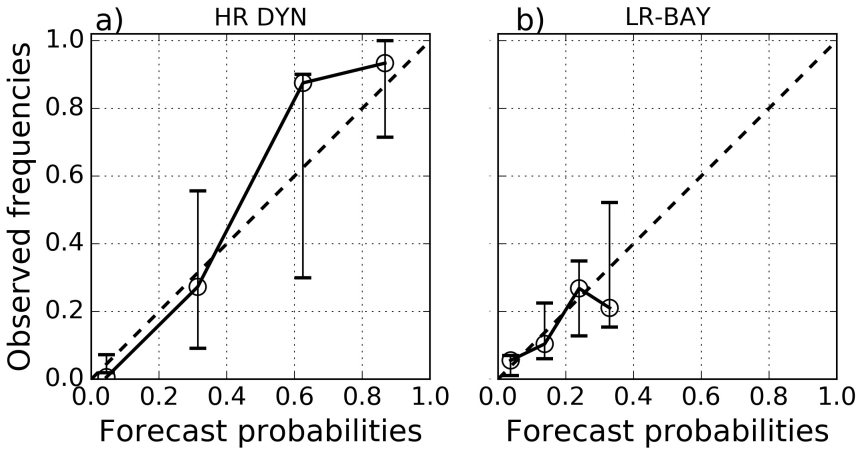


FIGURE 5.9: Reliability diagram at T+12 hours for the two probabilistic forecasts with the high-resolution control member taken as verification. Error bars are the same as in figure 5.1.

convection-permitting forecast has a higher AUC ($\gtrsim 0.2$ with respect to previous verification), whereas the Bayesian forecast has a slight higher AUC forecast compared to the other verification. This might be due to the fact that the frequency bias has not been completely removed since the high-resolution control member is taken as verification and not the low-resolution control member. This implies that in this case the AUC difference

between the convection-permitting and the Bayesian forecast is even higher.

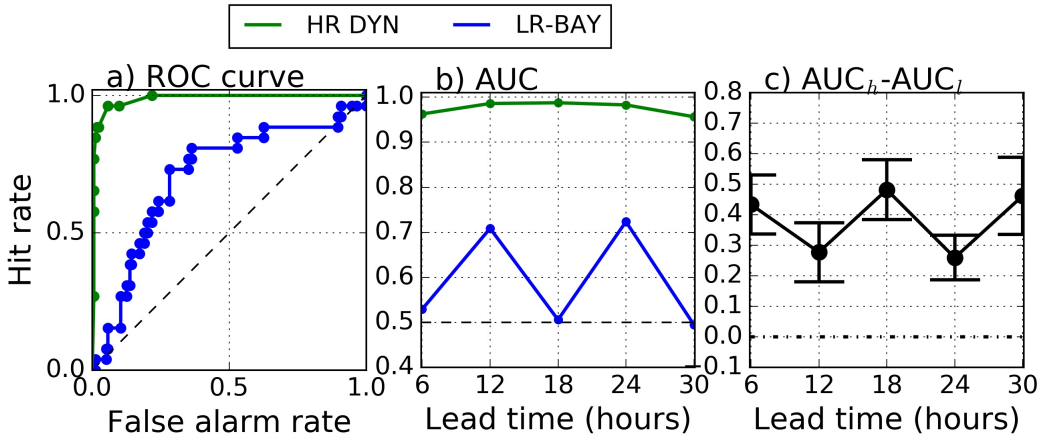


FIGURE 5.10: Same as in figure 5.2 with the high-resolution control member taken as verification.

5.7 Onset time and propagation speed forecast verification

Here the binary probabilistic forecasts calculated in the subsection 4.6.1 are verified against station observation. In particular, the area under the ROC curve is calculated for all the lead times. The results are shown in figures 5.11 and 5.12. It can be seen in general, for the propagation speed there are not significant differences between **HR-DYN** and **LR-BAY**. In general the probabilistic forecast are performing poorly in terms of resolution, especially for higher thresholds of the speed and for lead times longer than 12 hours ahead. In regard to the onset time, **HR-DYN** is performing significantly better than **LR-BAY** at 6 hours ahead for the early onset (< 12 UTC) and at 12 hours ahead for the late onset. In general, none of the forecast is able to predict the sea-breeze onset time from 18 hours ahead.

5.8 Sensitivity of Bayesian model verification to kernel bandwidth

In Chapter 4 the Bayesian model was introduced. One step in its construction is the kernel density estimation. One parameter to choose for the kernel density estimation is the *bandwidth factor*, filtering smoothing of the occurrence density. In this analysis the factor based on Scott (1979)'s rule has been selected. Here the area under the ROC curve of the Bayesian model has been calculated for different bandwidth factors in the interval $(0, 5]$. As shown in figure 5.13 the area under the ROC curve decreases with the bandwidth factor. This is because, as the kernel bandwidth increases, so does the overlap between the densities of the occurrences and non-occurrences. This implies that the probabilistic forecast becomes less sharp and so less able to discriminate between events and no-events, which is measured by the area under the ROC curve. It can be also

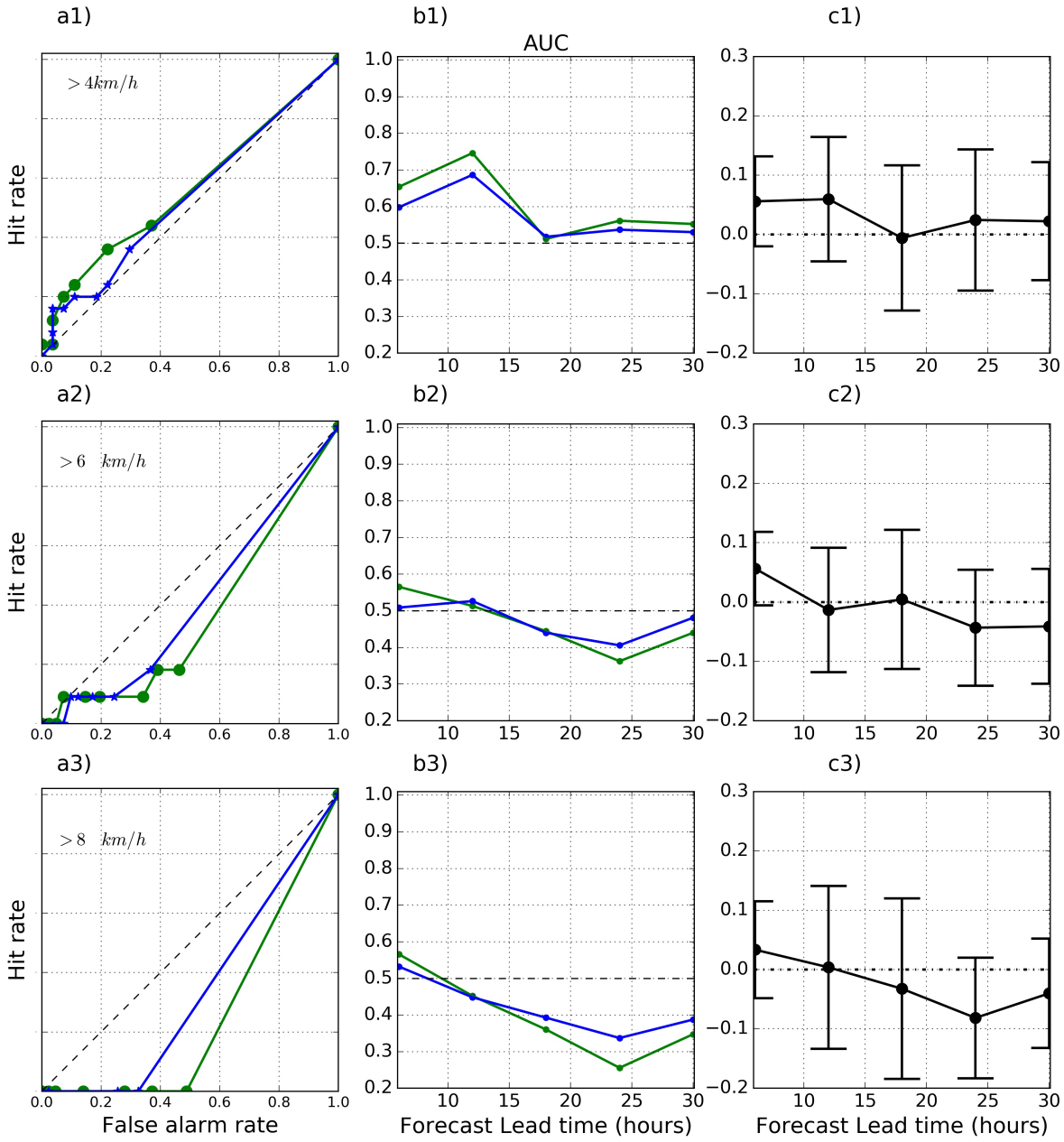


FIGURE 5.11: Same as in figure 5.2 for the probabilistic forecasts of sea-breeze propagation speed above different thresholds (4, 6 and 8 km/h from top row to bottom row). ROC curves in a1), a2) and a3) have been computed at 24 hours ahead.

noticed that the Scott's factors used in this analysis correspond to the maximum value of the area under the ROC curve.

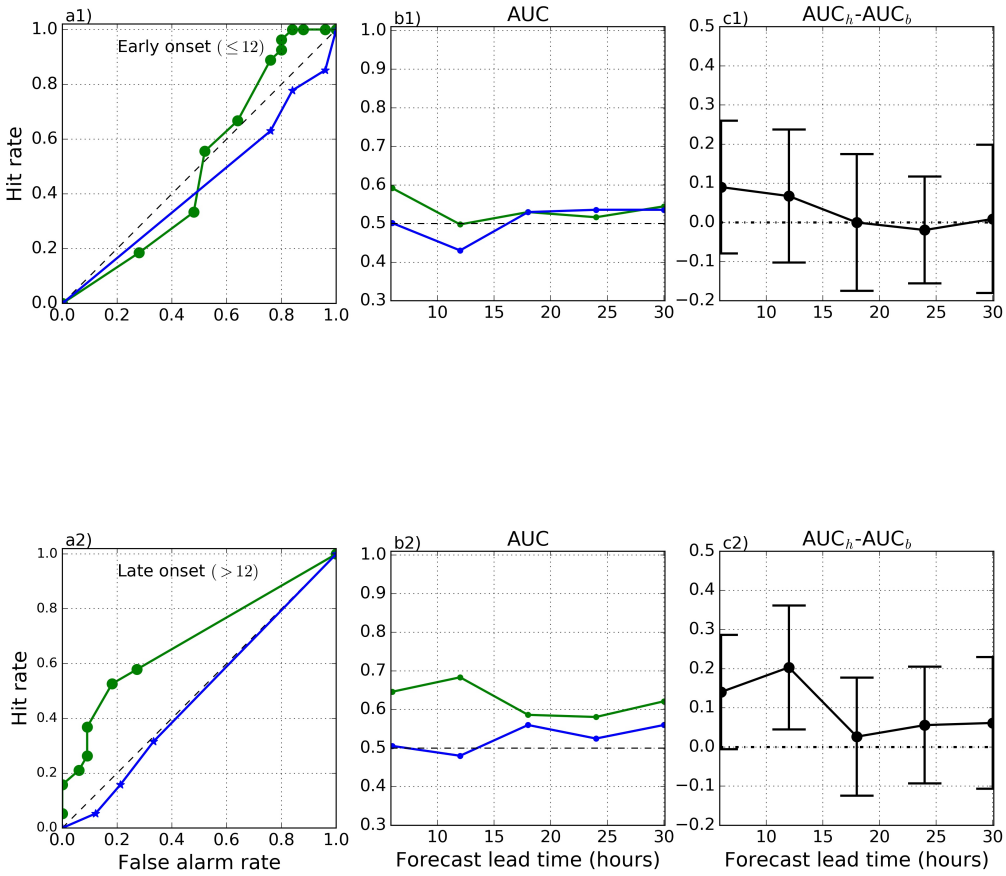


FIGURE 5.12: Same as in figure 5.2 for the probabilistic forecasts of onset times below (top row) and above (bottom row) a specified threshold (12 UTC).

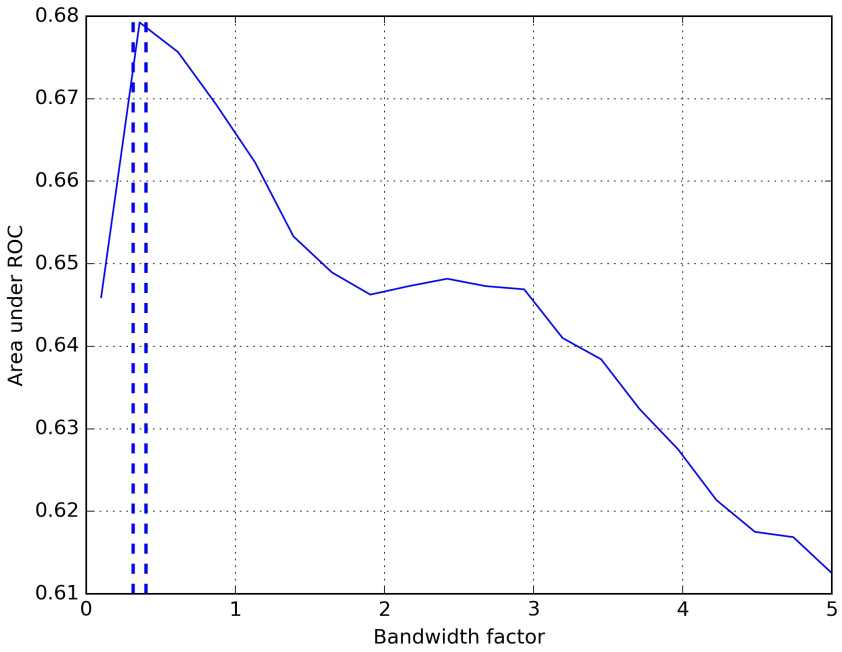


FIGURE 5.13: Area under the ROC curve of the Bayesian model against the bandwidth factor at 12 hours ahead. The dashed vertical lines represent the Scott's factors for the estimation of occurrences and non-occurrences respectively.

5.9 Conclusions

Presented here is the quantification of the added value provided by a convection-permitting ensemble forecast (**HR-DYN**) compared to a Bayesian forecast (**LR-BAY**) driven by the same global model ensemble in regard to the occurrence of sea breezes. The forecasts have been verified against observations over three different summer periods (2013-2015), for lead times from 6 hours up to 30 hours.
The alternative simple forecast for the sea-breeze occurrence and sea-breeze propagation speeds and onset times forecast were verified against station observations.
Reliability diagrams and ROC curves have been computed also taking the high-resolution control as verification. The conclusions from the probabilistic forecast verification against the observational dataset can be summarised as follows:

1. The two large-scale predictors used to train the Bayesian model are shown to contain more information (in terms of resolution) about the sea breeze occurrence predictability than the climatological forecast. The convection-permitting forecast **HR-DYN** has been shown however to provide more information than the Bayesian forecast **LR-BAY**. This could be due to the convection-permitting model's ability to explicitly represent the sea-breeze circulation and fine-scale topography which affects its dynamics. Also, the result is suggestive that the sea breeze occurrence depends on more than the two large-scale predictors used which may have limited the Bayesian forecasts skill.

2. The convection-permitting forecast is significantly more skillful than the Bayesian forecast, with respect to Brier Score and Information Gain score. The classic score decomposition in reliability and resolution highlights that the convection-permitting model is more reliable and has more resolution for all lead times.

3. When verifying against the high-resolution control member both the probabilistic forecasts have an higher AUC than when verifying against observations. However the AUC difference is higher in this case, implying that the convection-permitting forecast is more able to discriminate between events and no-events and that its skill relative to the Bayesian forecast does not depend on the fact the verification was performed against observations.

4. The convection-permitting forecast has significant skill also with respect to the climatological forecast, with higher skill relative to the model than observational climatology. On the other hand, the fact that the Bayesian model has better resolution than climatology is not sufficient to compensate for its lack of reliability.

5. Forecast skill does not decrease with lead time, showing that there is not a dependence on the initialisation time. A negligible dependence of the forecast skill on the lead time has been recently found also by **?** for short-range convective precipitation forecasts. In the case presented here, T+18h forecast is performing relatively poorly compared to other lead times. Ascertain the precise reasons for this would need more investigation. However, it is worth to note that forecasts of different lead

times are initialized at different initial times (see again figure 4.2). The T+18h forecast is initialized at 18UTC for the coarse resolution global model and 21UTC for the convection-permitting model. At 06UTC and 18UTC there are fewer observations, since radiosonde ascents data are assimilated only at 00UTC and 12UTC (https://www.metoffice.gov.uk/learning/making-a-forecast/first-steps/making-observations/upper-air).[4] Therefore, since sea breeze forecast are sensitive to stability of the atmosphere (among other factors, see Lombardo et al. (2016) and references therein), it might be that the more information about stability given by upper air observations could have an impact on the skill of sea breeze forecast.

6. The convection-permitting forecast is not significantly better than a statistical model based on large-scale predictors in the prediction of sea-breeze propagation speed for all lead times, whereas it is more skillful for the onset time, up to 12 hours ahead.

7. The alternative forecast (**WF**) is unreliable since it is underconfident in the lowest bin and overconfident in the highest bin. On the other hand it showed to have higher AUC values than the Bayesian forecast (**LR-BAY**).

[4] Accessed online on 24 November 2018.

Chapter 6

Probabilistic forecasts of convective wind gusts

In this chapter the information gain from a convection-permitting model is quantified for probabilistic forecasts of wind gust occurrence in a summer season. More precisely, the surface (10m) wind speed output from a convection-permitting model is compared to the gust parametrization from a coarser resolution model. The aim is to use these two variables to estimate the probability of high wind speeds occurrence, by counting the fraction of ensemble members exceeding a high percentile of the two distributions respectively.
Since the summer season is analysed, it is expected many of the high wind speeds event are related to convective outflows. Therefore two other probabilistic forecasts of high wind speeds are estimated, both using the Bayesian method: one conditioned on a gust parameter only and another one including also a convective parameter. The gust parameter alone accounts only for unresolved turbulent mixing. Since direct outflow associated with convection is not represented by the gust parameter parameter and neither is convection directly, the convective parameter is needed to assess whether it can provide useful information about the convective "wind gust" occurrence.
Therefore the two probabilistic forecasts estimated directly from ensemble members and the two Bayesian forecasts are computed, compared and verified against station observations and a control member, used as "truth".

6.1 Introduction

"Wind gust" is defined as the peak wind speed in a 10 min period based on a 3 s running mean measurement (World Meteorological Organization (2008)). Suomi and Vihma (2018) provide an overview of the state-of-art wind gust measurement techniques. In general wind gust measurements require a high temporal resolution device. This means that the whole measurement chain including the anemometer, data acquisition, processing, recording and reporting has to support this.
This brief and intense peak in wind gust can pose a threat for structures such as buildings, bridges and wind turbines. Therefore the knowledge of local gusts is important for those involved in designing structures, for instance. The Association of British Insurers estimates that averaged annual insured loss from wind-related damage to domestic property is in excess of £ 340 m (ABI (2005)). Moreover, damage from wind gusts are not proportionally related to the peak gust of a storm (Hewston and Dorling (2011)). Hawker

(2007) reports that an increase of 25% in peak gust speed can result in 650% increase in damage.

In (Sheridan, 2018, Table 1) a list of mesoscale or boundary-layer phenomena associated with high or gusty winds is presented. In this study the summer period is investigated and therefore it is expected that majority of gusty events are associated with convective turbulence.

