ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute of Environmental Engineering (IFU),
Chair of Hydrology and Water Resources Management
Institute for Atmospheric and Climate Sciences (IAC),
Chair of Dynamical Meteorology

# Tropical Cyclone Forecast Verification:

## Three Approaches to the Assessment of the ECMWF Model

**Master Thesis by**
Franziska Aemisegger

**Supervision:**
Dr. Olivia Martius (IAC)
Dr. Marc Wüest (Swiss Re)

**Advisors:**
Professor Dr. Paolo Burlando (IFU)
Professor Dr. Huw Davies (IAC)

Zürich, June 2009

*L'avenir est un présent que nous fait le passé.*

André Malraux

Franziska Aemisegger
Ch. du Grand-Praz 10
1012 Lausanne
aemisegf@student.ethz.ch

**Abstract**

Various aspects of feature specific forecast verification of tropical cyclones (TCs) in the Atlantic basin are the topic of this master thesis. The global atmospheric deterministic (operational) and probabilistic (ensemble) model suite of the European Centre for Medium-Range Weather Forecasts (ECMWF) is assessed. The forecast verification problem is approached from three different points of view, illustrating the various stakeholder perspectives behind numerical weather prediction.

To measure the performance of the ECMWF TC ensemble forecasts, methods of traditional forecast verification described in the literature and used by operational centres to assess the improvement of their forecasting system are applied in a one-dimensional storm tracking approach. A statistical evaluation of track and intensity (central pressure) forecasts of 29 hurricanes in the years 2005 to 2008 is carried out. The limited sample size of events that can be assessed for the region of interest with a consistent model setting is a general problem in this work. Nevertheless, some interesting tendencies and hypotheses can be formulated. Positive skill is found for long forecast ranges in intensity and track. A mean bias in central pressure of +21.7 hPa for the ensemble prediction system (EPS) and +15.3 hPa for the operational forecast is detected. The magnitude of the average bias decreases with improvements in resolution. For the current model set up, the mean bias is +18.1 hPa for the EPS and +11.3 hPa for the operational forecast. Furthermore, the EPS is found to be overconfident. Various effects are studied, like differences in forecasts over land or sea, as well as storm intensity influence on the forecast error statistics.

Questions concerning the ability of the ECMWF deterministic model to represent TC features in structure and dynamics call for a multidimensional quality assessment framework. To this end, a novel verification measure based on a series of feature specific components is adapted to tropical cyclones. Such an object-based verification approach has a complementary potential to the storm tracking procedure and is targeted at revealing specific problems and inabilities of the model to represent certain processes. The proposed framework is flexible and can be adapted to user-specific needs. The resulting quality measure is easy to communicate, while still containing a maximum of process-based information. As opposed to traditional grid point based verification methods, it avoids double-penalty effects by clearly separating the different properties into individual error components.

A further evaluation of ECMWF TC forecasts is performed, which focuses on the user needs. The insurance sector is particularly interested in meteorological forecasts of TCs for their short range management of potential losses. This knowledge advantage is important in the context of a tight business competition, potentially very profitable market and high level exposure. Insurance loss predictions are performed in collaboration with Swiss Re for 11 chosen case studies of particularly devastating events in the last four hurricane seasons. The losses are generally strongly underestimated, when using the ECMWF raw data as a meteorological input. Bias correction of pressure forecasts improve the results substantially. The use of a probabilistic forecasting system for loss predictions, when compared to a single deterministic forecast, helps to estimate the system inherent uncertainty of the meteorological input.

The complementary multi-viewpoint on the forecast verification problem of the ECMWF model with respect to TCs is found to be very valuable for a holistic quality assessment. The adopted threefold perspective leads not only to a better overview of the model performance, but also to an improvement of the methods used to assess it.

# Contents

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| ACE | accumulated cyclone energy |
| CatMos | catastrophe loss modelling system of Swiss Re |
| cdf | cumulative distribution function |
| CLIPER | Climatology and persistence track forecasting system |
| COSMO | MeteoSwiss operational local area numerical weather prediction system |
| CTL | ECMWF control run in the EPS (low) resolution deterministic forecast |
| DSHIFOR | SHIFOR with decay component for TCs over land |
| (D)SHIFOR | SHIFOR error data for hurricane events in 2005, DSHIFOR for events in 2006-2008 |
| ECMWF | European Center for Medium-Range Weather Forecasts |
| EIA | U.S. Energy Information Administration |
| ENS | Ensemble forecast |
| EPS | Ensemble Prediction System |
| ETHZ | Swiss Federal Institute of Technology in Zürich |
| GCM | General Circulation Model |
| GDP | Gross Domestic Product |
| GFS | Global Forecasting System by NWS |
| GOES | Geostationary Operational Environmental Satellite |
| IKE | Integrated Kinetic Energy |
| IPCC | Intergovernmental Panel on Climate Change |
| L91 | number of vertical levels in ECMWF model |
| NCDC | U.S. National Climatic Data Center |
| NCEP | U.S. National Center for Environmental Prediction |
| NHC | U.S. National Hurricane Center |
| NWP | numerical weather prediction |
| NWS | U.S. National Weather Service |
| OPER | ECMWF operational forecast with the high resolution deterministic model |
| pdf | probability density function |
| SAL | Structure-Amplitude-Location, object-based verification concept |
| SHIFOR | Statistical hurricane intensity forecast model |
| SV | singular vectors |
| Swiss Re | Swiss Reinsurance Company |
| T799 | triangular truncation to 799 waves, spectral horizontal resolution of the ECMWF model |
| TC | tropical cyclone |
| U.K. | United Kingdom |
| U.S. | United States |
| WMO | World Meteorological Organisation |
| 4D-Var | ECMWF data assimilation system |

# List of Symbols

| | | |
|---|---|---|
| $A$ | - | amplitude component of SAL object-based verification measure |
| $AT$, $ATE$, $ATB$ | km | along track component, error, bias |
| $CP$, $CPE$, $CPB$ | hPa | central pressure of a TC, error, bias |
| $CT$, $CTE$, $CTB$ | km | cross track component, error, bias |
| $DPE$ | km | direct position error |
| $DX$, $LONE$, $LONB$ | km | longitudinal component, error, bias |
| $DY$, $LATE$, $LATB$ | km | latitudinal component, error, bias |
| $IKE$ | TJ | integrated kinetic energy |
| $L$ | - | location component of SAL object-based verification measure |
| $M$ | - | number of subobjects in the TC object of interest |
| $N$ | - | number of grid points in a TC object |
| $R_c$ | km | average radius of a TC |
| $SS$ | - | skill score, between 1 and 0 |
| $Sc_0$ | - | score of the benchmark forecast |
| $Sc_1$ | - | best possible score |
| $Sc$ | - | score of the model to be verified |
| $S1, 2, 3$ | - | structure component(s) of SAL object-based verification measure |
| $U$ | m/s, kt | azimuthal wind |
| $V$ | $m^3$ | volume of a TC object serving the computation of $IKE$ |
| $V_{\text{adv}}$ | m/s | advection by the steering flow |
| $Vs_o$ | - | scaled volume of a wind subobject $o$ in a TC object |
| $V_M$ | - | weighted mean of all subobjects' scaled wind volume |
| | | |
| $a$ | $m^2$ | surface of a TC object |
| $\text{cov}(x, \hat{x})$ | | covariance of observation $x$ and forecast $\hat{x}$ |
| $o$ | - | wind subobject in TC pressure object |
| $nx, ny$ | - | number of grid points in subobjects inside a TC pressure object in longitudinal and latitudinal direction |
| $p$ | hPa | pressure |
| $p(x)$ | - | probability distribution of observations $x$ |
| $\hat{p}_f,\ q(x)$ | - | probability distribution of forecasts |
| $\vec{r}$ | $^\circ$ | position vector of a TC |
| $x,\ \hat{x}$ | | observation, forecast |
| $t$ | s | time |
| $\text{var}(x)$ | | variance of observation $x$ |
| | | |
| $\rho$ | - | Pearson correlation coefficient |
| $\rho_a$ | $kg/m^3$ | density of air |
| $\omega$ | - | weights between 0 and 1 |
| $\lambda,\ \phi$ | $^\circ$ | latitude, longitude |

# Chapter 1

# Introduction

Tropical cyclones (TC) are amongst the most impressive weather systems of Earth's atmosphere, not only as extraordinary natural phenomenon but also in terms of damage they can cause. In the USA, in the region of the Caribbean Sea as well as in many Asian countries, TCs are ranked as the costliest and deadliest natural disaster. Damage figures from TCs have been increasing within the last few decades. Climate change might have played a role in this trend. The latest IPCC (Intergovernmental Panel on Climate Change) assessment report (IPCC, 2007) mentions that there is observational evidence for an upward trend in intense TC activity in the North Atlantic and quantifies the probability that there has been an increase in strong hurricanes since 1970 in some regions as likely ( 66% chance). Furthermore, the growth of both population and wealth have made the local economies more vulnerable to natural disasters such as TCs (Pielke and Landsea, 1998).

Costs of weather-related natural disasters in general have been rising in recent years. However, the capacity of the insurance industry to absorb the induced losses has been declining in the time window 1980-2004, shifting more of the burden to governments and individuals (Mills, 2005). Changes in nature, magnitude, timing, duration or location of hazards challenge the insurance system, which bears great responsibility in elaborating proactive measures in order to face these extreme events that individuals cannot manage independently.

In order to assess natural hazard risks, when facing rapid global changes such as climate change, techniques like probabilistic catastrophe loss modelling have been developed and are widely accepted as an appropriate way to manage catastrophe risk (RMS, 2008). Instead of relying only on historical event frequencies and impacts to estimate future potential losses, probabilistic risk assessment models simulate many physically realistic, probabilistic scenarios, thus extending historical datasets. These comprehensive probabilistic event sets provide a reliable picture of exposure of a certain region to natural hazard risks. In addition to risk and vulnerability characterisation, real-time prediction and early warning systems are needed in order to face an imminent hazard and minimise losses (Hoff et al., 2003). For the insurance industry, early loss estimates are important in order to allocate capital and be prepared for last-minute trade of risk transfer solutions. Even more so, in the case of high impact events, like TCs, which have potential to cause damage as high as the Gross Domestic Product (GDP) of certain regions.

The performance in forecasting such extreme weather has improved significantly in the last few years with the continuously evolving computational possibilities. The European Centre for Medium-Range Weather Forecasts (ECMWF) computes twice a day 51 so-called ensemble predictions of the global atmospheric state as a part of its operative forecasting system. These ensembles represent atmospheric conditions, which are considered to be feasible given the uncertainty in the initial conditions and the uncertainty in the numerical and physical model. The first uncertainty stems from the limited number of observations of the variables and their representativeness, the latter from limited model grid and resulting sub-grid parametrisations. The motivation for the set up of an ensemble prediction system (EPS) is in essence to gain information in advance about the error of a forecast. The ensemble of forecasts can thus be interpreted as a probabilistic prediction of future weather (Woods, 2006) and can be used as a quantitative tool for risk assessment. According to ECMWF, the potential economic value of the EPS is much higher, than that of a forecasting system based only on a single deterministic prediction (Richardson,

2000). Indeed, forecast value depends not only on the quality of the forecasting system but also on the weather sensitivity of the user (Jolliffe and Stephenson, 2003). Thus, conceptually, the main benefit of probabilistic forecasts over deterministic ones lies in its adaptability to the specific needs of the different users.

Forecast verification implies testing the implicitly made hypothesis that a forecast of a future state using a certain model provides an added value compared to the case, in which no information is available. Weather forecast verification was first introduced in a precise and formal framework by Finley (1884) in the end of the 19th century, a few decades after the beginning of operational weather forecasting in the USA and Europe. After half a century of very modest activity in this research field the number of publications on the topic rapidly increased with the advent in the 1950s of numerical weather forecasting and the subsequent multiplication of available products (Muller, 1944a,b,c). Brier and Allen (1951) set up a three-way classification of the motivations behind forecast verification:

- **Administrative reason:** illustrates the need for a numerical measure of the quality of forecasts in order to assess the improvement of the forecasting system over time, justifying funding for research, equipment and providing a basis for the elaboration of resource investment strategies.
- **Scientific reason:** in order to improve the forecasting model in its structure and formulation, the strengths and weaknesses of the forecasts have to be analysed. Forecast verification can also lead to scientific questioning about the underlying physical processes, which if investigated, may lead to an improved understanding and thus in turn better models.
- **Economic reason:** includes the user-side point of view and the need for tailored verification schemes, aiming at satisfying specific interests and at simplicity in communication.

The selected verification scheme should be informative and specific in answering the question of interest (Jolliffe and Stephenson, 2003). The three aims of forecast verification formulated by Brier and Allen (1951) will all be explored in this work. The point of view of operational forecasting is adopted in the part on traditional forecast verification of the ECMWF TC ensemble forecasting system using a storm tracking approach. An object-based verification technique is proposed that can be attributed to the side of a scientific investigation of the performance of the model in representing the weather system. A user-oriented assessment is undertaken with the insurance loss predictions using ECMWF data. For each of these TC verification aspects specific questions were formulated:

1. **Storm tracking verification:**
   How well does the ECMWF TC ensemble forecasting system reproduce the observed data in terms of track and intensity? Can the model capture the magnitude of the system inherent uncertainty?
2. **Object-based verification:**
   How could we assess the model in a way that might answer following questions: *How does a TC look like in a global atmospheric high resolution deterministic model like ECMWF? Which properties are represented? What are the reasons for potential inabilities to render the feature accurately?*
3. **User-oriented verification:**
   What is the performance of the ECMWF TC ensemble forecasting system, when it is coupled to a catastrophe modelling system in order to estimate insurance losses? What is the added value for a reinsurance company like Swiss Re to use such a system instead of a deterministic forecast like the one provided for free by the U.S. national warning centre (National Hurricane Center, NHC)?

In this master thesis the traditional question of the benefit of ensembles, when compared to deterministic predictions is raised in the specific case of TC forecasts used for insurance loss predictions. The primary aim of this project is to evaluate the possibility of using ECMWF TC ensemble forecasts for predictions of induced damages and insured losses. The added value of using such a system instead of a deterministic forecast is characterised. The focus is laid upon northern Atlantic hurricanes of the last couple of years (2005-2008). In a first step the ECMWF TC ensemble prediction system is compared to observations of hurricanes for a set of recent events. For this purpose traditional verification measures are used and the model is assessed in a storm tracking framework. In a second step a new object-based quality measure is proposed, which aims at pointing out structural and dynamical deficiencies of TC representation in

the ECMWF model. A quality measure of this kind is desirable for a more holistic description of model performance. Finally, the quality of the resulting insurance loss predictions is assessed using a chain modelling approach.

In the following chapter, some theoretical and fundamental aspects of the physical structure and the dynamics of TCs are presented. A review of possible ways to quantify the destructive potential of TCs is made and the main attributes of forecasts are defined. Then, the thematic of forecast verification is addressed. Chapter 3 presents the different ECMWF forecasting systems, of which the output data is used in this work. The observational and forecast data from the U.S. National Hurricane Center (NHC) is described and the climatological and persistence track and intensity benchmark forecasts are introduced. A description of the damage assessment model CatMos from Swiss Re closes the data chapter. In chapter 4 the chosen hurricane events for the different approaches are introduced. In chapter 5, the storm tracking comparison procedure is explained, the framework of the object-based verification is presented and the methods for loss modelling with Swiss Re's catastrophe loss modelling system using ECMWF forecasts as an input are described. Chapter 6 relates and discusses the results. Finally, in chapter 7 the conclusions are drawn.

# Chapter 2

# Theories and Concepts

## 2.1 Fundamentals of Tropical Cyclones

TCs can be defined by a closed circulation around a centre of low pressure, which is induced by latent heat release in the core (Emanuel, 1991). This driving mechanism represents the main difference with extratropical cyclones, which draw their energy from the meridional temperature gradient (Wallace and Hobbs, 2006). TCs have an approximately symmetric structure with a wind reinforcement of the right side of their track in the Northern Hemisphere (Shapiro, 1983). The centre of the TC is characterised by a cloud-free warm core eye, in which the air descends. Convection is induced by solar heating and evaporation of ocean air. Thus air convergence takes place at the surface and divergence at the top, on altitudes usually slightly above 10 km (Emanuel, 2009). The centre of the storm is surrounded by a rapidly rotating wall of clouds. This part of the TC causes most wind damages. Behind the eyewall successive Cumulonimbus towers, merging together into a Cirrus overcast at high altitudes, induce heavy rain at the earth's surface (Houze, 1993). These so called rain bands are separated by gaps with only light rain and no wind.

The extremely low sea level pressure in the centre of the storm is due to the low density of the (warm) overlying air (Wallace and Hobbs, 2006). The azimuthal wind field is in cyclostrophic balance, the centripetal acceleration of the circulation around the low pressure centre being several orders of magnitude larger than the Coriolis force.

The dimensions of TCs are very variable. A large storm is not necessarily indicative of high intensity. But its size still has an influence on the observed damage as will be seen in section 2.2. The diameter of a storm being identified by the extent of the upper level Cirrus cloud cover can vary between 50 km and 1200 km. The size of the eye can amount up to 200 km in diameter (Merill, 1984).

TC naming is closely related to the intensity classification. TCs found in the North Atlantic with wind speeds exceeding 120 km/h are called hurricanes and given a proper name from pre-existing lists, alphabetically orderer and alternating feminine and masculine forenames. This practice is primarily done for warning purposes, in order to avoid confusion between basins if different storms occur simultaneously (`http://www.nhc.noaa.gov`).

There are 6 main formation ingredients for TCs (Lohmann, 2009):

1. Warm ocean water with temperatures above 27°C
2. Potentially unstable atmosphere to allow moist convection
3. Moist mid troposphere
4. Latitudes equatorwards from the baroclinic zone and 5°C polewards in order to have a strong enough Coriolis effect to induce cyclonic circulation
5. Pre-existing disturbance with sufficient vorticity and convergence
6. Low vertical shear in wind

Dissipation occurs if any of the above conditions is not fulfilled any more, in particular due to transition into the extratropics, landfall, stationarity over the ocean inducing significant water cooling, too cold waters or if the cyclone enters a zone with strong vertical wind shear.

## 2.2 Tropical Cyclone Destructive Potential

The Saffir-Simpson hurricane damage potential scale is the official scale used by the U.S. National Weather Service (NWS) to rate the size, intensity and destructive potential of North Atlantic hurricanes (Longshore, 2008). Developed by the construction engineer Robert Simpson, who was director of the NHC at that time and the meteorologist Herbert Saffir in the 1970ties, this scale is used for sustained winds above the uppermost Beaufort scale force ($>$ 120 km/h), thus exceeding intensities reached by simple tropical depressions and tropical storms. The Saffir-Simpson scale was primarily designed for the NWS to alert civil defence agencies to the destructive potential of a hurricane that lies within 72h of landfall (Longshore, 2008). It is a simple scale, which is easy to communicate to the general public as can be seen from its definition summarised in table 2.1.

**Table 2.1:** *Saffir-Simpson scale categories of hurricane destructiveness. Adapted from Simpson and Riehl (1981).*

| Cat. | Pressure at center [hPa] | Wind [km/h] | Surge [m] | Damage |
|---|---|---|---|---|
| 1 | above 980 | 120 - 153 | 1.2 - 1.5 | Minimal |
| 2 | 965 - 980 | 154 - 177 | 1.8 - 2.4 | Moderate |
| 3 | 945 - 965 | 178 - 209 | 2.7 - 3.7 | Extensive |
| 4 | 920 - 945 | 210 - 249 | 4.0 - 5.5 | Extreme |
| 5 | below 920 | 210 - 249 | 5.5 + | Catastrophic |

As a direct consequence of its generality, simplicity and conciseness its adequacy to describe destructive power of a hurricane has been intensely discussed in the last few years. Suggestions for new warning measures emerged in the last few years like the integrated kinetic energy (IKE) (Powell and Reinhold, 2007), the accumulated cyclone energy ACE (Bell et al., 2000), Emanuel (2005)'s index of the potential destructiveness of hurricanes based on total dissipation of power integrated over the lifetime of the cyclone as well as hurricane intensity and hazard indices (Kantha, 2006). One of the fundamental critiques on the Saffir-Simpson scale is that it only depends on a punctual measure of intensity and gives no information about the storm size (Powell and Reinhold, 2007). The storm development as well as its size are however highly relevant for estimating the storm's damage potential. Another point of critique is that two very different kinds of damage are confounded in one single scale, namely the wind induced damage and the one induced by storm surge. This problem became very pronounced in the case of last year's hurricane Ike. At the time of landfall the hurricane winds were only Category 2, however the storm surge in Galveston was equivalent to a Category 4 to 5. Following a report by the NHC (NHC, 2009) many residents would not evacuate, because the storm was only ranked as a Category 2-3. Other similar examples are mentioned and based on these, the NHC experimentally changed the Saffir-Simpson scale definition for the hurricane season 2009, removing the storm surge effects. Problems of this type illustrate the difficulty to set up a scale that is scientifically consistent, easy to communicate and does not cause any problems of liability.
Kinetic energy is relevant to the wind destructive potential, because it scales with the wind load acting on a structure (ASCE, 2005) as well as with the storm surge and waves generated by the shear stress of the wind on the ocean surface (Powell and Reinhold, 2007). Mahendran (1998) found that the damage induce by TC winds depended on the radius of maximum wind, the storm translation speed, central pressure and surface maximum wind gust. Liu et al. (2007) showed that hurricane-induced wave fields are heavily influenced by storm asymmetry, translation speed, intensity as well as background winds. Li and Rego (2009) identified the storm's forward motion as the most important factor in the wave generation.

The main physical parameters that are known from the above literature to influence TC induced damage may be summarised in a local maximum intensity information, translational speed and an indication on the surface wind structure. Integrated kinetic energy depends on both, the surface wind maximum and on the surface wind structure. The translation speed of the storm is however not explicitly present. Powell and Reinhold (2007) propose the following formulation for the integrated kinetic energy $IKE$:

$$IKE = \int_V \frac{1}{2}\rho_a U^2 dV \tag{2.1}$$

| | | |
|---|---|---|
| where | $V$ | volume of object with 1m thickness at 10 m above ground (sea) [$m^3$] |
| | $\rho_a$ | density of air ($\sim$ 1 kg/$m^3$) |
| | $U$ | wind velocity at 10 m above ground (sea) [m/s] |

In Powell and Reinhold (2007) $IKE$ is computed from a gridded wind field over a storm-centred 8° latitude/longitude domain, extending 1 m in the vertical and centred at the 10 m official wind measurement level. In Maclay et al. (2008) kinetic energy is integrated over a disk with constant radius and depth. Thus computed values are dependent on the employed method. Usually values of the order of magnitude of 1 TJ are obtained for hurricanes.

The destructive effect of a wind load acting on a structure is a highly non stationary process. Once a part of the building envelope fails the damage can rapidly increase. Especially when the storm intensity is not exceptionally high, the duration of wind loads and thus the translation speed of the storm become important. In order to account for this Powell and Reinhold (2007) propose to segregate $IKE$ into low, moderate and high wind speed ranges and multiply these by a representative damage factor.

The $IKE$ concept is difficult to implement as a warning scale, especially because kinetic energy is not a conservative property of the storm (Simpson and Saffir, 2007) and thus may fail to predict rapid and significant changes. However, it is very well suited for a loss-oriented verification procedure, as it relates the processes generated by dynamic sources to the character, magnitude and scale of the wind and wave damage.

## 2.3 Theoretical Aspects of Forecast Verification

### 2.3.1 Attributes of Forecasts

Until the mid-1980ties the conventional approach to forecast verification involved the reduction of information from a set of forecasts and observations into a number of measures (Jolliffe and Stephenson, 2003). The forecasting performance was evaluated based on the correspondence between forecasts and observations. The focus was laid on general aspects of forecast quality like accuracy, association or skill. This "measure-oriented" verification approach lead to the development of a high variety of scores like the root mean square error in the case of continuous variables, the hit rate, the false alarm rate, the Heidke skill score in the case of binary events (based on occurrence/non-occurrence or obtained by applying a threshold to continuous variables). All these scores were plausible, however all had some weaknesses. Murphy and Winkler (1987) proposed a general framework for so called diagnostic verification involving the use of the joint distribution of forecasts and observations. This framework provided a basic theoretical background to guide the selection of appropriate measures for particular verification problems and prepared the ground for systematic discussions on the properties of particular measures.

The "distribution oriented approach" sees forecast verification as the process of assessing the statistical characteristics of the joint distributions of forecasts and observations. In the case of probabilistic forecasts four main attributes can be defined following Murphy (1973). Probability forecasts should be able to reliably distinguish among situations for which the probability distribution functions (pdf) of the corresponding verifying observation are distinctly different.

In other words, a probabilistic forecasting system is **reliable** if the conditional pdf of the verifying observations $p(x|\hat{p_f} = q)$ given any forecast pdf $q(\hat{x})$ is equal to the latter pdf (Jolliffe and Stephenson, 2003):

$$p(x|\hat{p_f} = q) = q(\hat{x}) \quad \forall \; q(\hat{x}) \tag{2.2}$$

A system is said to have **resolution** if it is able to resolve the forecast problem in a probabilistic sense (Jolliffe and Stephenson, 2003). If resolution describes the variability of the *observed* pdf associated with different forecast scenarios around the climatological pdf, **sharpness** measures the variability of the *forecast* pdf around the climatological pdf. In the case of a perfectly reliable forecast, the forecast probability values are identical to the observed frequencies. Thus, in the latter case, sharpness and resolution describe the same forecast property. A fourth characteristic, is **uncertainty** of the underlying observations (i.e. the mean spread of the observation marginal distribution). This is an important aspect of a forecasting system, over which, however, the forecaster has no control.

All the above mentioned properties of forecasts cannot be verified by one single measure. Different measures, emphasising different aspects and attributes of forecast performance should be employed (Murphy and Winkler, 1987). In this study, only verification measures for continuous variables and summary statistics were applied. Verification scores using binary events definitions (i.e. reliability diagram, Brier skill score) or verification based on decision probability thresholds were not used here. This choice is partly due to time limits and partly because decision related verification techniques are based on much preciser situations (e.g. the definition of coastal breakpoints, meaning points on the coast at which the probability of landfalling is computed) as in Santos et al. (2009)) than the one defined as being the scope of this study, namely to assess ECMWF TC ensemble forecasts in the North Atlantic region.

### 2.3.2 Definition of Accuracy, Association and Skill

A verification score measures the relative quality of different forecasts. It is thus a function of the forecast and the observed value. **Accuracy** is a measure of correspondence between individual pairs of forecasts and observations. The mean absolute error is an example of a measure of accuracy. It can be summarised by the average distance of forecasts and observations, characterising the magnitude of the random error made by the forecasting system and depends on the bias, the resolution and uncertainty attributes (see section 2.3.1). The bias is ameasure of reliability and represents a systematic type of error made by the forecasting system. **Association** is the overall strength of the relationship between pairs of forecasts and observations. A measure of linear association is given by the correlation coefficient.

The quantification of the relative forecast quality (**skill**) involves the definition of an unskillful benchmark forecast. Climatology, persistence, a random walk forecast or a combination of these is generally used as a reference. Such a comparison helps to take into account the non-stationarities in the system to be forecast and can for example explain improved scores during specific periods, when the atmosphere is in a more persistent state (Jolliffe and Stephenson, 2003). A **skill score** allows to compare an absolute forecast quality score (e.g. measure of accuracy) with the score of a trivial forecast and is usually based on the following expression:

$$SS = \frac{Sc - Sc_0}{Sc_1 - Sc_0} \tag{2.3}$$

| | | |
|---|---|---|
| where | $SS$ | skill score of the forecast |
| | $Sc_0$ | score of the benchmark forecast |
| | $Sc_1$ | best possible score |
| | $Sc$ | score of the model to be verified |

Generally skill scores lie in the range 0 to 1, with 1 being the best score.

### 2.3.3 Quality Measures for Ensemble Forecasts

One way to measure the reliability (or statistical consistency) of an ensemble forecast is to compute the ensemble mean error $m$ and spread $s$. The verifying analysis is statistically indistinguishable from the ensemble members if the ensemble mean error is equal to the mean distance of the individual members from their mean (i.e. **spread** or ensemble standard deviation, Buizza (1997) ). For perfectly reliable forecasting systems, spread can also be considered as a measure of resolution (Jolliffe and Stephenson, 2003). The following terminology is used if the mean error and the spread are not equivalent:

$$m > s \rightarrow \text{EPS is underdisperdive, the model is overconfident}$$
$$m < s \rightarrow \text{EPS is overdisperdive, the model is underconfident}$$

This property ensures that on average the range of forecast values given by the ensemble members contain the observation.

The ECMWF TC forecasting system uses the same technique for generating each member of the ensemble (same NWP model and same initial perturbation technique). The ensemble members are thus assumed to be equally likely and can be considered independent realisation of the same random process. This assumption has to be verified in order to be able to define forecast probabilities for specific values as for example if strike probabilities are used or if predictions of insurance loss distributions are computed.

### 2.3.4 Object-Based Forecast Quality Verification

Object-based verification has emerged in the last few decades as a means of better capturing deficiencies of numerical weather predication (NWP) models in representing small scale ($< 100$ km) complex structures. An overview of the literature of the last few years can be found in Wernli et al. (2008). This new approach, which took on various forms, emerged from the necessity to address the "double penalty" problem of grid-point based error measures. The latter would not reward a perfect forecast of a precipitation intensity pattern, which is slightly shifted in terms of location. The aim of the object-based approach is to identify the nature and scale of the error through a more differentiated rating scheme. It can be seen as a way of getting closer to the subjective visual judgement of forecast accuracy, without loosing the quantitative and summarising aspect of a numerical score.

# Chapter 3

# Data

TC track and intensity forecasts are provided by many different institutions, which use different types of models. These models vary in their complexity, the physical and mathematical principles, on which they are based and the computational time they require. A complex numerical model (e.g. numerical weather prediction (NWP) model) is computationally more expensive than a simple linear regression between storm behaviour and certain relevant atmospheric variables. An overview of different models for TC forecasting available today and used at the NHC is given in Rhome (2007).

In this study, outputs of one of the most sophisticated and computationally expensive global models are used, namely the ones from the ECMWF atmospheric model. From a scientific point of view, this model offers the possibility to investigate the specific TC forecasting system in the framework of the synoptic atmospheric dynamics, in which it is embedded. From a more applied perspective, a reinsurance company like Swiss Re is interested in medium-range (3-4 days) forecasts of intensity and track, including an estimation of the uncertainty of the forecasts. The ECMWF TC ensemble prediction system suits these requirements very well.

Furthermore, ECMWF has been a reliable modelling partner for ETH since years. MeteoSwiss, which is a close research associate of the Institute for Atmospheric and Climate Sciences at ETH, collaborates with ECMWF and nests its regional COSMO model into the ECMWF prediction. The advantage of a global model over a nested limited area model of the tropical Atlantic region is firstly the availability of good quality ensemble forecasts and secondly the use of one and the same model formulation for different regions in the world. The second argument is not only important from a global company point of view. This also allows comparative studies on model performance in different regions to be done and may provide interesting hints of differences in the dynamics.

## 3.1 European Centre for Medium-Range Weather Forecasts (ECMWF) Data

The European Centre for Medium-Range Weather Forecasts (ECMWF) as an international organisation supported by many European states, provides state-of-the-art weather forecast data and products. Details on the centre's history can be found in Woods (2006), which depicts an interesting summary of the political backgrounds of the centre's foundation and its activity. In this project different ECMWF products were used. The ECMWF TC ensemble forecasts were verified against NHC best track data (see section 2.2.2 below). The ECMWF TC ensemble forecasts consist of both deterministic and probabilistic information on movement and intensity of individual TCs around the world (ECMWF[1]). The same forecasting system was used as an input for insurance loss estimations. For the object-based forecast verification, deterministic forecast fields were used and the analysis fields were selected as verifying observations.

[1] http://www.ecmwf.int/services/dissemination/3.1/TROPICAL_CYCLONE_trajectory_forecast_products.html#pgfId-385560

### 3.1.1 Data Assimilation System

Global data assimilation systems such as the one ECMWF operates, use a large amount of data from various instruments and observation times and interpolate them onto a regular grid applying an automatic objective analysis. In order to obtain a spatially and temporally coherent description of the actual state of the atmosphere, they run a general circulation model (GCM) in hindcast mode (Schär, 2008). The ECMWF data assimilation system is based on a 4D VAR data initialisation technique, which allows the assimilation of observations that are distributed over a given time interval (e.g. satellite data which are not available at synoptic times) as can be seen in figure 3.1. The initial conditions obtained from a 4D-VAR data assimilation system represent a best-fit of the forecast to the observations within the assimilation time interval (Kalnay, 2003). This means that a perfect model has to be assumed. The data assimilation cycle is run every six hours at ECMWF. A detailed description and comparison of the existing data assimilation techniques can be found in Kalnay (2003).

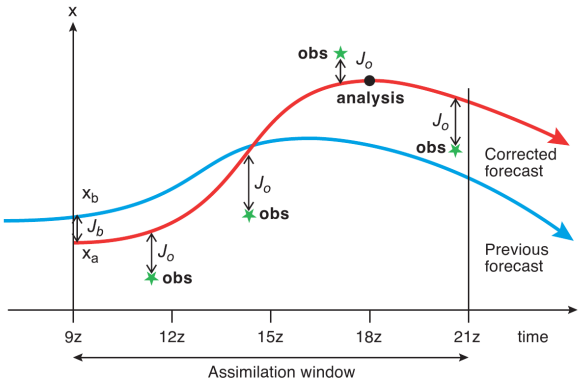



**Figure 3.1:** *Conceptual illustration of the 4D-VAR analysis done by ECMWF (Source: Persson and Grazzini (2005)).*

### 3.1.2 Deterministic Atmospheric Forecasts

The ECMWF deterministic, operational general circulation model is a hydrostatic spectral model, where the linear terms are triangularly truncated to 799 waves (T799), which corresponds to a horizontal grid resolution of 25 km at equator. The spectral horizontal model formulation transforms the variation of the model variables with latitude and longitude into a series of waves. The vertical coordinate system is a hybrid 91 levels system, which uses terrain following sigma coordinates in the boundary layer and purely isobaric vertical coordinates in upper, tropopause-near levels (see f. ex. Kalnay (2003) for details on discretisation techniques). The forecasts are issued up to 10 days, twice daily.

The ECMWF deterministic model is constantly updated to new standards in atmospheric modelling and all changes done in the period of interest 2005 to 2008 can be found in the ECMWF technical documentation and ECMWF model cycle reports[2]. One important change, however has to be pointed out, namely the model resolution change on the 1st of February 2006 from a T511 (40 km), 60 levels to a T799 (25 km), 91 level global model (ECMWF, 2007). This was a significant increase in resolution of the medium-range forecast system, especially for the representation of relatively small scale features like TCs. A strong hurricane's eye with dimensions above 100 km in diameter can now potentially be resolved by the deterministic model, following a rule of thumb, indicating that a process can be represented, when it can be characterised by more than a minimum of four model cells.

### 3.1.3 Ensemble Prediction System

The ECMWF TC EPS produces 51 ensemble members from two forecast runs per day using a model resolution of T399L62 (50 km transform grid spacing at equator, 62 pressure levels). Before the model resolution update in February 2006, the EPS model resolution was T255L40 (80 km). The 51 ECMWF ensembles comprise 1 control run using the ECMWF analysis at the EPS resolution as initial condition and 50 ensemble perturbations, which take into account the uncertainty in initial conditions and in parametrised processes.

[2] `http://www.ecmwf.int/products/data/technical/model_id/index.html`

The reason for conducting ensemble forecasts can be well illustrated by the following statement of Jules-Henri Pointcaré in 1890, who anticipated modern chaos theory (Lorenz, 1963):

> "*A tenth of a degree more or less at any given point, the cyclone will burst here and not there, and extend its ravages over districts it would otherwise have spared. If* [*meteorologists*] *had been aware of this tenth of a degree, they could have known it beforehand, but the observations were neither sufficiently comprehensive nor sufficiently precise, and that is the reason why it all seems due to the intervention of chance.*" (Poincaré, 1908)

Since the initial conditions can never be measured with infinite precision, rapid non-linear growth in initial errors fundamentally limits our ability to forecast the behaviour of the atmospheric system (Thompson, 1957). The core of an EPS is the generation of perturbations to the initial analysis used to start the forecast. The perturbations are not generated at random, as it would be the case for example with a pure Monte Carlo technique. They are a combination of 25 modes, which have the largest impact on the forecast for the Northern Hemisphere in the short-range (Molteni et al., 1996). ECMWF uses the singular vector (SV) technique, which is based on a sampling of the directions in the phase space that are characterised by the maximum amplification rate (Magnusson et al., 2009). The SVs are obtained by solving an eigenvalue problem, in which the total energy growth is maximised for a 48 h forecast (Persson and Grazzini, 2005). In an ideal set-up of initial perturbations, the differences obtained are a measure for the predictability of synoptic scale atmospheric flow, or in other words of the chaos inherent in the corresponding atmospheric conditions (see figure 3.2). Thus, an EPS provides an a priori information on the reliability of forecasts.

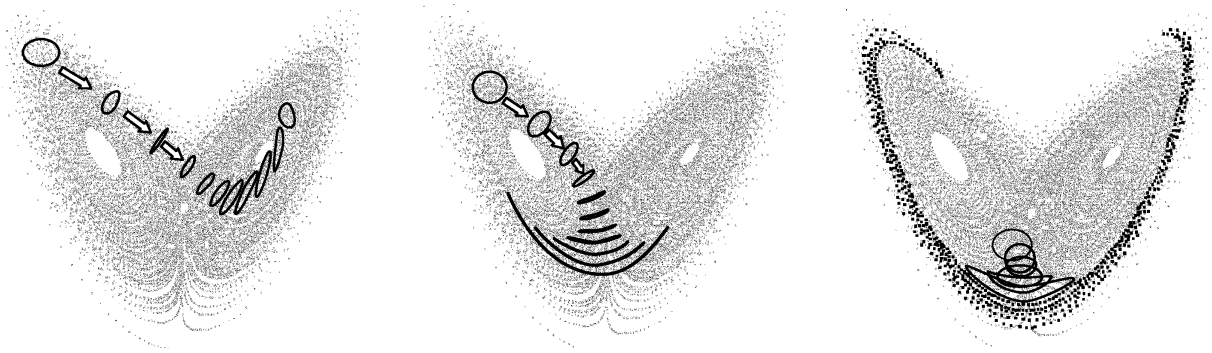


**Figure 3.2:** *Illustration of dependence of predictability on starting conditions after Buizza et al. (1999). Points represent the Lorenz Attractor (phase space), bold ellipses indicate initial regions of the same uncertainty. Ensembles of the figure on the left show consistency during the whole time of evolution. The development is predictable for a certain time before the attractor splits and predictability is lost in the middle. On the right, the predictability completely breaks down shortly after initialisation.*

### 3.1.4 Tropical Cyclone Forecasts

The ECMWF TC tracker has been introduced as an operational application in October 2004 (Van der Grijn et al., 2004). The implementation of 4D-Var data assimilation and the activation of satellite observations like seawinds from QuickSat (2002) as well as dropsondes (1999) in the assimilation system have improved the ECMWF analysis fields. Increase in horizontal and vertical resolutions (e.g. Feb. 2006) as well as changes to the microphysics and convective schemes have impacted the quality of TC tracking.

#### Track Forecasting System

TC features are followed by the ECMWF TC tracker, only if first NHC observations are already available in the 6 hours around the analysis time (Van der Grijn, 2002). No automatic TC detection algorithm during the forecasting time is yet implemented. Thus the tracker is a diagnosis system and does not interact with the model's dynamics (Van der Grijn et al., 2004). The tracking algorithm is illustrated in figure 3.3.

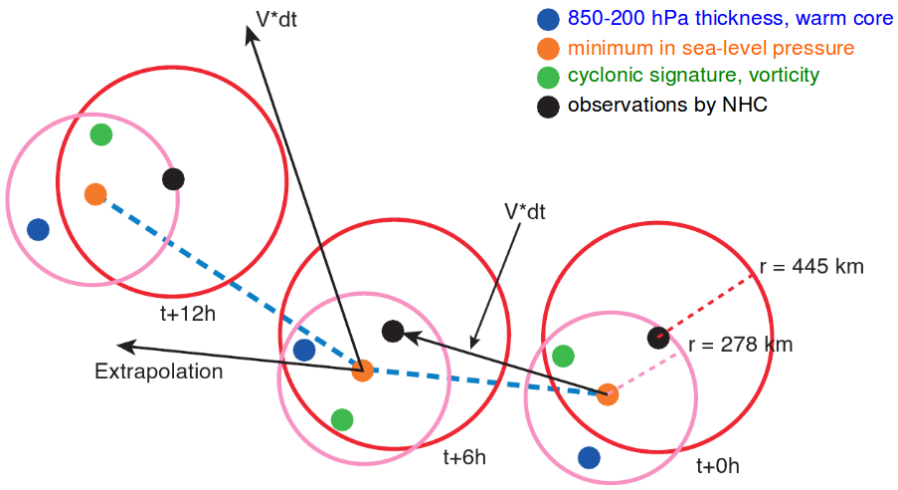



**Figure 3.3:** *TC tracking algorithm. (Source: Van der Grijn et al. (2004))*

When a TC observation is available (black dot in figure 3.3) different conditions are tested to identify the location of the cyclone in the analysis field (from Van der Grijn et al. (2004)):

- minimum sea-level pressure ($<$ 1015 hPa) within a radius of 445 km around the observed TC position
- cyclonic signature, maximum in relative vorticity at 850 hPa within a radius of 278 km around the minimum sea-level pressure
- warm core signature, maximum in 850-200 hPa layer thickness within a radius of 278 km around the minimum sea-level pressure (to distinguish from cold core extratropical cyclones)
- wind speeds over land exceeding $U > 8$ m/s at the 10 m level
- geographical and orographical limitation: the tracker stops if the TC is over high orography and more than 278 km away from the tracker first guess.

If the TC is successfully identified in the analysis, the tracker will follow it in the subsequent forecast steps. Because observations are no longer available, a first-guess is computed and used in order to detect the TC feature in the forecast fields. The same conditions as listed above apply again and if not fulfilled the TC is not further tracked and assumed to have dissipated. This point is important in the case of the ensemble tracks, because it implies that all the 51 members do not necessarily have the same length (see figure 3.4). Furthermore, unrealistically rapid displacements were filtered out in post-processing. The first guess position $r_{\mathrm{fg}}$ used to identify TCs in forecast fields is obtained from a weighted average of an extrapolation of the past movement $r(t) - r(t - \delta t)$ and of an advection by the steering flow $V_{\mathrm{adv}}$ (Van der Grijn, 2002):

$$\vec{r}_{\mathrm{fg}}(t + \delta t) = \vec{r}(t) + w(\vec{r}(t) - \vec{r}(t - \delta t)) + (1 - w)V_{\mathrm{adv}}\delta t \tag{3.1}$$

The steering flow $V_{\mathrm{adv}}$ is determined from a weighted mean of the zonal and meridional winds at 850 hPa, 700 hPa, 500 hPa and 200 hPa. The weight $w$ can assume values between 0 and 1.

An important element for issuing warnings or using the forecasts in reinsurance is the dissemination schedule for the ECMWF TC EPS forecast products. The forecast is normally available 8.5 hours after initialisation time. In figure 3.4 an example of a TC ensemble forecast is shown. The blue squares indicate dates at which forecasts were started.

The ECMWF TC tracking system output consists of 52 ensemble forecasts of position and intensity in terms of central pressure for a forecast range of 120 h with a forecast every 12 hours for the 51 members from the EPS and 1 operational high resolution forecast for the same forecast range but with steps of 6 h.

Figure 3.4 shows an example of an ECWMF tropical cyclone ensemble track forecast. In the following the high resolution deterministic forecast will be called "operational" (OPER) forecast, it represents the ensemble member 0. The lower, EPS resolution deterministic forecast, member 1 will be referred to as the control run (CTL). The other 50 members result from the perturbed initial conditions of the EPS model.

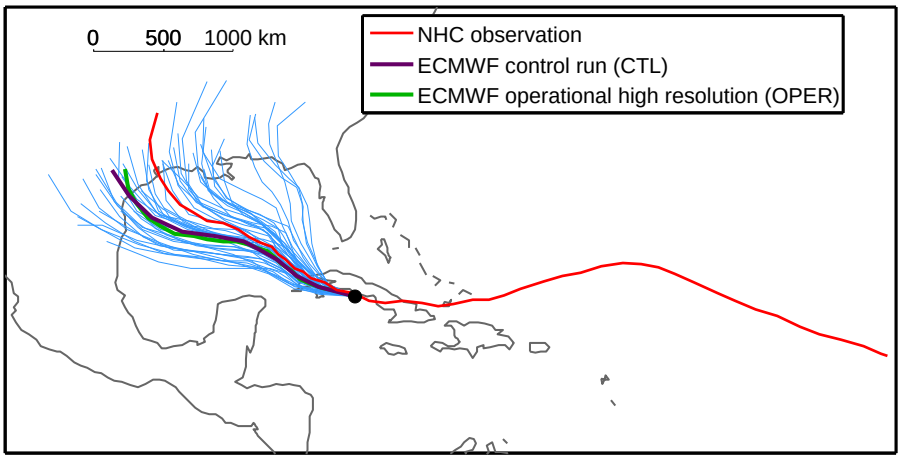



**Figure 3.4:** *Example of a TC ensemble track forecast. The operational (OPER) high resolution forecast is shown in green (ensemble member 0). The lower resolution control run (CTL) from the EPS is shown in violet (ensemble member 1). The ensemble members 2 to 51 are shown in blue.*

#### Ensemble Perturbations Targeted at Tropical Cyclones

TC ensemble forecasts were made possible by the introduction in 2002 of EPS perturbations targeted at observed TCs (Puri and Palmer, 2001). The SVs normally used in the ECMWF EPS system are only optimised for the region poleward of 30°. These SVs need a linearised model version in order to compute fast growing structures, which is not optimal for the tropics, because of the presence of strong diabatic processes. The set of linear physical parametrisations introduced by Mahouf (1999) made it possible to compute SVs for situations as in the tropics, where physical processes may play an important role in perturbation growth (Puri and Palmer, 2001). The derivation of tropical SVs and their properties can be found in Berkmeijer et al. (2000). Unlike the extratropics, where the optimisation of unstable modes is done globally, the SVs in the tropics have to be optimised for the vicinity of the feature of interest, here the cyclone, in order to obtain SVs associated with the cyclone dynamics (Berkmeijer et al., 2000). The SV computation scheme targeted on TCs implemented at ECMWF is described in ECMWF (2007).

## 3.2 National Hurricane Center (NHC) Data

The NHC, as a component of the U.S. National Center for Environmental Prediction (NCEP) and a unit of the NWS, issues observations, warnings, forecasts and analyses of hazardous tropical weather. Furthermore, through international agreement, the NHC has responsibility within the World Meteorological Organisation (WMO) to generate and coordinate TC analysis and forecast products for a certain number of defined countries in the Americas, the Caribbean and for the waters of the North Atlantic Ocean, Caribbean Sea, Gulf of Mexico and the Eastern North Pacific Ocean. For the above named countries and waters the NHC has a legal function as a warning and observation centre.

In this thesis the official NHC forecast was used as a qualitative comparison basis for the ECMWF forecasts. The best track data were used as verifying observations for the storm tracking approach to verification. These two NHC products will be shortly described in the following.

### 3.2.1 Official Tropical Cyclone Forecasts

The NHC official forecasts comprise a forecast of the cyclone's centre location and maximum 1 min surface wind speed. Forecasts are issued every 6 hours for lead times 12h, 24h, 36h, 48h, 72h, 96h and 120h after initial time, which are 0000 UTC, 0600 UTC, 1200 UTC or 1800 UTC. The official NHC forecast is derived from a combination of several models. An extensive summary of the NHC track and intensity models is given in Rhome (2007).

### 3.2.2 Observed Tropical Cyclone Track and Intensity Data

The NHC "best-track" data contains the official NWS historical record of TC tracks as well as intensities and is commonly used as a basis for verification of NWP model outputs for the Atlantic and the Eastern Pacific basins. The NHC best-track data of TCs is obtained from a subjectively smoothed representation of the TC's location and intensity generally at 6 h intervals over its lifetime. The center is defined by the location of the TCs minimum pressure at the surface. The data is based on post-storm assessments of all available measurements (Avila, 2002).

Surface observations of TCs over the ocean are rare. Near the coast lines observations from buoys are available. Drifting buoys can be deployed, but usually do not cover a sufficiently representative area in and around the storm. When TCs make landfall, conventional meteorological station reports can be used, even if such data has to be used with care, as the intensities are well above the range of measurement, for which the instruments are designed.

Upper-air observations, measurement from ships and weather radar can be used. Over the free ocean, the primary source of information is Geostationary Operational Environmental Satellite (GOES) and polar-orbiting weather satellite imagery, from which position and intensities are deduced using the Dvorak (1984) technique. Several other satellite-based remote sensors yield relevant information on the TC location and structure. The seawinds scatterometer on the QuickSCAT satellite for example provides a basis to determine the extent of the TC force winds and the identification of closed surface circulation. Relatively precise pressure and wind information can be gained from reconnaissance flights from their flight levels or using dropsondes. Concerning the comparative quality of direct and remote sensors, Brown and Franklin (2002) showed that nearly half of Dvorak satellite-based intensities fall within 7 kt of the reconnaissance-based best-track values and only 10% of the data contained differences of more than 20 kt.

On first order, it can be said that synoptic scale atmospheric circulation patterns determine the track of a TC. This means a first good guess would be to say that the storm follows the stream of the environment, in which it is embedded. However, the innermost portion of the storm that is used to fix the track is subject to small-scale (when compared to the size of the storm) oscillations, which usually have a magnitude of less than 40 km (Murnane, 2004). These so-called trochoidal motions are not representative of the large-scale motion of the entire storm and are thus excluded from the best track analysis (Jarvinen et al., 1984).

It is important to emphasise that best track data aim at reflecting the best available data describing the storm, however these data are not perfect and are sometimes subject to substantial uncertainties. The quality of present best track datasets are discussed in Murnane (2004). The most difficult parameter to handle are sustained wind speeds, which strongly depend on the averaging time span for surface observations or the estimation technique used to gain maximum sustained winds from satellite observations. To reduce the effect of this uncertainty in maximum wind estimation, minimum central pressure values are used for TC intensity verification in this work. Furthermore, central pressure is functionally related to both thermodynamic and dynamic sources of forcing and is more steady than the often ephemeral reported maximum wind speeds (Simpson and Saffir, 2007).

One final aspect is the uncertainty component given by the time dependency of data coverage density. Indeed point measurements, as well as reconnaissance flights are casual and not constant for the different events and in time for the same event. Satellite data are more reliable, but mostly yield preciser results if ground truth is available. Thus it can be said that best-track data do probably not exhibit a constant quality. This implies that verification of the ECMWF data is done using a reference dataset that is of variable severity. The same remark applies to the object-based verification case where ECMWF analysis

data are used as an observational benchmark. Because the observational data basis is the same, variations in quality over time should be similar in the NHC best track and the ECMWF analysis, which allows a consistent comparison of the two datasets to be done. Furthermore, it has to be emphasised that the NHC composite dataset is the best available information on TC track and position currently available.

## 3.3 Climate and Persistence (CLIPER) Tropical Cyclone Track Forecasts

CLIPER is a 5-day statistical CLImatology and PERsistence TC track forecast, which is considered a no-skill forecast and generally taken as a benchmark for forecast verification of a NWP model (Neumann, 1972; Merill, 1980; Aberson, 1998). The model was originally developed for 3-days lead time and was extended to 5-days by Aberson (1998) due to increasing prediction timescales of NWP models.

The CLIPER model predicts the current TC position from a regression equation. Two different models are used for the Atlantic and the Gulf of Mexico region. The predictors used in the CLIPER forecast are initial longitude, latitude, intensity, day number, zonal motion and meridional motion. A least square fit is operated using historical TC track data (1931-2004, status in 2008) and retaining only those predictors chosen from the linear terms and from all possible cross terms that explain at least 0.5 % of the variance of the predictand (see Aberson (1998) for model formulation). Prediction with CLIPER can be done at 12 h intervals up to 120 h.

CLIPER track errors in units of n mi with respect to best track data for the chosen hurricanes of the seasons 2005-2008 can be found in the respective seasonal verification reports by NHC.

## 3.4 Statistical Hurricane Intensity Forecast Model (SHIFOR)

SHIFOR is a 5-day Statistical Hurricane Intensity FORecast model analogous to CLIPER (Jarvinen and Neumann, 1979). It is used to assess the skill of other prediction schemes with respect to intensity. Seven basic predictors, namely day number, initial latitude, longitude, average zonal and meridional speed in the past 12h, current maximum sustained wind speed and previous 12h change in maximum sustained wind speed as well as these predictors' respective cross-product terms are used in the regression equation (Jarvinen and Neumann, 1979).

While the SHIFOR model can be a useful benchmark, it is not optimised for landfall situations (Franklin, 2006). Thus, the SHIFOR model was adapted in order to include a decay component, when the hurricane reaches land. The so-called DSHIFOR forecast is produced by adjusting the output of a SHIFOR forecast, using a decay rate introduced by DeMaria et al. (2006). This requires a forecast track, which is taken from CLIPER.

The (D)SHIFOR errors are given in terms of differences of maximum sustained wind (in kt). For the purpose of the verification in this report, they have to be converted into a corresponding minimum central pressure. This should normally be done using detailed storm information and a dynamic model. However for the sake of simplicity, the Saffir-Simpson scale correspondence was used (Simpson and Riehl, 1981). Figure 3.4 shows the relationships that can be used for the different categories.

The conversion rule can be assumed to be linear over the pressure $p$ and wind intensity $U_{\max}$ domain of interest:

$$p = -1.13 \cdot U_{\max} + 1073.70 \tag{3.2}$$

Intensity *errors*, meaning differences in central pressure were thus approximated as follows, using the NHC maximum wind speed differences:

$$\|\Delta p\| = 1.13 \cdot \Delta U_{\max} \tag{3.3}$$

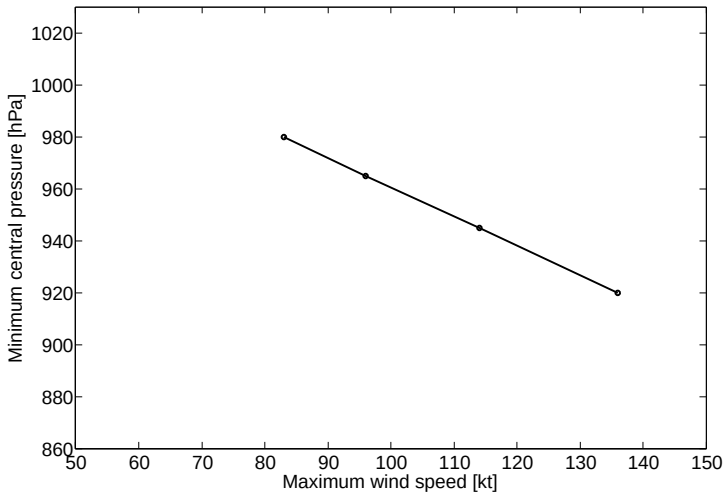



**Figure 3.5:** *Relation between maximum wind and minimum central pressure using the Saffir-Simpson scale (Simpson and Riehl, 1981).*

DSHIFOR intensity errors for the chosen hurricanes of the seasons 2005-2008 can be found in the respective seasonal verification reports by NHC. Note that for the year 2005 only SHIFOR intensity errors are available, as the DSHIFOR model adjustment for interaction with land was implemented only in 2006. This introduces a slight inconsistency into the analysis, which might, however, be negligible, when compared to the error introduced by converting the wind speed maximum errors in kt as given by NHC into central pressure error in hPa. Franklin (2009) mentions that on average DSHIFOR errors are 5-15% lower than SHIFOR errors for lead times 12 h to 72 h and about the same for longer lead times. Because DSHIFOR forecasts are used for the time window 2006 to 2008 and SHIFOR forecasts for 2005. (D)SHIFOR will be employed in the following, when refering to the statistical intensity benchmark forecasts.

## 3.5 Swiss Re Insurance Loss Data

For the insurance loss modelling using ECMWF TC ensemble track forecasts, Swiss Re's catastrophe loss modelling system was used. Natural hazards are generally assessed using a probabilistic modelling approach, by which the risk is determined using probabilistic event sets. The whole spectrum of physically realistic probabilistic scenarios is taken into account in order to give a reliable picture of the exposure of a certain region to a specific natural hazard risk. This is a major advantage over " deterministic", scenario-based modelling or underwriting experience, out of which any prediction of expected annual losses or occurrence frequencies would remain very uncertain (Zimmerli, 2003). The basis for setting the policy strategy for an insurance or reinsurance company is the estimation of the loss frequency curve (exceedance probability of a certain loss value) for all regions of interest.

The model used by Swiss Re in order to estimate the local risk exposure can be summarised as shown in figure 3.6. The four modules represent the fundamental building blocks of loss modelling (Meyer et al., 1997):

1. **Hazard:** representative selection ("event-set") of all possible events that have to be taken into account. For this, historical event catalogues as well as scientific knowledge about the physical characteristics of TCs are used. Wind intensities are computed from central pressure values using the so called Holland formula (Holland, 1980) relating central pressure to the maximum wind velocity by using additionally an assumption for the radius of maximum wind.

2. **Vulnerability:** takes the location specific building type and quality into account. The mean damage ratio is a measure describing the average amount of expected losses as a percentage of total insured value. Vulnerability curves are designed for different groups of objects with similar characteristics. Authentic loss data as well as complementary engineering data are used in this module.

3. **Value distribution:** in order to estimate the expected loss the value of the insured object has to be known. Thus information on the whole insured portfolio has to be acquired.

4. **Insurance conditions:** the extent of insured loss dependends on the conditions defined in the treaties. The estimation of these can be simple in the case of deductibles and liability limits, but it can get very difficult in cases, in which for example business interruption, power failure or consequential losses like accommodation costs are covered. For the case studies presented in this report, this module was not used. So called "from ground up" losses were estimated instead.