Damaging winds from deep convective systems can be classified according to their origin: those associated with strong sub-scale vortices (e.g. tornadoes), the so-called straight-line winds associated with the descent of cold air from a storm in the form of *downburst* or *microburst* (for a review see Wakimoto (2001)) or just the outflow from a shower associated with evaporative cooling leading to negative buoyancy.

Wind gusts resulting from convective thunderstorms can lead to severe damage and may be a particular concern for outdoor activities. They are more frequent during Summer, very common in the United States but also in central Europe, causing significant damage and also fatalities. Examples include the *Pukkelpop* storm (`https://www.bbc.com/news/world-europe-14586001`)[1], with wind speed between 29 and 37 ms^{-1}, 5 deaths and 140 injuries due to a stage collapse (De Meutter et al. (2015)). Another is the Pentecost storm on 9 June 2014, causing 6 fatalities and total losses for EUR 650 million (Mathias et al. (2017), Barthlott et al. (2017)).

In the UK these severe events are rarer, although when they occur they can be very damaging. An example was the combination of a *microburst* and tornado in York on 3 August 2011 (Smart et al. (2012)). Other examples of a microburst in the UK are reported by Waters and Collier (1995) and investigated by Gray (2006).

It is therefore undoubtedly important to forecast the likelihood of these events in order to give early warning.

Sheridan (2018) offers a review of the current gust forecasting techniques. These may be classified into three main categories: *machine learning* techniques, *statistical methods* and *physically-based approaches*.

In regard to the first ones, Mercer and Dyer (2014) show the potential of the kernel principal component analysis (KPCA), which is an extension of the principal component analysis when a linear relationship between the input variables cannot be deduced, used to reduce the dimensionality of the predictors set.

Statistical methods have been applied by Thorarinsdottir and Johnson (2012), using non-homogeous Gaussian regression to postprocess ensemble forecasts of maximum wind speed and gust occurrence and more recently by Pantillon et al. (2018) show the limits of predictability of windstorms in convection-permitting ensemble forecasts using statistical post-processing.

Finally, physically-based approaches refer to methods based on physical reasoning concerning boundary layer turbulence, using Monin-Obukhov similarity theory following Panofsky et al. (1977). One study by Brasseur (2001) proposes a new method based on physical principles, namely the deflection of air parcels which are brought down by turbulent eddies.)

There have been also studies selecting sounding-derived parameters associated with the

[1] Accessed online on 18 November 2018.

most severe wind gusts (e.g. Kuchera and Parker (2006) and Taszarek et al. (2017)) with the aim of improving the operational forecasting of damaging convective gusts.
Two crucial recent developments present new opportunities and challenge for gust forecasting :

1. increase in resolution of local area NWP models, which leads convective-scale phenomena and fine scale topographic features to be being better resolved;

2. the shift from deterministic forecasts to ensembles and probabilistic forecasts

In regard to the shift to convection-permitting models with horizontal grid-length of order of 1-2 km, they resolve some convective overturning on the grid scale, whereas shear-driven turbulence is still not properly resolved (Mylne and Roberts, 2017). For example, Harvey et al. (2017) show an example of this issue of high-resolution models not properly resolving shear-driven turbulence at cold fronts and that NWP models do not converge to observations with increasing resolution. Therefore the use of a shear-driven gust parameter in convective situations leads to an over-estimation of the gusts, because it adds to the explicitly represented convective outflows. The challenge is thus to deal with this "grey-zone", where resolved (convection) and unresolved (turbulence) features are not clearly separated (Mylne and Roberts (2017)).
In regard to the use of ensembles, Sheridan (2018) points out that "given this stochastic nature, a probabilistic approach seems to offer some advantages", despite challenges in the verification of probabilistic forecasts against sparse observations.
In this chapter a high percentile of 10m wind speed output (e.g. 95%) from a convection-permitting ensemble forecast and the wind gust parametrization from a coarser resolution ensemble are compared. In section 6.2 the methodology and aims are presented, in section 6.3 the data used are presented. Then in section 6.4 the comparison between the forecast field is discussed with relative probabilistic forecasts computed in section 6.5. Spatial verification of probabilistic forecasts is performed in section 6.6. In section 6.7 an application of the Bayesian approach used for sea breezes is applied to the wind gust prediction. Finally, conclusions of this chapter are summarized in section 6.8.

6.2 Aims and methodology

Here, a similar methodology to the sea-breeze case, is applied to high wind speeds forecasts. It is useful to assess whether resolved winds at high resolution provide greater forecast information about the high wind speed occurrence than the gust parameter from a coarser resolution model. It is expected that the dynamics associated with convective wind gusts is reasonably resolved only by convective-scale models. Therefore in order to extract the information about the same event (wind speeds exceeding a high percentile) from much coarser resolution models another method is implemented, exploiting as much as possible the information contained in the coarser resolution parameters. Therefore, as for the sea breeze case, four probabilistic forecasts are compared:

1. based on the direct information coming from the convective-scale members;

2. based on the direct information coming from the coarse resolution gust parametrization;

3. a Bayesian forecast conditioned on a gust parameter (described later);

4. a Bayesian forecast conditioned on the gust parameter and convective parameter.

The questions that will lead the study of this chapter are:

1. What is the relationship between the coarser resolution parameter and convective-scale wind speed outputs ? Are the convective cases, as defined by the convective parameter, significantly different from the non-convective ones ?

2. How do the probabilistic forecast differ ? Does the Bayesian forecast conditioned on gust and convection parametrization have more information than the forecast conditioned on the gust parametrization only ? In other words, does the convective parameter add any value to the probabilistic forecasts of both models ?

3. Does the probabilistic skill vary with geographical location, e.g. due to varying topography ? Can this be used to attribute an appropriate amount of convective-scale forecast skill to better resolved topography ?

6.3 Wind gusts outputs

The data come from the same ensemble models as described in chapter 4. The wind gust is parametrized in MOGREPS-G as follows (Lock et al. (2018)):

$$g = U_{10m} + \sigma_u \frac{1}{k} \log \left(\frac{5e^{kc_{ugn}} + z_{0m(eff)}}{5 + z_{z0m(eff)}} \right), \tag{6.1}$$

where U_{10m} is the 10 metres wind, σ_u is the standard deviation of the horizontal wind, k the von Kármán constant, c_{ugn} is determined from universal turbulence spectra of a 25% exceeding probability of the three second wind gust, $z_{0m(eff)}$ is the effective roughness length to avoid unrealistic high gust values u_* of the friction velocity over mountainous region. The stability dependence of σ_u is estimated on the basis of the similarity relation from Panofsky et al. (1977):

$$\sigma_u = \begin{cases} A_{gust} u_* (1 - z_h/(24L))^{1/3}, & \text{for} \quad L < 0 \\ A_{gust} u_* & \text{for} \quad L > 0 \end{cases}$$

with $A_{gust} = 2.29$, z_h is the height of the lowest inversion, L is the Obukhov length. For $L = 0$ the wind gust diagnostic is undefined and therefore it is set $g = U_{10m}$. Note that the friction velocity, u_*, must use the implicitly calculated surface stress components the because the explicitly calculated u_* can be erratic, particularly over mountains. Then, for consistency with the implicit u_*, L must be calculated implicitly and, to avoid potential numerical problems in very light winds, the unstable ($L < 0$) case is rewritten as:

$$\sigma_u = A_{gust} (u_*^3 + k w_*^3/24)^{1/3},$$

where w_* is the vertical component of the friction velocity. From MOGREPS-UK the 10m wind speed is used instead as a proxy for the full wind. This is because, the inherent dynamics leading to the generation of convective wind gust is expected to be reasonably resolved by the convective-scale model and hence generates convective outflows with higher peak values. For smaller-scale turbulence, when there is no convection, some parametrization is still needed for convection-permitting models.

6.4 Comparison of gusts

Let $g(X, Y, d, m)$ be the parametrized gust from from MOGREPS-G (defined in equation 6.1) and $|\mathbf{v}|(x, y, d, m)$ the 10 metre wind speed from MOGREPS-UK, where X, Y, x, y are the spatial coordinates on the respective model grids, d is the day of the year and m is the ensemble member.
Firstly, for each MOGREPS-G grid box (33×33 km), the maximum of the 10m wind speeds from MOGREPS-UK within that larger grid box is computed.
g is output every 3 hours and is the maximum gust over this period, whereas $|\mathbf{v}|$ is output every hour.
Therefore, in order to compare the two quantities, the temporal resolution is matched. In order to do this, the maximum $|\mathbf{v}|$ over the same 3 hours window is taken.
The wind gusts are detected and verified in the time window 12-18 UTC for the summer 2015. Only this period was chosen since the main aim of this chapter is to provide a proof of the concept of the application of the methodology for the sea-breeze case. Also, this time window has been chosen since it is when the highest gusts tend to occur in the summer period (see (Hewston and Dorling, 2011, Fig. 7)), when atmospheric instability due to surface heating facilitates thermally driven mixing, which in turn leads to transfer of momentum downward resulting in surface gusts.
In this time window there are two 3-hour windows (12-15 UTC and 15-18 UTC). If the day is defined as *convective* (meaning that the convective rainfall accumulation from MOGREPS-G is non-zero), then the period with the highest rainfall accumulation is chosen. If the day is not convective, then the time window corresponding to the highest MOGREPS-G gust is selected.
Thus the following variable is defined:

$$\mathrm{MRW}(X, Y, T, m) := \max_{x,y,t} |\mathbf{v}|(x, y, t, m) \tag{6.2}$$

$g(X, Y, T, m)$ and $\mathrm{MRW}(X, Y, T, m)$ (*maximum resolved wind*) can now be compared for each MOGREPS-G grid box. In this study only forecasts initialized at 00UTC for MOGREPS-G (and 03 UTC for MOGREPS-UK) on the same day are analysed. This is because this study serves as a proof of the concept. The subdomain chosen for this analysis is represented by the black box shown in figure (6.1).
Figure 6.2 shows the comparison between g and MRW for three different grid boxes, in the upper, centre and lower part of the subdomain considered in this study and represented by the smaller grid boxes in figure 6.1. This domain was chosen since it has a low and uniform elevation above sea level. This is to exclude cases where skill might depend

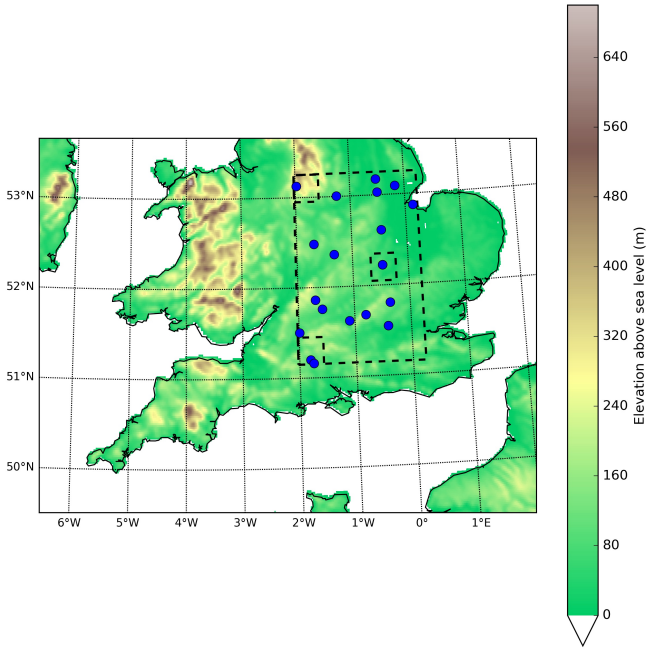


FIGURE 6.1: UK orography map from MOGREPS-UK. The black bigger dashed box indicates the domain considered in this study and the blue dots the station observation used for verification, whereas the three small dashed boxes indicate the three grid boxes referred after as upper, central and lower.

on orographic effects. However, only the upper left corner presents an higher elevation and later in the chapter it will be shown how this affects the skill of the convective-scale forecast relative to low resolution ones.

In order to quantify the differences between the convective and non-convective cases a two dimensional Kolmogorov-Smirnoff (KS) test has been performed on the pair $((g_{conv}, MRW_{conv})$,$(g, MRW))$ and a one dimensional KS test on the two frequency distributions of the differences with respect to the best fit line. Firstly, it can be noticed that there is a high positive correlation between the two variables. This is quantified by ρ, the Pearson correlation coefficient.

Also, it can be noticed that MRW $< g$ in most of the cases. However, for the convective cases, MRW is higher and closer to g. This is quantified by calculating the difference between the convective and non-convective data from the line of best fit trough all the points. Looking at the frequency distributions of the differences in figure 6.2, it can be seen that the convective cases are more shifted towards negative values of the differences between the best fit line and the observed values.

The two dimensional KS test gives a p-value of $6.07 \cdot 10^{-119}$, $1.54 \cdot 10^{-131}$, $6.20 \cdot 10^{-145}$ for the upper, central and lower grid box respectively. The test on the frequency distributions gives a p-value of $4.89 \cdot 10^{-26}$, $4.48 \cdot 10^{-15}$, $2.5 \cdot 10^{-20}$ for the the three grid boxes.

Therefore two preliminary conclusions can be drawn from this first comparison:

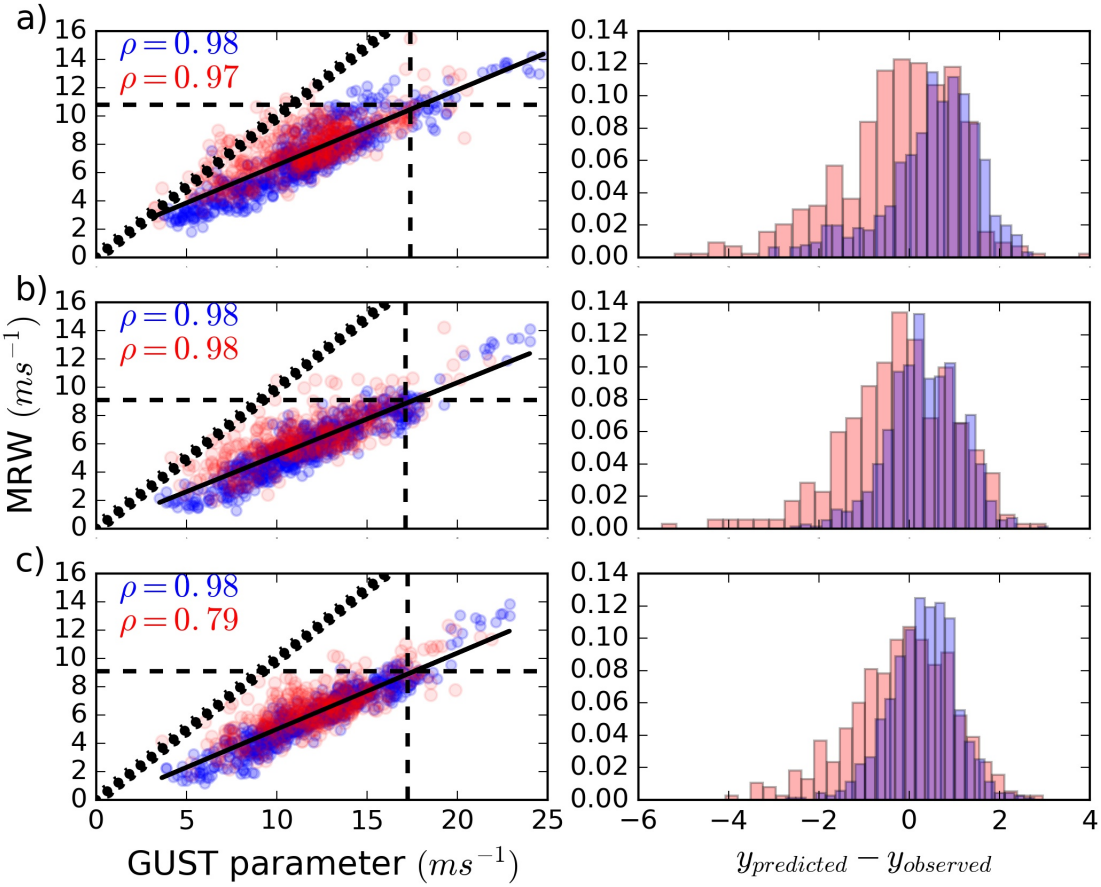


FIGURE 6.2: First column shows $g(X, Y, T, m)$ against $MRW(X, Y, T, m)$ for a) upper, b) central and c) lower grid box within the spatial domain considered. Red dots represent convective cases as defined by the convection parametrization, whereas blue dots are the no-convective cases. Horizontal and vertical dotted lines represent the 95^{th} percentile of the two distributions respectively. The solid line is the best linear fit through all the data, the dash-dotted line is the 1:1 line, ρ is the Pearson correlation coefficient. The second column represent the frequency distribution of the differences between the best fit line and the data for convective (red) and non-convective (blue) cases.

- There is high correlation between the MOGREPS-G parameter gust and the 10m wind speed from MOGREPS-UK.
- The convective cases are significantly different from the non-convective cases.

In the next sections the implications of these preliminary conclusions are investigated in terms of the probabilistic forecasts. For instance the fact that there is an high correlation between the two forecast fields, does it imply that the two probabilistic forecasts are performing in a similar way ? In other words, is the gust parameter a good predictor of high wind speed occurrence ? Also, does the convective parameter add additional information with respect to the gust parameter alone ? These will be addressed in the next sections.

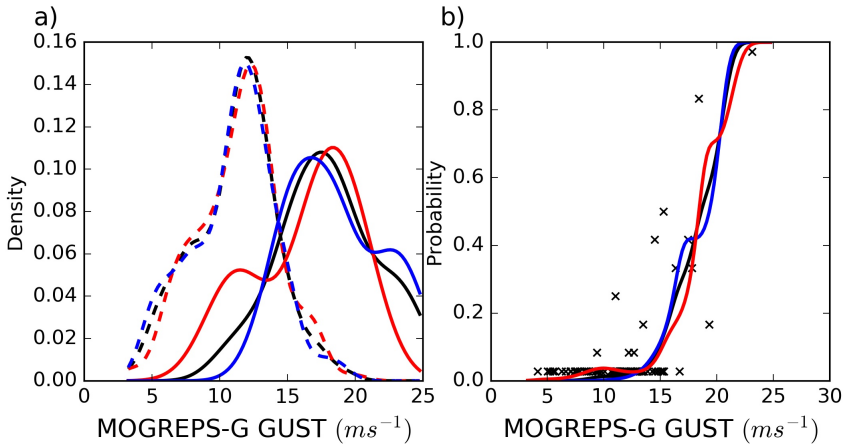


FIGURE 6.3: a) kernel density estimation of the distribution of wind gust events(solid lines) and no events (dashed lines). Events are defined as the occurrence of exceeding the 95^{th} percentile of the distribution. The distributions are plotted against the gust parameter for all (black), convective (red) and non convective cases (blue). b) shows the respective probabilities, obtained by applying the Bayes' formula which involves the ratio between the solid and dotted lines of the panel a). Black crosses represent the MOGREPS-UK probability values against the ensemble mean of the gust parameters, for each day of our period of evaluation.

6.5 Probabilistic forecasts

In this section the probabilistic forecasts of wind gusts are defined and calculated. These forecasts will be verified in the next section. The aim is to assess whether using only MOGREPS-G parameters (gust and convective) is sufficient to convey information about the wind gust occurrence or whether the 10m wind speed output from MOGREPS-UK can bring additional useful information.

The event to be predicted is the occurrence of high wind speed. This is defined here by the gust exceeding the 95^{th} percentile of the relevant distribution as will be defined shortly. Once that the event is defined, the probabilistic forecasts of this event can be estimated based on the ensemble prediction systems used.