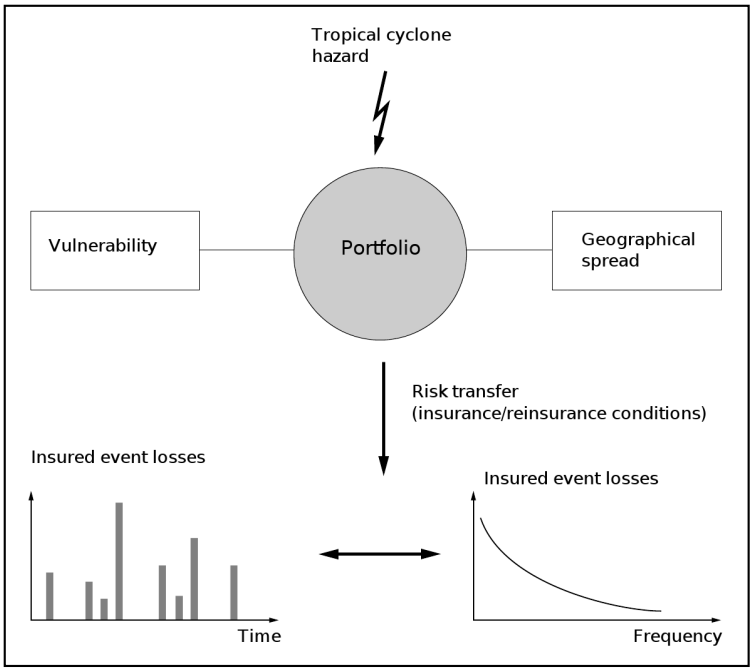



**Figure 3.6:** *Concept for assessing natural hazards (Swiss Re).*

All four modules impact the outcome of a loss assessment. The uncertainty of each individual module enters in the final uncertainty of the loss frequency relationship and combines into a final loss uncertainty. The main sources of uncertainty are the representativity of an event set and the estimation of the mean damage ratio.

# Chapter 4

# Case Studies

## 4.1 Tropical Cyclones in the Gulf of Mexico

In the Atlantic basin, hurricane activity is highest during the summer and major hurricanes generally occur in August or September. From the pattern in the hurricane track climatology of the years 2000 to 2007 (Figure 4.1) the tracks can be classified into four general paths groups, depending primarily on their recurving point: (1) along the U.S. Eastern coast, (2) through the Caribbean Sea into the Gulf of Mexico, (3) over Florida and then into the Gulf of Mexico, (4) from the Caribbean Sea, recurving at the edge of the Gulf, over Florida and then along the Eastern U.S. coast.

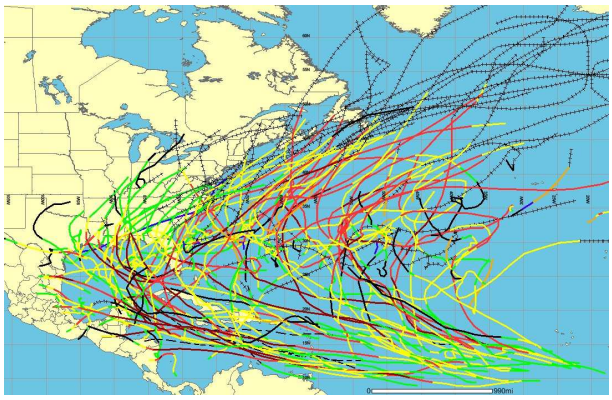


**Figure 4.1:** *Hurricane track climatology of the years 2000 to 2007 (Source: NHC). Light and dark red represent storm intensities of the Saffir-Simpson Categories 1-2 and 3-5 respectively. Yellow, green, black and grey segments show intensities below hurricane force. Other colours represent sub- or extratropical storms*

Approximately one third of the Atlantic hurricanes eventually pass through the Gulf of Mexico (Hodge, 2006). In the last couple of years several major hurricanes passed through the region. The Atlantic hurricane season 2005 was the most active season since the beginning of accurate record keeping in 1944 (Beven et al., 2008). NHC recorded 27 tropical storms, of which 15 became hurricanes. Five hurricanes made landfall in the U.S. causing well over 100 billion Dollars in damages in the U.S. alone and nearly 1700 deaths (Beven et al., 2008). As can be seen in figure 4.2, particularly hurricane Rita and Katrina passed through the heart of the Gulf oil and natural gas production, resulting in widespread disruption of activity. Following the U.S. Energy Information Administration (EIA) only approximately one fourth of the annual average oil and natural gas production could be achieved by the offshore fields in the year following the 2005 hurricane season (Hodge, 2006).

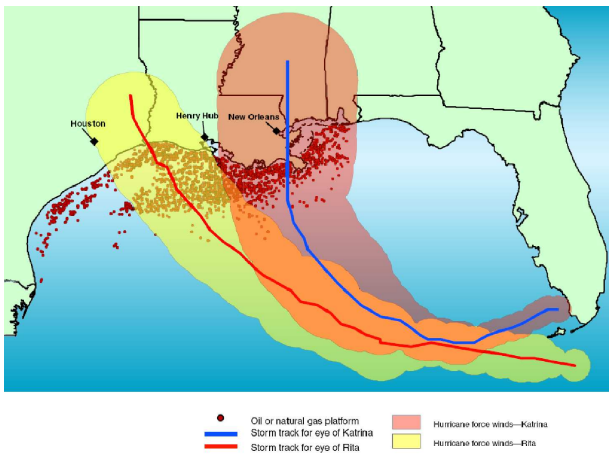



**Figure 4.2:** *Hurricanes Katrina and Rita, season 2005 in the context of natural gas and oil platforms in the Gulf of Mexico (Source: Wells (2006)).*

In addition to these on-sea damages to the Gulf oil and gas production, the infrastructure on land and thus the whole regional economy can be severely struck. Considering the high magnitude of overall losses that can occur in the Gulf region, insurers are very cautious about the capacity they expose. After major hurricanes like Ike, Gustav in 2008 or Katrina and Rita in 2005 several reinsurance companies were forced out of business. However, precisely because of the elevated damage potential, the Gulf region is still an attractive and potentially very profitable market for the insurance and reinsurance industry. The actual exposure is far from being totally covered. Thus, there is a high interest from the side of the insurance sector to improve the way the hurricane risk is handled in order to better understand the exposure and adapt their policy strategy.

Medium to short-range forecasts of TCs are used in a very different way by a state government or a global reinsurance company. For evacuation planning a lead time of at least 24h is needed and false alarms may cost several hundred thousand dollars and affect the trust of the population in the authorities. In the (re)insurance domain such forecasts can influence last minute trading of risk-linked securities, like catastrophic risk (CAT) bonds. Theses innovative financial vehicles known as "Insurance-Linked Securities" (ILS) emerged in the 1990ties and play an important role in today's financing of mega-catastrophes (see Cummins (2007) and Forchaux (2006) for an overview). The general idea is to transfer catastrophic risks to capital markets, thus easing industry capacity constraints (Sigma, 2006). This development has created a large source of new capacity, involving investors in the broad capital markets, who are better suited than reinsurance companies to assume the high severity/low frequency risk profile of natural catastrophes (Sommerville, 2009). Trading of CAT bonds continues until a few hours before landfall and information on probable landfall locations and intensities for three to four days lead time can be very useful. Short-time insurance loss modelling based on the available forecasts can give a good picture of the overall losses that might be expected and can help on a strategic decision making basis (trading) as well as on an communication level (top management information, press release).

The TCs used in this report were selected with the above summarised background of the insurance sector and potential damage for society in general in mind. Because of the quality and consistency of available forecast data and because of the specific interest for extreme weather, only TCs ranked as hurricanes (central pressure $\leq$ 989 hPa) were considered. In the following sections the hurricane events used in the three foci of analysis of this report are presented.

## 4.2 Selected Hurricanes 2005-2008 for the Storm Tracking Verification Approach

A total of 29 hurricanes were used for the statistical analysis of forecast errors of the ECMWF TC forecasting system as shown in figure A.1 in the appendix. Table A.1 summarises the main properties of these hurricanes. In principal all hurricane force TCs of the North Atlantic were used. This choice is due on one hand, because these are the damage effective events, which are most interesting for damage assessment. On the other hand, the data availability was best for these events. There were 70 TCs in the period 2005 to 2008 in total. The number of forecasts for the tropical storms is however significantly lower, than the ones for hurricane force storms, because of their mostly shorter lifetime. For four of them the ECMWF forecasts were not available. These were thus left out (Cindy (2005), Dennis (2005), Epsilon (2005), Lorenzo (2007)).

The sensitivity of the computed verification scores towards the hurricane characteristics like intensity and landfalling was investigated using the group definition shown in table 4.1. Of the 29 hurricanes chosen, 17 passed through the Gulf of Mexico once in their lifetime. All the hurricanes with Saffir-Simpson Category higher or equal to 3 making landfall (H3Land), passed through the Gulf of Mexico. The percentage of hurricanes that occured in the year 2005 when compared to the total over the whole period of analysis is emphasised in table 4.1. As mentioned in section 3.1 the model resolution changed in the beginning of 2006, which is expected to have improved the forecast quality substantially in subsequent years. This has to be borne in mind, when performing a sensitivity analysis of the forecast errors towards these hurricane groups. If higher errors are found for the group of landfalling Category 3 hurricanes (H3land) for example than for the other groups, this result might have been affected by the higher 2005 hurricane proportion in this group.

**Table 4.1:** *Hurricane groups for investigation of forecast quality dependence on landfalling and intensity. AllH comprises all selected hurricanes. H3 represents only hurricanes having sustained a central pressure classified as Category 3 or more on the Saffir-Simpson scale at least once during their lifetime. AllHLand encompasses all hurricanes having made landfall. The group H3Land combines the Category 3 hurricanes that made landfall.*

| Group | 2005 | 2006 | 2007 | 2008 | Total |
|---|---|---|---|---|---|
| AllH | 11 (38%) | 5 | 5 | 8 | 29 |
| H3 | 6 (40%) | 2 | 2 | 5 | 15 |
| AllHLand | 7 (37%) | 1 | 4 | 7 | 19 |
| H3Land | 5 (45%) | 0 | 2 | 4 | 11 |

## 4.3 Hurricane Event for Object-Based Forecast Verification

For the object-based verification of tropical cyclone forecasts in the ECMWF deterministic model only one hurricane was used, partly due to the time constraint and because the focus was intentionally laid on the development and construction of such a new verification measure. The analysis of other events using this technique and its application to ensemble forecasts are beyond the scope of this master thesis, but would be interesting in future work.

Hurricane Ike was selected for this analysis because it was the most destructive one of the 2008 season, for which the latest ECMWF model state could be used. Important elements concerning the dynamics of this hurricane are shortly summarised below. The detailed synoptic history and meteorological statistics of hurricane Ike can be found in Berg (2009). The track with the intensity evolution is shown in figure 4.3.

Hurricane Ike emerged from a tropical wave moving off the west African coast on the 1st of September 2008. Favourable low vertical wind shear conditions as well as a strong low over the North-western Atlantic and an upper-level trough favoured a quick intensification up to a Category 4 storm with a minimum recorded central pressure of 935 hPa and wind peak intensities of 233 km/h. Strong north-westerly shear weakened Ike to a Category 3 storm as it moved westward towards Cuba. Haiti was touched by Ike's outer rainbands, which caused floods and mudslides. Loosing 2 Categories in strength, while crossing Cuba, where it caused important wind and flooding damage, hurricane Ike moved further into the Gulf of Mexico. The warm Gulf waters helped a new strengthening into a Category 2 hurricane. Ike made landfall in Galveston (Texas) on the 13th of September 2008. After merging with a cold front on the 14th of September, Ike weakened to a tropical storm.

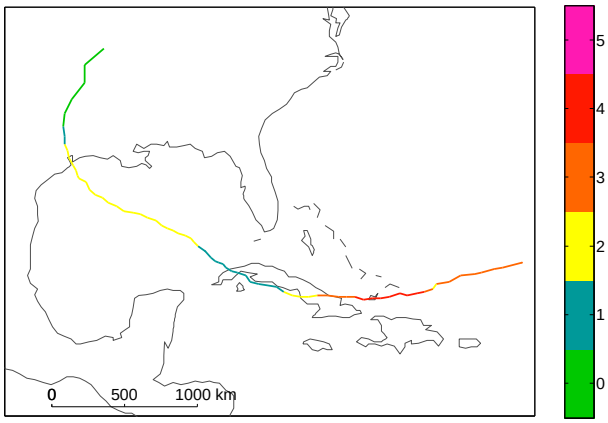



**Figure 4.3:** *Hurricane Ike track and evolution of intensity. The colorbar indicates the intensity in terms of Saffir-Simpson Categories. Category 0 corresponds to the tropical storm level. For the corresponding wind speeds and central pressure see table 2.1 in section 2.3 of chapter 2.*

## 4.4 Hurricane Events for Insurance Loss Predictions

For loss predictions, the most destructive and costliest hurricanes, which made landfall in the Gulf region in the period 2005-2008 were used. The selected storms are shown in figure 4.4. All the hurricane track types mentioned in section 3.1 above are represented except type 1 track, which follows the U.S. East coast.

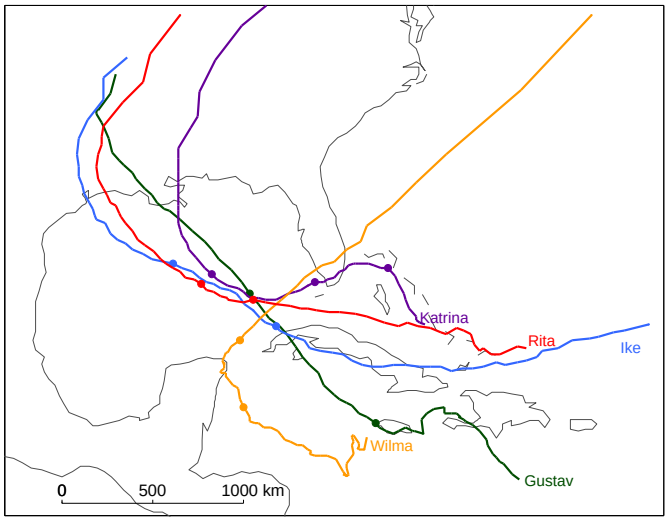



**Figure 4.4:** *Selected hurricanes in the time period 2005-2008. The location of the name indicates the starting point of the hurricane.*

At least two different lead times were chosen for each hurricane, 1 and 3 days before landfall. The first one is intended as a reference for public measures like evacuation planning, or information of top management in the case of a company like Swiss Re. The second one allows strategic measures to be taken both by governments (issuing a storm watch) and trading in (re)insurance. The chosen lead times are listed in table 4.3. For hurricane Katrina, 3 lead times were chosen because de first landfall in Florida already induced considerable damage. The 24h ECMWF forecast for hurricane Rita was not available, thus the two used forecast lead times only have 1 day difference.

**Table 4.3:** *Landfalling dates and times for the hurricanes selected for loss predictions.*

| Hurricane | Landfall date | Landfall location | F1 | F2 | F3 |
|---|---|---|---|---|---|
| Ike | 08/09/13 07:00 | Galveston Island, Texas | 79h | 31h | - |
| Gustav | 08/01/09 12:00 | Cocodrie, Louisiana | 72h | 24h | - |
| Wilma | 08/10/21 22:00 | Cozumel, Mexico | 22h | - | - |
| | 08/10/22 03:30 | Puerto Morelos, Mexico | 27.5h | - | - |
| | 08/10/24 10:30 | Cape Romano, Florida | 82.5h | 22.5h | - |
| Rita | 08/09/24 07:30 | Texas/Louisiana Border | 67.5h | 43.5h | - |
| Katrina | 08/08/25 22:30 | Southeastern Florida | 22.5h | - | - |
| | 08/08/29 11:00 | New Orleans, Louisiana | 107h | 71h | 23h |

In the following the main damage figures and most important aspects of each hurricane are presented. The indicated losses and number of deaths are drawn from the U.S. National Climatic Data Center (NCDC), which is responsible for monitoring and assessing climate events that have great economic and societal impacts in the U.S. and globally. The focus is laid on insurance damage occuring in the U.S., because no portfolio information on the Gulf Islands is available. Furthermore, the societal damage information contained in the computed losses is only representative for the U.S. because the the insured value in the hispanic countries around the Gulf of Mexico does probably not comprehend a wide enough range of activities.

**Ike:** 01.09.2008 - 14.09.2008
Hurricane Ike made landfall in Galveston (Texas) with wind peaks ranking it as a Category 2 hurricane. In terms of storm surge it was rather a Category 4 to 5 hurricane. Storm surge was considerable in Texas, near the point of landfall, wind and flooding damage occured in Texas and downwind up to Illinois and Pennsylvania. Hurricane Ike enters present history as being the largest size Atlantic hurricane on record. As mentioned in section 3.1 it severely struck the oil and gas industry of the region. The total damage is estimated at 27 billion U.S. Dollars and approximately 200 fatalities.

**Gustav:** 25.08.2008 - 04.09.2008
Hurricane Gustav made landfall as a Category 2 hurricane in Louisiana a week before Ike and caused significant wind, storm surge and flooding damage in Louisiana, Mississippi, Arkansas and Alabama. Total damage is estimated at more than 5 billion U.S. Dollars and 150 deaths.

**Wilma:** 15.10.2005 - 26.10.2005
Hurricane Wilma was the last major hurricane of the record season 2005 and made landfall in Mexico before recurving north, where it struck Southern Florida as a Category 3 hurricane. Its minimal pressure of 882 hPa is estimated to be the lowest pressure ever measured in the Atlantic. Its quick intensification of 169 km/h wind speed increase in 24 h also established a new record. Total losses are estimated at 30 billion U.S. Dollars and 60 deaths.

**Rita:** 18.09.2005 - 26.09.2005
Hurricane Rita hit the Texas-Louisiana border region as a Category 3 hurricane. It was responsible for signification storm surge and wind damage along the coast as well as inland flooding in Alabama, Mississippi, Louisiana, Arizona and Texas. Its minimum pressure of 897 hPa is the third lowest ever recorded. Total damage is estimated at 17 billion Dollars and 119 deaths.

**Katrina:** 23.08.2005 - 31.08.2005
Hurricane Katrina made a first landfall in Florida, near Miami as a Category 1 hurricane and a second landfall as a strong Category 3 hurricane at the Louisiana/Mississippi coastline, near New Orleans. It induced major storm surge as well as wind damage and severe flooding through levee system failure.The total damage is still not definitely established but is estimated around 125 billion U.S. Dollars, which makes it the most expensive natural disaster in U.S. history. The fatalities amounted to 1800 deaths, which also figure amongst the highest number in history.

Detailed discussions of these events can be found on the NHC web page (`http://www.nhc.noaa.gov/pastall.shtml#tcr`). It should be pointed out that the above numbers have to be considered with care. The differences between individual reports can be substantial. Thus, they should only be taken as indicative of the order of magnitude of the real damage.

# Chapter 5

# Methodology

In order to compare different forecasting systems and to assess their performance an adequate framework for forecast verification has to be set up. For deterministic track forecasts a semi-automatic verification procedure has been proposed by Heming (1994) and applied at the U.K. Met Office. Several case studies have assessed the value of short and medium-range deterministic and ensemble forecasts in the domain of hydro-meteorology (e.g. McCollor and Stull (2008b) and McCollor and Stull (2008a), Roulin (2007)). Feature specific quality measures have been proposed, as for example the object-based "**S**tructure-**A**mplitude-**L**ocation" quality measure (Wernli et al., 2008) for verification of quantitative precipitation forecasts.

This chapter describes the methods that were applied to investigate the three ways of looking at the verification problem presented in the introduction. First, the interpolation method used in order to compare the ECMWF forecasts to the NHC observations as well as the scores used in the traditional storm tracking approach to verification of tropical cyclones are presented. In the second section, the object-based verification procedure is introduced. Finally, an overview is given of Swiss Re's catastrophe modelling system, which was used for the loss predictions.

## 5.1 Methods of Traditional Tropical Cyclone Forecast Verification

Next to the routinely produced verification statistics of NWP from their operational model, several forecasting institutions have implemented a specific verification procedure for TC forecasts. Heming (1994) describes a semi-automatic track verification method, which has been applied since 1994 at the UK Met Office. Verification statistics of TC forecasts by ECMWF are to be found in Van der Grijn et al. (2004). The aim behind such a feature specific evaluation is primarily to assess the model's ability to handle these weather systems. In the case of a global NWP models such as ECMWF, to improve the skill of TC forecasts can also help reducing the downstream development of forecast errors, when cyclones originating from tropical regions enter the mid-latitudes (Van der Grijn, 2002). Indeed, cyclones that move from the tropics into the extratropics were found by Froude (2009) to be affected by larger forecast errors in their predicted intensities than storms that originate in the extratropics.

In this part, the verification procedure is based on 2 variables, namely the storm's position and intensity, which are compared for every available forecast with the NHC best track observation. The TC is thus represented by its minimum pressure location and the value of this pressure as an indication of intensity. In figure 5.1 the verification approach is illustrated. The different positional error components that are introduced in section 5.1.2 are shown as well. The data interpolation onto the same time grid is described in section 5.1.1. In the two subsequent sections the different error statistics that were computed in order to assess the performance of the ECMWF TC forecasting system are presented. Two types of statistics can be derived from the model output. The cyclone's forecast track and the predicted intensity can be verified. A simple skill score for the position and intensity forecast, the computation scheme for the ensemble mean and spread, a measure for linear association and the equal likelihood analysis of ensemble members are introduced thereafter.

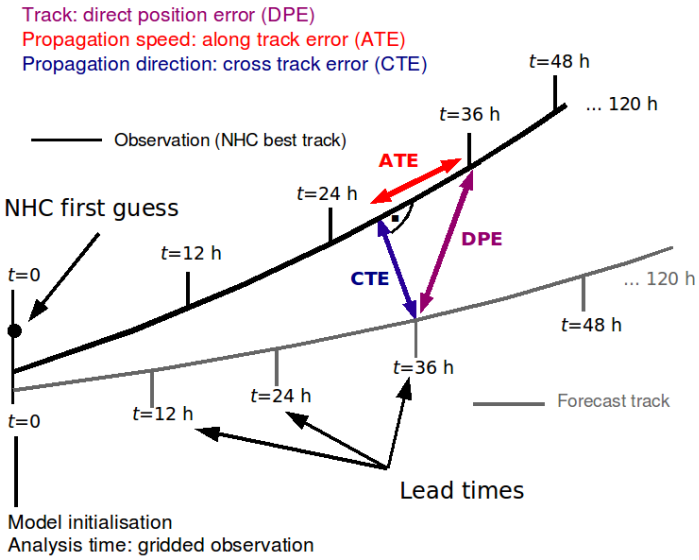



**Figure 5.1:** *Forecast and observed track. Adapted from Froude et al. (2007)*

### 5.1.1 Interpolation of the Observational Time Grid

The observed and forecast positions and intensities are not given on the same time grid. Observations generally have a time resolution of 6h and are given for UTC times 03:00, 09:00, 15:00, 21:00. However, the time grid is not regular and for some dates, data is available on 3 h resolution. Forecasts are available at 00:00 UTC and 12:00 UTC. Although it could be argued that the good quality of the observed data should not be compromised by performing an interpolation, the observations were chosen to be interpolated onto the forecast time grid. The reasons for this choice are, firstly, that the forecast time grid does not vary in resolution as the observational time grid does and secondly that the quantity of data to be interpolated is lower when selecting the observations. There are many forecasts along the track with each time 52 members, whereas there is only 1 observational track. The sensitivity of the computed error statistics with respect to the interpolation scheme was analysed for one hurricane (Ike, 2008) and no significant differences were observed (see appendix, section A.3.1).

The observations that are available to ECMWF at the time of forecast initialisation may be different from the best-track data available today. The latter data has been improved by measurements that can be used only in a post-assessment of the event like satellite imagery or aircraft observations for example. To perform a thorough model verification the best available dataset of the observed event should be used. This is reflected in an initial positional error of the cyclone in the verification results. The initial forecast position at lead time 0 and the observation interpolated to the same time grid, do not necessarily overlap (see schematic in figure 5.1). Figure 5.2 shows an example of a track with the original observed track data, the interpolated track data. The distance between the forecast starting points (blue squares) and the interpolated observations (red circles) represents the initial position error. Figure 5.3 shows the magnitude of these errors.

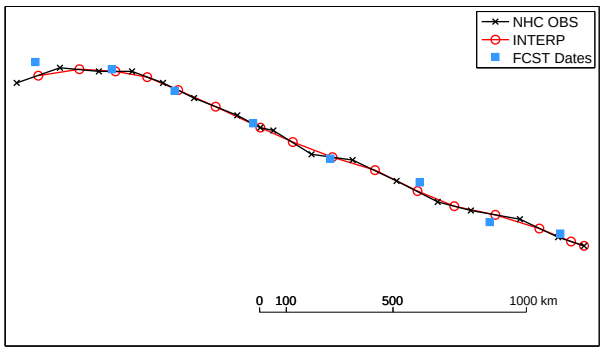



**Figure 5.2:** *Interpolation of the observed track (red line). The difference in the initial forecast position (blue squares) and the corresponding interpolated observation (red circles) is apparent.*

The median value of the initial longitudinal error of the forecast is 63.4 km to the East. The median value of the initial latitudinal error is 41.1 km to the North. These values, as well as the surprisingly high spread of these errors (figure 5.3), are similar to the ones found in the literature. In an early assessment of the performance of the ECMWF TC forecasting system for the months of February, March, April and May 2002 values around 100 km were found as initial positioning errors (Van der Grijn, 2002). In a subsequent study by ECMWF, the initial direct position error amounted again to slightly more than 100km (Van der Grijn et al., 2004). Neumann and Pelissier (1981) indicate a value of 38 km over the period 1970-1979, which is significantly less than the values obtained above. In this latter study official NHC forecasts were assessed. In the NHC forecast advisories the accuracy of the location of the cyclone centre is around 20 to 30 n mi ($\sim$ 35 to $\sim$ 55 km).

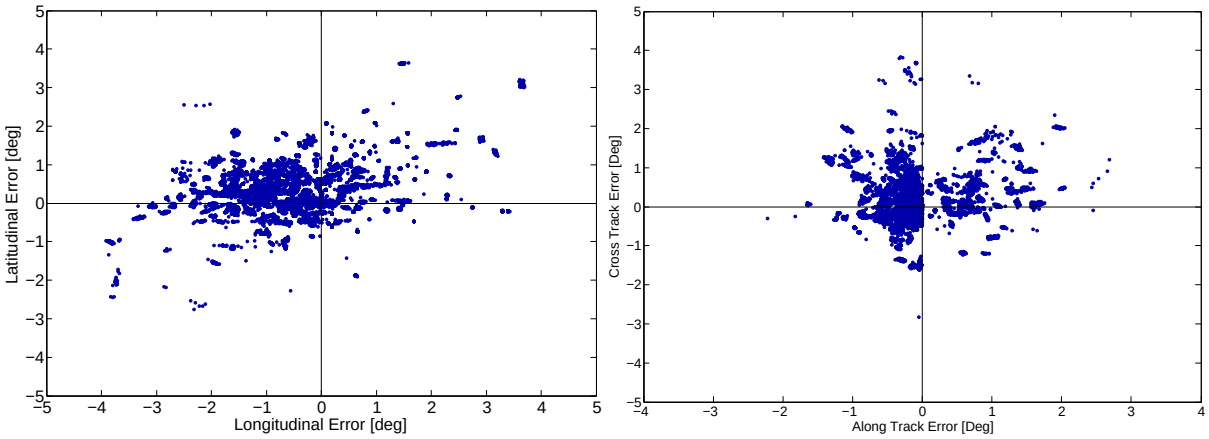



**Figure 5.3:** *Initial positional errors at the forecast time for the selected hurricanes of the years 2005 to 2008.*

Considering the magnitude of the above discussed initial position error, other model related reasons than the mentioned post-event correction of observations by the NHC have to be found. One explanation is that the ECMWF TC forecasting system does not use the observed position given by the NHC as a starting point for the tracking but the corresponding minimum in mean sea level pressure detected in the analysis. Furthermore, an important contribution is surely given by the model grid spacing of 50 km at equator for the ensemble forecasts and 25 km at equator for the operational forecasts. Finally, the definition of the TC center can vary slightly, depending on the forecasting institution. NHC uses the minimum pressure location at the surface, which is also what ECMWF identifies in the analysis fields and follows in subsequent forecasts.

### 5.1.2 Definition of Positional Errors and Biases

The NHC defines the track error as the great-circle distance between a cyclone's forecast position and the best track position at the forecast verification time (NHC, 2008; Neumann and Pelissier, 1981). In this report, this distance will be referred to as direct position error $DPE$. For small distances between two geographical points on an idealised spherical earth, the haversine formula can be used:

$$DPE = 2 \arcsin\left(\sqrt{\sin^2\left(\frac{\Delta\phi}{2}\right) + \cos\phi_1 \cos\phi_2 \sin^2\left(\frac{\Delta\lambda}{2}\right)}\right) \tag{5.1}$$

| | | |
|---|---|---|
| where | $DPE$ | direct position error [km] |
| | $\phi_{\text{obs}}$, $\lambda_{\text{obs}}$ | latitude, longitude of the observed point [km] |
| | $\phi_{\text{fct}}$, $\lambda_{\text{fct}}$ | latitude, longitude of the forecast point [km] |
| | $\Delta\phi$, $\Delta\lambda$ | difference in latitude, longitude between the observed and the forecast position [km] |

The direct position error gives an indication about the general quality of the track forecast, but gives no information as to whether the forecast errors result from a slow/fast bias or from an early/late recurvature

towards the pole. The direct position error can be split into two distinct variables, which account for the above mentioned characteristics. This separation of $DPE$ into an along track error $AT$ component and a cross track component $CP$ is shown in figure 5.4 and figure 5.1. The two components are computed using the observation OB2 corresponding to the forecast point FC and the previous observational point OB1.