In particular the following probabilities are calculated:

$$\mathrm{P}_{hrdyn}(d) := \mathbf{P}\left(\mathrm{MRW}(X,Y,d) > 95^{th}\mathrm{MRW}(X,Y)\right) \tag{6.3}$$

$$\mathrm{P}_{gust}(d) := \mathbf{P}\left(\mathrm{MRW}(X,Y,d) > 95^{th}\mathrm{MRW}(X,Y) | g(X,Y,d) > 95^{th} g(X,Y)\right) \tag{6.4}$$

$$\mathrm{P}_{bayfgust}(d) := \mathbf{P}\left(\mathrm{MRW}(X,Y,d) > 95^{th}\mathrm{MRW}(X,Y) | g(X,Y,d,m)\right) \tag{6.5}$$

$$\mathrm{P}_{bayfgustconv}(d) := \mathbf{P}\left(\mathrm{MRW}(X,Y,d) > 95^{th}\mathrm{MRW}(X,Y) | g(X,Y,d,m), \mathrm{conv}(X,Y,d,m)\right) \tag{6.6}$$

where conv(X, Y, d, m) is the convective parameter output from MOGREPS-G for each day d and for each member m. More precisely the convective parameter is a binary variable which is set equal to 1 if the convective rainfall accumulation output from MOGREPS-G is non-zero, otherwise equal to 0.
The probabilistic forecasts calculated in equations (6.3-6.4) are directly estimated from the ensemble members of MOGREPS-UK and MOGREPS-G respectively, by counting the fraction of the ensemble members exceeding the 95 % percentile. On the other hand, the probabilistic forecasts in equations (6.5-6.6) are the conditional probabilities of the same event, based on the parametrized gust only and with the parametrized convection respectively. These probabilities are calculated using the Bayes' formula, as done in the sea breeze case as discussed in chapter 4. Also in this case MOGREPS-UK data are used to train the Bayesian model. In the calculation of these conditional probabilities, the first step is the estimation of the likelihood function, i.e. the distribution of the predictors given that the event of exceeding the 95^{th} percentile has occurred or not.
This function is shown in figure 6.3. In the left panel the three cases of all, convective and non-convective cases are plotted for upper left grid box. First of all, it can be seen that there is a discrimination between events (solid lines) and no-events (dashed lines), with the overlap of the two densities more noticeable for the convective cases (red lines). In the right panel, the corresponding conditional probabilities are shown, against the parametrized gust. It can be noticed the sharp increase of probability values at about $15ms^{-1}$. Also, for lower values of the gust, the conditional probability based on convective parameter is slightly higher than the others, indicating that gust parameter alone does not capture information about gusts occurring during convection, as expected.
Before computing probabilistic scores associated with these four probabilities, a qualitative comparison is offered in the remainder of this section.
First of all, figure 6.4 shows a single case example of the spatial map of the four probabilistic forecast calculated. The first thing to notice is that P_{hrdyn} has the highest probability value. P_{gust} has lower probabilistic values with a different structure. $\mathrm{P}_{bayfgustconv}$ and $\mathrm{P}_{bayfgust}$ have a very similar structure and comparable probability values.
Figure 6.5 shows the frequency distribution of the four probabilities in the whole spatial domain and for the whole period considered in this study. It can be noticed that P_{hrdyn} and P_{gust} have more occurrences near $p \approx 0$ or $p \approx 1$, whereas $\mathrm{P}_{bayfgustconv}$ and $\mathrm{P}_{bayfgust}$ have more occurrences also for values of $0.125 < p \leq 0.375$ and fewer values for $p \approx 1$. In other words, P_{hrdyn} and P_{gust} are sharper than $\mathrm{P}_{bayfgustconv}$ and $\mathrm{P}_{bayfgust}$. As last comparison, figure 6.6 shows the map of the Spearman's correlation coeffient for different pairs of the probabilistic and Bayesian forecasts. It can be seen that the pair ($\mathrm{P}_{bayfgust}$,$\mathrm{P}_{bayfgustconv}$) show the highest correlation, in the bottom part of the spatial domain, whereas P_{hrdyn} and P_{gust} have an high correlation in the upper part of the spatial domain.

6.6 Spatial verification

In this section the relative skill of the different forecasts defined by equations (6.3-6.6) is assessed. Firstly, using the MOGREPS-UK control forecast in place of observations, in order to estimate the maximum possible information gain and avoid complications from

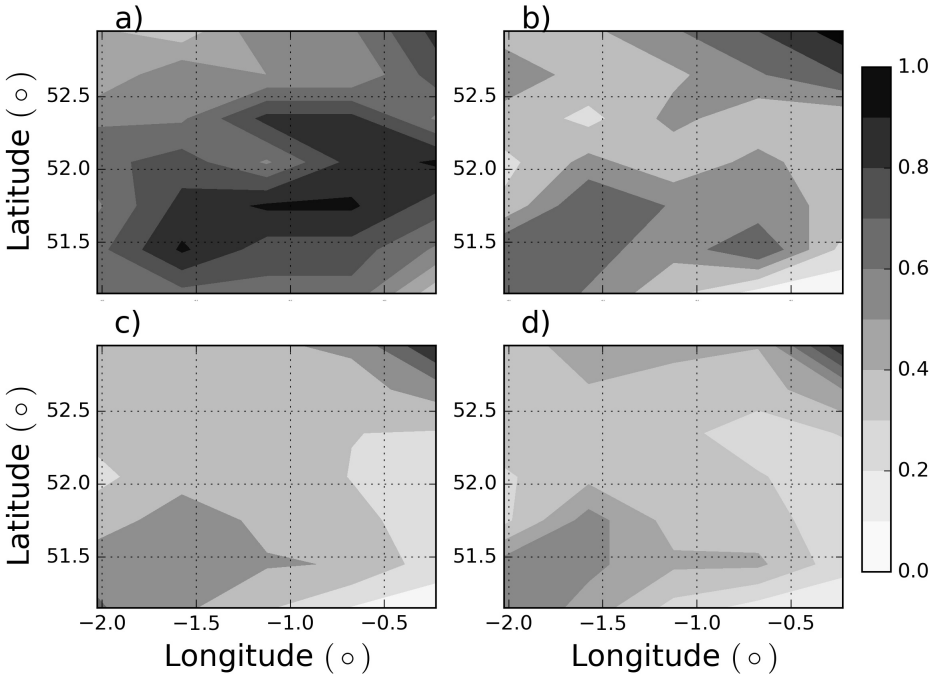


FIGURE 6.4: Probability values in the spatial domain considered here for a) P_{hrdyn}, b) P_{gust}, c) $P_{bayfgustconv}$ and d) $P_{bayfgust}$ at 12 hours ahead valid on 01/06/2015.

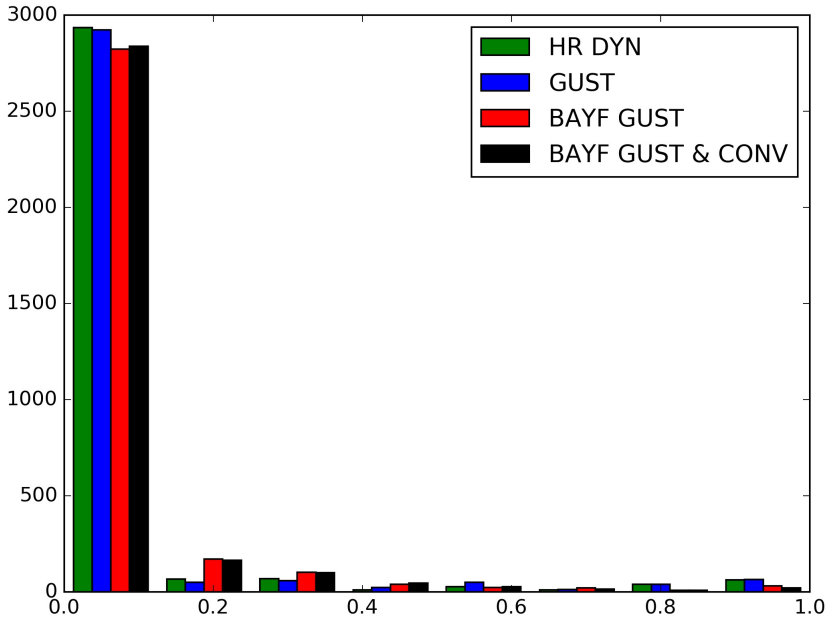


FIGURE 6.5: Frequencies of the maxima probability values within 33 km for the different probabilistic forecasts at 12 hours ahead.

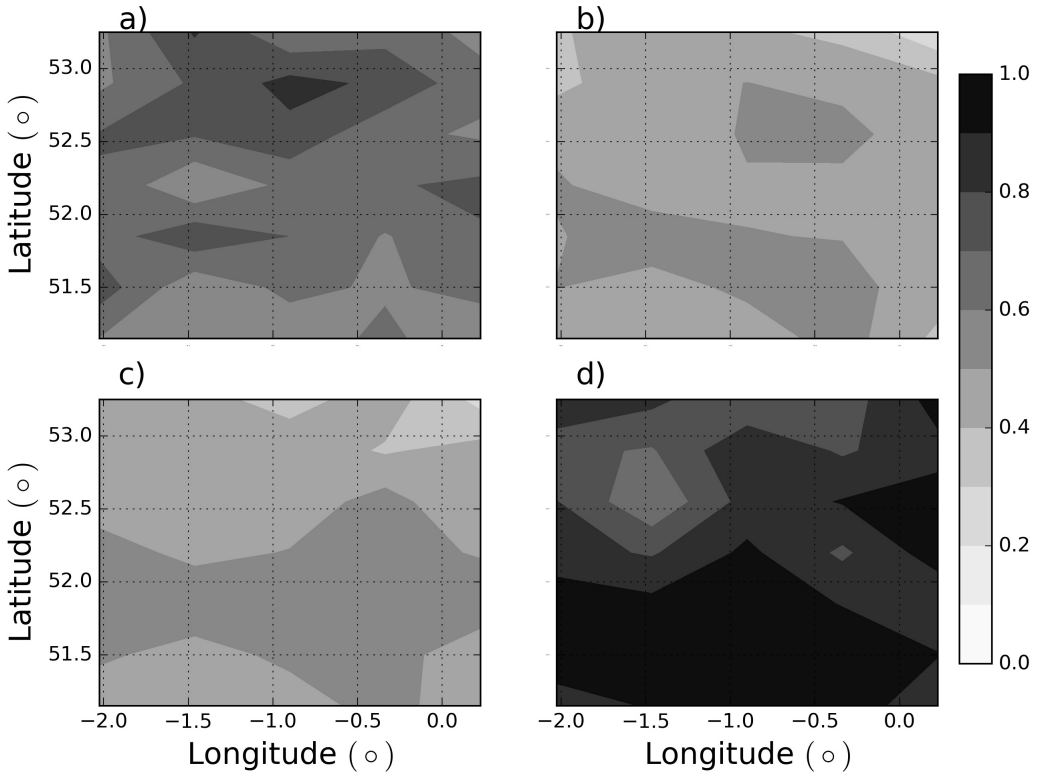


FIGURE 6.6: Spatial map of the Spearman's correlation coefficient for different pairs of the probabilistic forecasts. a) (P_{hrdyn},P_{gust}), b) (P_{hrdyn},$P_{bayfgust}$), c) (P_{hrdyn},$P_{bayfgustconv}$), d) ($P_{bayfgust}$,$P_{bayfgustconv}$).

limited observations and representativity errors. The event is defined to be observed on day d if MRW$(X, Y, d) > 95^{th}$MRW(X, Y), where MRW(X,Y) is the time distribution of the control member at the grid point X, Y. Secondly, the surface wind gust data from synoptic stations will be used.
The verification is performed for each MOGREPS-G grid box, i.e. the scores are not aggregated. This allows to avoid considering spatial correlations into the calculation of the probabilistic scores.
Given the distribution of the stations (see figure 6.1), it can be seen that stations are not uniformly distributed. There are some grid boxes not occupied by any station. Also, since convective gusts are very localized phenomena (on a spatial scale smaller than the distance between the stations), it is likely that some stations will miss the detection of the highest convective gusts.
This motivates the implementation of a neighborhood approach for the generation and verification of probabilistic forecasts.
The results of this section are cross-validated. More precisely, to verify the Bayesian forecasts, the leave-one-out cross-validation is performed with training the model on all the days of the summer period excluding the day which it is validated on, in each cross-validation run.

6.6.1 Wind gusts in observations

As mentioned in chapter 3, the maximum gusts are recorded every hour and it represents the 10 minutes max gust speed in the period between 20 and 10 minutes before each hour. The event to be predicted is the strong wind gust. This is defined as the occurrence to exceed the 95^{th} percentile of the probability distribution. Similarly, in the station observations, the event is defined as the occurrence to exceed the 95^{th} percentile of the wind speed distribution for each individual station observation. Then it is checked that in the time window period considered here (12-18 UTC), the wind speed recorded at any hour during this period exceeds the 95^{th} percentile of the distribution.

6.6.2 Neighborhood

Schwartz and Sobash (2017) provide a review of the state of the art of the neighborhood techniques (see also references therein). Here the neighborhood approach is implemented, otherwise the observations are badly undersampling in space, due to their sparsity.
In particular Ben Bouallègue and Theis (2014) examines two different methodologies to implement a neighborhood approach. They differ both in the implementation and in the interpretation: one, which is called "fuzzy probabilistic forecast" (or *smoothing*) and the other called *upscaling*. The *smoothing* method calculates the grid-scale probabilities, whereas the *upscaling* is used to calculate non-grid-scale probabilities. There is a common step for both the methods: the probabilities are first calculated for each grid point, as described in the previous section. Let $\mathbf{P}_i$ be the probability at grid point i. Then three different neighborhoods are defined, as described in figure 6.7. The square geometry is then used. Each neighborhood can be identified either by the number N_b of grid boxes or

by the full width of grid. For instance, the neighborhood comprising 3×3 grid boxes can be also indicated by the distance (33 km) from the grid i at the centre of the neighborhood.

- **smoothing**: The smoothing techniques consists in averaging the probability values over the N_b points of the neighborhood. Let S_i be the set of grid points within the neighborhood i. Then:

$$SP_i = \frac{1}{N_b} \sum_{k=1}^{N_b} P_k, k \in S_i \tag{6.7}$$

  SP_i is interpreted as the smoothed probability of the event at the grid point i given a smoothing length scale which defines the neighborhood size.

- **upscaling**:

$$\mathrm{MP}_i = \max_{S_i} P_k, k \in S_i \tag{6.8}$$

  MP_i is interpreted as the probability of the occurrence of the event in any point of the neighborhood, thus on a scale larger than the native grid.

It is important to notice that the same neighborhood approach is applied to station observations. For the smoothing method, the number of stations where the event has been observed is divided by the total number of stations within the neighborhood i. In regard to the upscaling method, if the event is observed at least in one station of the neighborhood the event is said to have occurred within the neighborhood. As a result, as also Schwartz and Sobash (2017) points out, when verifying smoothing probabilities the observations can be fractional, whereas they are necessarily binary for the verification of the upscaling method.
Here, three different neighborhoods are defined, as shown in figure 6.7. The the verification is performed at each MOGREPS-G grid box independently.
For neighbourhoods which extend outside the analysis domain, i.e. that ones whose centre is located near the edge of the domain, the neighbourhood is defined to comprise only the grid boxes inside the domain.

6.6.3 Results

After defining the two neighborhood methods, the verification attributes, computed in the previous chapter for the sea breeze prediction, are here computed for each MOGREPS-G grid box. In particular ROC curves are computed for MP_i (since ROC curves are defined for binary observations) and probabilistic scores calculated for both the methods. More precisely, Brier score is well defined also for the fractional observations (see Equation 5.7), whereas the Ignorance score only for binary observations. Then, also the potential forecast skill is evaluated, considering the MOGREPS-UK control member as truth (Buizza (1997)) [2]
. This has been done in order to avoid representativity error issues and to remove the frequency bias.

[2] In this case the control member has not been considered for the calculation of probabilistic forecasts.

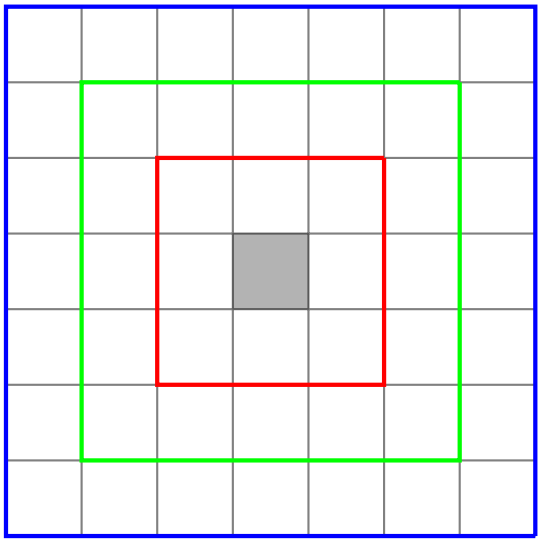


FIGURE 6.7: Schematic diagram of the MOGREPS-G native grid box (shaded in grey), with the three neighbours comprising 3×3 (red), 5×5 (green) and 7×7 (blue) grid boxes.

Bootstrap resampling is applied to test any significance in the score differences, analogously with the sea-breeze forecast verification. In this case the entire neighbourhood has been resampled (not the individual grid boxes) in order to retain spatial dependence.

Figure 6.8 shows the ROC curves for three different grid boxes in the domain for the different probabilistic forecasts evaluated both against the MOGREPS-UK control member and stations observations.

Not surprisingly, it can be immediately noticed that the probabilistic forecasts verified against the MOGREPS-UK control member perform better than the ones verified against the station observations. Also, when verified against observation $P_{bayfgust}$ has the highest AUC value and this is pretty consistent across the different grid boxes and neighborhoods.

Also, the grid box in the upper part of the domain is shown to perform better than the one in the centre, especially for $P_{bayfgustconv}$ and $P_{bayfgust}$. In order to better quantify this, areas under the ROC curve are calculated for each grid point, against the control member and station observations.

Figure 6.9 shows the area under the ROC curve (AUC) for each grid point for the probabilistic forecasts verified against the control member. P_{hrdyn} has the highest values of AUC, between 0.9 and 1, whereas P_{gust} has the lowest scores. $P_{bayfgustconv}$ and $P_{bayfgust}$ show slightly higher values than P_{gust}, especially in upper part of the domain. Also $P_{bayfgustconv}$ is shown to have higher values than $P_{bayfgust}$ in the same part of the domain.

Figure 6.10 shows the AUC values with the forecasts verified against stations observations. The situation is in this case reversed. $P_{bayfgustconv}$ and $P_{bayfgust}$ have higher AUC values than P_{hrdyn} and P_{gust}. This can be explained by the low density of station observations which could miss some gusts. That is the reason why the more conservative Bayesian forecasts (see again figure 6.5) perform better.