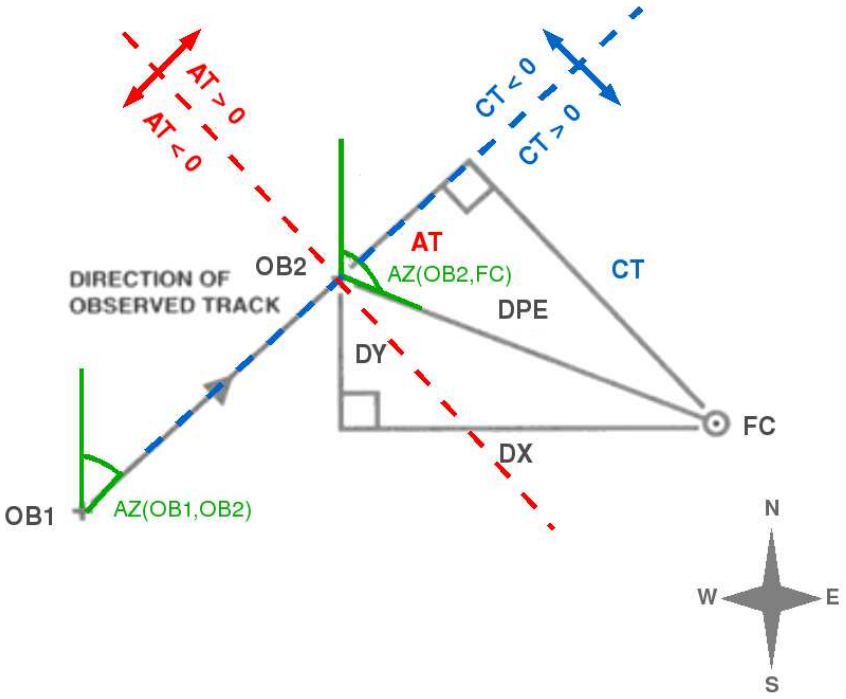



**Figure 5.4:** *Types of positional forecast errors. DPE represents the direct positional error, CT is the cross track component, AT the along track component. DX represents the longitudinal component and DY is the latitudinal component. (Adapted from Heming (1994))*

The positional error components can all be expressed as absolute **errors** or as **biases**. If the model is well calibrated that is if it is reliable as defined in section 2.3 in chapter 2, the long term mean of the ensemble as well as of the operational forecast bias should be zero.

The along track bias $ATB$ occurs, if the forecast storm moves at a different speed than the analysed storm. It is negative when the forecast cyclone is too slow. $ATB$ is defined as follows, when refering to the notation in figure 5.4:

$$ATB = \frac{\left(\overrightarrow{OB1\,OB2} \cdot \overrightarrow{OB1\,FC}\right)}{\parallel \overrightarrow{OB1\,OB2} \parallel} - \overrightarrow{OB1\,OB2} \tag{5.2}$$

The along track error $ATE$ is defined as $ATE = |ATB|$.

The cross track bias $CTB$ is defined using the same conventions:

$$CTB = \sqrt{\parallel \overrightarrow{OB1\,FC} \parallel^2 - \frac{\left(\overrightarrow{OB1\,OB2} \cdot \overrightarrow{OB1\,OB2}\right)^2}{\parallel \overrightarrow{OB1\,OB2} \parallel^2}} \tag{5.3}$$

$CTB$ is positive when the forecast position lies right of the observed track, which is known up to OB2 and is linearly extrapolated using point OB1. This definition is artificial and the differentiation is only made in order to detect recurvature biases in one of the two directions relative to the track. The along track error $ATE$ is defined as $ATE = |ATB|$.

The direct position error can also be separated into two track independent components by using a conventional zonal/meridional coordinate system. This leads to the longitudinal position error $LONE$

and bias $LONB$, as well as the latitudinal position error $LATE$ and bias $LATB$. Such a separation allows an investigation of possible biases in North/South and East/West directions of the track forecast as well as the evaluation of the model error in representing each component. $LATB$ is positive, if the forecast has a bias towards the north. $LONB$ is positive, when the forecast has a bias towards the east.

### 5.1.3 Intensity Error and Bias

The ECMWF intensity forecast of tropical cyclones is given in terms of the estimated central pressure. The central pressure error $CPE$ is defined as the absolute difference between the observed intensity interpolated on the same time grid as the forecast and the forecast intensity.

$$CPE = |CP_{\mathrm{OBS}} - CP_{\mathrm{FCT}}| \tag{5.4}$$

The central pressure bias $CPB$ corresponds to

$$CPB = CP_{\mathrm{OBS}} - CP_{\mathrm{FCT}} \tag{5.5}$$

The pressure bias is thus negative if the forecast underestimates the intensity (i.e. overestimates the pressure).

### 5.1.4 Simple Skill Score for Track and Intensity Forecasts

A simple skill score for the position forecast can be defined as follows:

$$SS_T = \frac{DPE_{\mathrm{CLIPER}} - DPE_{\mathrm{ECMWF}}}{DPE_{\mathrm{CLIPER}}} \tag{5.6}$$

Similarly for the intensity forecast, the following equation can be used to compute the skill:

$$SS_I = \frac{CPE_{\mathrm{(D)SHIFOR}} - CPE_{\mathrm{ECMWF}}}{CPE_{\mathrm{(D)SHIFOR}}} \tag{5.7}$$

This definition allows a direct and simple assessment of the forecasts, when compared to CLIPER data for the track and to (D)SHIFOR data for the intensity. In theory, for all the error measures ($ATE$, $CTE$) defined above, a skill score of the general form can be formulated. This was not done in this thesis, because the benchmark forecast CLIPER was only directly available for $DPE$.

### 5.1.5 Ensemble Mean Error and Spread

In order to assess the statistical reliability of track and intensity forecasts, the ensemble mean error and spread were compared. If the EPS is reliable in the sense defined in section 2.3 of the chapter on Theories and Concepts, the mean error and the mean spread are equivalent. The mean error represents the geodesic separation/absolute intensity difference between the mean tracks/intensities and the corresponding analysis tracks/intensities. The ensemble spread can be obtained by computing the mean ensemble deviation in position/absolute intensity of the ensemble members from the ensemble mean.

### 5.1.6 Measure of Linear Association

The Pearson correlation coefficient is a very practical score, because of its invariance properties with respect to scale. It is used to assess linear association between forecasts and observations:

$$\rho = \frac{\mathrm{cov}(x, \hat{x})}{\sqrt{\mathrm{var}(x) \cdot \mathrm{var}(\hat{x})}} \tag{5.8}$$

| | | |
|---|---|---|
| where | $\mathrm{var}(x)$ | variance of the observations |
| | $\mathrm{var}(\hat{x})$ | variance of the forecasts |
| | $\mathrm{cov}(x, \hat{x})$ | covariance between the observations and the forecasts |

It should be pointed out that two uncorrelated variables are not necessarily independent since they can be non-linearly rather than linearly related to one another (Jolliffe and Stephenson, 2003).

### 5.1.7 Likelihood of Ensemble Members

In order to verify the assumption that all members of the ECMWF TC ensemble forecast are independent realisation of the same random process and that the members can be treated in an indistinguishable fashion, a slightly adapted form of an equal likelihood frequency plot (Jolliffe and Stephenson, 2003) was used. For each forecast in the verification period, the best member was sorted out. The frequency at which each member was the forecast closest to the verifying diagnostic was then analysed for each lead time. This check of independence between members has to be done in order to be able to use the ensemble correctly in the insurance loss prediction part, where forecast probabilities are defined.

## 5.2 Object-Based Verification Approach

### 5.2.1 Demands on the Novel Object-Based Verification Measure

In order to provide an assessment framework that is as targeted as possible, specific demands on the new verification measure were formulated, with respect to form and substance. They are shortly listed in the following.

**Formal demands:**

- Problems of double-penalty of grid-point based measures should be avoided
- The new measure should be easy to communicate
- Important problems should be pointed out visually

**Substantial demands:**

- Verify feature specific properties
- Identify problems in TC structure and dynamics
- Help explaining inabilities of the model to render the feature accurately

The **S**tructure-**A**mplitude-**L**ocation (SAL) framework introduced by Wernli et al. (2008) for precipitation features is found to be adequate for the needs formulated above and was thus adapted for its use in the case of TC verification. The transcription of the SAL concept to the case of TC features should allow to point out structural and dynamical deficiencies in the representation of these weather systems in the ECMWF deterministic high resolution model. For setting up and testing the object-based verification approach, the ECMWF deterministic forecasts for hurricane Ike (2008) are used (see section 3.1.2 for details about the model and section 3.3 for a description of the chosen hurricane).

In the following the object identification procedure is explained. Thereafter three object-based verification components are proposed.

### 5.2.2 Object Identification

The object identification procedure is based on the 3 following steps:

1. The structure is identified by successively filling contours of increasing pressure. The procedure begins at a pressure $p$ of 989 hPa (upper limit of the Saffire-Simpson hurricane Category 1) and reaches 1015 hPa by increments of 1 hPa:

$$989\text{hPa} \leq \text{p} \leq 1015\text{hPa} \tag{5.9}$$

   This implies a variable threshold for one particular date, when analysing different forecast ranges, as can be seen in figure 5.5 for hurricane Ike in 2008. The differences between the forecasts for a certain date are of maximum 5 hPa around the landfall date. It is at the same point in time, as can be expected, that the area differences are also maximum (see figure 5.7). The intensification of the hurricane can be perceived indirectly by the decreasing pressure threshold and has to do with the minimum area and maximum area growth thresholds introduced and explained below.

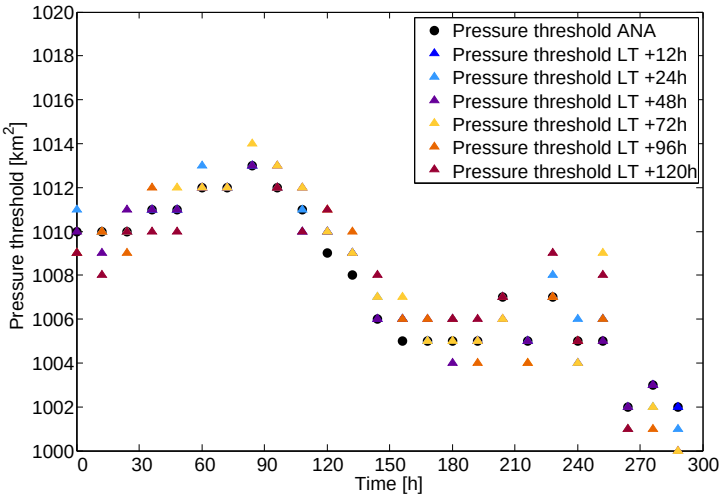



**Figure 5.5:** *Pressure threshold applied to identify TC objects for hurricane Ike ECMWF analysis and forecasts.*

2. The filling of a contour is done by using a 4 direction recursive seed fill (Soille, 1999). The starting point is given by the NHC observation. In order to obtain only objects with an approximately symmetrical structure and concentric isobars, the growth of the object is limited to 25 % of its area $a(p)$ for the pressure threshold $p$ with respect to its area $a(p-1)$ for the pressure threshold $p-1$:

$$\frac{a(p) - a(p-1)}{a(p)} < 0.25 \tag{5.10}$$

The area growth thresholds especially avoids the inclusion of extratropical features because of a too high pressure threshold as would have been the case in the situation shown in the figure 5.6.

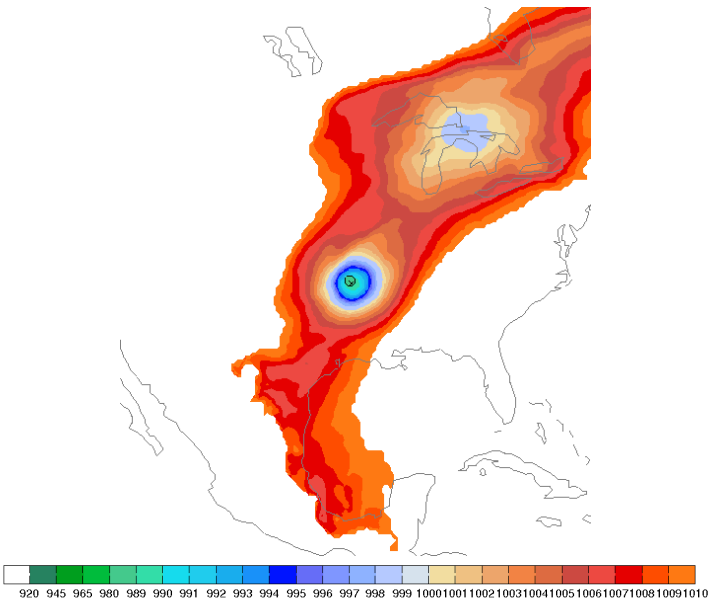



**Figure 5.6:** *Object identification without area growth threshold. The colorbar indicates the pressure scale in hPa. The example shows hurricane Ike shortly after landfall.*

3. Condition 2 is only applied, if the area of the TC exceeds $2 \cdot 10^5$ km$^2$ (i.e. an equivalent 500 km radius) to account for the strong pressure gradients surrounding the center of the storm. This minimum size is not a stringent threshold. The object can be smaller by the area of one 1 hPa contour ring, if the latter represents an area bigger than the 25 % areal growth that was fixed in step 2. This means that the identified objects have a size around $2 \cdot 10^5$ km$^2$ and can be much bigger only if the isobars enclosing the so far identified object are close together (implying an increase of less than 25 % of the area at each 1 hPa step). Figure 5.7 shows the area of the identified objects as a function of time for hurricane Ike. The chosen "soft" size threshold implies an object size with respect to the model resolution of approximately 200 grid points.

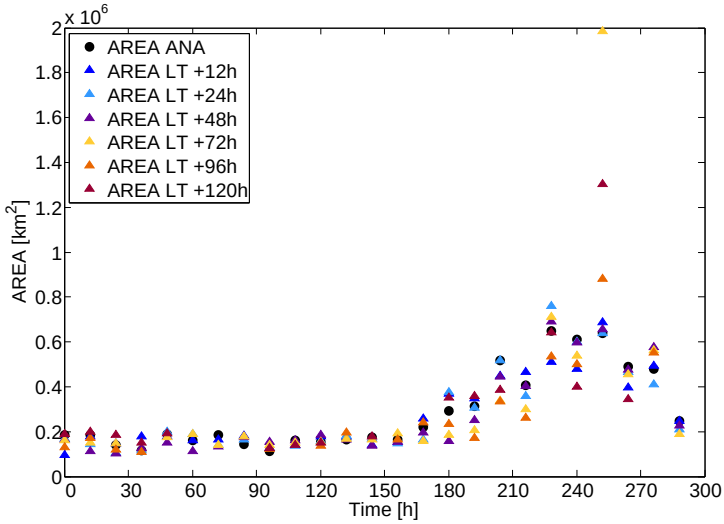



**Figure 5.7:** *Area of the identified objects for hurricane Ike ECMWF analysis and forecasts.*

The presented object identification scheme does not necessarily find an object for every pressure field. If the conditions above are not fulfilled, there is no identified object for the given date and time. A comparison using the components that will be defined in the next section is done only if there is an identified object in both the forecast and the observation field.

### 5.2.3 Definition of the Components

Three components were defined to assess the cyclone objects' properties in a holistic way. All components are normalised by an adequate range of expected value of the considered error measure. The values of the components are set within the range $[-2, +2]$, with 0 indicating a perfect forecast. Values greater than 0 indicate an overestimation by the model.

#### Location Component

The location component $L$ was defined using the point of minimum pressure in the defined object. This characteristic was chosen instead of the barycentre of the object, as it was done in Wernli et al. (2008). The location component is normalised by an average cyclone radius $R_c$ of 250 km:

$$L = \frac{\vec{r}_{\text{fct}} - \vec{r}_{\text{ana}}}{R_c} \tag{5.11}$$

The location component is furthermore artificially separated into positive and negative by setting negative values for $L$, if the forecast lies south of the observation and positive, if it lies north.

### Amplitude Component

The amplitude component $A$ is defined using the concept of integrated kinetic energy (IKE) (Powell and Reinhold, 2007) in the TC object (see chapter on Theories and Concepts, section 2.2 for a definition of $IKE$.):

$$A = \frac{IKE_{\text{fct}} - IKE_{\text{ana}}}{\frac{1}{2}\left(IKE_{\text{fct}} + IKE_{\text{ana}}\right)} \tag{5.12}$$

This component can be adapted depending on the needs and goals of a specific model assessment. For example $IKE$ can be computed using only winds above a certain threshold. This can either be done as presented in Powell and Reinhold (2007), where wind speed thresholds are chosen, which correspond to the limits of the Saffire-Simpson scale. The alternative consists in using a quantile threshold from the wind fields of each particular object. The latter way of proceeding ensures that $IKE$ can always be computed for an identified object.

The time series of obtained $IKE$ for hurricane Ike analysis and forecast objects is shown in figure 5.8. The relative importance of the wind speed and the area in the resulting $IKE$ can be qualitatively assessed by comparing the graphs in figure 5.8 with figure 5.7. The peaking $IKE$ around time 250 h is given partly by the big TC object area and partly by the peaking maximum winds. The increase in $IKE$ from time 0 h to time 120 h is on the other hand induced by the increasing wind speeds. The slight decrease in $IKE$ after passing over Cuba is mainly due to the momentary decrease in wind speeds after landfall. The sharp decrease after landfall in the U.S. is again a combined effect from wind and TC area.

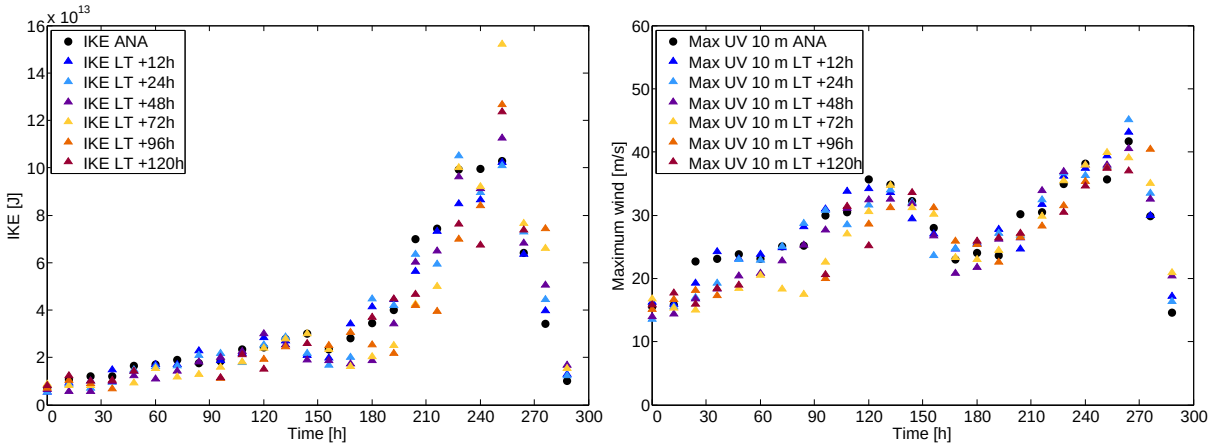



**Figure 5.8:** *IKE and maximum wind of the identified objects for hurricane Ike ECMWF analysis and forecasts.*

### Structure

The structure component $S$ aims at verifying the correct representation of the wind field in the identified object. This can be done in different ways:

1. Verify the location of the peak wind using a radial ($R$) and an azimuthal ($AZ$) information:

$$S1_R = \frac{R_{\text{fct}} - R_{\text{ana}}}{\frac{1}{2}\left(R_{\text{fct}} + R_{\text{ana}}\right)} \tag{5.13}$$

$$S1_{AZ} = \frac{AZ_{\text{fct}} - AZ_{\text{ana}}}{\frac{1}{2}\left(AZ_{\text{fct}} + AZ_{\text{ana}}\right)} \tag{5.14}$$

2. Verify the whole wind field:

$$S2 = \frac{Vs_{\text{fct}} - Vs_{\text{ana}}}{\frac{1}{2}\left(Vs_{\text{fct}} + Vs_{\text{ana}}\right)} \tag{5.15}$$

The scaled volumes $VS_{\text{fct}}$ and $VS_{\text{ana}}$ are defined as follows

$$Vs_{\text{fct}} = \frac{1}{N} \sum_{i,j} \frac{U_{\text{fct},i,j}}{U_{\text{fct,max}}}$$

$$Vs_{\text{ana}} = \frac{1}{N} \sum_{i,j} \frac{U_{\text{ana},i,j}}{U_{\text{ana,max}}}$$

| | | |
|---|---|---|
| where | $Vs$ | scaled average wind volume of the object [-] |
| | $N$ | number of grid points in object [-] |
| | $U_{i,j}$ | wind velocity at 10 m above ground (sea) [m/s] |

3. Verify individual wind subobjects inside the identified pressure object. The structure component is then defined in a similar way as in Wernli et al. (2008). The wind subobjects can be defined using the 80% percentile threshold value from the wind field inside the pressure object. Then a "scaled wind volume" (similarly as the scaled precipitation object in (Wernli et al., 2008)) $V_o$ can be calculated as:

$$Vs_o = \sum_{i,j}^{nx,ny} \frac{U_{o,i,j}}{U_o^{\max}} \tag{5.16}$$

| | | |
|---|---|---|
| where | $Vs_o$ | scaled wind volume of the subobject $o$ [-] |
| | $nx \cdot ny$ | number of grid points in subobject [-] |
| | $U_o^{i,j}$ | wind velocity field of subobject $o$ |
| | | at 10 m above ground (sea) [m/s] |
| | $U_o^{\max}$ | wind velocity maximum of the field $U_o^{i,j}$ [m/s] |

The weighted mean $V_M$ of all $M$ subobjects' scaled wind volume is then computed for both the analysis and the forecast fields:

$$V_M = \frac{\sum_{o=1}^{M} Vs_o \cdot U_o}{\sum_{o=1}^{M} U_o} \tag{5.17}$$

The weights $U_o$ are defined by the total wind speed summed over all grid points inside the different subobjects. The normalised difference between the forecast and the analysis field is computed in order to obtain the $S$-component:

$$S3 = \frac{V_{M_{\text{fct}}} - Vs_{M_{\text{ana}}}}{\frac{1}{2}\left(V_{M_{\text{fct}}} + Vs_{M_{\text{ana}}}\right)} \tag{5.18}$$

Note that the third and the second method proposed above for the structure component are very similar. Method 3 uses an average scaled value instead of the total volume with only one object and a wind threshold of 0 m/s as it is done in the second method. The last two methods to compute the structure component of the proposed object-based verification scheme rather aim at revealing problematic aspects that should help to improve the model representation. The first method is more straight forward in terms of interpretation and thus targets at a model user oriented public, like a reinsurance company.

## 5.3 Damage Assessment with CatMos

Figure 5.9 summarises the procedure applied for insurance loss predictions with Swiss Re's catastrophe modelling scheme (CatMos), using ECMWF TC forecasts. In order to obtain probabilistic insurance loss predictions, the ECMWF probabilistic position and intensity forecasts from the EPS model are used as input data to CatMos. Conceptually, a so called wind "footprint" is first obtained by transforming the central pressure information into a wind distribution using the Holland formula (Holland, 1980) and a constant radius of maximum wind. A wind "footprint" example is shown for hurricane Katrina in figure A.24 of the appendix. Then the different loss modelling steps are performed in order to obtain a predicted insurance loss distribution. The loss predictions are normed by the observed loss. The latter is computed from the footprint of the NHC best track data.

One problem is that the observed track is sometimes longer or much shorter than the forecasts. The option of computing the losses by unit of track length was not adopted, because the insurance sector in general is less sensitive towards the exact track or track length. This drawback has to be kept in mind, while analysing the results and track as well as intensity plots have to be consulted.

The insured value data is taken from SwissRe's 2008 portfolio. The fact that the 2008 portfolio is used for the hurricanes chosen from the 2005 season has a negligible influence on the result. Furthermore, the losses computed only correspond to wind damages. No estimations were made for the consequences of storm surge. An adequate modelling framework for this component is not yet ready. For details on Swiss Re's model the reader is referred to the Data chapter, section 3.5.

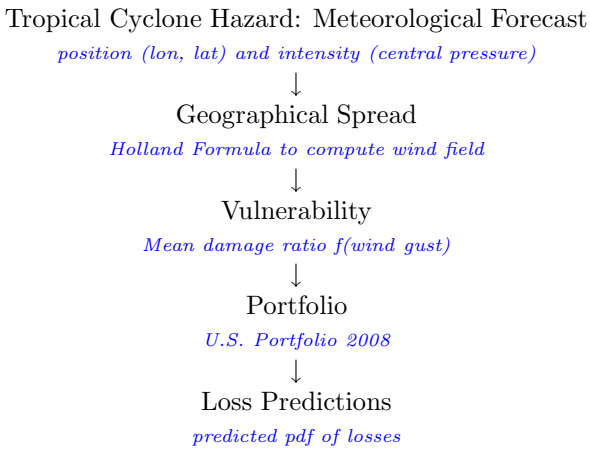



**Figure 5.9:** *Methods of insurance loss predictions using ECMWF TC ensemble forecasts.*

Because of the difficulties in the representation of correct pressure information, different loss modelling cases were performed for the selected hurricanes. In a first step the ECMWF pressure forecasts were taken as they were. In a second step, the ECMWF pressure forecasts were corrected for bias, using the pressure bias found from the application of the storm tracking verification approach. Two lead times were chosen per hurricane event. If available, the forecast ranges of 1 and 3 days were chosen. Especially the second one is a highly relevant time for Swiss Re in terms of short-term management of losses (see chapter Case Study, section 4.1). The hurricanes chosen for loss modelling are described in the Case Study chapter, section 4.4.

# Chapter 6

# Results and Discussion

In this chapter, the results of the storm tracking verification of the ECMWF TC ensemble forecasting system are presented first. Then, preliminary results of the application of an object-based quality measure to the ECMWF deterministic atmospheric model data are shown. Finally, the performed loss predictions for 5 selected TCs are assessed.

## 6.1 Storm Tracking Verification of the ECMWF Tropical Cyclone Ensemble Track Forecasting System

The aim for this first results section is to assess the ECMWF forecasting system of TCs with methods commonly used in forecast verification by operational centres (see section 5.1). Forecast error statistics were computed for hurricane events of the years 2005 to 2008 (see section 4.2). In this period some model updates have occured as described in section 3.1.

In the following, the ECMWF track and intensity forecasts are discussed in terms of accuracy, reliability, statistical consistency, skill and linear association (see chapter on Theories and Concepts, section 2.3 for definitions). An analysis of sensitivity of the forecast errors and biases towards the interpolation technique, the model resolution change, the location over land or sea and the intensity of the hurricanes is then presented. Finally, the question of equal likelihood of ensemble members is addressed.

### 6.1.1 Accuracy and Reliability

#### Position Errors and Biases

Figure 6.1 shows the direct position error $DPE$ of ECMWF TC forecasts as a function of lead time for the years 2005-2008. The boxplots represent the ensemble, which is compared to the operational forecast (in green), the deterministic run of the EPS (control run CTL in violet) as well as to the NHC forecast (in orange). The climatological and persistence prediction CLIPER (in grey) serves as a baseline for forecast skill. The shift at lead time 0 between observations and initial conditions of the forecast can be attributed primarily to the model resolution in the case of the ECMWF model and the post-storm adjustment of best-track data in the case of NHC warnings and CLIPER forecasts. This problem was already discussed in the methodology chapter, in section 5.1.1. The yellow and grey stars corresponding to NHC and CLIPER errors respectively, indicate the value of the direct position error $DPE$ for these forecasts, if the whole predicted track is shifted by the difference between the respective forecast (CLIPER or NHC) and ECMWF mean initial error. This can be seen as a way of assessing the uncertainty of the direct position error $DPE$ with respect to the problems of resolution and post-storm data corrections.

The accuracy of the ECMWF track forecasts decreases linearly with increasing lead time. The slope in the $DPE$ of the operational forecast is smaller ($\sim$ 33 km for 12h additional forecast range) than the slope of the median error of the ensemble ($\sim$40 km/12 h). The increase in the mean error of the ensemble is even larger with 50 km for 12 h additional forecast range. In an assessment of the forecasts in the

years 2002 to 2004 by Van der Grijn et al. (2004), a slope of 45 km per 12h additional forecast range was found for the operational forecast, as well as the EPS mean and the control run. In their study all TCs of the mentioned time window were considered, which probably explains the slightly smaller slope for the ensemble (not only hurricane force TCs as here). However, the operational error increase per additional 12h forecast range is found to be much lower in figure 6.1 than in Van der Grijn et al. (2004), indicating the importance of the resolution improvement from 40 km to 25 km for the track representation.

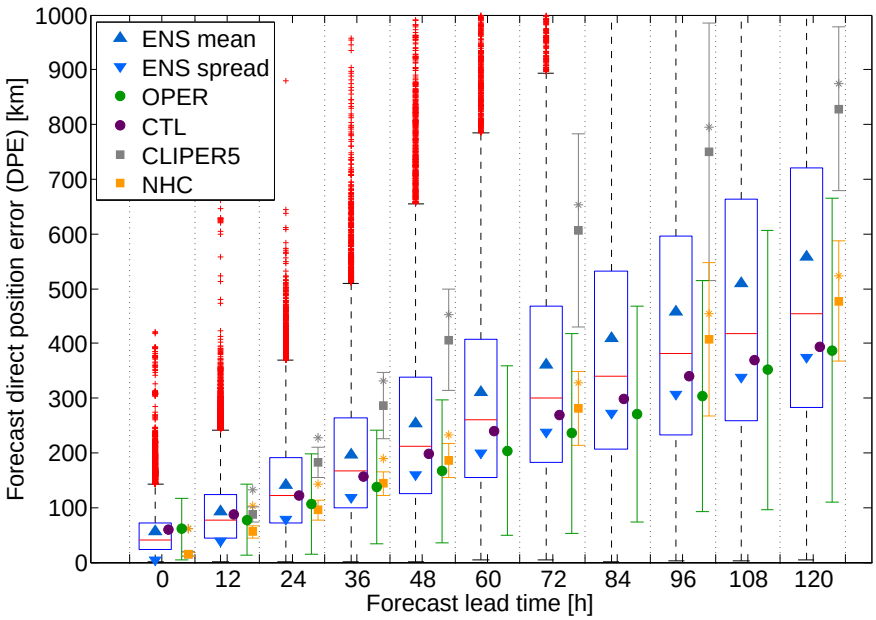



**Figure 6.1:** *Direct position error of ECMWF operational and ensemble forecasts, climatological and persistence CLIPER data and NHC official forecasts for the selected hurricanes of the years 2005 to 2008. Stars indicate uncertainty introduced by initial position error.*

The increasing difference between the ensemble mean DPE and the median $DPE$ indicates that the error distribution of the ensemble becomes more and more skewed with increasing forecast range. Deviations from normality of the ensemble forecasts lead to biases in the ensemble mean (Stephenson and Doblas-Reyes, 2000). For positively skewed error distributions, as it is the case here, the mean is drawn towards the large positive values, whereas the median is more "robust" to such extremes. If one is interested in tendencies, a notable advantage of the median over the mean is that there is equal chance of forecasts falling either above or below a specific value. The interquartile range of the ensemble becomes larger as a result of a widening error distribution. This may be due to decreasing predictability for longer forecast ranges as indicated by the strongly increasing average CLIPER error. Another problem is the decreasing sample size for longer lead times as shown in figure 6.2. Indeed for the ensemble, the sample size is reduced by 70% of the initial sample size at 5 days lead time. This is considerable and induces a higher uncertainty in the long-range forecast error computation. The decrease in sample size can be explained by the fact that, when reaching the "end" of the available best track data, meaning when the storm dissipates, the forecast will contain the storm only in the short-range lead times. In the early phase of the storm, long-range forecasts are available only after the storm's age has reached the respective forecast range. Furthermore, as mentioned in section 3.1.4, certain forecast members indicate early dissipation and are thus missing in the statistic.

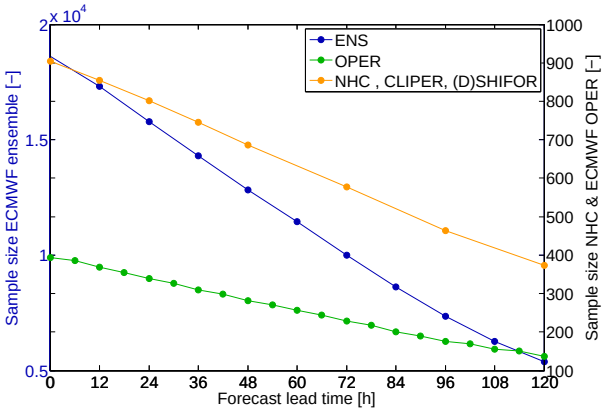



**Figure 6.2:** *Sample size of ECMWF ensemble, CLIPER and NHC official forecasts for the selected hurricanes of the years 2005 to 2008.*

The operational forecast is more accurate than 50% of the ensemble members between lead times 24 h and 120 h. However, on average the operational forecast is affected by an error, which is included in the interquartile range of the ensemble error distribution, meaning that on average there are individual ensemble members that are as good as, or event better that the operational forecast. The control forecast yields a higher error, than the operational forecast for lead times between 24 h and 108 h with a maximum difference at around 60 h lead time. However, for 12 h and 120 h the errors of the two deterministic runs are comparable. This gives an idea about the forecast time scale for which resolution probably plays the most important role, namely around 1-4 days forecast range. When comparing the ECMWF TC forecasting system as a whole, considering the ensemble and the operational forecasts, with the NHC "human synthese" forecast (see Data chapter, section 3.2.1), the NHC performs better in the short-range. From 36 h lead time up to 120 h, the ECMWF performs better. At 96 h to 120 h lead time, more than half of the ECMWF ensemble members perform better than the NHC forecast.

Overall, considering the uncertainty indicated by the orange star coming from the system inherent error, partly due to the "inexact" representation of the initial position in the ECMWF analysis and partly due to post-storm corrections of best track data, the two forecasting systems NHC and ECMWF are comparable in terms of accuracy of the track position. Both ECMWF forecasting systems show some skill, when compared to the baseline forecast CLIPER between lead time days 1 to 5. The good performance of the climatological forecast for the 12 h forecast range has to do again with the uncertainty of the ECMWF model initialisation.

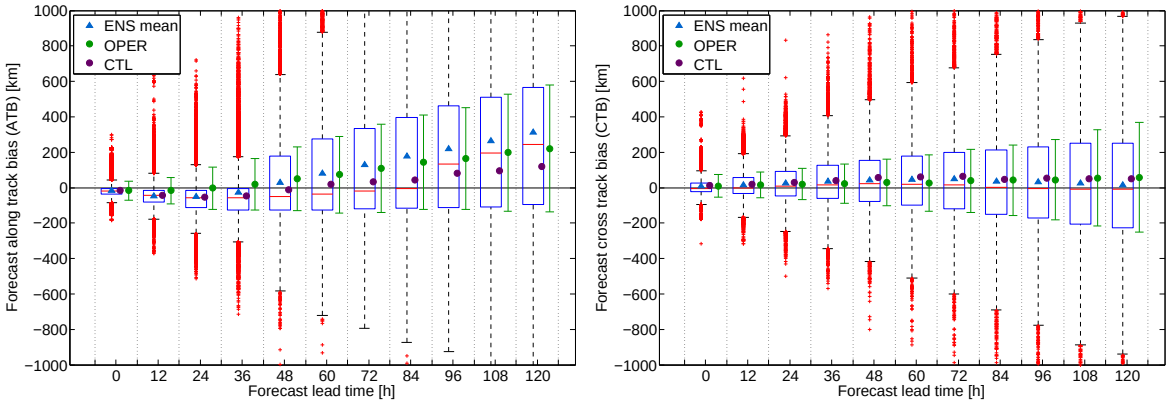
The classical underdispersivity problem of today's EPS is clearly apparent in figure 6.1. As mentioned in the introduction, for an EPS to be statistically reliable the ensemble mean, when compared to the observations should exhibit the same error as the average difference between the individual members and the ensemble mean. In other words, the ensemble mean error should be reflected in the ensemble spread, which should indicate the level of uncertainty inherent in the forecasting system. The underestimation of the spread in this study augments for increasing forecast ranges and reaches 200 km for 120 h lead time. The model is thus less and less capable of covering the whole range of physically consistent and realistic evolutions. In Froude (2009), the difference between the ensemble mean error and spread was found to be smaller. The latter study, however, analyses the performance of the ECMWF EPS model in forecasting extratropical cyclones. This apparent difference in reliability of the ECMWF ensemble forecasting system in handling TCs and extratropical cyclones comes probably mainly from the difference in the perturbation method used to generate the ensemble members. Apparently the tropical perturbations targeted at TCs do not sample initial errors as well as perturbation done in the mid-latitudes (see section 3.1.4 of the Data chapter for details about the perturbation technique used by ECMWF in the tropics). Underdispersivity can be seen as the symptom of the fact that today's NWP models do not account for all possible sources of error and especially underestimate the structural uncertainty caused by the model parametrisation being inaccurate. Fovell and Su (2007) found evidence for significant track sensitivity to cloud microphysics details, which supports the concept of conducting ensemble forecasts using not only initial perturbations but also different model physics parametrisation in order to obtain an objective measure of forecast uncertainty (Buizza et al., 1999; Pellerin et al., 2003; Krishnamurti et al., 1999).

Figure 6.3 shows the along track and cross track biases $ATB$ and $CTB$. The corresponding error statistics $ATE$ and $CTE$ are shown in the appendix, figure A.2. The latter figure as well as figure A.7 in the appendix show that $ATE$ is the dominant error component for forecast lead times between 12h and 120h. For short forecast ranges the $ATE$ dominance is not as pronounced as for longer ones, for which $ATE$ explains $\sim 65\%$ of $DPE$ . The underdispersivity problem is present in both error components.

The cross track component does not exhibit a pronounced bias. For short lead times the along track component shows a small slow bias of 40 km on average for the ensemble, which is negligible, because it is at the limit of the grid spacing of the EPS. In 2004, Van der Grijn et al. (2004) mentioned a strongly negative $ATB$. This bias has been nearly removed, through the model changes that have occured since 2004. For long forecast lead times the TCs propagate on average too fast now. This may be explained by two factors. Firstly, the significantly diminishing sample size has to be kept in mind. Secondly, the long range forecast times contain a higher proportion of TC positions over land than short lead times (see figure 6.10). Landfalling changes the way the TC evolves drastically, especially in terms of energy source (Lohmann, 2009). The first negative, then positive bias in propagation speed might be indicative

of difficulties in the representation of TCs' vertical structure, intensity and energy budget. The non-interaction between the atmospheric and the oceanic components may be one important reason for the named structural errors.

It can be said that the propagation direction (resulting in a $CTE$) is more easy to handle for the model, because it is more affected by the synoptic situation, than by the TC structure itself. The propagation speed, however, seems to be more closely linked to the inner-core dynamics and thus processes that take place at smaller spatial scales (Froude, 2009; Fovell and Su, 2007). This is how the dominance of the along track component in the direct position error might be explained.




(a) Along Track Bias (b) Cross Track Bias

**Figure 6.3:** *Bias in along (a) and cross track (b) components of the ECMWF operational and ensemble forecasts for the selected hurricanes of the years 2005 to 2008.*

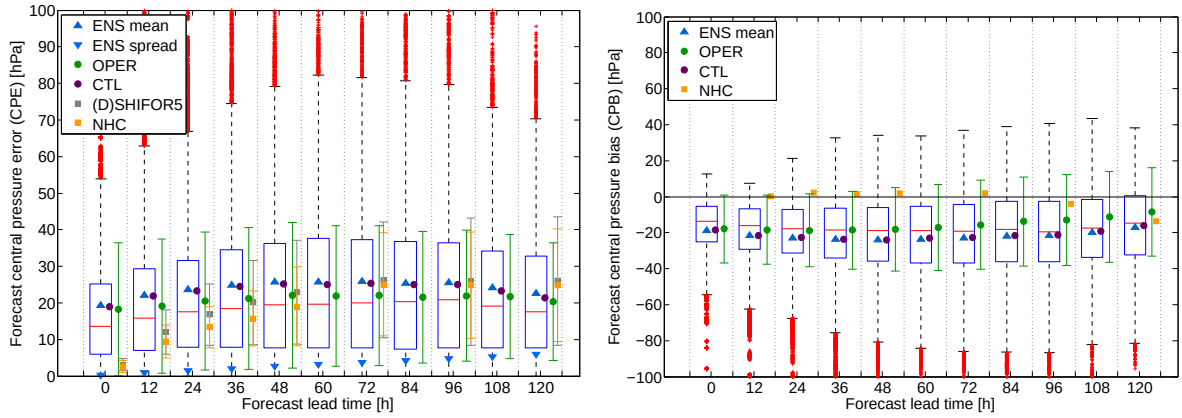
The analysis of the longitudinal and latitudinal errors in track forecasts is similar as the one above (see figures A.4 in the appendix), because the definition of these components simply involve a rotation of the reference axis from the direction of propagation of the storm to an axis pointing north.

### Intensity Error and Bias

The growth rate of the intensity error is not linear. This can be seen in figure 6.4. In the case of the ECMWF ensemble and operational forecasts, there is only a slight 5 hPa increase between initialisation and lead time 72h. The general form of the central pressure error curve for ECMWF is concave, with a high initial error, when compared to the NHC or even to the (D)SHIFOR baseline, of approximately 20 hPa. This initial error probably comes from the smoothing of extremes by the NWP model gridding. The minimum pressure in the center of the storm is a localised phenomena that is present over very small scales of a few tens of kilometres with a strong radial gradient leading to it. Even a 25 km resolution does not allow such a strongly peaking intensity to be built up in the model. The reason, why the NHC performs much better, is given partly by the fact that they can extrapolate directly the observed pressure and do not need to perform an analysis like ECMWF, which smoothens the observational data and leads to a high initial error. The statistical "no skill" intensity forecast also performs better than ECMWF up to lead time 48 h for the same reason. Furthermore, NHC uses many models and especially also local area models, which can resolve the small scale high intensity characteristic better.

The concave form of the ECMWF curve/levelling off of the NHC curve with decreasing/constant central pressure errors for long-term forecast ranges is not significant. The following explanations are merely suppositions. As already mentioned in section 6.1.1, lead times 84 h to 120 h are affected by proportionally more cyclones over land (see also figure 6.10). Storms are generally less intense over land, because they are practically deprived of their main energy source (Lohmann, 2009). Moreover, the described drastic decrease in sample size (see figure 6.2) may also contribute to the levelling off in intensity error, because storm dissipation plays an increasing role at these long lead-times. A forecast that forsees early dissipation is not further penalised because it does not correctly represent the intensity. Even if, in fact

such a dissipation forecast is even worse than a forecast that badly reproduces the observed intensity. This artefact is difficult to remove from the present analysis, but it has to be kept in mind as a significant factor influencing the error statistic. The central pressure error curve is different in the case of the NHC and (D)SHIFOR forecasts. The sample sizes of these two models do not decrease as strongly as the one of the ECMWF on the one hand, on the other hand it could also be interpreted as a sign that these forecasting systems do not capture long term changes (strengthening or weakening of the cyclone) very well and that given its physically based numerical integration the ECMWF is better at handling these. The latter comment does also apply to the direct position error, for which the ECMWF model was also the more accurate model for long-range forecasts.




(a) Central pressure error

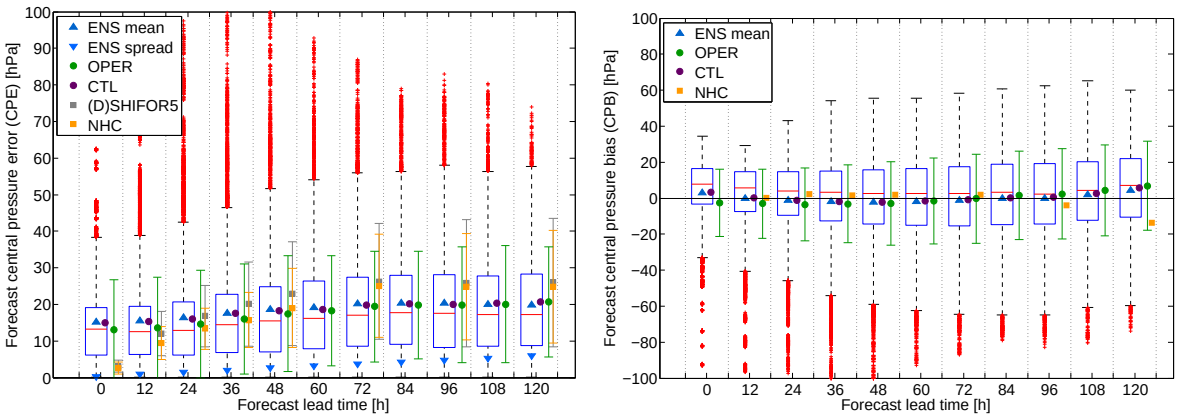
(b) Central pressure bias

**Figure 6.4:** *Central pressure error (a) and bias (b) of the ECMWF operational and ensemble forecasts, statistical (D)SHIFOR intensity data and NHC forecasts for the selected hurricanes of the years 2005 to 2008. On the right (a) negative values mean that the forecast pressure is too high.*

There is a substantial average bias in the ECMWF ensemble as well as operational forecasts, with little dependence on the lead time. Posterior model calibration could be done for practical applications like insurance loss predictions to improve results from further modelling, which depend strongly on the correct representation of intensity. Because loss modelling at Swiss Re is done using only an information on position and central pressure, the correction of the intensity bias can be done easily. Figure 6.5 illustrates the implication of such a bias correction on the central pressure error and bias.

The NHC bias shown in panel (a) of figures 6.4 and 6.5 represents an average over all TC forecasts. Numbers stem from the yearly verification reports. The long term NHC bias is very small and shows a tendency for negative values in long-range forecasts (3-4 days, intensity underestimation). Beven et al. (2008) discusses the long term intensity biases in NHC forecasts in detail.

For the insurance loss predictions presented in section 6.3.2, a central pressure bias correction was applied by removing the ensemble average pressure error over 4 years from all ensemble members except for the operational track (ECMWF deterministic, high resolution model). The operational intensity was corrected by using a slightly lower value, also obtained form a 4 year statistic of the pressure error of the operational forecast. As the pressure error does not depend strongly on the lead time, the applied correction can be set constant. The magnitude of the correction is +21.7 hPa for the ensemble (members 1 to 51) and +15.3 hPa for the operational forecast (member 0). In this thesis insurance loss case studies were taken from the seasons 2005 and 2008. The model resolution change may have a measurable effect on central pressure forecast accuracy and bias (see section 6.1.5). For future applications thus lower values of +18.1 hPa for the ensembles and +11.3 hPa for the operational forecasts, which were computed for the events of the seasons 2006-2008 (after resolution change in February 2006) should be applied. Investigations on the central pressure bias dependency on observed and forecast central pressure and central pressure change between 2 time steps were conducted. These are presented in the appendix in section A.4.




(a) Central pressure error, after bias correction

(b) Central pressure bias after bias correction

**Figure 6.5:** *Central pressure error (a) and bias (b) of ECMWF operational and ensemble forecasts for the selected hurricanes of the years 2005 to 2008 after bias correction (calibration). On the right (a) negative values mean that the forecast pressure is too high.*

After bias correction, the ECMWF EPS and operational forecasts are approximately equivalent in terms of forecast accuracy, as each of the two systems were corrected individually for their bias. In terms of accuracy, the ECMWF system improves after bias correction, when compared to NHC forecasts. The ECMWF forecasts are on average equivalent or even more accurate than the NHC forecasts from the forecast range of 2 days on and beyond. For the lead times 72 h to 120 h ECMWF is on average more accurate than NHC by 5 hPa. The average error of the intensity forecast by the ECWMF EPS after bias correction is around 18 hPa, which represents approximately one Category on the Saffir-Simpson scale. For short lead times the error is slightly lower but there is still no clear dependency of the error on the lead time as it is the case for the NHC forecast.

It seems that with a bias corrected ECMWF model, the latter should be prefered for lead times above 2-3 days, whereas for short lead times the NHC forecast is better. This observation makes sense from a warning perspective. The role and primary goal of NHC is precisely to provide excellent short-range forecasts. From an insurance point of view the longer range forecasts by ECMWF might be more adequate.

### 6.1.2 Statistical Consistency

In figure 6.6 the percental underdispersivity ($\frac{m-s}{m}$) of the ECMWF EPS found in the position and intensity forecasts is shown. The underdispersivity relative to the mean error $m$ is greater for the central pressure forecast than for the track components. As in the case of accuracy, the propagation speed component dominates underdispersivity in track forecast.

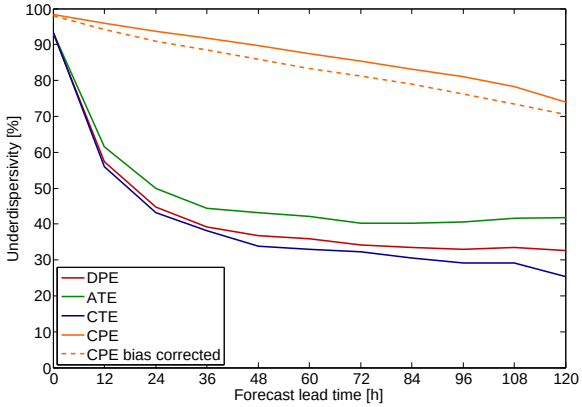



**Figure 6.6:** *Underdispersivity in track components and intensity forecasts.*

### 6.1.3 Forecast Skill

The skill of the track and intensity forecasts is shown in figure 6.7 and 6.8 respectively. Skill is computed as a relative difference between the forecast error and the error of the trivial forecast (CLIPER, (D)SHIFOR), normalised by the error of the trivial forecast (see section 5.1.4 in the methodology chapter for detailed description). Skill of track and intensity forecasts summarises the statements of the above two sections 6.1.1 and 6.1.1. Because of the high initial error in terms of track and intensity of the ECMWF model, its skill is negative for the analysis time as well as the 12 h lead time forecast. This is especially due to the very low climatological errors of the CLIPER and (D)SHIFOR models respectively and might be attributed to the gridding and smoothing affecting the ECMWF model accuracy.

For both track and intensity forecasts all the examined models have skill for lead times beyond 24 h. The NHC has more skill for short-range forecasts, whereas the ECMWF system is better on longer time scales. Considering only the ECMWF forecasting system, the operational forecast has the highest skill in terms of track forecast. The skill obtained from the ensemble median error in track and intensity is always higher than the ensemble mean, implying that more than 50% of the members perform on average better than indicated by the skill of the forecast mean.

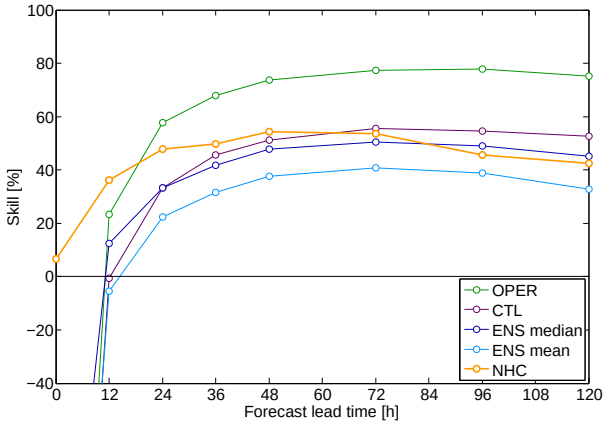



**Figure 6.7:** *Skill of the position forecast by ECMWF and NHC for the selected hurricanes of the years 2005 to 2008 with respect to CLIPER.*

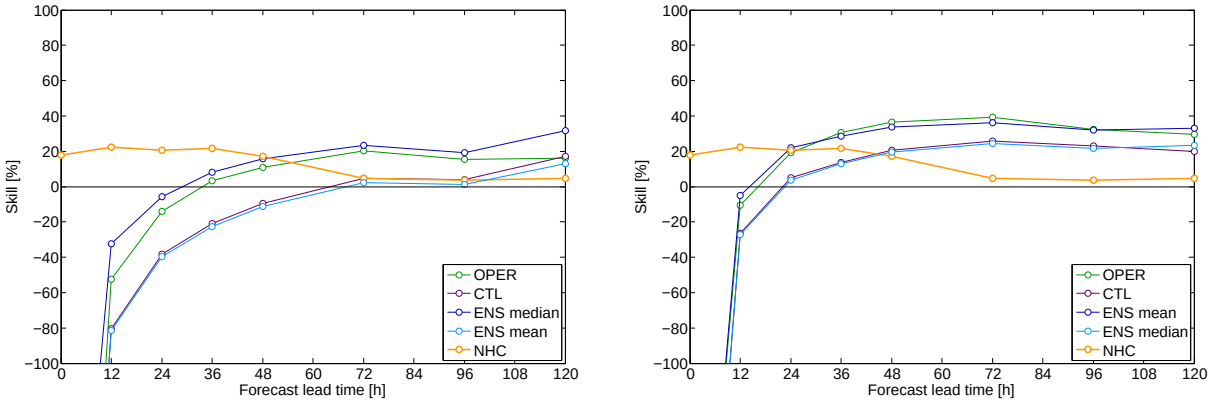



**Figure 6.8:** *Skill of the intensity forecast by ECMWF and NHC for the selected hurricanes of the years 2005 to 2008 with respect to (D)SHIFOR data. Left: Skill of ECMWF central pressure raw forecasts. Right: Skill of ECMWF bias corrected central pressure forecasts.*

The computation of a skill measure for both track and intensity does not only allow to measure the quality of forecasts with respect to baseline forecasts, but it also provides a framework to compare these two principal forecast elements of TCs. As mentioned in other scientific studies (Davis et al., 2008; Franklin, 2006, 2009) and even in the American press (Spinner, 2009), the present modelling efforts in NWP show satisfying skill in TC track forecasts, but provide only limited performance in terms of intensity forecasts

with skill values around 20% (e.g. Froude (2009), Froude et al. (2007), Davis et al. (2008)). The annual verification reports by NHC also show that progress in the last decade in this matter has been significantly slower than in the case of track errors (Franklin, 2009). The increase in track forecasting skill has been substantial in the last few years, starting from values around 0 in the 1990, following Franklin (2009) and reaching values up to 50-60%, even 80% in the case of the ECMWF operational track forecast in 2005-2008, as found here. However, the decrease in track error for short lead times has been only of a few tens of km as shown in Franklin (2009). Following NHC yearly forecast verification reports (e.g. Franklin (2009)) in the case of intensity forecast errors the situation since 1990 has evolved only slightly by 5 hPa decrease at maximum. Skill has, however, increased by around 10 %, still remaining very low, with values found in this work of around 20%, maximum 30% in the case of the bias corrected ECMWF forecast in 2005-2008. The increase in skill of track and intensity forecasts, which is not reflected in a corresponding decrease in errors may indicate an improvement in predictability in the last decade. This tendency might be induced by the better spatial coverage of measurements. To our knowledge the influence of changes in observational density has not yet been thoroughly studied in the case of TC forecasts. Froude et al. (2007) have investigated the impact that observations of different types have on the prediction of the extratropical cyclones. They used forecasts integrated from analyses that were constructed using reduced observing systems. Since the skill of TC forecasts, at least in the short term, strongly depends on the NWP model initialisation, observation systems effects should be investigated as well in the case of TC forecasts.

The lower skill of intensity forecasts, when compared to the skill of track forecasts can be attributed to the fact that intensity description involves a range of scales. It especially also depends on very small scale microphysical processes (Fovell and Su, 2007; Davis et al., 2008). Froude (2009) found that TCs that move into the extratropics are affected by higher average intensity errors than storms that originate in the extratropics. This shows the importance of the current "intensity" problem in forecasting TCs and their effects. NHC has set a goal of improving the hurricane intensity forecasts by at least 20% in 5 years and 50% in 10 years (Spinner, 2009). Thus developments can be expected in the next few years in this area.

### 6.1.4 Correlation Analysis between Forecasts and Observations

A good summary of linear association between forecast and observations for both track and intensity is given by the Pearson correlation shown as a function of lead time in figure 6.9. On average over all lead times the best correlation is reached by the longitudinal track component forecast. It is important to point that this is not a contradiction to the dominance of the longitudinal error in terms of accuracy of the track position (see figure A.7 in the appendix). Correlation is a relative measure and can indicate good performance independently of the accuracy, as it is normed by the variances of the forecast and observation variable of interest. Moreover, the difference in correlation between latitude and longitude is negligible (0.99 on average over all lead times for longitude and 0.96 for latitude). However, the correlation of forecast and observation in the case of pressure is clearly weak (0.33 on average). This might be due to the bad representation of pressure changes over time, meaning the model seems to be affected by a certain inertance concerning pressure changes. Correlation of pressure changes in the model, when compared to the changes found in the observations is very low (-0.01 on average). The simulation of several hurricane examples (hurricanes used for loss predictions, see appendix figures A.17 to A.23) showed that the model might indeed be too slow in the intensification phase but also too reluctant in the attenuation.