In order to establish whether these differences in the AUC scores are significant, figure 6.11 reveals that P_{hrdyn} is performing better than P_{gust}, almost everywhere in the domain, significantly in the upper part of the domain. P_{hrdyn} is worse than $P_{bayfgustconv}$ and more

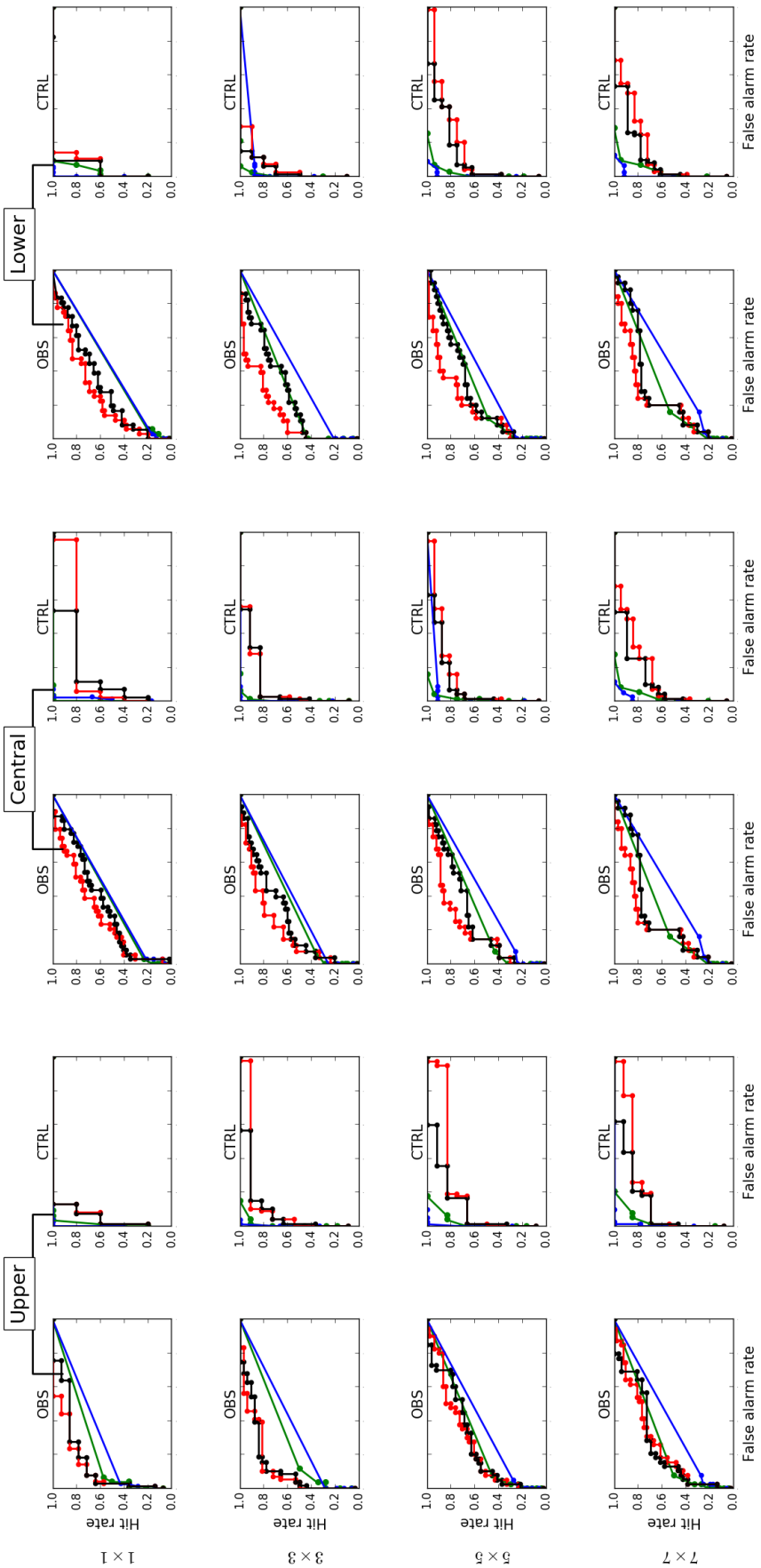


FIGURE 6.8: ROC curves for P_{hrdyn} (green), $P_{gust}(blue)$, c) $P_{bayfgustconv}$(black) and $P_{bayfgust}$(red) evaluated against station observations (first, third and fifth column) and MOGREPS-UK control member (second, fourth and sixth column) for a upper (first and second column), central (third and fourth column) and lower (fifth and sixth column) grid box in the spatial domain. The rows indicate the different neighborhoods.

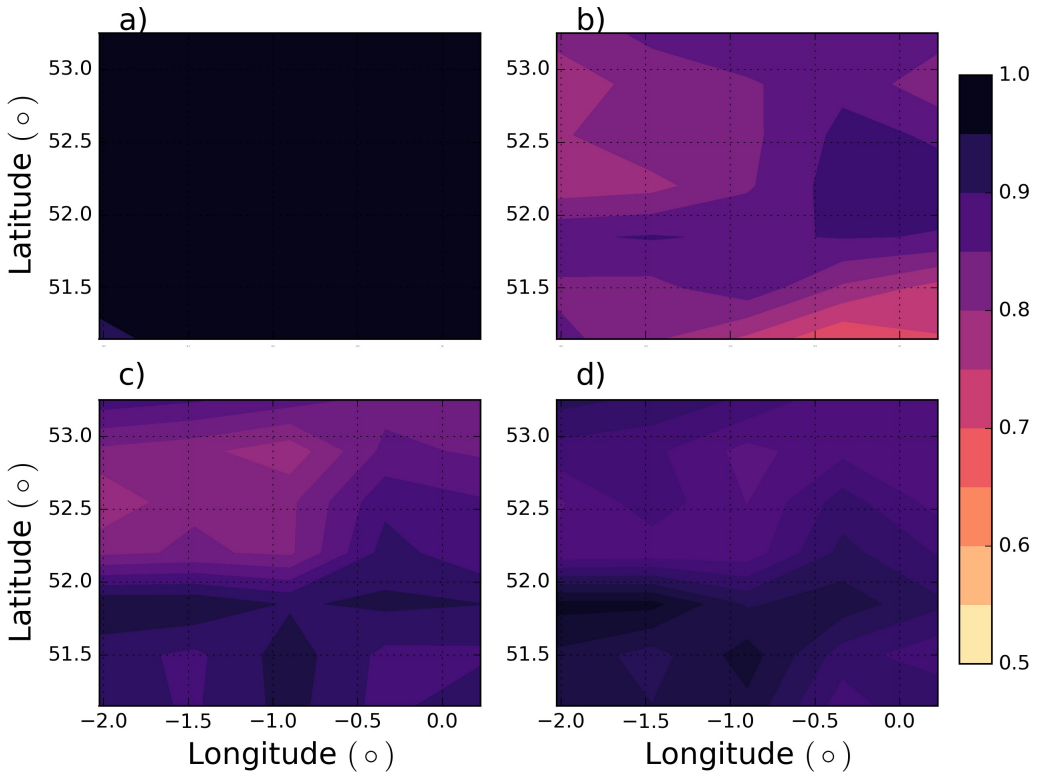


FIGURE 6.9: Spatial map of the AUC values for a)P_{hrdyn}, b) P_{gust}, c) $P_{bayfgust}$ and d) $P_{bayfgustconv}$ at 12 hours ahead, with verification performed against the MOGREPS-UK control member using the 3×3 neighbourhood.

significantly than $P_{bayfgust}$. This means that the convective parameter is not adding additional information (in terms of resolution) with respect to the gust parameter only.
When quantifying the differences in AUC against the MOGREPS-UK control member, P_{hrdyn} is shown to have significant higher AUC values than the other forecasts, but not everywhere in the spatial domain. The differences between the two Bayesian forecast are not significant anywhere in the domain (figure 6.12d). Then Brier score (BS) and ignorance score (IG) are computed, at each grid point, verified against the control member and station observations.

Score differences between different pairs of the probabilistic forecasts defined are shown in figure 6.13 and 6.14. They have been calculated using equations (5.9-5.10). P_{hrdyn} performs better than any other probabilistic forecast, significantly from the neighborhood containing 3×3 grid boxes or more (for the central and lower grid box). In the neighborhood containing only 1 grid box, the differences between P_{hrdyn} and the other forecasts are not significant. The last column shows the differences between $P_{bayfgust}$ and $P_{bayfgustconv}$. Generally $P_{bayfgust}$ perform better than $P_{bayfgustconv}$. In terms of BSD, differences on the smallest neighborhood are significant only in the central grid box (figure 6.13), whereas they become significant in the other parts of the domain with the neighborhood containing at least 3×3 grid boxes. In terms of IG, same conclusions can be drawn.
In order to have a clearer idea of spatial structure of probabilistic scores, figure 6.15 shows

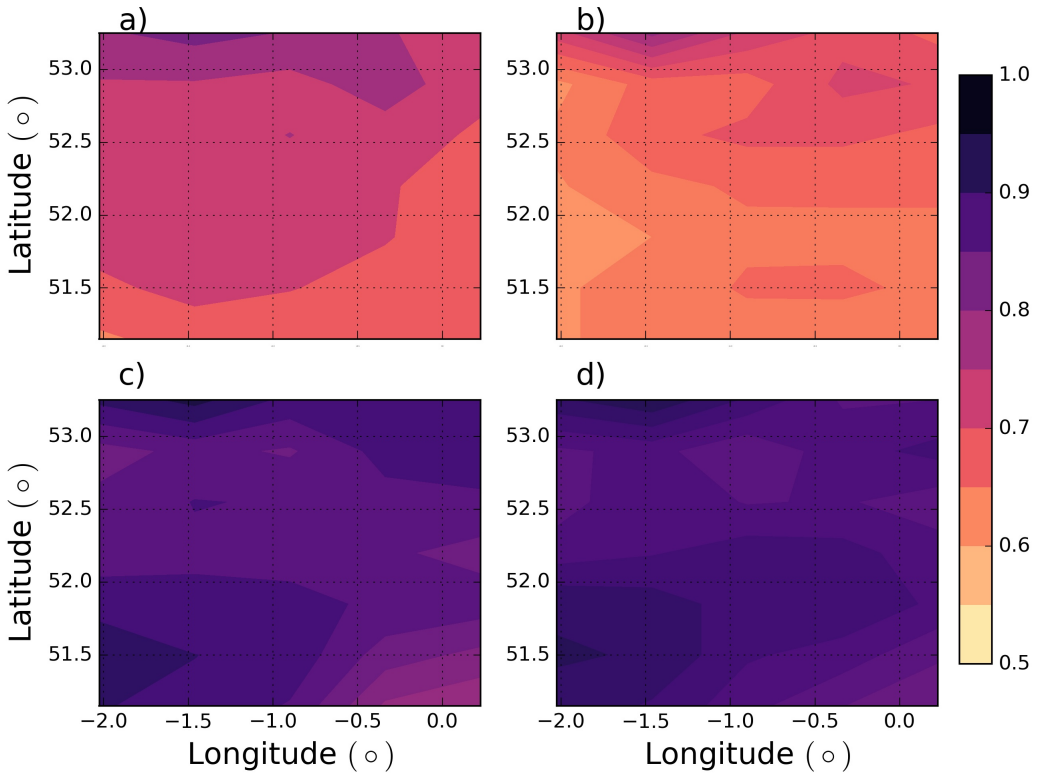


FIGURE 6.10: Same as in figure 6.9, with forecast evaluated against stations observations.

the spatial map of BS, for the *smoothing* forecast on the 3×3 neighborhood. It is worth to notice the variation of BS with latitude for all the probabilistic forecasts: the probabilistic forecasts perform better at high latitudes than lower latitude. The same is true also with respect to IGN (not shown here). More precisely, BS is lowest (indicating more skill) in the upper left part of the domain. By looking at figure 6.16, it can be seen that in the upper left part of the domain (indicated by the black box) the elevation is higher. Therefore, it is plausible that topography plays a role in the wind gust prediction. In figure 6.17 the Brier Score differences between each pair of the probabilistic forecast are shown, evaluated against observations on the 3×3 neighborhood P_{hrdyn} is performing better than any other probabilistic forecast, almost everywhere significantly. The difference between the two Bayesian forecasts is shown in figure 6.17d. The difference is higher in the central and upper part of the domain than in the lower part. Everywhere, the Bayesian model based on the gust parameter only is providing more information than the one based also on the convective parameter.

Figure 6.18 shows the same as the previous figure, but with probabilistic forecast evaluated against the control member. In this case, it can be appreciated that the differences are almost everywhere not significant, indicating that none of the probabilistic forecast has a significant superior potential skill for the wind gust occurrence.

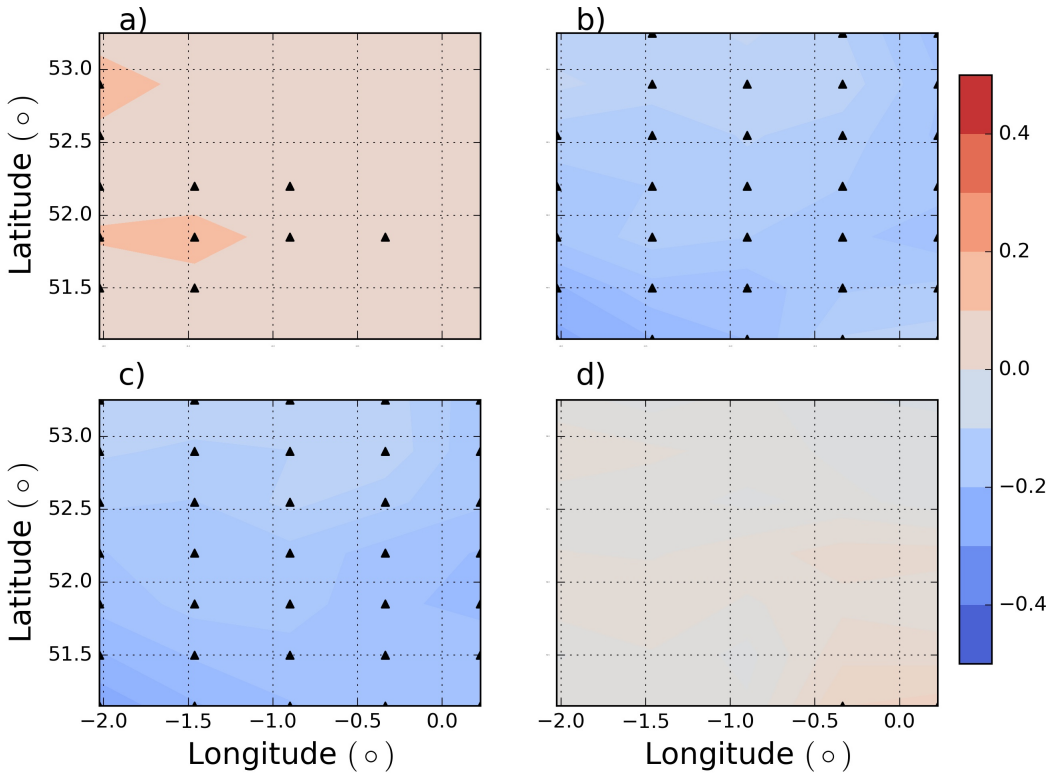


FIGURE 6.11: AUC differences for a)P_{hrdyn}-P_{gust}, b)P_{hrdyn}- $P_{bayfgustconv}$, c)P_{hrdyn}-$P_{bayfgust}$, $P_{bayfgustconv}$-$P_{bayfgust}$ for the 3×3 neighborhood, verified against station observations. Black triangles indicate significant differences at 5% level.

6.7 The relationship between wind gusts and large-scale conditions

In this section the relationship between wind gusts and large-scale conditions is investigated. Similarly as done with the sea breeze prediction, two predictors are selected and computed by MOGREPS-G. Any parameter should reflect a physical process involved in the generation of wind gusts. Two main mechanisms are associated with wind gust occurrence: wind shear and buoyancy (Suomi et al. (2013)). More precisely, "wind gusts observed at the surface seem to originate from air parcels flowing at higher levels and being deflected downwards to the surface" (Brasseur (2001)). In order to reach the surface, the stability of the layers is a very important factor. A stable layer prevents the deflection of an air, while instability favours the vertical transport and is able to deflect air parcels. Therefore as predictors the magnitude of wind shear and static stability at two geopotential heights (500 and 850 hPa) are used. These are defined in the following way:

$$\mathrm{WS}^2 = \frac{(\Delta U)^2 + (\Delta V)^2}{(\Delta z)^2} \tag{6.9}$$

$$\mathrm{N}^2 = \frac{g}{\theta_{w_0}} \frac{\Delta \theta_w}{\Delta z}, \tag{6.10}$$

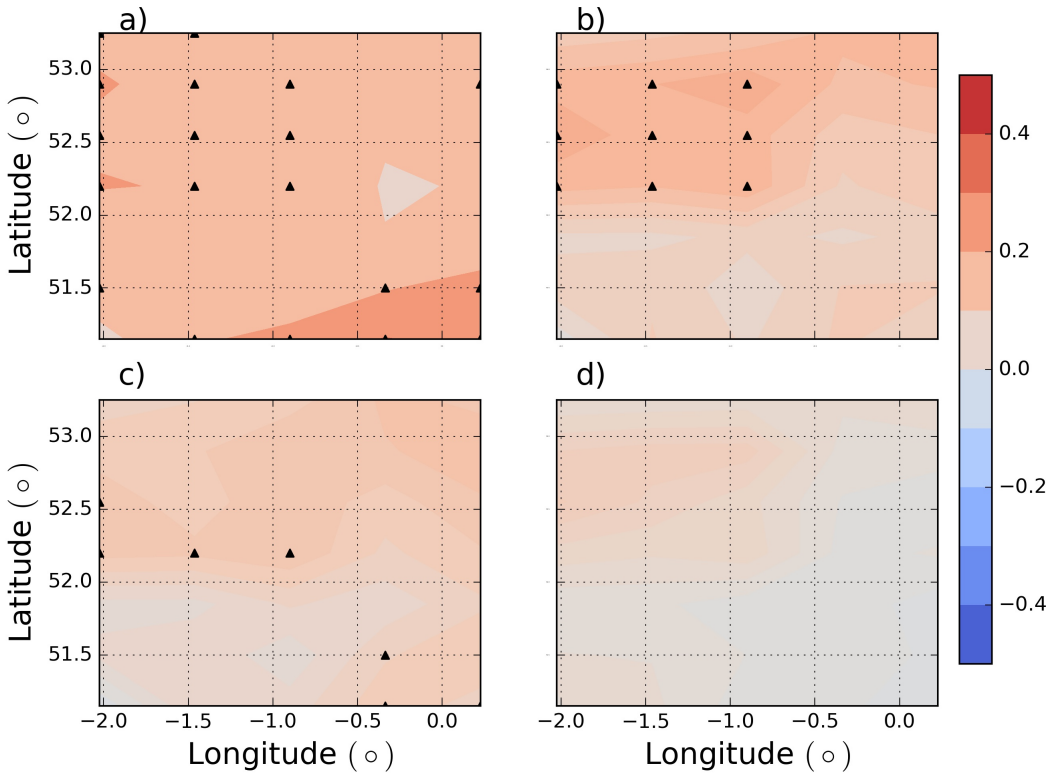


FIGURE 6.12: Same as in figure 6.11, with the MOGREPS-UK control member taken as verification.

where g is the gravitational acceleration, θ_{w_0} is the wet-bulb potential temperature and $\Delta\theta_w$ is the wet-bulb potential temperature difference and Δz is the height difference between the two geopotential heights.

These predictors are valid at 12UTC, for each MOGREPS-G grid box. The ratio N^2/WS^2 is known as *bulk Richardson number*.

The aim is to estimate the distribution of high wind speeds events (as defined in the previous sections by exceeding the 95^{th} percentile) and no-events in the parameter space defined by N^2 and WS^2. These distributions, estimated using again a Gaussian kernel density estimation technique, are shown in figure 6.19. for three different MOGREPS-G grid boxes, located in three different regions of the spatial domain considered in this study.

The distributions of all the cases (events and no-events) and no-events only (middle row) look very similar. This is because most of the cases comprise no-events. The distribution of the events is different and show interesting features. Two of the three locations show two peaks of the distribution. One peak is located in the portion of the parameter space corresponding to high values of shear and positive values of static stability. The other peak corresponds to low values of wind shear and vanishing values of static stability. High values of shear are found to be indicative of severe wind gust also by Taszarek et al. (2017). The other peak associated with near-neutral stable conditions ($N^2 \approx 0$) may be indicative of deeper mixed layer and thus of surface-induced convection. Therefore these two peaks may correspond to two different dynamical situations leading to wind

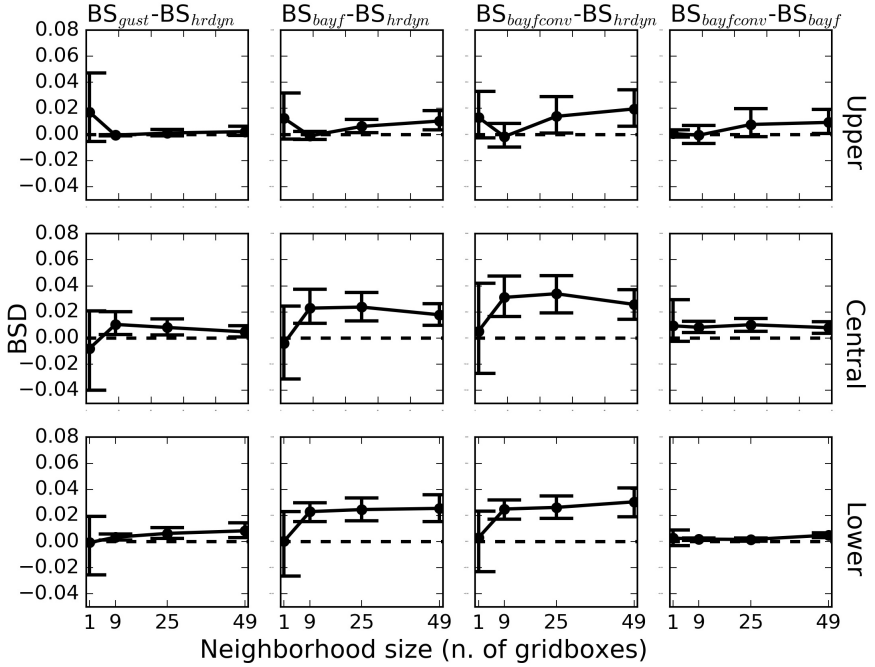


FIGURE 6.13: Brier score differences of P$_{hrdyn}$ relative to P$_{gust}$ (first column), to P$_{bayfgust}$ (second column) and to P$_{bayfgustconv}$(third column. The fourth column shows the BS of P$_{bayfgust}$ relative to P$_{bayfgustconv}$. Probabilistic forecast are produced with the upscaling method and verified against observations. Upper, central and lower refer to the three grid boxes as in the spatial domain of interest. Error bars show the 95% confidence interval calculated using the bootstrap method.

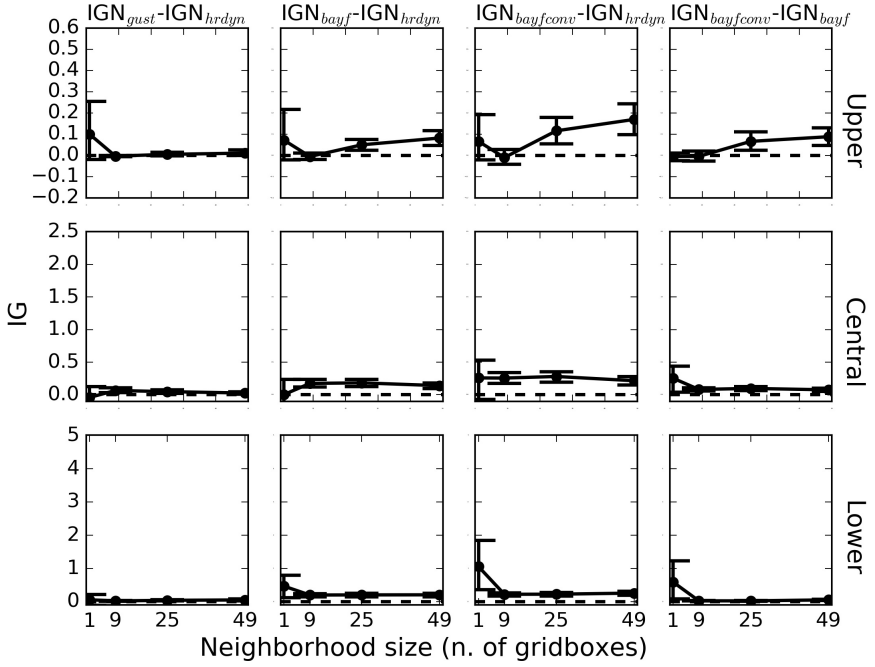


FIGURE 6.14: Same as in figure 6.13 for information gain score.

gusts: the first is connected to turbulent motions, the second to convective cases. In regard to the first peak, this corresponds to high values of wind shear and thus to high

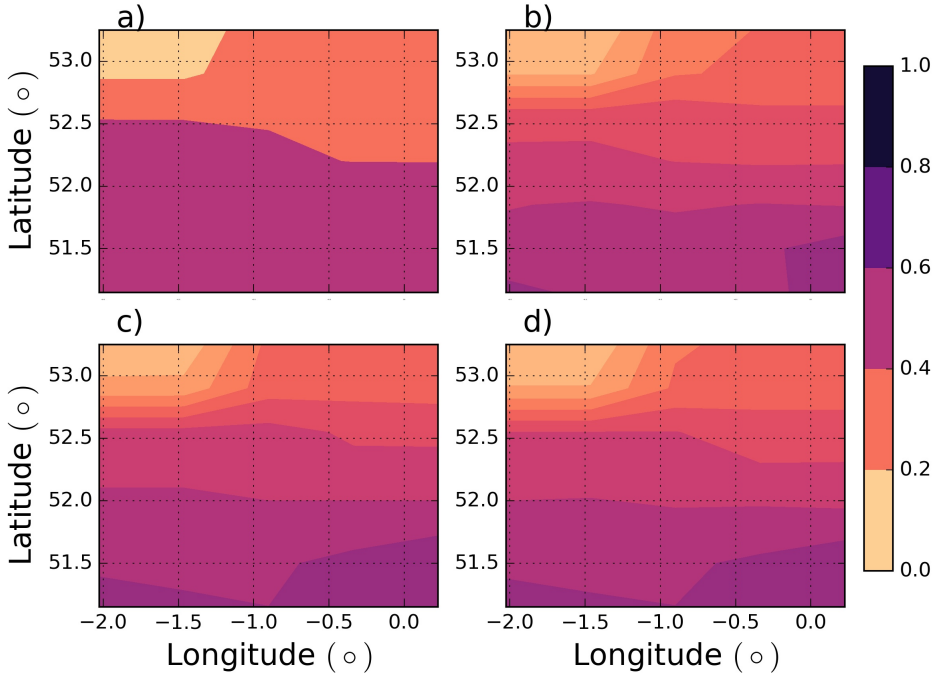


FIGURE 6.15: Map of the Brier Score for the *smoothing* probability forecast in the 3×3 neighborhood evaluated against observations.

values of *turbulent kinetic energy*, since the two quantities are strongly related. If then the mean turbulent kinetic energy is greater than the buoyant energy (measured by the static stability), the parcel of air is able to reach the surface generating the gusts.

Similarly as for the sea-breeze cases, the densities of wind gust occurrence and non-occurrence presented here can be used to calculate the conditional probabilities of the wind gust occurrence (as defined in the previous sections by exceeding the 95^{th} percentile of the MOGREPS-UK distribution). Then the spatial verification performed in the rest of this chapter can be applied to these probabilistic forecasts to calculate the skill scores relative to other probabilistic forecasts defined in section 6.5.

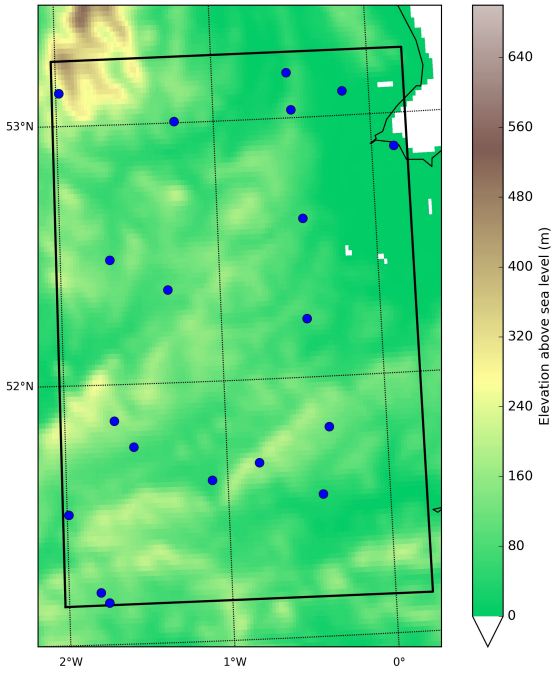


FIGURE 6.16: A closer look on the spatial domain of the figure 6.1.

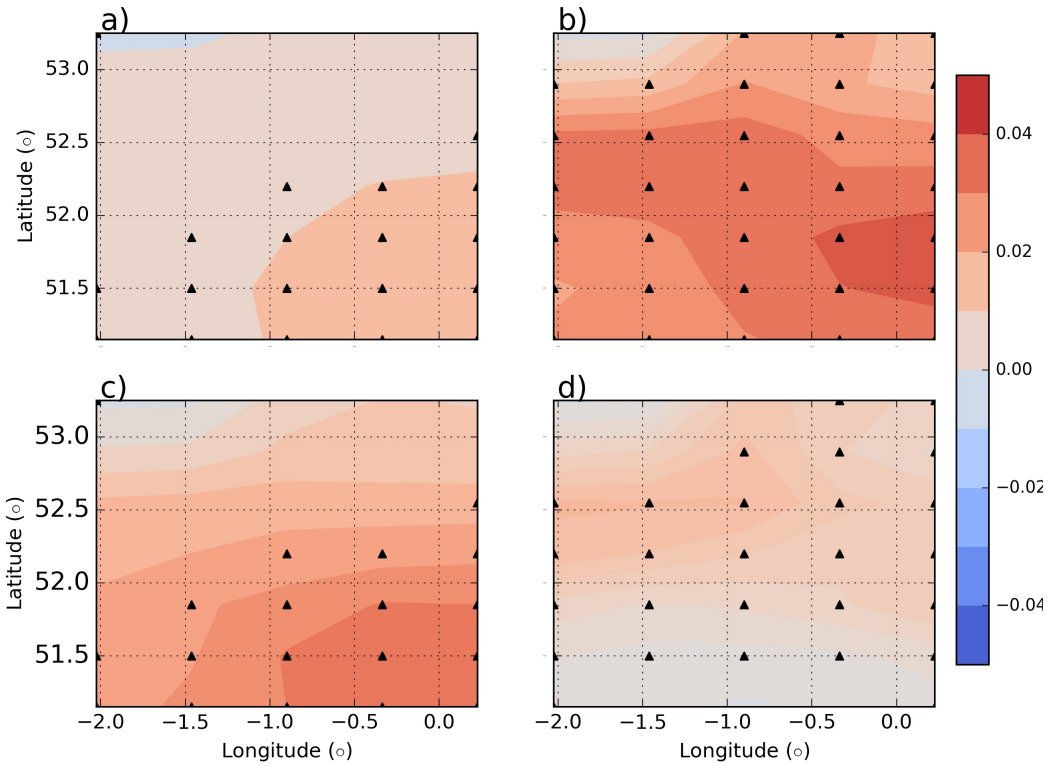


FIGURE 6.17: Map of the Brier Score difference for the *smoothing* probability forecast and observations within the smallest neighbourhood (3×3 grid boxes.). a) $BS_{gust} - BS_{hrdyn}$, b) $BS_{bayf\,gust\,conv} - BS_{hrdyn}$, c) $BS_{bayf\,gust} - BS_{hrdyn}$,d) $BS_{bayf\,gust\,conv} - BS_{bayf\,gust}$. Black triangles indicate significant differences at 5% level.

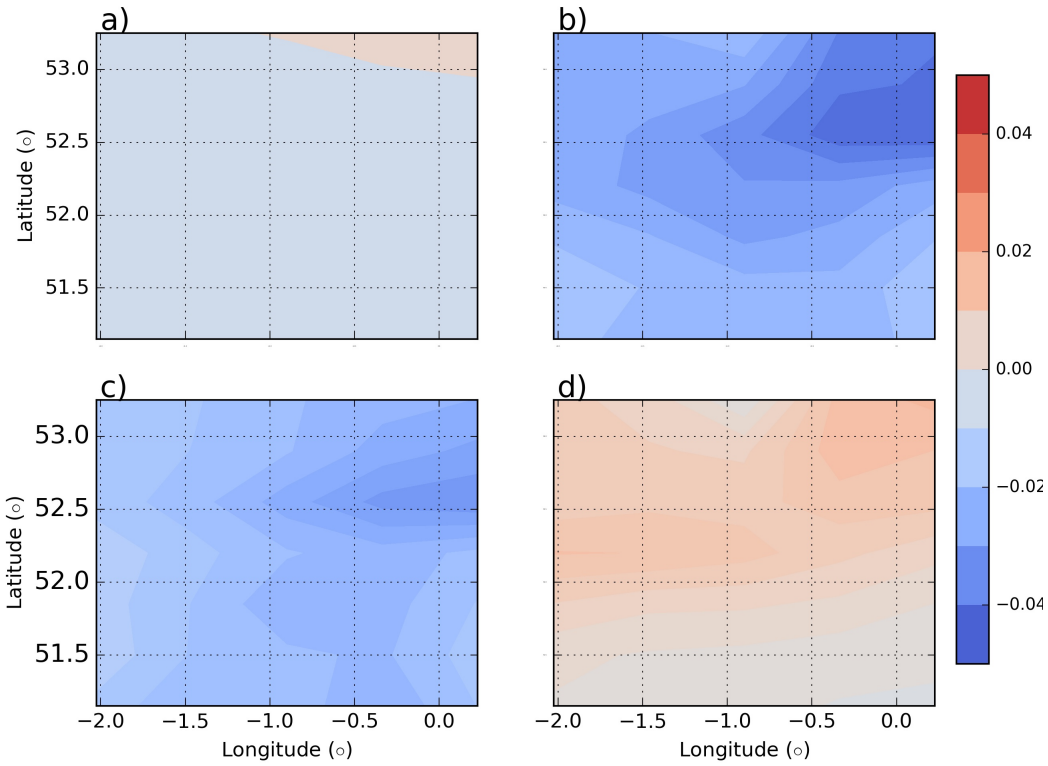


FIGURE 6.18: Same as figure 6.17, with probabilistic forecasts compared against the control member.

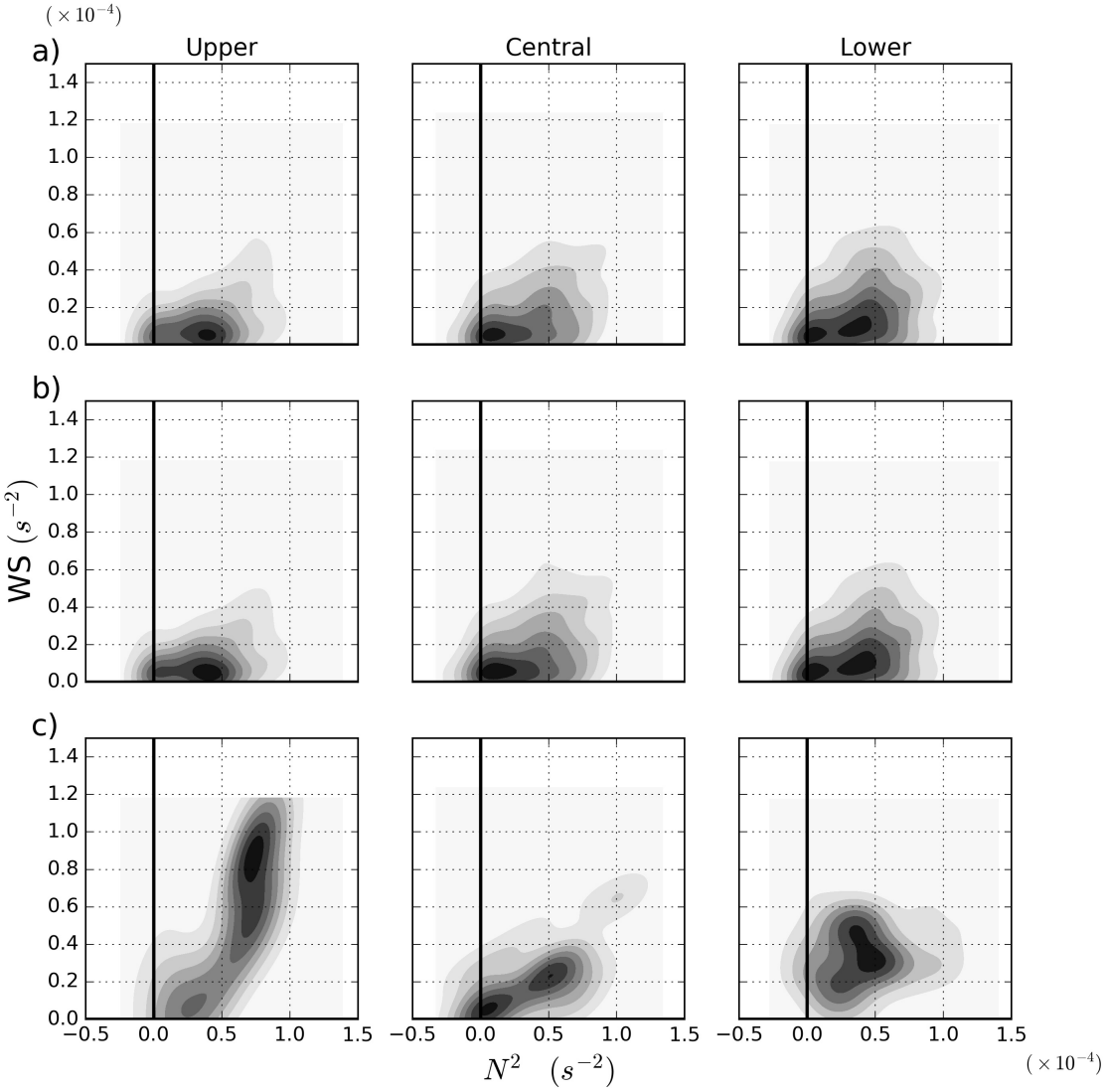


FIGURE 6.19: Kernel density estimation of the distribution of the high wind speeds for a) all cases, b) no-events and c) events in the parameter space defined by the wind shear and static stability for the upper, central, lower grid box. Events (i.e. the exceedance of the 95^{th} percentile of wind speed distribution) are defined according to MOGREPS-UK forecasts.