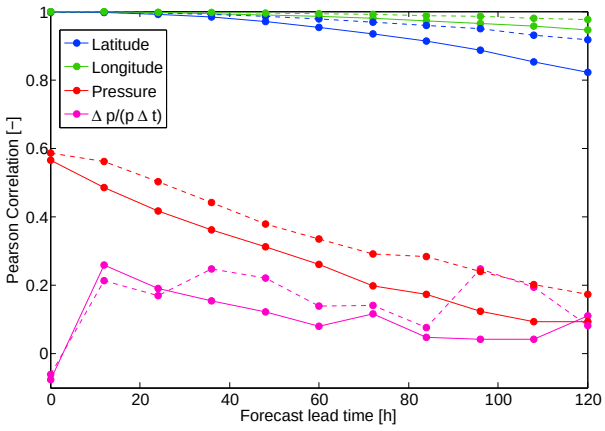



**Figure 6.9:** *Correlation analysis for latitude, longitude, central pressure and central pressure change as a function of lead time. The full lines represent the Pearson correlation between ensemble forecasts and the observations, the dashed lines represent the correlation between the operational forecasts and the observations.*

### 6.1.5 Sensitivity Analysis of Forecast Errors

In the following sections the sensitivity of the above results towards different elements is analysed, such as the used interpolation method, the model resolution change, the hurricane intensity and landfalling properties as well as towards effects of the cyclone's location over land or sea. For all these assessments one common problem will always be present, namely the problem of difference in sample size between the groups to be compared. The decrease in sample size for longer forecast ranges has to be kept in mind as well. Figure 6.10 summarises the sample sizes as a function of lead time for the ensuing sensitivity analyses.

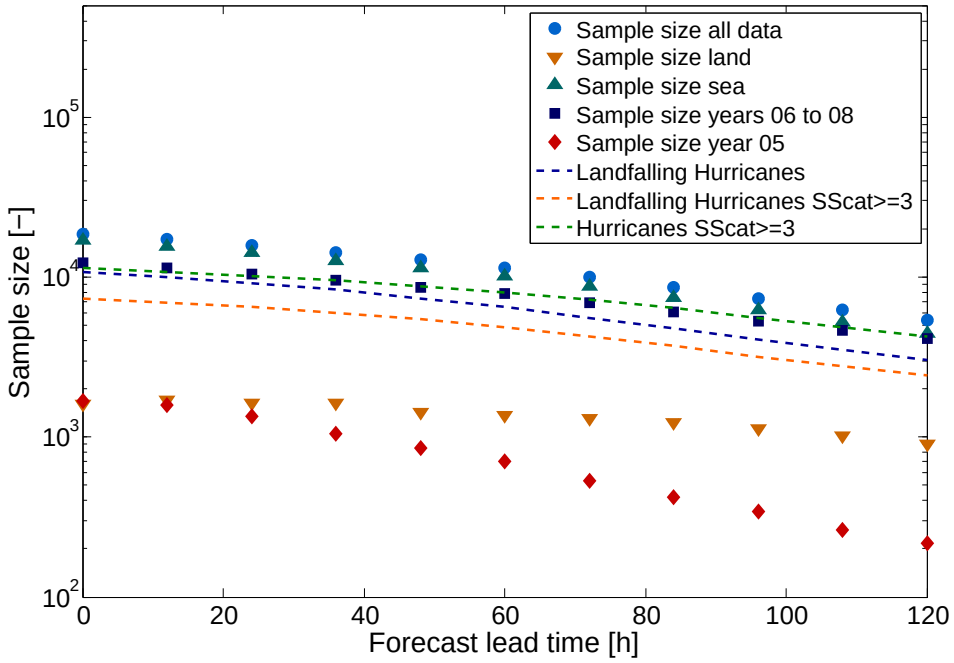



**Figure 6.10:** *Sample sizes for investigated years, sensitivity towards ECMWF model resolution change in February 2006, land/sea effects and the intensity/landfalling groups.*

#### Sensitivity towards Interpolation Method

The method used to interpolated the data in time does not affect the results obtained for the track and intensity statistics. A detailed description of the sensitivity of the results towards the interpolation method for one hurricane example event can be found in the appendix in section A.3.1.

#### Sensitivity towards Model Resolution Change

The ECMWF model horizontal and vertical resolution change appears to have a positive effect on both positional and intensity forecast accuracies. The $DPE$ diminishes for long lead times, which goes in hand with a weaker error growth for increasing lead times (see figure 6.11).

The results are affected by differences in the sample size. Further on, the season 2005, in which many strong hurricanes occured, is compared to 3 other years of lower activity.

#### Sensitivity towards Hurricane Intensity and Landfalling

The influence of hurricane event sampling on the error statistics was investigated and the results are shown and described in the appendix in section A.3.2. Properties of the whole TC track were used as a classifying element. The different analysed groups are presented in section 4.2 of the case study chapter. Track and intensity errors and biases are not found to be highly influenced by the intensity and landfalling properties of the TC.

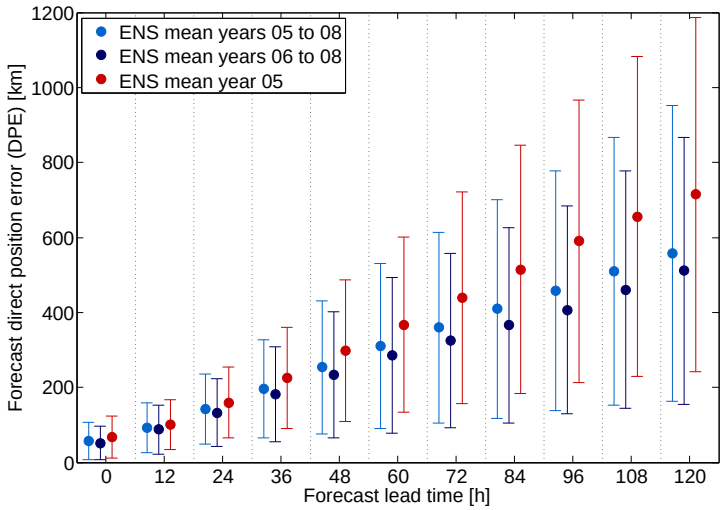



**Figure 6.11:** *Direct position error for different model resolutions. ECMWF model resolution changed in February 2006 (see Data chapter, section 3.1).*

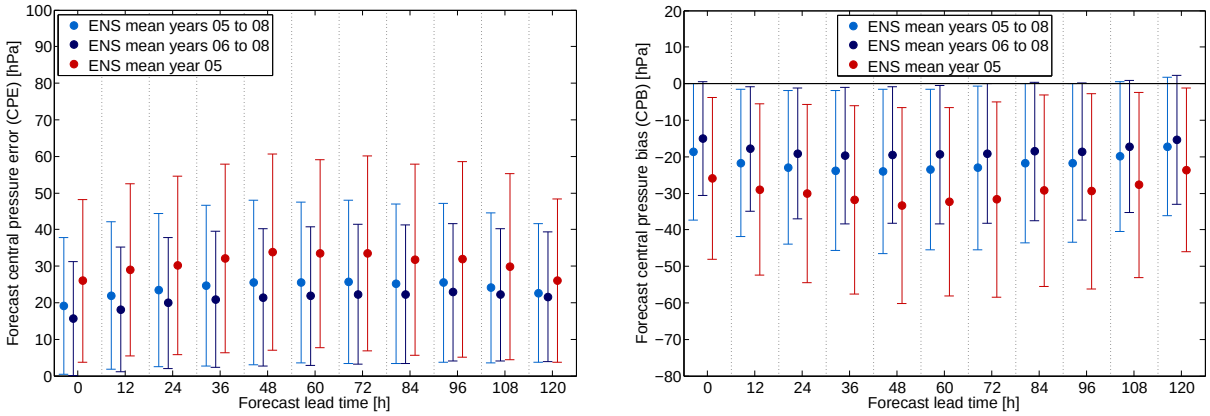



**Figure 6.12:** *Error in central pressure forecast, differences potentially due to ECMWF model resolution change in February 2006. On the right negative values mean that the forecast pressure is too high.*

### Sensitivity towards Location over Sea/Land

When analysing the track and intensity error dependency on land/see effects, differences in available observations have to be kept in mind. Land regions are generally better covered by observations, which improves the analysis and short term forecasts. However, it also helps providing better best track data, which in turn implies a harder benchmark. These arguments may apply in the case of the track and intensity errors shown in figures 6.13 and 6.14. In all these graphs, the error over land is lower for short-range forecasts, when the analysis step was probably made around landfall time. On longer time scales the error however is much higher, which might be due to cases just after landfall, from forecasts that were initialised over the open sea and are then compared to the harder observation benchmark. These suppositions are only hypotheses and have to be verified by computing the proportion of forecasts initialised over sea and over land for all lead times. The smaller pressure error over land for short lead times might be reinforced by the fact that storms are stronger over the sea and thus the forecasts are generally affected by higher errors. The fact that the land $DPE$, $CPE$ and $CPB$ curves are less smooth than the ones for TC positions over sea is probably due to the much smaller sample size of TC positions over land.

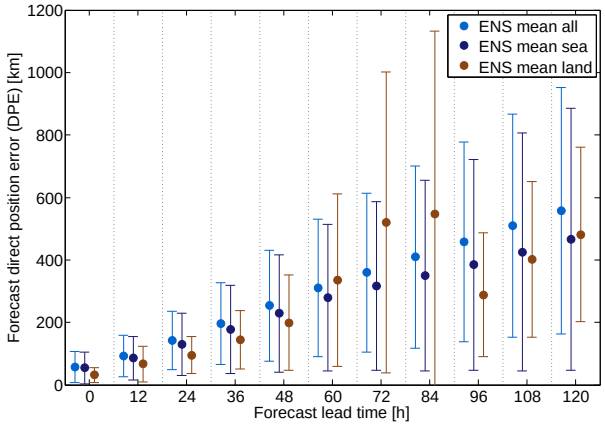



**Figure 6.13:** *Error in track forecast, direct position error, separated into land and sea locations of the cyclone over land/sea.*

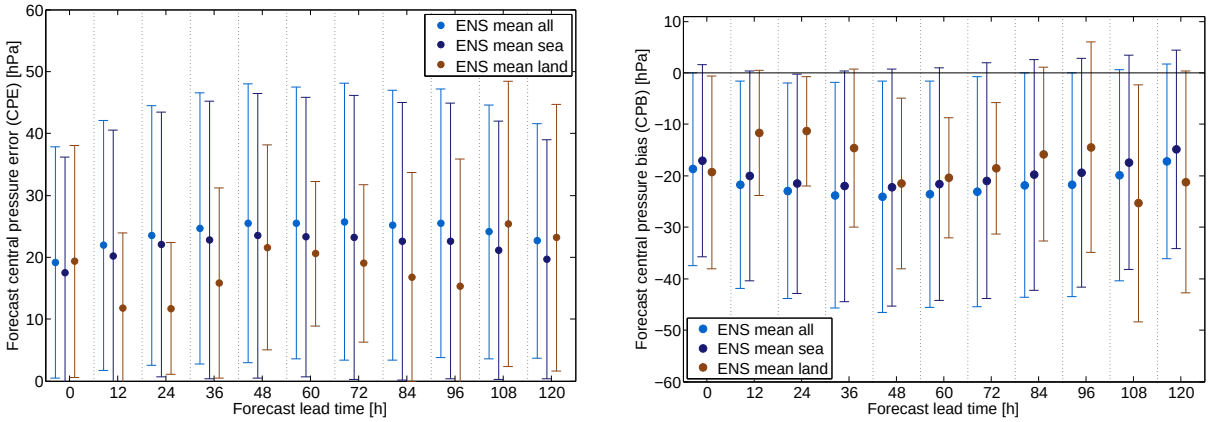


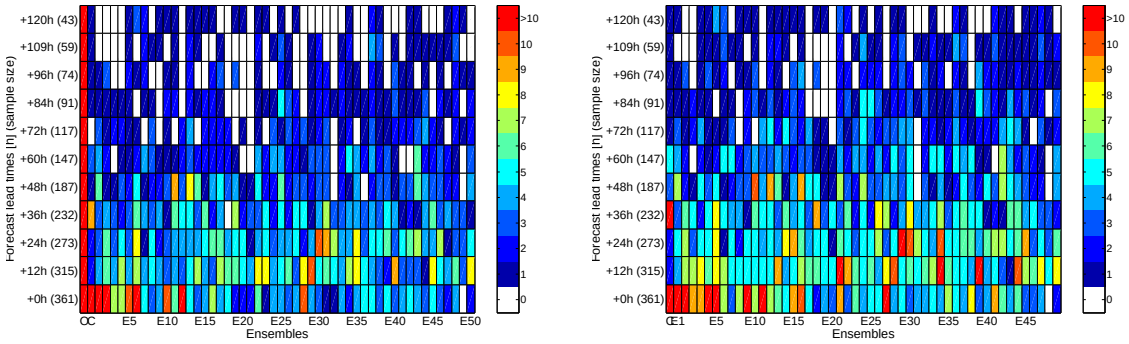

**Figure 6.14:** *Error in central pressure forecast, sensitivity towards location over land/sea. On the right negative values mean that the forecast pressure is too high.*

### 6.1.6 Equal Likelihood Analysis of Ensemble Members

Figures 6.15 and 6.16 show the number of times each ensemble forecast member was the best in terms of track or pressure respectively, when compared to the observation for each lead time. Figures 6.15 and 6.16 panel (a) show the statistics for all the members including the operational forecast. On the right the same analysis is made but including only ensemble members at the EPS model resolution. The sample size decreases with longer forecast range for the reasons mentioned in section 6.1.1.

No statistical test of significance of the results shortly described in the following was made, only tendencies and hypotheses are presented. From the point of view of the track the best forecast seems generally to be given by the operational run for all lead times. There is no clear preferential member for long forecast ranges. On short time windows, there seem to be certain members that perform regularly better than others. Some individual members stick out up to lead time 48 h. The control run is clearly not regularly the best ensemble member, which emphasises the added value of an ensemble forecast over a deterministic one at the same resolution.

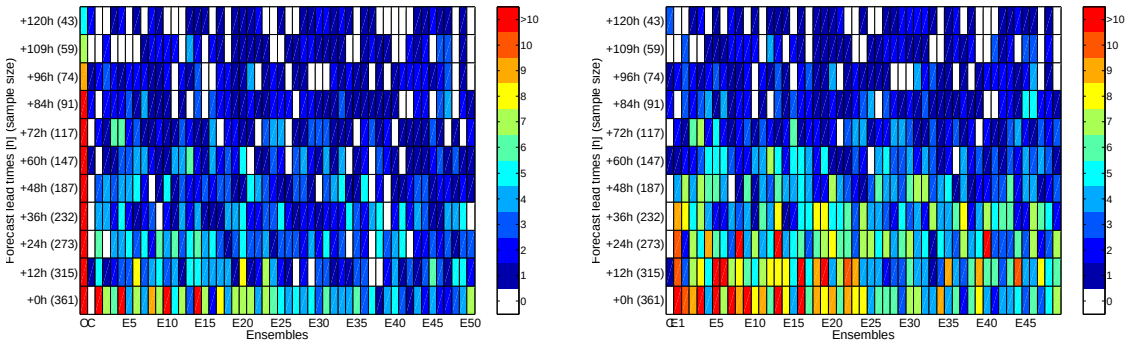
In the case of intensity the best forecast is generally given by the operational run, however with a less clear dominance than in the track forecasts, especially for long forecast ranges. The first few members up to member 20 seem to be more frequently the best available. The control run however is astonishingly rarely the best member in terms of central pressure. The bias in central pressure forecast may be a reason for this, giving advantage on average to members, which forecast lower central pressure than the control run.




(a) EPS with operational track forecast

(b) EPS without operational track forecast

**Figure 6.15:** *Likelihood of ensemble members in ECMWF **track** ensemble forecasts. Frequency at which individual ensemble members yield the best track forecast for the indicated lead times (vertical axis). The numbers in parenthesis show the sample size for each lead time. The graph on the left (a) shows the ensemble including the operational forecast. The first column from the left corresponds to the operational track, the second to the control run. The graph on the right (b) shows the same statistic for the ensemble alone. The first column from the left corresponds to the control run, the following to the ensemble members with increasing number.*




(a) EPS with operational intensity forecast

(b) EPS without operational intensity forecast

**Figure 6.16:** *Likelihood of ensemble members in ECMWF **central pressure** ensemble forecasts. Frequency at which individual ensemble members yield the best central pressure forecast for the indicated lead times (vertical axis). The numbers in parenthesis show the sample size for each lead time. The graph on the left (a) shows the ensemble including the operational forecast. The first column from the left corresponds to the operational track, the second to the control run. The graph on the right (b) shows the same statistic for the ensemble alone. The first column from the left corresponds to the control run, the following to the ensemble members with increasing number.*

## 6.2 Object-Based Quality Assessment of ECMWF deterministic Tropical Cyclone Forecasts

In this section, first results of an adaptation of the SAL framework (Wernli et al., 2008) to TCs are presented. The primary goal of this master thesis with respect to object-based verification was to formulate a proposition for a new method. A thorough testing and application to different case studies and for different models is out of the scope of this project. Only the event of hurricane Ike of the season 2008 could be investigated here. The sample size thus consists of 25 forecasts along the hurricane track and 6 lead times for each (12 h, 24 h, 48 h, 72 h, 96 h, 120 h). Sometimes no object is identified, because one of the established conditions described in the methods chapter, section 5.2.2 could not be fulfiled. The results are preliminary, they may, however, point out some interesting aspects.

In figure 6.17, the track of hurricane Ike is shown with indications of changes in intensity in terms of the Saffir-Simpson Scale classification. The three dates indicate times, from which examples of identified objects were chosen and presented in section 6.2.2. In the following section, the quality of the ECMWF analysis as a verifying observation is discussed. First verification results are shown in section 6.2.3.

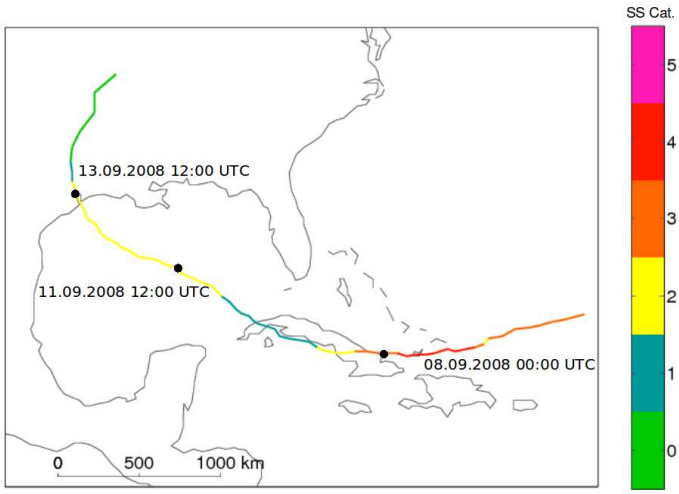



**Figure 6.17:** *Track of hurricane Ike 2008. The colors indicate the intensity in terms of the Saffir-Simpson Scale sategories see chapter Theories and Concepts, section 2.2. The 3 dates indicate times from which examples of identified objects were chosen and discussed in section 6.2.2*

### 6.2.1 Quality of Analysis as Verifying Observation

In order to be aware of the quality of the ECMWF analysis as a verifying observation, the minimum sea level pressure found in the objects identified in the ECMWF analysis was thus compared to the NHC central pressure observations. The outcome of this investigation is shortly presented here. Further aspects, like the comparison of the location property are shown in the appendix in figures A.14 to A.16.

With respect to position, the ECMWF analysis exhibits almost perfect correlation with NHC observations. Furthermore, the figures A.14 to A.16 in the appendix show that the correspondence is substantially better in the case of the analysed position than in the forecast position. The central pressure values in the objects from the ECMWF analysis, however, is different from the best track data of NHC. A correlation of 0.55 is found in the case of hurricane Ike. In terms of pressure, the 24 h forecast is better compared to NHC data than the analysis (see figure A.16 in the appendix). This shows the limitations of the ECMWF analysis as a verifying observation.

Figure 6.18 shows the analysed central pressure by ECMWF as a function of observed central pressure values by NHC. Perfect correspondence is indicated by the diagonal line. Good correspondence is found for NHC values above 975 hPa. Below this pressure threshold the model seems to have greater difficulties in accurately representing central pressure. The dependency on the location on the track of the different values reveals an interesting pattern. It seems that good correlation is obtained in the beginning and the end of the track (dark blue and dark red points). In the parts of the track, when intensity changes rapidly due to landfall in Cuba, the analysis data do not correlated well with NHC observations. It might be argued that this is a confirmation of the previously emitted hypothesis that the model is too slow in adaptation and exhibits a certain inertance (see section 6.1.4).

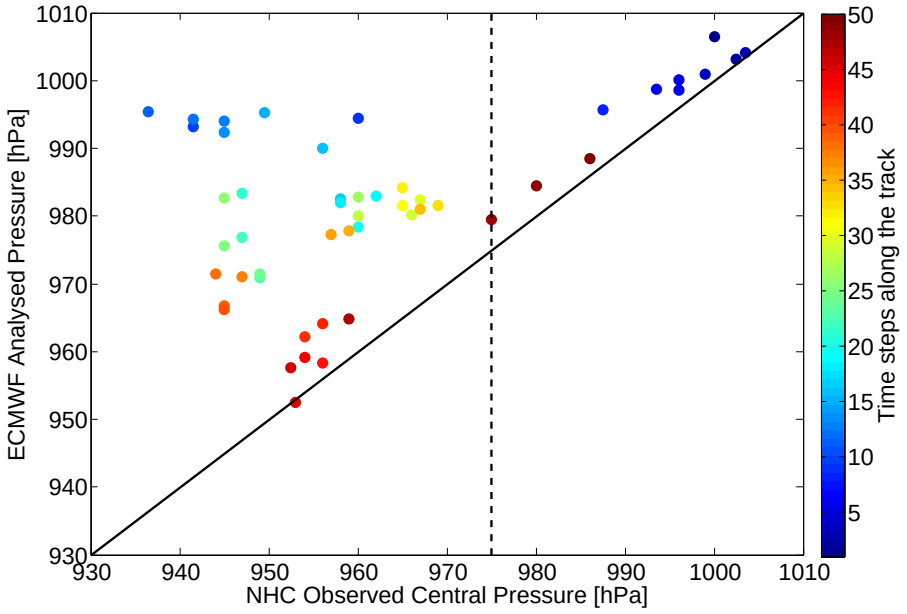



**Figure 6.18:** *Analysis minimum pressure compared to NHC minimum pressure. The colours indicate the time in the progression of the storm. Dark blue points stem from the beginning of the track and dark red points from the end of the track.*

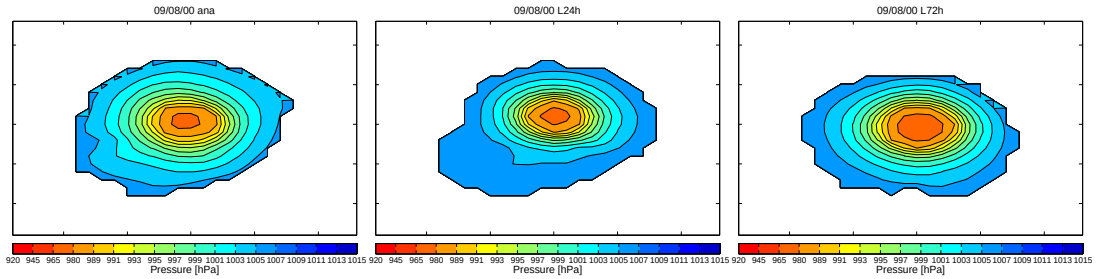
### 6.2.2 Examples of Pressure and Wind Distributions in Identified Objects for Hurricane Ike

In this section, 3 examples of TC objects identified with the procedure presented in the methodology chapter, section 5.2.2 are presented. The pressure and wind distributions in the objects are shown in figures 6.19 to 6.21 for the respective times in the life of hurricane Ike (see figure 6.17). It has to be emphasised that these figures only fulfil a qualitative purpose of giving an idea of the structures found at 3 key points in the progression of the storm. The first example is taken just before/at landfall in Cuba. The second example is taken, when the storm is over sea in the Gulf of Mexico and the last one during landfall on the continent. Table 6.1 lists a few important properties of the selected objects.

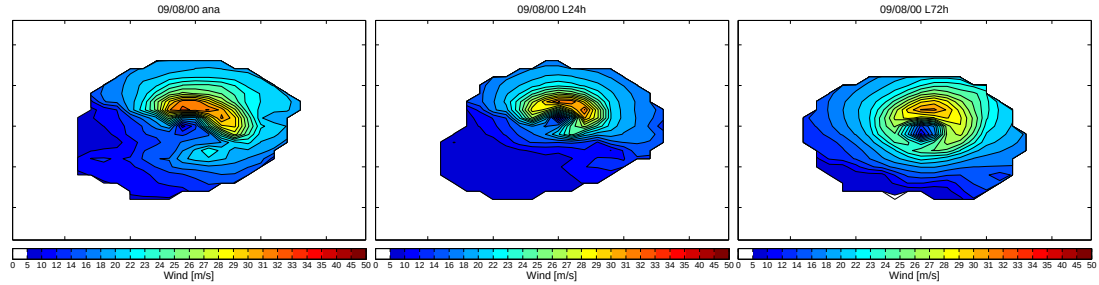
Differences in form, in area, in maximum wind speed, in wind and pressure distribution in the object and in destructive potential (kinetic energy) can be observed. What is not shown here, but also of great importance, is the difference in location of the low pressure center. The wind reinforcement on the right of the storm's center is due to the cumulative effect of the storm's displacement and the counter-clockwise rotation of air around the low pressure center. The effects of wind structure and object area on the resulting destructive potential is pronounced in the last example. The wind ring is much more peaked in the 72 h forecast than in the analysis or the 24 h forecast. Combined with a larger object area this induces a higher integrated kinetic energy and thus points towards a greater destructive potential.

**Table 6.1:** *Summary of properties of hurricane Ike TC objects for the chosen dates. LT stands for lead time, A for area, $U_{\max}$ stands for the maximum wind found in the object, $IKE$ stands for integrated kinetic energy and "land" indicates the percentage of cells containing land.*

| | LT | $A$ [$10^5$ km$^2$] | $U_{\max}$ [m/s] | $IKE$ [TJ] | land [%] |
|---|---|---|---|---|---|
| 08.09 00:00 UTC | 0h | 1.76 | 32.3 | 29.9 | 21 |
| | 24h | 1.69 | 32.1 | 21.9 | 23 |
| | 72h | 1.63 | 31.2 | 29.9 | 11 |
| 11.09 12:00 UTC | 0h | 6.49 | 34.9 | 99.3 | 4 |
| | 24h | 7.58 | 36.7 | 104.9 | 5 |
| | 72h | 7.12 | 35.4 | 100.2 | 3 |
| 13.09 12:00 UTC | 0h | 4.81 | 29.8 | 34.3 | 77 |
| | 24h | 4.10 | 33.5 | 44.3 | 70 |
| | 72h | 5.63 | 35.1 | 65.9 | 64 |

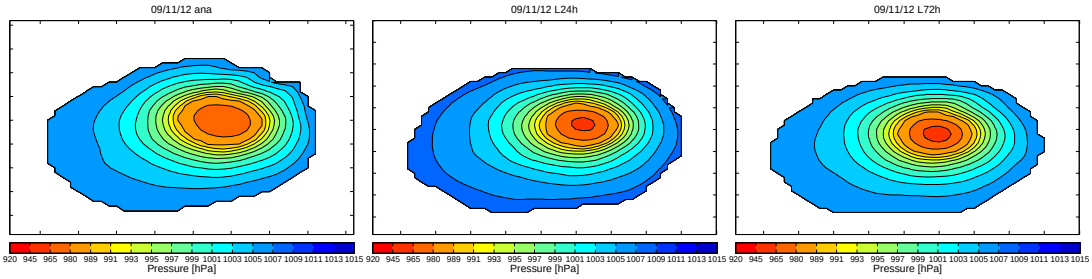



(a) Pressure distribution

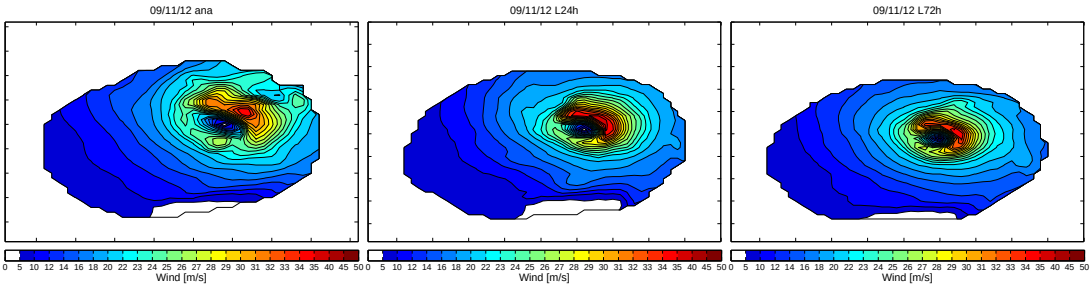



(b) Wind distribution

**Figure 6.19:** *Identified object for the 08.09.2008 00:00 UTC. Pressure distribution in the upper row (a) and wind distribution in the lower row (b). Propagation direction is to the West. Left: analysis. Middle: 24 h lead time forecast. Right: 72h lead time forecast. The objects are not georeferenced. Hot colours show low pressure/strong wind speeds, cold colours show high pressure/low wind speeds.*

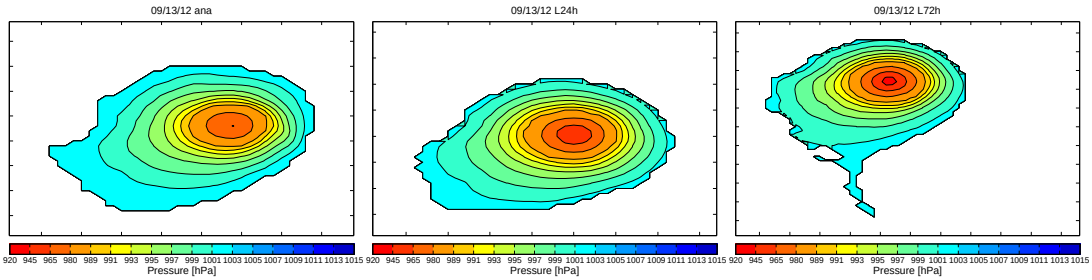



(a) Pressure distribution

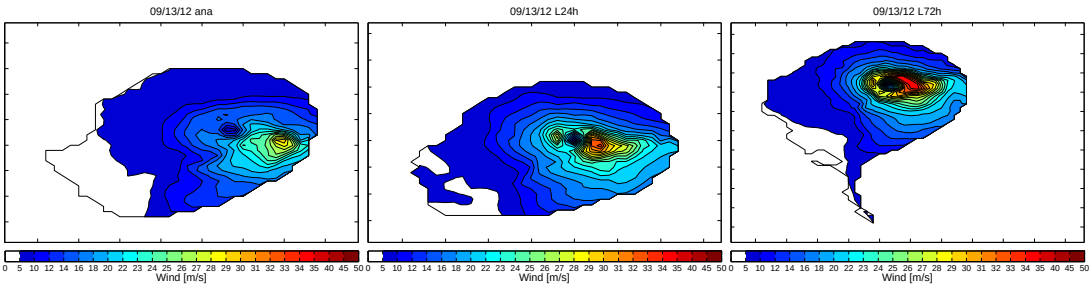



(b) Wind distribution

**Figure 6.20:** *Identified objects for the 08.09.2008 00:00 UTC. Pressure distribution in the upper row (a) and wind distribution in the lower row (b). Propagation direction is to the North-West. Left: analysis. Middle: 24 h lead time forecast. Right: 72 h lead time forecast. The objects are not georeferenced. Hot colours show low pressure/strong wind speeds, cold colours high pressure/low wind speeds.*




(a) Pressure distribution




(b) Wind distribution

**Figure 6.21:** *Identified objects for the 08.09.2008 00:00 UTC. Pressure distribution in the upper row (a) and wind distribution in the lower row (b). Propagation direction is to the North/North-West. Left: analysis. Middle: 24 h lead time forecast. Right: 72h lead time forecast. The objects are not georeferenced. Hot colours show low pressure/strong wind speeds, cold colours high pressure/low wind speeds.*

### 6.2.3 Structure-Amplitude-Location Components for Hurricane Ike

Preliminary results for one single case study are shown in this section. The aim here is to present a few advantages of the novel assessment method and show qualitative aspects of the end product.

Figure 6.22 is a 3-dimensional illustration of the performance of the model for 6 chosen lead times with respect to the 3 error components structure, amplitude and location. Plots with only 2 components at a time are shown in figure 6.24. Figure 6.23 shows the empirical cumulative distribution of the individual error components.

In general 2 observations can be made. These results seem to confirm the fact that the position error (location component) is lead time dependent and augments with forecast range. This is well illustrated by the regular colour progression from blue to red of the median error with lead time in the legend of figure 6.22. This interpretation is confirmed by figure 6.23 (a). The second result from this single case study is that the structure and amplitude error do exhibit a slight lead time dependency in the short-range up to 48 h. Beyond this time window structure and amplitude errors are comparable to errors from shorter ranges as can be seen in panel (b) and (c) of figure 6.23 and from the evolution of the black forms representing the median error of the different lead times in amplitude and structure in figure 6.22.

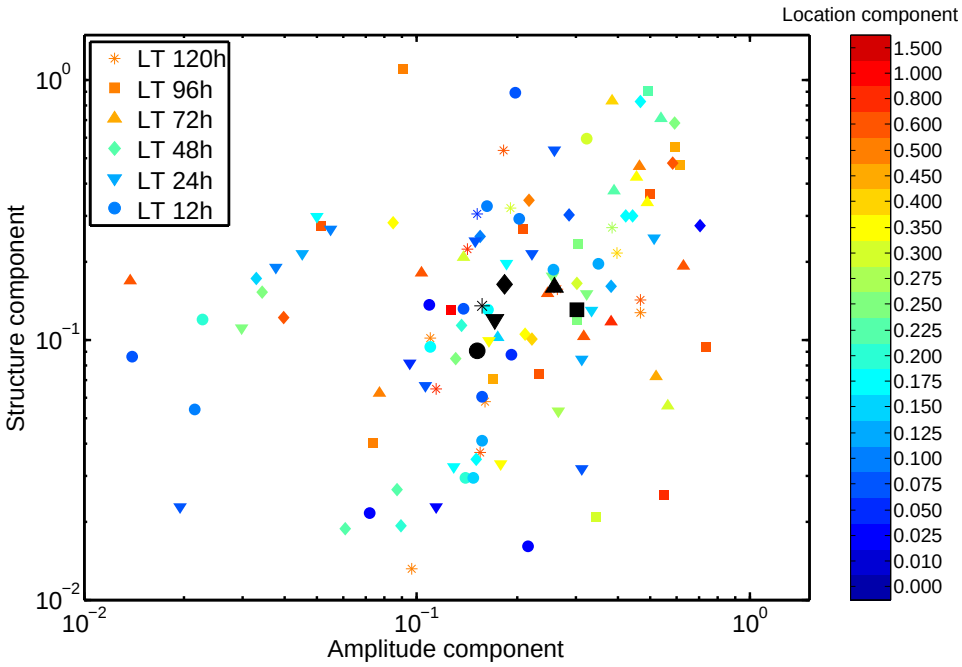
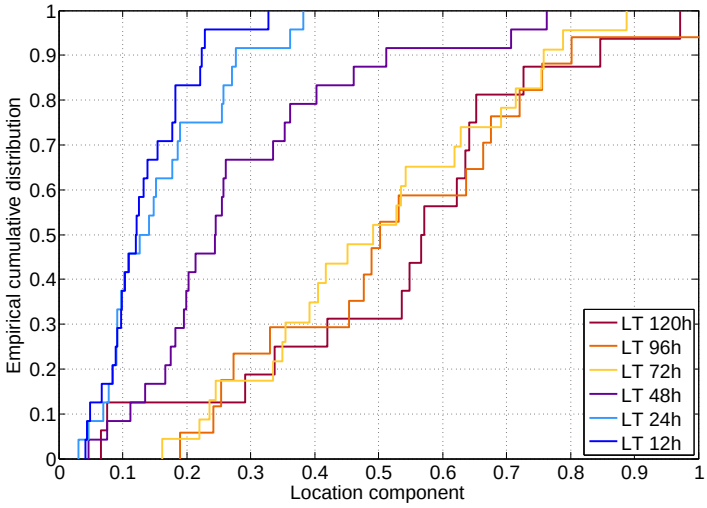
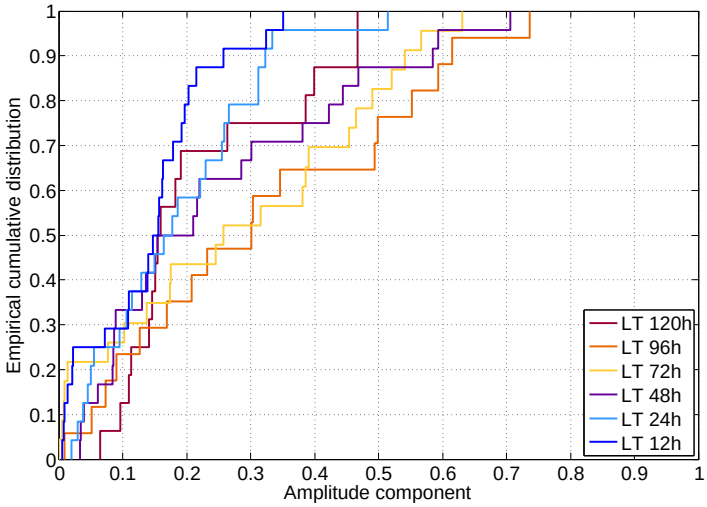



**Figure 6.22:** *SAL summary for hurricane Ike. Different forms are used for each lead time. The black forms indicate the respective median errors for each lead time in terms of structure and amplitude. The median error in location is shown in the corresponding colour in the legend.*

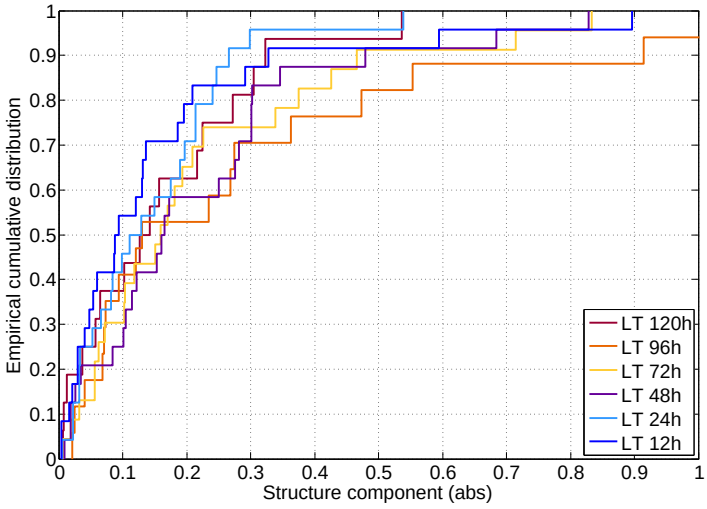
From the 3 panels in figure 6.24 a tendency of amplitude underestimation can be detected. Structure and location however does not seem affected by a systematic error. A slight correlation between amplitude and structure can be deduced from figure 6.24 panel (b), which indicates a certain redundancy of information in the 2 components.




(a) Empirical cdf for location error component.

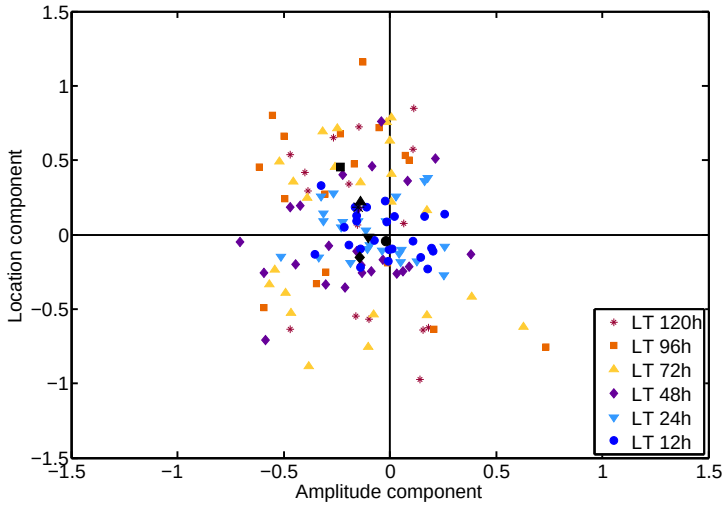



(b) Empirical cdf for amplitude error component.

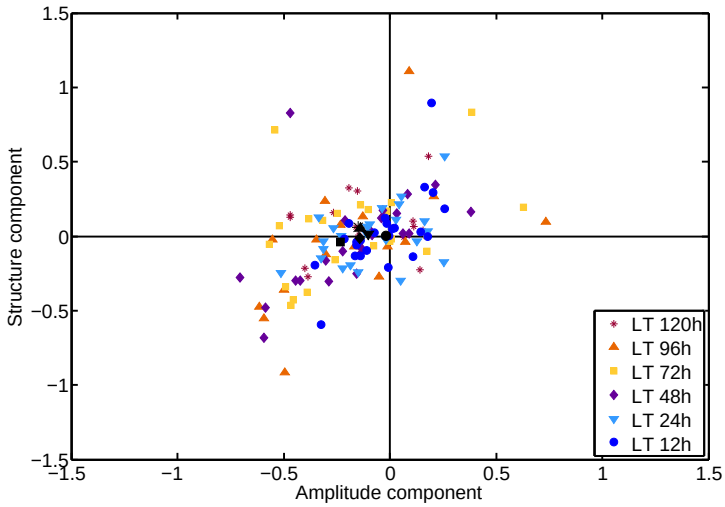



(c) Empirical cdf for structure error component.

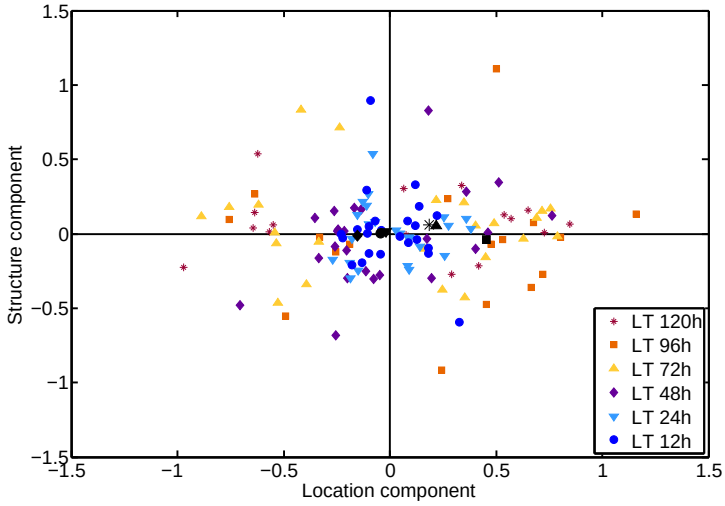
**Figure 6.23:** *Empirical cumulative distribution functions (cdf) for the error components of the SAL quality measure for hurricane Ike.*




(a) Amplitude and location error components.




(b) Amplitude and structure error components.




(c) Location and structure error component.

**Figure 6.24:** *Error components of the SAL quality measure for hurricane Ike.*

## 6.3 User-Oriented Assessment: Verification of Insurance Loss Predictions

In order to assess the value of ECMWF ensemble forecasts for the reinsurer Swiss Re, loss predictions were made using the meteorological input data and the company's loss modelling system for 5 extreme events of the past few years. Two forecast ranges, 1 and 3 days, were considered. The chosen events are described in section 4.4. Following the results from section 6.1 a bias correction in central pressure derived from the average bias over 4 years (2005-2008) in North Atlantic hurricane data (SS cat $\geq 1$) was applied to the ECMWF forecasts and the impact of this correction on the insurance loss predictions was studied. The forecast track and intensity fields for all events are shown in the appendix figures A.17 to A.23 as well as the probabilistic insurance loss prediction results in figures A.25 to A.29. The different events are not discussed individually here, but single aspects are drawn from each one of them.

In the following the assessment framework is shortly discussed and in section 6.3.2 the results of the performed chain modelling case studies of insurance loss predictions using ECMWF TC forecast data are presented.

### 6.3.1 Assessment Framework for Loss Predictions

Loss predictions were performed as described in the methodology chapter section 5.3, applying the traditional procedure of Swiss Re. Observed losses are obtained from NHC best track information. Predicted losses are computed from track and intensity forecasts. Following problems are associated with such an assessment framework:

1. The observed and forecast tracks are not equivalent in length, which highlights the need for a more equitable measure. Damage per track length is not fair because the value and vulnerability is not uniformly distributed in space. A categorical framework with certain loss threshold values could help solving this problem. Another possibility to avoid track length dependence of triggered losses would be to limit the assessment only to the forecast range. This would be honest for the model, but not useful for Swiss Re. For the reinsurance company the total loss is important.

2. The observed losses do not represent the total amount actually paid by the company. The computations correspond to "from ground up" losses, without insurance conditions. Furthermore there is a high uncertainty in the value distribution and the mean damage ratio.

3. The assessment done in this report consists of 5 strong landfalling storms of the last few seasons. The number of events is not representative to attribute general validity to the statements below. In the best case, they sketch important tendencies.

### 6.3.2 Loss Prediction Case Studies

The under/overestimation of losses with respect to observations are presented for each event and forecast range in table 6.3. For ECMWF, results are shown for raw data (biased) $b$ and bias corrected data $c$. Loss prediction errors from the NHC and characteristic errors from the predictions using ECMWF data are shown. Average values are indicated only for better overview and qualitative purposes. They are not estimated in a consistent way, as different lead times are combined. Furthermore, the uncertainty in these values is high as indicated by the standard deviations.

Generally, the loss predictions are underestimated by $\sim 80\%$ on average, when using the ECMWF EPS raw forecasts. The operational raw forecast performs better with only $\sim 30\%$ average underestimation. When forecast central pressure values are bias corrected the underestimation problem is less pronounced with values around $\sim 10\%$ to $\sim 40\%$ depending on the ECMWF error characteristic (control run, median or mean error). The error distribution in predicted losses has a negative skew. The median of the error distribution is higher than the mean error. Together with the underestimation tendency, this indicates that less than 50% of all members point towards the high observed losses. The operational forecast, on average overestimates damages after bias correction mainly due to the case of hurricane Ike 31 h before landfall, where the losses are overestimated by $\sim 400\%$.

**Table 6.3:** *Summary of loss prediction performance by NHC, ECMWF TC EPS mean error, the median error, the ECMWF operational run as well as the control run. The numbers indicate percentage of underestimation (negative) or overestimation (positive) when compared to the losses generated by the observed track.*

| Forecast | NHC | EPS MEAN | | EPS MEDIAN | | OPER | | CTL | |
|---|---|---|---|---|---|---|---|---|---|
| | | *b* | *c* | *b* | *c* | *b* | *c* | *b* | *c* |
| Ike -79 h | -90 | -40 | +78 | -64 | +12 | -16 | +21 | -20 | +236 |
| Ike -31 h | -73 | -28 | +120 | -49 | +53 | +205 | +441 | -50 | +47 |
| Gustav -72 h | -58 | -53 | -38 | -93 | -43 | -65 | -1 | -93 | -38 |
| Gustav -24 h | -34 | -58 | +59 | -58 | +61 | -30 | +47 | -48 | +81 |
| Katrina -107 h | -96 | -100 | -86 | -100 | -91 | -97 | -74 | -100 | -82 |
| Katrina -71 h | -88 | -84 | -63 | -86 | -67 | -51 | -14 | -84 | -51 |
| Katrina -23 h | -53 | -77 | -45 | -80 | -53 | -26 | +37 | -80 | -58 |
| Rita -68 h | -66 | -95 | -67 | -98 | -82 | -77 | -35 | -99 | -91 |
| Rita -44 h | -14 | -77 | -1 | -80 | -15 | 0 | +142 | -88 | -55 |
| Wilma -83 h | -89 | -98 | -93 | -100 | -100 | -100 | -99 | -100 | -100 |
| Wilma -23 h | -69 | -91 | -62 | -92 | -66 | -77 | -43 | -89 | -53 |
| MEAN | −66 | −73 | −18 | −82 | −36 | −30 | 38 | −77 | −7 |
| STD | 26 | 25 | 72 | 18 | 56 | 85 | 149 | 26 | 96 |

More cases of loss overprediction are found in 2008 than in 2005. This might be a confirmation of the tendency shown in section 6.1.5, where the sensitivity of the track and intensity forecast errors to model resolution change in 2006 is discussed. The bias corrected pressure forecasts for hurricane Ike is indeed 1-2 Saffir-Simpson Scale Categories too low, which obviously increases the loss estimates substantially. In the case of hurricane Rita at 44 h before landfall the overestimation by the operational forecasts is due to an unusual recurving of the track (see figure A.23, green line in the appendix). Loss overprediction for hurricane Katrina 23 h before landfall by the operational forecast is induced by underestimation of pressure after landfall. The same is true for the operational and EPS forecasts in the case of hurricane Gustav 24 h before landfall. The results of section 6.1.5 already suggested that the bias correction is dependent on whether the hurricane is over land or sea. But the sample size over land is not large enough to estimate a robust, specific correction value.

No clear changes with the introduction of bias correction were observed for the cases of hurricane Katrina 107 h before landfall in New Orleans and hurricane Wilma 83 h and 23 h before landfall. Hurricane Rita at 68 h before landfall can also be classified into this group, in which no improvement is reached with bias correction. In these cases the track forecast plays an important role. The forecast for hurricane Katrina at 107 h before landfall does not reach land. Hurricane Wilma is a "binary" high loss or no loss case because landfall does not seem certain in Florida at both 83 h and 23 h lead time. It finally passed over Florida as a Saffir-Simpson Category 3 hurricane. Hurricane Rita is a classical example of a track misforecast. At 83 h before landfall the underestimation of losses is probably due to the fact that no major city and industrial region is hit.

If the question of landfall is clear, the track does not play a major role any more for loss estimations as can be well illustrated, when comparing results of the bias corrected cases of hurricane Katrina 71 h and 23 h before landfall and hurricane Rita 44 h before landfall. The misforecast in the track of hurricane Rita does not affect the goodness of the loss predictions. Except for the operational track, which is much longer and induces an overestimation of losses, the ECMWF EPS forecast performs very well in terms of loss predictions. On the contrary, the 2 Katrina track forecasts are good, but the pressure is too high even after bias correction, which results in an underestimation of observed losses. From this we can conclude that if landfall is certain, and the value distribution of the region 500 km around is approximately uniform, the pressure information is more important for good loss estimations.

NHC forecasts underestimate observed losses for all case studies, which might be due to the slight negative bias that is found in pressure for long forecast ranges in the case of the chosen hurricanes. This would mean that it does not trigger high enough damage further inland from the landfalling point. On the long term mean, however, NHC intensity forecasts do not need a bias correction for their use in loss prediction (Franklin, 2006, 2009).

In table 6.5 a few summary statistics with respect to the performance of the ECWMF EPS forecast members are shown. The number of members is indicated in absolute value (out of 52 members), which are included in a certain range of the observed losses. Indications are made in the exponent if the operational high resolution forecast or the control run is included. In general the number of members in the respective margin of the observed losses increases with bias correction. The same observations that are made above for the individual cases hold here as well. The ECMWF operational forecast in regularly amongst the best members. The control run seems to perform much better in the 2008 cases, which might be linked to the resolution change.

**Table 6.5:** *Summary of performance in loss predictions by the ECMWF TC EPS. The number of ensemble members that yielded loss values within ± 10%, 20%, 50% of the observed loss are indicated. The presence of a small* $^{O}$ *in the exponent means the operational high resolution forecast is included,* $^{C}$ *means the control run is included.*

| Forecast | ENS 10% | | ENS 20% | | ENS 50% | |
|---|---|---|---|---|---|---|
| | *b* | *c* | *b* | *c* | *b* | *c* |
| Ike -79 h | 4 | 5 | $10^{O,C}$ | 7 | $19^{O,C}$ | $18^{O}$ |
| Ike -31 h | 3 | 5 | 9 | 10 | $21^{C}$ | $25^{C}$ |
| Gustav -72 h | 0 | $6^{O}$ | 0 | $8^{O}$ | 0 | $31^{O,C}$ |
| Gustav -24 h | 0 | 1 | 0 | 9 | $13^{O,C}$ | $21^{O}$ |
| Katrina -107 h | 0 | 0 | 0 | 0 | 0 | 0 |
| Katrina -71 h | 0 | 1 | 0 | $3^{O}$ | 0 | $8^{O}$ |
| Katrina -23 h | 0 | 3 | 0 | 5 | $1^{O}$ | $23^{O}$ |
| Rita -68 h | 0 | 1 | 0 | 3 | 0 | $10^{O}$ |
| Rita -44 h | $1^{O}$ | 5 | $2^{O}$ | 10 | $4^{O}$ | 19 |
| Wilma -83 h | 0 | 0 | 0 | 0 | 0 | 0 |
| Wilma -23 h | 0 | 3 | 0 | 4 | 0 | $14^{O}$ |
| MEAN | 1 | 3 | 2 | 5 | 5 | 15 |
| STD | 1 | 2 | 4 | 4 | 8 | 10 |

# Chapter 7

# Conclusions and Outlook

In this report the performance of the ECMWF model with respect to TC forecasts was assessed in three different ways. First, the TC ensemble forecasting system, which is embedded in the ECMWF operational deterministic and EPS model was analysed using a storm tracking approach to forecast verification. Then, the framework for a novel object-based verification measure was set up with the aim of characterising the representation of TC features in the operational ECMWF deterministic model. Finally, the user-oriented aspect of chain-modelling using meteorological ensemble forecasts for medium-range catastrophe loss predictions in the insurance sector was investigated.

## 7.1 Strengths and Weaknesses of the ECMWF Tropical Cyclone Ensemble Forecasts

The one dimensional verification of TC ensemble forecasts by ECMWF involved the statistical analysis of track and intensity errors for 29 hurricanes of the seasons 2005 to 2008. The position and the central pressure were verified. The best available verifying observation of these variables are given by NHC. Different forecast properties like accuracy, skill, reliability in terms of calibration (bias), correlation and statistical consistency of the ECMWF TC EPS were assessed. Furthermore, the performance of the model was compared to NHC warnings. The following conclusions can be drawn from this analysis.

1. **Uncertainty and significance of results**
   The uncertainty in the obtained results is high. No statistical test was performed in order to prove the significance of the estimations and statements. On the one hand the data availability is low. Only 4 seasons with approximately consistent model state could be analysed here. Because of the stakeholder's perspective of a reinsurance company, only high impact events were chosen. On the other hand regular model improvements in the last few years make it difficult to analyse a homogeneous sample of forecast events. It is thus important to recall that only tendencies could be sketched in this report. Theoretically, for a consistent and thorough analysis of the model performance a series of hindcasts of historical events with the most recent model formulation should be performed by ECMWF. An alternative is to include weaker tropical storm events into the investigation. The outcome with respect to intensity forecasts would probably change in this case.

2. **Accuracy**

   - **Initial errors:** Initial positioning and intensity errors in the ECMWF operational and EPS data of around 60 km and 20 hPa were found. The initial central pressure error is probably mainly due to the inability of the ECMWF model to resolve the small scale features in the present grid spacing. The spatial scale of the storm's eye is of approximately the same size as the model resolution. Thus, a number of crucial processes need to be parametrised. Present parametrisations apparently are not able to fully capture such strong events as hurricane force TCs.

- **Track forecast error:** The accuracy of track forecasts is lead time dependent and decreases linearly with forecast range by 40 km per 12 h for the ensemble median error. It reaches $\sim$ 250 km at 3 days, the business relevant lead time for Swiss Re. For long forecast ranges the track error distribution is positively skewed, more than 50% of all members perform better than indicated by the mean error. Further investigations of the track errors with respect to the influence of propagation speed and direction showed that the propagation speed error constitutes up to $\sim$ 70% of the direct position error.
- **Intensity forecast error:** The mean intensity error of ECMWF forecasts is approximately lead time independent and amounts to about $\sim$ 22 hPa. The independence of the intensity error on forecast range is expected to be due to the inability of the model to resolve the peaking low pressures in intense storms. The pressure evolution being strongly linked to land/ocean-atmosphere interactions, the lack of coupling between these compartments might be one severe disadvantage. Another problem might be given by the microphysics parametrisation that influences the storm's development in terms of intensity.
- **Sensitivity of forecast errors:** The sensitivity of the track and intensity errors with respect to changes in resolution, land-sea effects and the influence of sampling of events with respect to intensity were conducted. All results exhibit high uncertainty, but resolution is suspected to play a major role as well as the effects of land and sea. The sample size is too small for a robust analysis of event sampling dependency of the results.

3. **Reliability**

   - **Track forecast bias:** In order to investigate biases in track a signed error had to be computed, using a decomposition of the position error into propagation speed and direction. No bias was found in the latter, however in the case of the propagation speed a pronounced rapid bias for long forecast ranges was detected.
   - **Intensity forecast bias:** An average intensity bias of $\sim$ 20 hPa for the ECMWF ensemble forecasts and $\sim$ 15 hPa for the ECMWF operational forecast were found.

4. **Statistical Consistency of the ensemble forecasts**
   The ECMWF ensemble forecasts are underdispersive in terms of intensity and track. This indicates that the EPS does not, on average, cover the whole possible range of developments. The model overconfidence is stronger for the intensity with a maximum relative underdispersivity of $\sim$ 80% against $\sim$ 50% for track forecasts.

5. **Skill**

   - **Track forecasts** by ECMWF are characterised by good skill for forecast ranges of 2 days and longer. Skill values of $\sim$70% for the operational forecast and 40-50% for the EPS were obtained for 4 to 5 days lead time.
   - **Intensity forecasts** by ECMWF were found to have much lower skill than track forecasts with values below $\sim$20%.
   - **Comparison of ECMWF with NHC:** In the very short range of 1 day, human made NHC forecasts were found to be more skillful than ECMWF forecasts. Furthermore, the NHC warnings essentially aim at giving a decision basis for administrative evacuation measures. These are mostly taken around 1-2 days before landfall.

6. **Correlation**
   High correlation between ECMWF ensemble forecasts and observations was found in the case of track forecasts ($\rho =\sim 1$) and low correlation in terms of intensity ($\rho =\sim 0.3$). The latter may be explained by the even lower correlation in pressure change, indicating that the model might be too inert, which prevents rapid evolutions to be forecast.