6.8 Summary and conclusions

In this chapter probabilistic forecasts of the occurrence of strong wind gust have been calculated and verified against station observations and the MOGREPS-UK control member for each grid point in the spatial domain using a neighborhood approach. Three different neighborhood sizes have been considered, consisting of 3×3, 5×5 and 7×7 grid boxes. Probabilistic forecasts of high wind speeds (above the 95^{th} percentile) are initialized at 00 UTC (03 UTC for the convection-permitting) and the 12-18 UTC time window is verified. From the analysis conducted the following conclusions can be drawn:

- There is high correlation between the gust diagnostic produced by the coarse resolution model and the 10m wind speed output. Furthermore, in convective situations the correlation is lower.
- The gust parameter produced by coarse resolution model and the Bayesian models based on this diagnostic only and also on convective parametrization have some information about the high wind gust occurrence (as defined earlier).
- Neverthless, the convection-permitting model is shown to provide still additional information (not in terms of AUC when verifying against observations), significantly for neighborhoods containing 3×3 grid boxes or more, when compared against station observations. If the MOGREPS-UK control member is taken as verification, the two probabilistic forecasts and the two Bayesian forecasts are shown to perform in a similar way and none of them is significantly superior.
- It seems that the geographical location plays a role for the performance of the probabilistic forecasts. More precisely, higher latitude regions perform better than the ones at lower latitudes. This can be attributed to the different elevation between higher and lower latitudes.
- Two large-scale predictors (static stability and wind shear) have shown to be able to discriminate between high wind speeds events and non events. In particular, the distribution of the events has two peaks which correspond to two different dynamical mechanisms (turbulent mixing and surface induced convection). leading to high wind speeds.

Chapter 7

Conclusions and Future work

7.1 Summary

Convection-permitting ensembles represent a new appealing forecasting technology . At the same time, due to very fine model grid size, they are computationally expensive. Therefore the aim of this thesis was to determine whether the resolved scales represent significant new forecast information or simply respond to large-scale forcing in a highly predictable way.
The other aim was also to understand the dynamical mechanisms leading to any improvement in the skill. Thus, two dynamical phenomena were selected to test the probabilistic forecasts' performance: sea breezes and convective wind gusts.
Sea breezes are atmospheric counterparts of the so called *density current*. Therefore a related aim was to investigate the dynamics of two colliding density currents, by means of numerical simulations in an idealized framework.
These two phenomena occur on spatio-temporal scales which are realistically represented only at convection-permitting resolutions. However, the mechanism of their occurrence is still influenced by the large-scale environmental conditions in which they are embedded. Therefore, the probability extracted from a convection-permitting model for the sea breeze and convective wind gust occurrence was compared to a conditional probability of the occurrence, based on large-scale conditions extracted from a lower resolution ensemble prediction system.
Probabilistic score differences between these forecasts were then calculated to quantify any additional value provided by the convection-permitting ensemble forecast.

7.2 Main results

The main results for each research questions are listed below the project described in chapter 2 and the remainder of the thesis:

7.2.1 Chapter 2

(a) Is a two-dimensional vorticity model able to capture the essential features of the collision dynamics, in an idealized framework similar to the laboratory experiments ? This can be seen first visually in the supplementary videos of (Cafaro and Rooney, 2018) and comparing with similar videos of (van der Wiel et al., 2017).

For a more quantitative analysis, the interface shape angle and the maximum collision height were calculated. In particular maximum collision height for both numerical simulations and laboratory experiments were compared (figure 2.8c), showing a similar behaviour with respect to the buoyancy ratio parameter.

(b) What is the role of baroclinically generated vorticity in determining and explaining some characteristics of the collision dynamics ?

For a steadily translating density current, the vorticity generation is balanced to first order by the horizontal current motion. It is conjectured that during a density-current collision, this balance is lost, and takes some time to be recovered. This implies that the dynamics during the collision are controlled by the motion of the current-head vortices which are already established.

(c) Can these characteristics be predicted using only the initial conditions, prior to the running of the numerical model ?

Yes, two predictive formula for the interface shape angle were presented, based on two different arguments. Both of them were found to agree well with the values of the numerical model (figures 2.7 and 2.12). This could have potential application in future parametrization of these processes in coarser resolution models.

7.2.2 Chapters 3, 4, 5, 6

a) How can the information about different weather events be extracted for an automated forecast production from both the convection-permitting model and its lower resolution counterpart ?

In chapter 4 methodologies to extract information about sea-breezes occurrence were described. For the convection-permitting model, using gridded data, a novel tracking algorithm of the sea-breeze front inland penetration was implemented. Detection methods used by other authors either used observational data, by thresholding the meteorological variables affected by the sea-breeze passage. This implies that different thresholds have to be used for different parts of the world. Hourly rate of change of surface wind and temperature variables were used to detect for each grid point, the position of the sea-breeze front. The main aim of the algorithm is to detect the spatio-temporal coherence of the sea-breeze front and to minimise the risk of spurious detection, given the resemblance between sea-breeze characteristics and cold fronts. This tracking algorithm is used to provide a definition of sea-breeze occurrence for each member and therefore to estimate the probability of sea-breeze occurrence. It could be implemented in an operational environmental system.

A different methodology was implemented in order to estimate the probability of sea-breeze occurrence from the coarser resolution ensemble members (section 4.3.3). Due to the inability to resolve the sea-breeze circulation, the former methodology is not applicable. A Bayesian approach was used instead. Two predictors (land-sea temperature contrast and large-scale wind), describing the large-scale

conditions and influencing the sea-breeze occurrence, define a two-dimensional parameter space. In this space, the two densities of occurrences and non-occurrences, defined with the previous tracking algorithm, are estimated. These densities are then used to estimate the probability of sea-breeze occurrence conditioned on the two-large scale predictors given by the coarse resolution model.
The same methodology was applied to the prediction of wind gusts. In this case different predictors were selected (static stability and wind shear), related to the wind gust occurrence.
Again, this Bayesian method could be implemented operationally.

b) How do forecast probabilities calculated from both the ensembles differ ?

In regard to the sea-breeze case, differences between the convection-permitting probabilistic forecast and the Bayesian forecast estimated from the coarse resolution ensembles, are established with respect to lead time. The higher resolution probabilistic forecast is sharper than the the Bayesian forecast for all lead times (figure 4.10). The Bayesian forecast exhibits a similar behaviour between T+12 and T+24 and T+18 and T+30 forecasts. At 12 and 24 hours ahead it does not have probability values higher than 0.5, whereas for the other two lead times, there are instances of values ≈ 1. Interestingly, these are the lead times when the Spearman's correlation coefficient is almost zero (figure 4.11), whereas it is ≈ 0.3 at 12 and 24 hours ahead. This coefficient measures to what extent there is a monotonic relationship between the two probabilistic forecast (ranging from -1 for a negative monotonic relationship to 1 for a positive relationship).
In regard to the wind gust variable, two probabilistic forecasts and two Bayesian forecasts, defined in equations (6.3-6.6), are compared, only for one initialization time and one valid time window for a whole summer season.
Once again, the two probabilistic forecasts estimated directly from the convection-permitting and coarse resolution ensembles are sharper than the Bayesian forecasts (figure 6.5).
In regard to correlation, given the probabilistic forecast were produced for each coarse resolution grid box, the Spearman's correlation coefficient was calculated for each grid box (figure 6.6). The probabilistic forecasts computed directly from the convection-permitting and coarse resolution members show correlations in the range $0.5 - 0.8$, with the peaks in the upper part of the spatial domain. The pair formed by the Bayesian forecasts conditioned on the gust parameter only and also on the convective parameter show the highest correlation (≈ 1), especially in the lower part of the domain.

c) What is the forecast skill of the convection-permitting ensemble relative to a coarser resolution ensemble ? How does it depend on the variable ? How does the skill evolves with forecast lead time ? Is any gain in the skill significant ?

In Chapter 5 the probabilistic forecasts of sea-breeze computed in chapter 4 were

verified against station observations. The skill of the convection-permitting ensemble was calculated relative to both the Bayesian forecast based on the large-scale predictors and the climatological forecasts, i.e. the percentage of occurrence in the stations observations.
Different verification metrics were used to quantify the added value of the convection-permitting ensemble: area under the ROC curve, Brier Score and Ignorance Score. It was found that the convection-permitting ensemble does have an higher area under the ROC curve for all lead times. This means that it is more able to discriminate between sea-breeze events and no-events than the Bayesian forecast (figure 5.2). The calculation of probabilistic scores above mentioned, indicate that the convection-permitting ensemble is significantly more skillful than the Bayesian forecast (figure 5.5). The difference in the Brier score is in the range $0.05 - 0.1$, whereas the Information gain in the interval $0.2 - 0.6$. Whether these values warrant the cost for running the convection-permitting ensemble forecast is however beyond the scope of this thesis.
Interestingly, it was found that the skill does not decrease with lead time. In particular, forecasts at 12 and 24 hours ahead show a similar performance, in terms of area under the ROC curve, reliability and resolution terms (figures 5.2, 5.5). Different initialization times with different assimilated observational data, boundary-layer and land-surface schemes may all have contributed to this evolution of the skill with lead time. Also, since sea breeze is a self-generating phenomenon if the conditions are right, there is no non-linear feedback from previous evolution.
For the wind gust prediction, in chapter 6, the probabilistic and Bayesian forecasts was verified against station observations (using a neighborhood approach) and also against the convection-permitting control member. The relative skill between different pair of the forecasts, calculated against station observations and for each grid box, shows that the convection-permitting ensemble is more skillful, almost everywhere significantly in the domain (figure 6.17). Instead, when the control member is taken as verification, the conclusion is different. Almost everywhere, none of the forecasts is significantly more skillful than the others.

d) Can be the potential information gain be predicted from the synoptic-scale flow ?
This was addressed by calculating the probability of the weather event, conditioned on some synoptic-scale predictors. In the sea-breeze case, the land sea temperature contrast and the ambient wind was shown to be able to discriminate between sea breeze events and no-events (figure 4.8). This implies that the synoptic-scale flow can convey some information about sea-breeze occurrence. After performing the verification against observation, it can be seen that in terms of resolution, the Bayesian model has significantly more resolution than the climatological forecast (which has zero resolution, by construction, figure 5.5). This means that having daily information about the large-scale conditions is more useful for the sea-breeze occurrence prediction than the climatological forecast.
Synoptic-scale predictors were chosen also for the wind gust prediction. The Bayesian

probabilistic forecasts based on gust parameter and the convective parameter, given by the coarse resolution ensemble, are shown to have potential information about wind gust occurrence, by performing not worse than the convection-permitting forecast (figure 6.18).

7.3 Future work

Here some limitations of this work along with some suggestions for extending it are offered.

7.3.1 Chapter 2

The work presented in chapter 2 although it offers interesting insights into the understanding of the dynamics of collision of two unequal density currents, has some limitations due to the numerical simulations carried in an idealized framework. Here some suggestions for future work for this specific project:

- Firstly, a stratified ambient fluid may be considered. This could represent other boundary layer types and the same features (maximum collision height and interface angle) can be calculated into this new framework.
- Secondly, the ambient wind at rest was imposed, as initial conditions. Therefore the effect of wind shear was not accounted.
- After incorporating these effects into the idealized numerical simulations, a 3D numerical simulation may be considered, maybe using large eddy simulations (LES) models.

7.3.2 Chapters 4,5,6

The work presented on the sea breeze and convective wind gusts has some limitations. Here these limitations are discussed some directions for extending it in the future are offered.

- The sea breeze tracking algorithm assumes a east-west oriented coast. In this case the maximum of the hourly changes of the variables involved was calculated in the meridional direction (in the equations 4.4-4.6), which is the direction perpendicular to the coast. For more general coastline orientation, the maximum on the gridded data should be calculated with respect to both the meridional and zonal directions.
- The sea breeze probabilistic occurrence forecasts are produced by calculating the probability of the sea breeze diagnostic exceeding a certain threshold. Sensitivity tests on how the probability values vary with this threshold can be carried, defining maybe the optimal threshold the one for which the highest probabilistic score is achieved.

- The Bayesian model presented in chapter 4 can be improved by adding other predictors, whose information is not already embedded by temperature contrast and large-scale wind. For instance, the stability parameter can be included, or other conditions which are not directly or indirectly included in the parameters used. However, in this case much more data is needed to obtain good kernel density estimates as the dimension increases from two. Also, other statistical methods can be used (e.g. multivariate logistic regression, model selection approach). However these methods could have not provided us a physical link between the higher density of sea breeze events and the values of the predictors used.

- The algorithm to detect sea breezes in observations is basic. More sophisticated methods, including fuzzy algorithms may be used (Coceal et al., 2018). Furthermore, observational errors are not taken into account in the verification process (e.g. as in Bowler (2008), Candille and Talagrand (2008)). This is because the main scope of this thesis is the inter-comparison of the two probabilistic systems and not assessing the quality of the systems themselves. However, taking into account observational error "helps to produce better forecasts of the truth and to ensure we favour forecasters who issue better forecasts of the truth" (Ferro (2017)).

- The work presented in chapter 6 is an application of the methodology introduced in chapter 4 to a severe weather event. Therefore it serves as a proof of concept of the methods. Thus only one forecast lead time and one validation year only were investigated, implementing the leaving one-day-out technique for training the Bayesian models. Also, the probability of exceeding a high percentile was used and not also the full probability distribution of the gust. Furthermore, the 10m wind speed was used as a proxy for the wind gust and compared with the gust parameter from the coarse resolution ensemble. This is reasonable in convective situations, when then convective outflow contribution is higher than the turbulent process contribution. However, in non-convective situations, there would still need to parametrize turbulence mixing for the convection-permitting model, providing a more direct comparison with the gust parameter output from the coarse resolution ensemble. The problem is, as pointed out by Mylne and Roberts (2017) is that in convective situations, the use of a parametrized gust would lead to an overestimation of the wind speed. Therefore in convective situation, the gust parametrization should be switched off or modified in convective situations. A clear discrimination between convective and no-convective events is needed to remove this overestimation.
  Also, the Bayesian model based on wind shear and static stability can be trained on a much larger sample and then used to calculate the conditional probability of high wind speeds based on the two predictors.

Finally the author would like to raise some other research questions, which are not directly related to the main research questions of this thesis but that could lead to some interesting work in the future:

- Looking at figure 4.5, some ensemble members show the sea breeze signal, whereas others not. It could be interesting to know what kind of physical differences are responsible for this behaviour.

- A general *thought* on reliability property of ensemble forecast. One assumption is that whenever a probability of sea breeze of 40% is issued, it is expected the sea breeze to occur 40% of the time, if the ensemble prediction system is reliable.
  However, this could not be true, not only because the system is indeed not reliable but also because when the same probability is issued, the atmospheric conditions are not necessary the same.
  In this study, sea breeze forecast were analysed in anticyclonic conditions, so this assumption should be verified. However, for instance, for variables like precipitation this assumption, in principle, is not true.
  That is why it is important to categorize events for different weather regimes, to have a more reliable analysis of the reliability property. So, once again, the verification of the phenomena themselves and not the conventional model variables is important in this respect.
  The hope is that with this thesis, such analysis of specific weather phenomena is encouraged and will become more widespread in the future.

Appendix A

Modified maximum likelihood estimator of sea breeze occurrence

In this appendix a modified maximum likelihood estimator (MLE) of the Bernoulli distribution is presented. In chapter 4, to generate the convection-permitting forecast from the ensemble members, the classic MLE was used (see equation 4.13). Here a modified maximum likelihood estimator is presented. This is to take into account the finite ensemble size, which can lead to the generation of probability values of 0, which are problematic in the calculation of the Ignorance score.
Let the functional $\phi(p)$ be defined as

$$\phi(p) := -\log \mathcal{L}(p) + \lambda D_{KL}(p, p_c), \tag{A.1}$$

where $\mathcal{L}$ is the classic likelihood function of the Bernoulli distribution, D_{KL} is the Kullback-Leibler divergence between p and p_c already defined in Chapter 5 and λ a real parameter to choose and p_c a climatological probabilistic forecast. More precisely

$$\mathcal{L}(p) := p^n (1-p)^{N-n} \tag{A.2}$$

$\mathcal{L}$ can be rewritten as:

$$\mathcal{L}(p) = \left(\frac{p}{\hat{p}}\right)^n \left(\frac{1-p}{1-\hat{p}}\right)^{N-n} \hat{p}^n (1-\hat{p})^{N-n}, \tag{A.3}$$

where $\hat{p} = n/N$ is the maximum likelihood estimator (MLE) for the Bernoulli distribution, the sample proportion of successes n and N is the ensemble member size. Therefore

$$\begin{aligned} -\log \mathcal{L}(p) &= -n \log\left(\frac{p}{\hat{p}}\right) - (N-n)\log\left(\frac{1-p}{1-\hat{p}}\right) + \log C(N,n) \\ &= N\left(-\hat{p}\log\left(\frac{p}{\hat{p}}\right) - (1-\hat{p})\log\left(\frac{1-p}{1-\hat{p}}\right)\right) + \log C(N,n) \end{aligned} \tag{A.4}$$

where $C(N,n) = \hat{p}^n(1-\hat{p})^{N-n}$ that does not depend on p.
Therefore $\phi(p)$ can be rewritten as:

$$\phi(p) = N D_{KL}(p, \hat{p}) + \lambda D_{KL}(p, p_c) \tag{A.5}$$

Thus, in order to calculate the MLE for ϕ:

$$\begin{aligned}\frac{\partial}{\partial p}\phi(p) &= N\left(-\frac{\hat{p}}{p}+\frac{1-\hat{p}}{1-p}\right)+\lambda\left(-\frac{p_c}{p}+\frac{1-p_c}{1-p}\right)\\ &= \frac{1}{p}\left(-N\hat{p}-\lambda p_c\right)+\frac{1}{1-p}\left(N(1-\hat{p})+\lambda(1-p_c)\right)\end{aligned} \tag{A.6}$$

Then it follows that:

$$\frac{\partial}{\partial p}(\phi(p))=0 \iff p=\frac{\hat{p}+\frac{\lambda}{N}p_c}{1+\frac{\lambda}{N}}, \tag{A.7}$$

which is the MLE for the functional ϕ. It is worth to notice that $p \to \hat{p}$ for $N \to \infty$ or if $\lambda = 0$, i.e. if the ensemble size becomes bigger and bigger or if the climatological probability has no weight.

Appendix B

Sufficiency property

In order to provide additional evidence related to the equation (5.1), here another property is examined. This is the sufficiency property explored by (De Groot and Fienberg, 1982) and reformulated by (Bröcker, 2009) in these mathematical terms:

$$\mathbb{P}(Y = 1|(p, q)) = \mathbb{P}(Y = 1|q), \tag{B.1}$$

where p and q are the convective-scale and the Bayesian forecast respectively. In other words, equation (B.1) means that the forecasting scheme q is sufficient for p. Therefore if equation (B.1) holds, there is not additional information provided by p.
Again, its validity is tested statistically, using the *log-likelihood ratio test* (Mood et al. (1974)), which has been already used to test resolution (section 5.3).
In particular p and q are divided into categories $\mathcal{C}_m$ and $\mathcal{D}_l$ $(1 \leq m \leq M, 1 \leq l \leq L)$ respectively. In this case $M = L = 2$. The two categories $\mathcal{C}$ and $\mathcal{D}$ are the two evenly spaced bins in the interval $[0, \max p]$ and $[0, \max q]$ respectively. This will define a *three-way contingency table*:

$$\#\{Y = k, p \in \mathcal{C}_m, q \in \mathcal{D}_l\} =: n_{k,m,l}, \tag{B.2}$$

with

$$\mathbb{P}(Y = k, p \in \mathcal{C}_m, q \in \mathcal{D}_l) =: P_{k,m,l}. \tag{B.3}$$

Categories $\mathcal{C}$ and $\mathcal{D}$ have been chosen in order to not have zero entries in the contingency table $n_{k,m,l}$. Therefore equation (B.1) can be rewritten as:

$$\mathbb{P}(Y = k|p \in \mathcal{C}_m, q \in \mathcal{D}_l) = \mathbb{P}(Y = k|q \in \mathcal{D}_l) \tag{B.4}$$

$$\iff \frac{P_{k,m,l}}{\sum_k P_{k,m,l}} = \frac{\sum_l P_{k,m,l}}{\sum_{k,l} P_{k,m,l}} (\mathcal{H}_0) \tag{B.5}$$

Equation B.5 represents the null hypothesis $\mathcal{H}_0$. The likelihood function and the likelihood ratio are thus defined as:

$$\mathcal{L} = \prod_{k,m,l} P_{k,l,m}^{n_{k,l,m}}. \tag{B.6}$$

$$\Lambda = \frac{\max \log \mathcal{L}}{\max_{\mathcal{H}_0} \log \mathcal{L}} = \frac{\lambda_1}{\lambda_0}. \tag{B.7}$$

$\lambda_0 = \max_{\mathcal{H}_0} \log \mathcal{L}$ has to be calculated. From equation (B.5):

$$P_{k,l,m} = \pi_{k,l} \rho_{l,m}, \tag{B.8}$$

where

$$\begin{aligned} \pi_{k,l} &= \frac{\sum_l P_{k,l,m}}{\sum_{k,l} P_{k,l,m}} \\ \rho_{l,m} &= \sum_k P_{k,l,m} \end{aligned} \tag{B.9}$$

Therefore

$$\log \mathcal{L}_{\mathcal{H}_0} = \log \left(\prod_{k,l,m} (\pi_{k,l}\rho_{l,m})^{n_{k,l,m}} \right), \tag{B.10}$$

where $\mathcal{L}_{\mathcal{H}_0}$ is the likelihood function under the null hypothesis $\mathcal{H}_0$.
Equation B.10 has to maximised with respect to both $\pi_{k,l}$ and $\rho_{l,m}$ There are two additional constraints to take into account:

$$\begin{aligned} &\sum_{l,m} \rho_{l,m} - 1 = 0 \\ &\sum_k \pi_{k,l} - 1 = 0 \qquad \forall 1 \leq l \leq L. \end{aligned} \tag{B.11}$$

Hence, in order to calculate $\max_{\mathcal{H}_0} \log \mathcal{L}$, the method of Lagrange multiplier is applied subject to (B.11).
The following expression needs to be maximised:

$$F(\pi_{k,l}, \rho_{l,m}, R, \Pi_l) = \sum_{k,l,m} n_{k,l,m} \left(\log(\pi_{k,l}) + \log(\rho_{l,m})\right) + R \left(\sum_{l,m} \rho_{l,m} - 1 \right) + \sum_{k,l} \Pi_l \left(\pi_{k,l} - \frac{1}{k} \right) \tag{B.12}$$

$$\begin{aligned} \frac{\partial F}{\partial \pi_{k,l}} &= \frac{\sum_m n_{k,l,m}}{\pi_{k,l}} + \Pi_l = 0 \\ \frac{\partial F}{\partial \rho_{l,m}} &= \frac{\sum_k n_{k,l,m}}{\rho_{l,m}} + R = 0 \end{aligned} \tag{B.13}$$

Combining equations (B.13-B.11) the following expressions for $\pi_{k,l}$ and $\rho_{l,m}$ are obtained:

$$\begin{aligned} \pi_{k,l} &= \frac{\sum_m n_{k,l,m}/N}{\sum_{m,k} n_{k,l,m}/N} \\ \rho_{l,m} &= \sum_k n_{k,l,m}/N, \end{aligned} \tag{B.14}$$

where N is the total length of the dataset.
It can be proved that $2(\lambda_1 - \lambda_0) = 2\log(\frac{\max L}{\max_{\mathcal{H}_0} L}) \approx \chi_d$, where $d = (K - 1 + M)L - 1$.[1]
The results of the *log-likelihood* ratio test are shown in table B.1.

It can be seen both from the *p-value* and from the value Λ that there is enough evidence to reject $\mathcal{H}_0$ at 5% significance level. In other words, this statistical test suggests

[1] In our case K=2

Forecast lead time (hours)	6	12	18	24	30
Λ	427.00	$8.43 \cdot 10^3$	$2.26 \cdot 10^6$	1.04×10^7	1.07×10^4
p	0.03	0.002	$2.05 \cdot 10^{-5}$	5.14×10^{-6}	0.023

TABLE B.1: Log-likelihood ratio test results for all lead times.

that the Bayesian forecast alone does not provide sufficient information.

Bibliography

ABI (2005). Financial risks of climate change. Technical annex, Association of British Insurers.

Arritt, R. W. (1993). Effects of the Large-Scale flow on Characteristics Features of the Sea Breeze. *Journal of Applied Meteorology*, 32:116 – 125.

Atkins, N. T. and Wakimoto, R. M. (1997). Influence of the synoptic-scale flow on sea breezes observed during CAPE. *Monthly Weather Review*, 125(9):2112–2130.

Azorin-Molina, C. and Chen, D. (2008). A climatological study of the influence of synoptic-scale flows on sea breeze evolution in the Bay of Alicante (Spain). *Theoretical and Applied Climatology*, 96:249 – 260.

Azorin-Molina, C. and Tijm, D. C. S. (2011). Development of selection algorithms and databases for sea breeze studies. *Theoretical and Applied Climatology*, 106:531 – 546.

Barrett, A. I., Gray, S. L., Kirshbaum, D. J., Roberts, N. M., Schultz, D. M., and Fairman, J. G. (2016). The Utility of Convection-Permitting ensembles for the prediction of stationary convective bands. *Monthly Weather Review*, 144(3):1093–1114.

Barthlott, C., Mühr, B., and Hoose, C. (2017). Sensitivity of the 2014 Pentecost storms over Germany to different model grids and microphysics schemes. *Quarterly Journal of the Royal Meteorological Society*, 143(704):1485–1503.

Beck, J., Bouttier, F., Wiegand, L., Gebhardt, C., Eagle, C., and Roberts, N. (2016). Development and verification of two convection-allowing multi-model ensembles over Western Europe. *Quarterly Journal of the Royal Meteorological Society*, 142(700):2808–2826.

Ben Bouallègue, Z. and Theis, S. E. (2014). Spatial techniques applied to precipitation ensemble forecasts: from verification results to probabilistic products. *Meteorological Applications*, 21(4):922–929.

Benedetti, R. (2010). Scoring rules for forecast verification. *Monthly Weather Review*, 138(1):203–211.

Benjamin, T. B. (1968). Gravity currents and related phenomena. *Journal of Fluid Mechanics*, 31(2):209–248.

Biggs, W. G. and Graves, M. E. (1962). A Lake breeze index. *Journal of Applied Meteorology*, 1(4):474–480.

Bigot, S. and Planchon, O. (2003). Identification and characterization of sea breeze days in Northern France using singular value decomposition. *International Journal of Climatology*, 23(11):1397–1405.

Birch, C. E., Roberts, M. J., Garcia-Carreras, L., Ackerley, D., Reeder, M. J., Lock, A. P., and Schiemann, R. (2015). Sea-breeze dynamics and convection initiation: The influence of convective parameterization in weather and climate model biases. *Journal of Climate*, 28(20):8093–8108.

Bouttier, F., Raynaud, L., Nuissier, O., and Benjamin, M. (2016). Sensitivity of the AROME ensemble to initial and surface perturbations during HyMex. *Quarterly Journal of the Royal Meteorological Society*, 142(S1):390–403.

Bowler, N., Arribas, A., Mylne, K. R., Robertson, K., and Beare, S. E. (2008). The MOGREPS short-range ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 134:703 – 722.

Bowler, N. E. (2008). Accounting for the effect of observation errors on verification of mogreps. *Meteorological Applications*, 15(1):199–205.

Bowler, N. E., Arribas, A., Beare, S. E., Mylne, K. R., and Shutts, G. J. (2009). The local ETKF and SKEB: Upgrades to the MOGREPS short-range ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 135(640):767–776.

Brasseur, O. (2001). Development and application of a physical approach to estimating wind gusts. *Monthly Weather Review*, 129(1):5–25.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1):1–3.

Bröcker, J. (2015). Resolution and discrimination – two sides of the same coin. *Quarterly Journal of the Royal Meteorological Society*, 141:1277 – 1282.

Bröcker, J. and Smith, L. A. (2007a). Increasing the reliability of reliability diagrams. *Weather and Forecasting*, 22(3):651–661.

Bröcker, J. and Smith, L. A. (2007b). Scoring probabilistic forecasts: The importance of being proper. *Weather and Forecasting*, 22(2):382–388.

Bröcker, J. and Smith, L. A. (2008). From ensemble forecasts to predictive distribution functions. *Tellus Series A: Dynamic Meteorology and Oceanography*, 60(4):663 – 678.

Bryan, G. H. and Rotunno, R. (2014). The optimal state for gravity currents in shear. *Journal of the Atmospheric Sciences*, 71(1):448–468.

Bröcker, J. (2009). Reliability, sufficiency, and the decomposition of proper scores. *Quarterly Journal of the Royal Meteorological Society*, 135(643):1512–1519.

Buizza, R. (1997). Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF ensemble prediction system. *Monthly Weather Review*, 125(1):99–119.

Buizza, R., Houtekamer, P. L., Pellerin, G., Toth, Z., Zhu, Y., and Wei, M. (2005). A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Monthly Weather Review*, 133(5):1076–1097.

Cafaro, C., Frame, T. H. A., Methven, J., Roberts, N., and Bröcker, J. (2019). The added value of convection-permitting ensemble forecasts of sea breeze compared to a bayesian forecast driven by the global ensemble. *Quarterly Journal of the Royal Meteorological Society*, 145(721):1780–1798.

Cafaro, C. and Rooney, G. G. (2018). Characteristics of colliding density currents: A numerical and theoretical study. *Quarterly Journal of the Royal Meteorological Society*, 144(715):1761–1771.

Candille, G. and Talagrand, O. (2008). Impact of observational error on the validation of ensemble prediction systems. *Quarterly Journal of the Royal Meteorological Society*, 134(633):959–971.

Carr, M., Franklin, J., King, S. E., Davies, P. A., Grue, J., and Dritschel, D. G. (2017). The characteristics of billows generated by internal solitary waves. *Journal of Fluid Mechanics*, 812:541–577.

Celisse, A. and Robin, S. (2008). Nonparametric density estimation by exact leave-p-out cross-validation. *Computational Statistics Data Analysis*, 52(5):2350–2368.

Clappier, A., Martilli, A., Grossi, P., Thunis, P., Pasi, F., Krueger, B. C., Calpini, B., Graziani, G., and van den Bergh, H. (2000). Effect of Sea Breeze on Air Pollution in the Greater Athens Area. Part I: Numerical Simulations and Field Observations. *Journal of Applied Meteorology*, 39(4):546–562.

Clark, P., Roberts, N., Lean, H. S. P., Ballard, S. P. C., and Charlton-Perez, C. (2016). Review: Convection-permitting models: a step-change in rainfall forecasting. *Meteorological Applications*, 23:165 – 181.

Coceal, O., Bohnenstengel, S. I., and Kotthaus, S. (2018). Detection of sea-breeze events around London using a fuzzy-logic algorithm. *Atmospheric Science Letters*, 19(9):e846.

Cochran, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics*, 10(4):417–451.

Corpetti, T. and Planchon, O. (2011). Front detection on satellite images based on wavelet and evidence theory: Application to the sea breeze fronts. *Remote Sensing of Environment*, 115(2):306 – 324.

Crosman, E. T. and Horel, J. D. (2010). Sea and lake breezes: A review of numerical studies. *Boundary-Layer Meteorology*, 137(1):1–29.

Davies, T. (2014). Lateral boundary conditions for limited area models. *Quarterly Journal of the Royal Meteorological Society*, 140(678):185–196.

Davies, T., Cullen, M. J. P., Malcom, A. J., Mawson, M. H., Staniforth, A., White, A. A., and Wood, N. (2005). A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Quarterly Journal of the Royal Meteorological Society*, 131:1759 – 1782.

De Groot, M. W. and Fienberg, S. E. (1982). Assessing probability assessors: calibration and refinement. *Statistical Decision Theory and Related Topics*, (1):291–314.

De Meutter, P., Gerard, L., Smet, G., Hamid, K., Hamdi, R., Degrauwe, D., and Termonia, P. (2015). Predicting small-scale, short-lived downbursts: Case study with the NWP Limited-Area ALARO model for the Pukkelpop Thunderstorm. *Monthly Weather Review*, 143(3):742–756.

DeLong, E., DeLong, D., and Clarke-Pearson, D. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 44(3):837—845.

Dey, S. R. A., Plant, R. S., Roberts, N. M., and Migliorini, S. (2016). Assessing spatial precipitation uncertainties in a convective-scale ensemble. *Quarterly Journal of the Royal Meteorological Society*, 142(701):2935–2948.

Dritschel, D. G. and Fontane, J. (2010). The combined Lagrangian advection method. *Journal of Computational Physics*, 229:5408–5417.

Droegemeier, K. K. and Wilhelmson, R. B. (1985). Three-dimensional numerical modeling of convection produced by interacting thunderstorm outflows. Part I: Control Simulation and Low-Level Moisture Variations. *Journal of the Atmospheric Sciences*, 42(22):2381–2403.

Ebert, E., Wilson, L., Weigel, A., Mittermaier, M., Gill, P. N. P. P., Göber, M., Joslyn, S., Brown, B., Fowler, T., and Watkins, A. (2013). Progress and challenges in forecast verification. *Meteorological Applications*, 20(2):130–139.

Egan, J. P. (1975). *Signal detection theory and ROC-analysis*. New York : Academic Press.

Ferro, C. A. T. (2017). Measuring forecast performance in the presence of observation error. *Quarterly Journal of the Royal Meteorological Society*, 143(708):2665–2676.

Findlater, J. (1964). The sea-breeze and inland convection: an example of their interrelation. *Meteorological Magazine*, 93:82 – 9.

Finkele, K. (1998). Inland and offshore propagation speeds of a sea breeze from simulations and measurements. *Boundary-Layer Meteorology*, 87(2):307–329.

Frysinger, J. R., Lindner, B. L., and Brueske, S. L. (2003). A statistical sea-breeze prediction algorithm for charleston, south carolina. *Weather and Forecasting*, 18(4):614–625.

Gallo, B. T., Clark, A. J., and Dembek, S. R. (2016). Forecasting tornadoes using convection-permitting ensembles. *Weather and Forecasting*, 31(1):273–295.

Gebhardt, C., Theis, S., Paulat, M., and Bouallègue, Z. B. (2011). Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation of lateral boundaries. *Atmospheric Research*, 100(2):168 – 177.

Gilleland, E. (2017). A new characterization within the spatial verification framework for false alarms, misses, and overall patterns. *Weather and Forecasting*, 32(1):187–198.

Gneiting, T., Raftery, A. E., Westveld, A. H., and Goldman, T. (2005). Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Monthly Weather Review*, 133(5):1098–1118.

Golding, B., Roberts, N., Leoncini, G., Mylne, K., and Swinbank, R. (2016). MOGREPS-UK convection-permitting ensemble products for surface water flood forecasting: Rationale and first results. *Journal of Hydrometeorology*, 17(5):1383–1406.

Golding, B. W., Ballard, S., Mylne, K., Roberts, N., Saulter, A., Wilson, C., Agnew, P., Davis, L. S., Trice, J., Jones, C., Simonin, D., Z, L., Pierce, C., Bennett, A., Weeks, M., and Moseley, S. (2014). Forecasting Capabilities for the London 2012 Olympics. *Bull. Amer. Meteor. Soc*, 95:883 – 896.

Gowan, T. M., Steenburgh, W. J., and Schwartz, C. S. (2018). Validation of Mountain Precipitation Forecasts from the Convection-Permitting NCAR ensemble and operational forecast systems over the Western United States. *Weather and Forecasting*, 33(3):739–765.

Grandpeix, J.-Y. and Lafore, J.-P. (2010). A density current parameterization coupled with Emanuel's convection scheme. Part I: The models. *Journal of the Atmospheric Sciences*, 67(4):881–897.

Gray, M. E. B. (2006). The use of a cloud resolving model in the development and evaluation of a probabilistic forecasting algorithm for convective gusts. *Meteorological Applications*, 10(3):239–251.

Hagelin, S., Son, J., Swinbank, R., McCabe, A., Roberts, N., and Tennant, W. (2017). The Met Office convective-scale ensemble, MOGREPS-UK. *Quarterly Journal of the Royal Meteorological Society*, 143(708):2846–2861.

Hamill, T. M. and Church, A. T. (2000). Conditional Probabilities of Significant Tornadoes from RUC-2 Forecasts. *Weather and Forecasting*, 15:461 – 475.

Harvey, B., Methven, J., Eagle, C., and Lean, H. (2017). Does the representation of flow structure and turbulence at a cold front converge on multiscale observations with model resolution? *Monthly Weather Review*, 145(11):4345–4363.

Hawker, M. (2007). Climate change and the global insurance industry. *The Geneva Papers on Risk and Insurance. Issues and Practice*, 32(1):22–28.

Hewston, R. and Dorling, S. R. (2011). An analysis of observed daily maximum wind gusts in the UK. *Journal of Wind Engineering and Industrial Aerodynamics*, 99(8):845 – 856.

Hohenegger, C. and Schar, C. (2007). Atmospheric predictability at synoptic versus cloud-resolving scales. *Bulletin of the American Meteorological Society*, 88(11):1783–1794.

Hohenegger, C., Walser, A., Langhans, W., and Schär, C. (2008). Cloud-resolving ensemble simulations of the August 2005 Alpine flood. *Quarterly Journal of the Royal Meteorological Society*, 134(633):889–904.

Horányi, A., Mile, M., and Szucs, M. (2011). Latest developments around the ALADIN operational short-range ensemble prediction system in Hungary. *Tellus A: Dynamic Meteorology and Oceanography*, 63(3):642–651.

Intrieri, J. M., Jr., A. J. B., and Hardesty, R. M. (1990). Details of colliding thunderstorm outflows as observed by Doppler lidar. *Journal of the Atmospheric Sciences*, 47(9):1081–1099.

Joliffe, I. T. and Stevenson, D. B. (2012). *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. John Wiley & Sons, Ltd.

Jones, E., Oliphant, T., Peterson, P., et al. (2001). SciPy: Open source scientific tools for Python. [Online; accessed <today>].

Justel, A., Peña, D., and Zamar, R. (1997). A multivariate Kolmogorov-Smirnov test of goodness of fit. *Statistics Probability Letters*, 35:251–259.

Kambezidis, H., Weidauer, D., Melas, D., and Ulbricht, M. (1998). Air quality in the athens basin during sea breeze and non-sea breeze days using laser-remote-sensing technique. *Atmospheric Environment*, 32(12):2173 – 2182.

Kay, A. (1992). Frontogenesis in gravity-driven flows with non-uniform density gradients. *Journal of Fluid Mechanics*, 235:529–556.

Keulegan, G. H. (1957). An experimental study of the motion of saline water from locks into fresh water channels. *13th Progress Report on Model Laws for Density Currents, Rep. 5168*.

Klasa, C., Arpagaus, M., Walser, A., and Wernli, H. (2018). An evaluation of the convection-permitting ensemble COSMO-E for three contrasting precipitation events in Switzerland. *Quarterly Journal of the Royal Meteorological Society*, 144(712):744–764.

Kroonenberg, P. M. and Verbeek, A. (2018). The tale of Cochran's Rule: My contingency table has so many expected values smaller than 5, what am I to do? *The American Statistician*, 72(2):175–183.

Kuchera, E. L. and Parker, M. D. (2006). Severe convective wind environments. *Weather and Forecasting*, 21(4):595–612.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Ann. Math. Statist.*, 22(1):79–86.

Kühnlein, C., Keil, C., Craig, G. C., and Gebhardt, C. (2014). The impact of downscaled initial condition perturbations on convective-scale ensemble forecasts of precipitation. *Quarterly Journal of the Royal Meteorological Society*, 140(682):1552–1562.

Lamb, H. (1945). *Hydrodynamics*. Dover Publications.

Lapworth, A. (2005). Collision of two sea-breeze fronts observed in Wales. *Weather*, 60(11):316–318.

Lee, B. D. and Wilhelmson, R. B. (1997). The numerical simulation of non-supercell tornadogenesis. Part i: Initiation and evolution of pretornadic misocyclone circulations along a dry outflow boundary. *Journal of the Atmospheric Sciences*, 54(1):32–60.

Leutbecher, M. and Palmer, T. N. (2008). Ensemble forecasting. *Journal of Computational Physics*, 227(7):3515 – 3539. Predicting weather, climate and extreme events.

Lin, L.-Y. (2007). *Mesoscale Dynamics*. Cambridge University Press.

Linden, P. (2012). Gravity currents—theory and laboratory experiments. In Chassignet, E., Cenedese, C., and Verron, J., editors, *Buoyancy-Driven Flows*, pages 13–51. Cambridge University Press.

Liu, C. and Moncrieff, M. W. (1996). A numerical study of the effects of ambient flow and shear on density currents. *Monthly Weather Review*, 124(10):2282–2303.

Liu, C. and Moncrieff, M. W. (2000). Simulated density currents in idealized stratified environments. *Monthly Weather Review*, 128(5):1420–1437.

Liu, J.-G., Wang, C., and Johnston, H. (2003). A fourth order scheme for incompressible Boussinesq Equations. *Journal of Scientific Computing*, 18(2):253–285.

Lock, A., Edward, J., and Boutle, I. (2018). The parametrization of boundary layer processes. Unified model documentation paper 024, Met Office, Exeter.

Lombardo, K., Sinsky, E., Jia, Y., Whitney, M. M., and Edson, J. (2016). Sensitivity of simulated sea breezes to initial conditions in complex coastal regions. *Monthly Weather Review*, 144(4):1299–1320.

Lorenz, E. N. (1969). Predictability of a flow which possesses many scales of motion. *Tellus*, 21:289 – 307.

Loughner, C. P., Tzortziou, M., Follette-Cook, M., Pickering, K. E., Goldberg, D., Satam, C., Weinheimer, A., Crawford, J. H., Knapp, D. J., Montzka, D. D., Diskin, G. S., and Dickerson, R. R. (2014). Impact of bay-breeze circulations on surface air quality and boundary layer export. *Journal of Applied Meteorology and Climatology*, 53(7):1697–1713.

Marsigli, C., Boccanera, F., Montani, A., and Paccagnella, T. (2005). The COSMO-LEPS mesoscale ensemble system: validation of the methodology and verification. *Nonlinear Processes in Geophysics*, 12(4):527–536.

Marsigli, C., Montani, A., and Paccangnella, T. (2008). A spatial verification method applied to the evaluation of high-resolution ensemble forecasts. *Meteorological Applications*, 15(1):125–143.

Mason, S. J. and Graham, N. E. (2002). Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Quarterly Journal of the Royal Meteorological Society*, 128(584):2145–2166.

Mathias, L., Ermert, V., Kelemen, F. D., Ludwig, P., and Pinto, J. G. (2017). Synoptic analysis and hindcast of an intense bow echo in Western Europe: The 9 June 2014 storm. *Weather and Forecasting*, 32(3):1121–1141.

Meir, T., Orton, P. M., Pullen, J., Holt, T., Thompson, W. T., and Arend, M. F. (2013). Forecasting the New York City urban heat island and sea breeze during extreme heat events. *Weather and Forecasting*, 28(6):1460–1477.

Mercer, A. and Dyer, J. (2014). A new scheme for daily peak wind gust prediction using machine learning. *Procedia Computer Science*, 36:593 – 598. Complex Adaptive Systems Philadelphia, PA November 3-5, 2014.

Miller, S. T. K., Kelm, B. D., Talbot, R. W., and Mao, H. (2003). Sea breeze: structure, forecasting and impacts. *Rev. Geophys.*, 41:1/1–31.

Montani, A., Cesari, D., Marsigli, C., and Paccagnella, T. (2011). Seven years of activity in the field of mesoscale ensemble forecasting by the COSMO-LEPS system: main achievements and open challenges. *Tellus A*, 63(3):605–624.

Mood, A. M., Graybill, F. A., and C. Boes, D. (1974). *Introduction to the theory of statistics*. McGraw-Hill.

Mueller, C. K. and Carbone, R. E. (1987). Dynamics of a thunderstorm outflow. *Journal of the Atmospheric Sciences*, 44(15):1879–1898.

Munsell, E. B., Sippel, J. A., Braun, S. A., Weng, Y., and Zhang, F. (2015). Dynamics and predictability of hurricane Nadine (2012) evaluated through convection-permitting ensemble analysis and forecasts. *Monthly Weather Review*, 143(11):4514–4532.

Murphy, A. H. (1993). What is a good forecast? An essay on the nature of goodness in weather forecasting. *Weather and Forecasting*, 8(2):281–293.

Murphy, A. H. and Winkler, R. L. (1977). Reliability of subjective probability forecasts of precipitation and temperature. *Applied Statistics*, pages 41–47.

Mylne, K. and Roberts, N. M. (2017). A challenge for wind gust forecasting with convection-permitting models. In *EMS Annual Meeting*. European Meteorological Society. http://meetingorganizer.copernicus.org/EMS2017/EMS2017-813.pdf.

Nasr-Azadani, M. M. and Meiburg, E. (2016). Gravity currents propagating into ambients with arbitrary shear and density stratification: vorticity-based modelling. *Quarterly Journal of the Royal Meteorological Society*, 142(696):1359–1370.

Neumann, J. (1973). The sea and land breezes in the classical Greek literature. *Bulletin of the American Meteorological Society*, 54(1):5–8.

Orf, L. G., Anderson, J. R., and Straka, J. M. (1996). A three-dimensional numerical analysis of colliding microburst outflow dynamics. *Journal of the Atmospheric Sciences*, 53(17):2490–2511.

Panofsky, H. A., Tennekes, H., Lenschow, D. H., and Wyngaard, J. C. (1977). The characteristics of turbulent velocity components in the surface layer under convective conditions. *Boundary-Layer Meteorology*, 11(3):355–361.

Pantillon, F., Lerch, S., Knippertz, P., and Corsmeier, U. (2018). Forecasting wind gusts in winter storms using a calibrated convection-permitting ensemble. *Quarterly Journal of the Royal Meteorological Society*, 144(715):1864–1881.

Papanastasiou, D., Melas, D., Bartzanas, T., and Kittas, C. (2010). Temperature, comfort and pollution levels during heat waves and the role of sea breeze. *Int. J. Biometeorol.*, 54(307):5–8.

Parker, D. J. (1996). Cold pools in shear. *Quarterly Journal of the Royal Meteorological Society*, 122(535):1655–1674.

Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076.

Pedgley, D. E. (2003). The Bracknell hailstorm of 7 May 2000. *Weather*, 58(5):171–182.

Peralta, C., Ben Bouallègue, Z., Theis, S. E., Gebhardt, C., and Buchhold, M. (2012). Accounting for initial condition uncertainties in COSMO-DE-EPS. *Journal of Geophysical Research: Atmospheres*, 117(D7).

Persson, A. and Grazzini, F. (2007). User guide to ECMWF forecast products. Technical report, ECMWF, Reading.

Planchon, O., Damato, F., Dubreuil, V., and Gouery, P. (2006). A method of identifying and locating sea-breeze fronts in north-eastern Brazil by remote sensing. *Meteorological Applications*, 13(3):225–234.

Porson, A., Steyn, D. G., and Schayes, G. (2007). Formulation of an index for sea breezes in opposing winds. *Journal of Applied Meteorology and Climatology*, 46(8):1257–1263.

Raftery, A. E., Gneiting, T., Balabdaoui, F., and Polakowski, M. (2005). Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review*, 133(5):1155–1174.

Rasp, S. and Lerch, S. (2018). Neural networks for postprocessing ensemble weather forecasts. *Monthly Weather Review*, 146(11):3885–3900.

Raynaud, L. and Bouttier, F. (2016). Comparison of initial perturbation methods for ensemble prediction at convective scale. *Quarterly Journal of the Royal Meteorological Society*, 142(695):854–866.

Rio, C., Grandpeix, J.-Y., Hourdin, F., Guichard, F., Couvreux, F., Lafore, J.-P., Fridlind, A., Mrowiec, A., Roehrig, R., Rochetin, N., Lefebvre, M.-P., and Idelkadi, A. (2013). Control of deep convection by sub-cloud lifting processes: the ALP closure in the LMDZ5B general circulation model. *Climate Dynamics*, 40(9):2271–2292.

Robinson, F. J., Patterson, M. D., and Sherwood, S. C. (2013). A numerical modeling study of the propagation of idealized sea-breeze density currents. *Journal of the Atmospheric Sciences*, 70(2):653–668.

Rooney, G. G. (2018). Similarity-based approximations for the evolution of a gravity current. *Quarterly Journal of the Royal Meteorological Society*, 144(716):2302–2310.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, 27(3):832–837.

Rottman, J. W. and Simpson, J. E. (1983). Gravity currents produced by instantaneous releases of a heavy fluid in a rectangular channel. *J. Fluid Mech.*, 135:95–110.

Rotunno, R., Klemp, J. B., and Weisman, M. L. (1988). A theory for strong, long-lived squall lines. *Journal of the Atmospheric Sciences*, 45(3):463–485.

Roulston, M. S. and Smith, L. A. (2002). Evaluating Probabilistic Forecasts Using Information Theory. *Mon. Weather Rev.*, 130(6):1653 – 1660.

Schellander-Gorgas, T., Wang, Y., Meier, F., Weidle, F., Wittmann, C., and Kann, A. (2017). On the forecast skill of a convection-permitting ensemble. *Geoscientific Model Development*, 10(1):35–56.

Schwartz, C. S., Kain, J. S., Weiss, S. J., Xue, M., Bright, D. R., Kong, F., Thomas, K. W., Levit, J. J., Coniglio, M. C., and Wandishin, M. S. (2010). Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Weather and Forecasting*, 25(1):263–280.

Schwartz, C. S., Romine, G. S., Sobash, R. A., Fossell, K. R., and Weisman, M. L. (2015). NCAR's experimental real-time convection-allowing ensemble prediction system. *Weather and Forecasting*, 30(6):1645–1654.

Schwartz, C. S. and Sobash, R. A. (2017). Generating probabilistic forecasts from convection-allowing ensembles using neighborhood approaches: A review and recommendations. *Monthly Weather Review*, 145(9):3397–3418.

Scott, D. W. (1979). On optimal and data-based histograms. *Biometrika*, 66(3):605 – 610.

Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley Sons, New York, Chicester.

Seigel, R. B. and van den Heever, S. C. (2012). Simulated density currents beneath embedded stratified layers. *Journal of the Atmospheric Sciences*, 69(7):2192–2200.

Shao, J. (1993). Linear model selection by cross-validation. *J. Amer. Statist. Assoc.*, 88(422):486–494.

Sheridan, P. (2018). Current gust forecasting techniques, developments and challenges. *Advances in Science and Research*, 15:159–172.

Shin, J. O., Dalziel, S. B., and Linden, P. F. (2004). Gravity currents produced by lock exchange. *Journal of Fluid Mechanics*, 521:1–34.

Simpson, J. and Britter, R. (1980). A laboratory model of an atmospheric mesofront. *Q. J. R. Met. Soc.*, 106:485 – 500.

Simpson, J. E. (1969). A comparison between laboratory and atmospheric density currents. *Q. J. R. Met. Soc.*, 95:758 – 765.

Simpson, J. E. (1997). *Gravity Currents in the Environment and the Laboratory. Second Edition*. Cambridge University Press.

Simpson, J. E. and Linden, P. F. (1989). Frontogenesis in a fluid with horizontal density gradients. *Journal of Fluid Mechanics*, 202:1–16.

Simpson, J. E., Mansfield, D. A., and Milford, J. R. (1977). Inland penetration of sea-breeze fronts. *Quarterly Journal of the Royal Meteorological Society*, 103(435):47–76.

Smart, D., Clark, M., Hill, L., and Prosser, T. (2012). A damaging microburst and tornado near York on 3 August 2011. *Weather*, 67(8):218–223.

Smith, L. A. (2006). Predictability past predictability present. In *Predictability of weather and climate*, pages 217–250. Cambridge University Press.

Sobash, R. A., Schwartz, C. S., Romine, G. S., Fossell, K. R., and Weisman, M. L. (2016). Severe weather prediction using storm surrogates from an ensemble forecasting system. *Weather and Forecasting*, 31(1):255–271.

Steele, C. J., Dorling, S. R., von Glasow, R., and Bacon, J. (2014). Modelling sea-breeze climatologies and interactions on coasts in the southern North Sea: implications for offshore wind energy. *Quarterly Journal of the Royal Meteorological Society*, 141(690):1821–1835.

Suomi, I. and Vihma, T. (2018). Wind gust measurement techniques—from traditional anemometry to new possibilities. *Sensors*, 18(4).

Suomi, I., Vihma, T., Gryning, S.-E., and Fortelius, C. (2013). Wind-gust parametrizations at heights relevant for wind energy: a study based on mast observations. *Quarterly Journal of the Royal Meteorological Society*, 139(674):1298–1310.

Swinbank, R., Kyouda, M., Buchanan, P., Froude, L., Hamill, T. M., Hewson, T. D., Keller, J. H., Matsueda, M., Methven, J., Pappenberger, F., Scheuerer, M., Titley, H., Wilson, L., and Yamaguchi, M. (2016). The TIGGE Project and its achievements. *Bull. Amer. Meteor. Soc*, 97:49 – 67.

Taszarek, M., Brooks, H. E., and Czernecki, B. (2017). Sounding-derived parameters associated with convective hazards in europe. *Monthly Weather Review*, 145(4):1511–1528.

Tennant, W. (2015). Improving initial condition perturbations for MOGREPS-UK. *Weather and Forecasting*, 141:2324 – 2336.

Thorarinsdottir, T. L. and Johnson, M. S. (2012). Probabilistic wind gust forecasting using nonhomogeneous gaussian regression. *Monthly Weather Review*, 140(3):889–897.

Torri, G., Kuang, Z., and Tian, Y. (2015). Mechanisms for convection triggering by cold pools. *Geophysical Research Letters*, 42(6):1943–1950.

Trier, S. B., Romine, G. S., Ahijevych, D. A., Trapp, R. J., Schumacher, R. S., Coniglio, M. C., and Stensrud, D. J. (2015). Mesoscale thermodynamic influences on convection initiation near a surface dryline in a convection-permitting ensemble. *Monthly Weather Review*, 143(9):3726–3753.

van der Wiel, K., Gille, S. T., Llewellyn Smith, S. G., Linden, P. F., and Cenedese, C. (2017). Characteristics of colliding sea breeze gravity current fronts: a laboratory study. *Quarterly Journal of the Royal Meteorological Society*, 143(704):1434–1441.

Vermeire, B. C., Orf, L. G., and Savory, E. (2011). A parametric study of downburst line near-surface outflows. *Journal of Wind Engineering and Industrial Aerodynamics*, 99(4):226–238.

von Kármán, T. (1940). The engineer grapples with nonlinear problems. *Bulletin of the American Mathematical Society*, 46(8):615–683.

Wakimoto, R. M. (2001). Convectively driven high wind events. In *Severe Convective Storms*, pages 255–298. American Meteorological Society, Boston, MA.

Walsh, J. E. (1974). Sea breeze theory and applications. *J. Atmos. Sci.*, 31:2012 – 2026.

Wang, X. and Bishop, C. H. (2003). A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *Journal of the Atmospheric Sciences*, 60(9):1140–1158.

Wang, Y., Bellus, M., Wittmann, C., Steinheimer, M., Weidle, F., Kann, A., Ivatek-Šahdan, S., Tian, W., Ma, X., Tascu, S., and Bazile, E. (2011). The Central European limited-area ensemble forecasting system: ALADIN-LAEF. *Quarterly Journal of the Royal Meteorological Society*, 137(655):483–502.

Wapler, K. and Lane, T. P. (2012). A case of offshore convective initiation by interacting land breezes near Darwin, Australia. *Meteorology and Atmospheric Physics*, 115(3):123–137.

Warren, R. (2014). *Quasi-stationary convective systems in the UK*. PhD thesis, University of Reading.

Waters, A. J. and Collier, C. G. (1995). The Farnborough storm—evidence of a microburst. *Meteorological Applications*, 2(3):221–230.

Weijs, S. V., van Nooijen, R., and van de Giesen, N. (2010). Kullback–Leibler divergence as a forecast skill score with classic reliability–resolution–uncertainty decomposition. *Monthly Weather Review*, 138(9):3387–3399.

Weisman, M. L., Skamarock, W. C., and Klemp, J. B. (1997). The resolution dependence of explicitly modeled convective systems. *Monthly Weather Review*, 125(4):527–548.

Wilks, D. (2010). Sampling distributions of the brier score and brier skill score under serial dependence. *Quarterly Journal of the Royal Meteorological Society*, 136(653):2109–2118.

Wood, N., Staniforth, A., White, A., Allen, T., Diamantakis, M., Gross, M., Melvin, T., Smith, C., Vosper, S., Zerroukat, M., and Thuburn, J. (2014). An inherently mass-conserving semi-implicit semi-Lagrangian discretization of the deep-atmosphere global non-hydrostatic equations. *Quarterly Journal of the Royal Meteorological Society*, 140(682):1505–1520.

World Meteorological Organization (2008). Guide to meteorological instruments and method of observation.

Xu, Q. (1992). Density currents in shear flows—a two-fluid model. *Journal of the Atmospheric Sciences*, 49(6):511–524.

Xu, Q. and Moncrieff, M. W. (1994). Density current circulations in shear flows. *Journal of the Atmospheric Sciences*, 51(3):434–446.

Xue, M., Kong, F., Weber, D., Thomas, K. W., Wang, Y., Brewster, K., Droegemeier, K. K., Kain, J., Weiss, S., Bright, D., Wandishin, M., Coniglio, M., and Du, f. J. (2007). CAPS realtime storm-scale ensemble and high-resolution forecasts as part of the NOAA hazardous weather testbed 2007 Spring experiment. In *22nd Conf. on Weather Analysis and Forecasting/18th Conf. on Numerical Weather Prediction*. American Meteorological Society.

Xue, M., Xu, Q., and Droegemeier, K. K. (1997). A theoretical and numerical study of density currents in nonconstant shear flows. *Journal of the Atmospheric Sciences*, 54(15):1998–2019.

Zhong, Q., Hussain, F., and Fernando, H. J. S. (2018). Quantification of turbulent mixing in colliding gravity currents. *Journal of Fluid Mechanics*, 851:125–147.

Zhong, S. and Takle, E. S. (1993). The effects of large-scale winds on the sea-land-breeze circulations in an area of complex coastal heating. *Journal of Applied Meteorology*, 32(7):1181–1195.

Zhou, X., Obuchowski, N. A., and McClish, D. K. (2008). Comparing the accuracy of two diagnostic tests. In *Statistical Methods in Diagnostic Medicine*, chapter 5, pages 165–194. Wiley-Blackwell.