## 7.2 Object-Based Tropical Cyclone Verification: Gained Insights and Possibilities

The main objective for a new multidimensional object-based quality measure was to gain a holistic insight into how the ECMWF model handles TCs. Pointing out the main problems in TC representation in the model and giving a basis for explaining possible inabilities were further motivations for suggesting a new assessment method. The SAL concept presented by Wernli et al. (2008) for precipitation fields was found to be very suitable. The formulation of different demands on such a new measure in the methods chapter, section 5.2.1 helped to identify important aspects in TC representation in the model. Moreover, identifying general desirable properties of a verification measure was very profitable in terms of questioning and reflection on the verification process.

The multidimensional perspective of forecast verification given by different error components depicts a more holistic overview of model performance and is more stringent, because it does not allow compensation of different error types. The different feature properties are treated separately in specific error components. The dynamical and structural elements of TCs are taken into account. The SAL concept applied to TCs is easily adaptable to the interest of a specific model user. Other physical variables than the ones used here may be used to formulate the error components. The formal aspect of clarity in communication is given by the fact that one summary descriptive component is formulated for each aspect.

The object-based multicomponent measure may provide an excellent framework to compare different models or to evaluate the performance of one specific model with respect to different events. Moreover, a basis for understanding problems in process representation is given. It is an informative quality measure in this sense. Certain problems found in the one dimensional storm tracking approach were confirmed in the case study of hurricane Ike. The misrepresentation of the surface pressure in the model is one confirming example. Finally, the new assessment framework is very versatile and seems promising. Further testing and case studies are of course needed in order to assess the full capabilities of this quality measure.

## 7.3 User-Oriented Verification

The conclusions drawn from the project collaboration with Swiss Re are summarised in two sections. In the first one, somewhat more general, personal conclusions are presented. The second section is addressed to the company.

### 7.3.1 User Needs and Specificities

In the user-oriented part of this project a few general observations about user needs were made that are here shortly summarised. The statements are based on personal insights, gained in this project and not on quantitative results.

Different types of answers and solutions are necessary for different users. User interests weight the various forecast aspects in a specific manner. The picture obtained from such an assessment is thus more one-sided but also reveals interesting questions. In the present case, the track information for insurance loss predictions is not as important as the intensity, if it is clear that the hurricane makes landfall. In the case of an official warning institution, it is however crucial to know, where the storm might pass, meaning where to evacuate.

When the model output data is used only in qualitative way and no "binary" threshold decision (evacuation yes or no) depends on it, it becomes difficult to measure the added value the model brings in a cost-loss value based framework. The added value of knowledge gain is not easily measurable in economic terms and might be quantified only if the company's behaviour could be simulated in both, the case of available and missing knowledge about the short-range development of a storm. Such a simulation seems however difficult to put into practice. The knowledge gained from the tight collaboration with research and operational centres also induces a longer term profit.

Finally, like in all other cases of chain modelling of weather impacts, the uncertainty margin is high. Uncertainties from the meteorological input, the loss modelling system and the estimation of observed losses flow into the analysis.

### 7.3.2 Use of ECMWF Data for Insurance loss Predictions: Recommendations for Swiss Re

The aim of the user-oriented assessment of the ECMWF model from the side of Swiss Re was to get a better idea of the added value of an ensemble forecast, when compared to a deterministic forecast without physically based information on spread.

The information contained in the ECMWF deterministic and ensemble forecast was shown in the storm tracking part to be complementary to the NHC forecast, which is freely available on the web. The ECMWF deterministic and ensemble track and intensity forecasts have more skill in the business relevant forecast range of 3 days, than the NHC forecast, which is more skillful in the short-range.

In terms of statistical consistency the ECMWF TC EPS is overconfident. This means that the whole physically possible range of developments is not covered and indicated by the forecast. This has to be kept in mind, when using the data for real-time insurance loss predictions. A clear advantage of using an ensemble forecasting system is the possibility of working in a probabilistic loss framework, which allows for example the formulation of striking probabilities.

If the question of landfall is solved by the track forecast, meaning if landfall is certain, the pressure information is more important for loss predictions than the track. Central pressure relates directly to the potential losses. The track is relevant if the geographical value distribution is highly variable. The central pressure forecast by ECMWF was found to be affected by a systematic error. As the intensity information is crucial in a damage-oriented loss modelling approach, this bias should be corrected as proposed in section 6.1.1.

Track and intensity give a simple but complete description of the storm that can be well used with simplified assumptions on the storm size in order to perform insurance loss predictions. Based on recent publications on the damage potential of TCs, the size of the storm is mentioned as an important factor, missing for example in an intensity indication on the Saffir-Simpson scale (see review in the chapter on Theories and Concepts, section 2.2). If an additional parameter is to be included, the radius of maximum wind for example or the diameter of the storm would be a valuable information. The first one might be preferable, because it contains an information on size and additionally allows a better wind field to be deducted from the central pressure information.

Finally, it has to be emphasised that ECMWF is a reliable modelling partner for many European institutions and has a good reputation worldwide. Being data user is also a way of staying informed about present modelling standards, problems and developments in this field.

## 7.4 Multiple-Perspective Approach to Verification

A multiple perspective view on the forecast verification problem can be very useful as illustrated in this report by many examples. The storm tracking approach helped to point out a bias in pressure forecasts, which could not have been unambiguously identified by conducting solely the user-oriented loss prediction verification. Furthermore, the case study in the object-based assessment gave some hints of possible explanations for the problems found in the storm tracking perspective. The setting up of the object-based verification measure implied that relevant physical properties for the storm's representation on the ground and the induced damage had to be identified. This allowed to link the scientific side of model verification to the user's interests. Finally, the gain is not only on the side of the user, as it would be in the case of a linear and one-way communication between the different assessment methods. A model can also be improved in a more targeted way, if the different user needs and problems are included.

# Appendix A

## A.1 Selected Hurricanes 2005-2008 for Traditional Forecast Verification

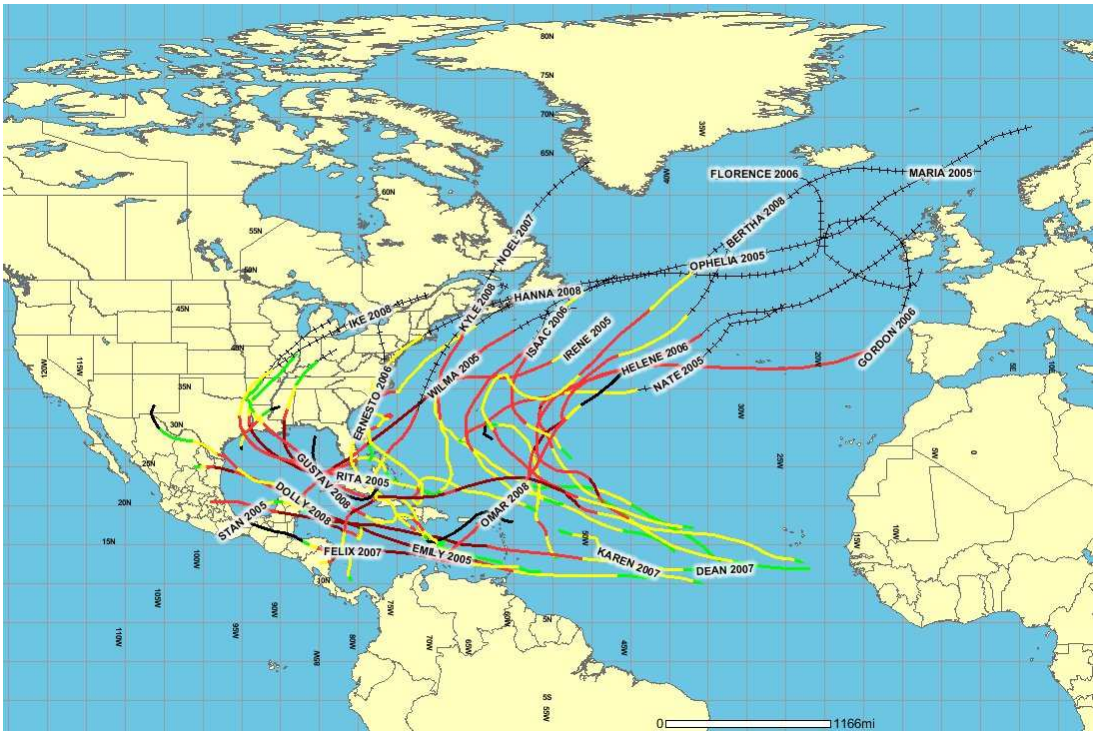



**Figure A.1:** *Selected hurricanes of the seasons 2005 to 2008 for the storm tracking verification of the ECMWF TC EPS.*

**Table A.1:** *Properties of selected hurricanes of the years 2005 to 2008 for the storm tracking verification of the ECMWF TC EPS.*

| **Hurricane** | **Year** | **SS Category** | **Landfall** |
|---|---|---|---|
| Emily | 2005 | 5 | 1 |
| Irene | 2005 | 2 | 0 |
| Katrina | 2005 | 5 | 1 |
| Maria | 2005 | 3 | 0 |
| Nate | 2005 | 1 | 0 |
| Ophelia | 2005 | 1 | 0 |
| Philippe | 2005 | 1 | 0 |
| Rita | 2005 | 5 | 1 |
| Stan | 2005 | 1 | 1 |
| Wilma | 2005 | 5 | 1 |
| Beta | 2005 | 1 | 1 |
| Ernesto | 2006 | 1 | 1 |
| Florence | 2006 | 1 | 0 |
| Gordon | 2006 | 3 | 0 |
| Helene | 2006 | 3 | 0 |
| Isaac | 2006 | 1 | 0 |
| Dean | 2007 | 5 | 1 |
| Felix | 2007 | 5 | 1 |
| Humberto | 2007 | 1 | 1 |
| Karen | 2007 | 1 | 0 |
| Noel | 2007 | 1 | 1 |
| Bertha | 2008 | 3 | 0 |
| Dolly | 2008 | 2 | 1 |
| Gustav | 2008 | 4 | 1 |
| Hanna | 2008 | 1 | 1 |
| Ike | 2008 | 4 | 1 |
| Kyle | 2008 | 1 | 1 |
| Omar | 2008 | 3 | 1 |
| Paloma | 2008 | 4 | 1 |

## A.2 Errors and Biases in Track Components

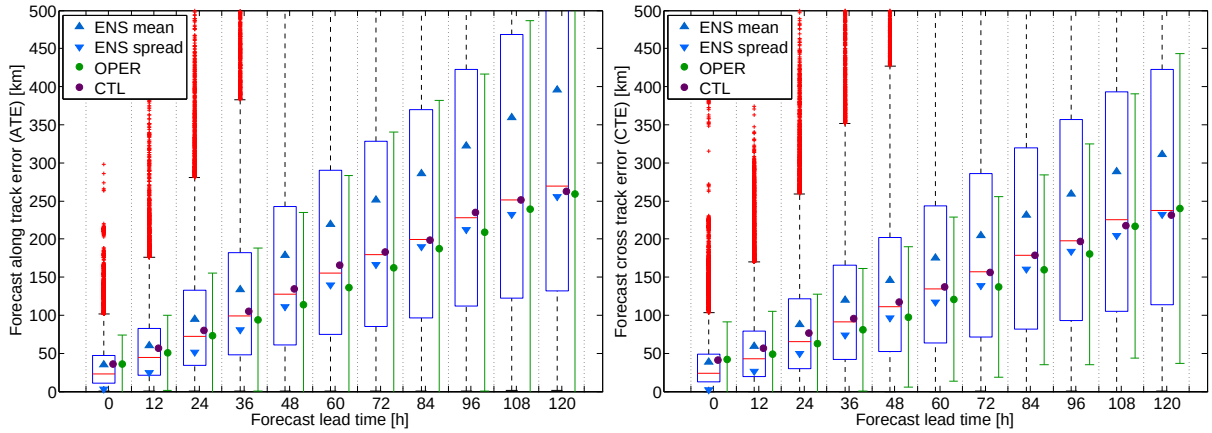



**Figure A.2:** *Along and cross track position error for the selected hurricanes of the years 2005 to 2008.*

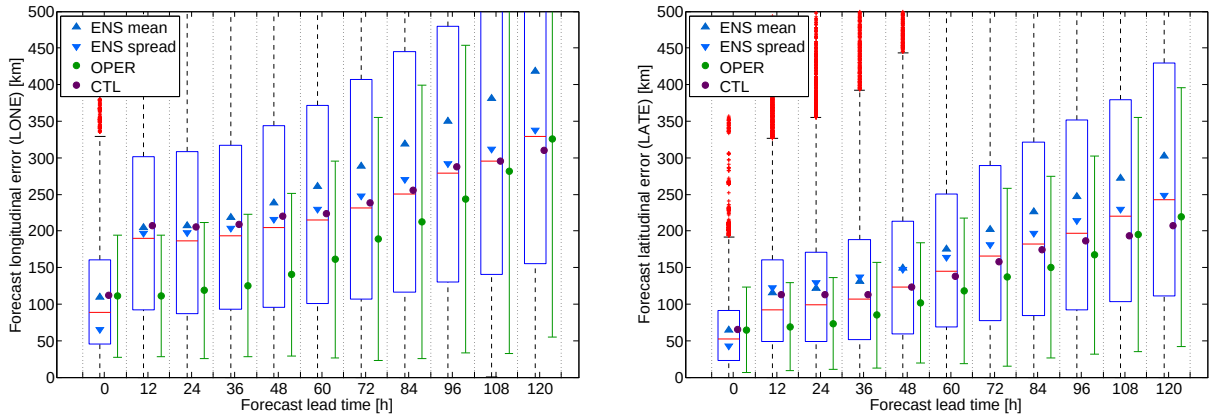



**Figure A.3:** *Longitudinal and latitudinal track position error for the selected hurricanes of the years 2005 to 2008.*

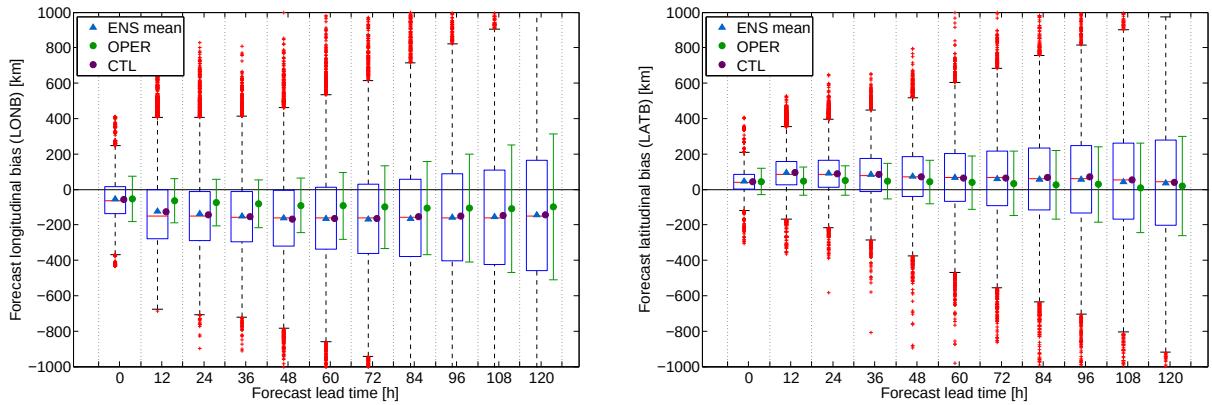



**Figure A.4:** *Bias in longitudinal and latitudinal track components for the selected hurricanes of the years 2005 to 2008. Positive bias in the longitudinal component means a bias to the East, positive bias in latitudinal component means a bias to the North.*

## A.3 Sensitivity Analysis Results for Forecast Errors

### A.3.1 Sensitivity towards the Interpolation Method

The interpolation method plays a negligible role in the results obtained for the track and intensity statistics. As an example, the direct position error and the central pressure errors are shown in figure A.6 for hurricane Ike. In section 4.1.1 the interpolation technique for observations is introduced and the advantages of the chosen method are described. The problem of irregularity in the observational time grid becomes apparent in figure A.6 and A.5. The sample size for example decreases perfectly linearly and smoothly with forecast lead time in the case of interpolated observations on the regular 12 h time grid, whereas the sample size of interpolated forecasts includes jumps. These discontinuities lead to the irregularities in the $CPE$ curve shown in figure A.6. Except the latter effect no sign of dependence of the results on the interpolation technique can be found in the case of hurricane Ike. The mean value for $DPE$ over all lead times amounts to 185±120 km in the case of interpolated observations and 165±107 km in the case of the interpolated forecasts. The errors are thus slightly lower in the case of forecast interpolation. This is partly due to the fact that different lead times are used, as the time grid is shifted by 3 hours. With the data available this implies that the last lead time in the interpolation of forecasts is 4 days and 6 h and not 5 days as when the raw forecasts are used with interpolated observations. For $CPE$ the average values are $31 \pm 17$ hPa for the interpolated observations and $31 \pm 18$ hPa for the interpolated forecasts. The errors in the case of central pressure are exactly the same as there is no lead time dependency.

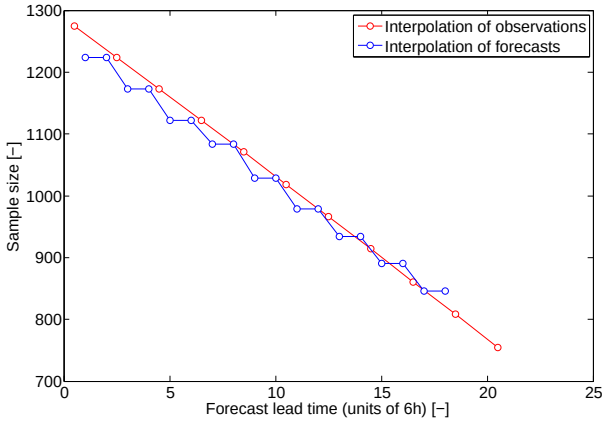



**Figure A.5:** *Sample size dependence on interpolation method for hurricane Ike (2008). The lead times for the two curves were shifted by 0.5 units in order to allow a comparison to be made.*

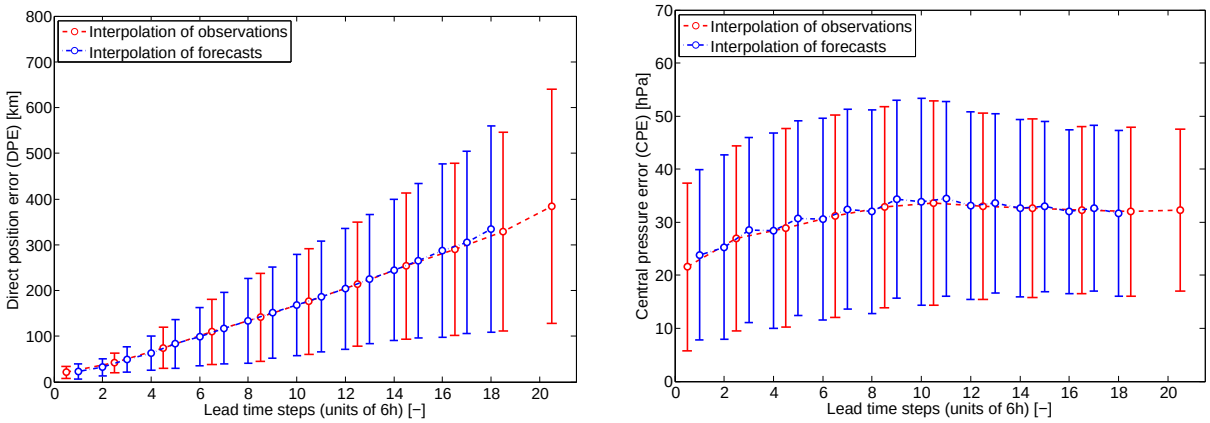



**Figure A.6:** *Error in direct position (left) and central pressure (right), sensitivity towards interpolation method for hurricane Ike (2008). The lead times for the two curves were shifted by 0.5 units in order to allow a comparison to be made.*

### A.3.2 Sensitivity towards Hurricane Intensity and Landfalling

The hurricane groups defined on the basis of their intensity and whether the TC hits land or not show slightly differing results for the position and intensity errors. Interestingly, in the case of track errors, ECMWF and NHC official errors show the same pattern, while the tendency in the case of CLIPER is different (see figures A.7 and A.8). The position error of strong landfalling hurricanes (orange) in CLIPER increases more strongly with augmenting forecast range than in the other groups, beginning with the lowest error for short lead times and reaching the highest error for longer lead times. NHC indicates highest track errors and uncertainty for the original hurricane group including all selected hurricanes. ECMWF shows slightly lower track errors for the strong landfalling hurricane and as in the case of NHC is affected by highest errors, when considering all hurricanes. The differences though are small and can probably mostly be attributed to the sample size difference. An other reason for the higher errors in the "all hurricanes" group in the ECMWF and NHC forecasts may be given by the Category 1-2 hurricanes, which stay over the ocean. The data scarcity over the free ocean make the analysis fields more uncertain than near the coasts, where terrestrial systems can be included for bias correction of remote sensing observations.

The partition between along and cross track errors is similar for all groups (see figure A.7). For long forecast times the influence of the propagation speed in the total error is of around 65%. Differences in partitioning of around 15% can be found at lead time 24 h. The strong landfalling hurricanes seem to be affected by a higher influence of the along track component in its position error. In combination with the higher intensity error (see figures A.9 and A.10), this can be seen as an indication that the propagation speed might be affected by small scale, microphysical and vortex dynamical elements, than the propagation direction.

All models exhibit a clear tendency of producing higher intensity errors for strong landfalling hurricanes (see figures A.9 and A.10). As already said in the Case Study chapter, section 4.2, this may also be influenced by the fact that the proportion of 2005 hurricanes in this group is slightly higher than in other groups. For the track error taking all hurricanes into account yields the highest errors. For the intensity errors, on the contrary, grouping hurricanes after the intensity and landfalling characteristics produces higher errors. Apparently the hurricanes staying over the ocean without making landfall are better forecast in terms of their intensity. Indeed the changes induced by passages over land are difficult to capture well in the model, because the magnitude of the weakening has to do with small scale inner core dynamics (Lohmann, 2009).

Some of the statements on the sensitivity towards hurricane intensity and landfalling characteristics can be verified in the section on the sensitivity towards location over land or sea in the results part, section 6.1.5. The latter analysis however considers forecast position as a classifying property and not the characteristics of the whole TC track, like in this section.

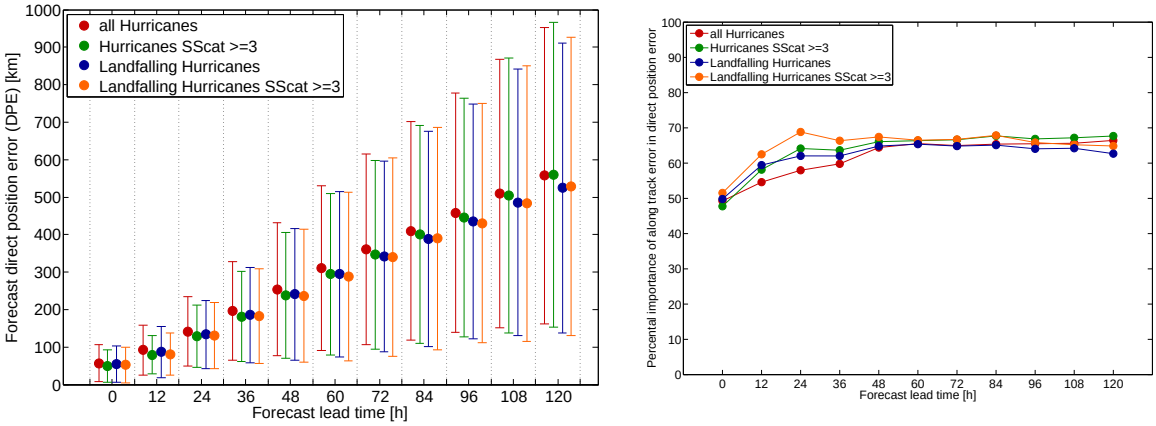



**Figure A.7:** *Positional errors for different hurricane groups. Left: Comparison of the ECMWF ensemble mean positional error and standard deviation for different hurricane groups. Right: Sensitivity of average importance of the along track error in the direct position error of the ECMWF EPS model for different hurricane groups.*

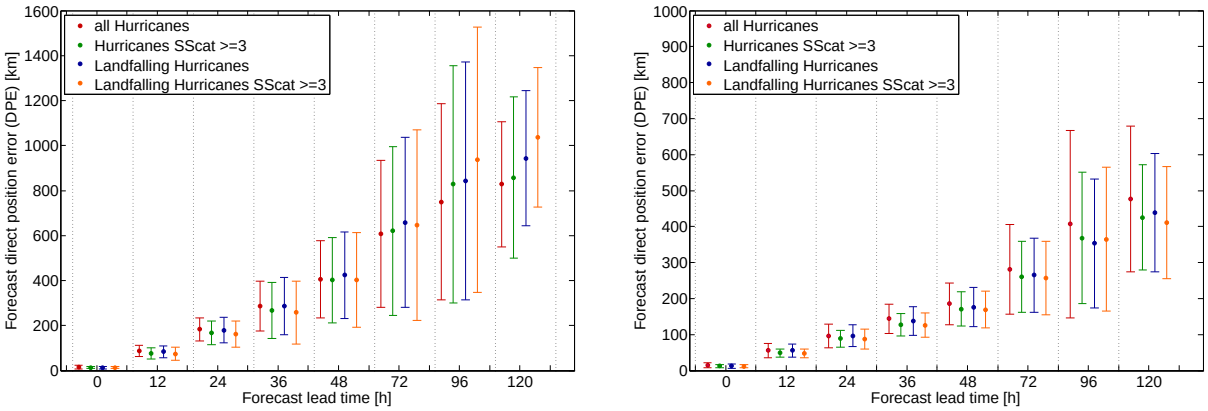



**Figure A.8:** *Comparison of CLIPER5 (left) and NHC official forecast (right) mean error and standard deviation for the different hurricane groups.*

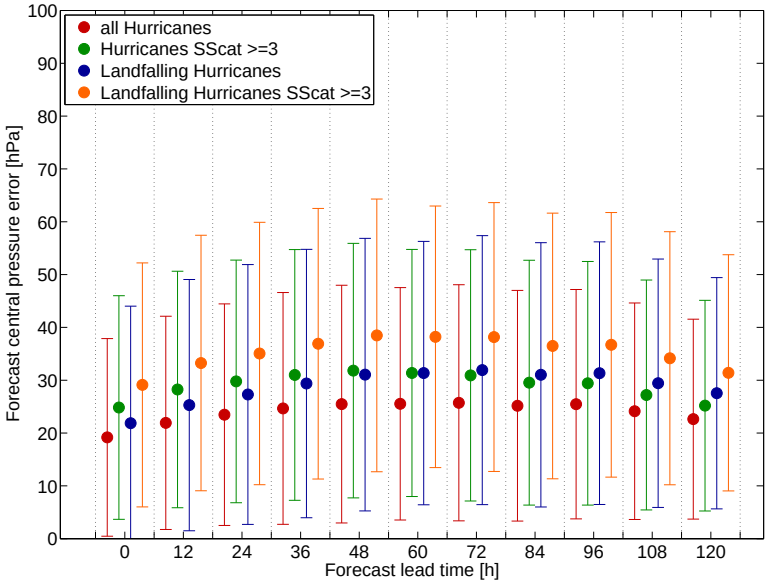



**Figure A.9:** *Comparison of the ECMWF ensemble mean central pressure error and standard deviation for different hurricane groups.*

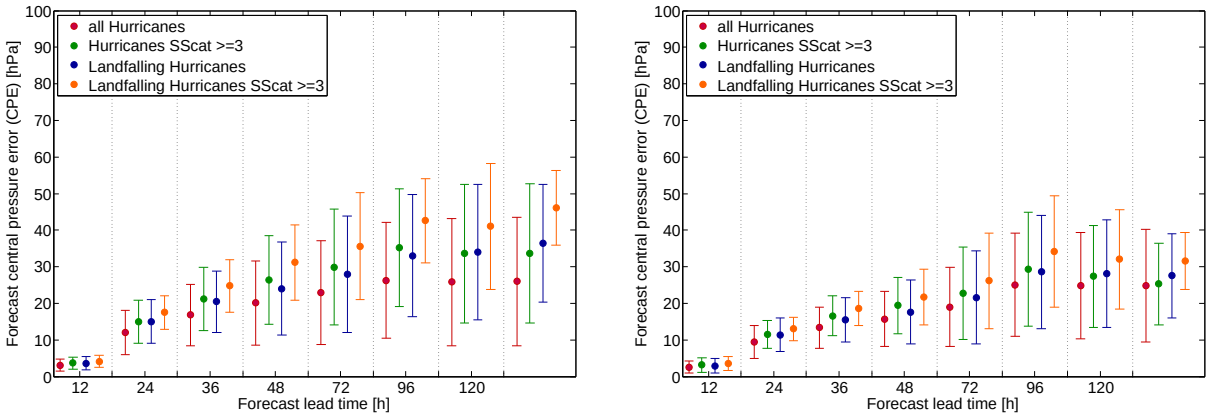



**Figure A.10:** *Comparison of CLIPER5 (left) and NHC official forecast (right) mean error and standard deviation for the different hurricane groups.*

## A.4 Correlation Analysis for Central Pressure Error

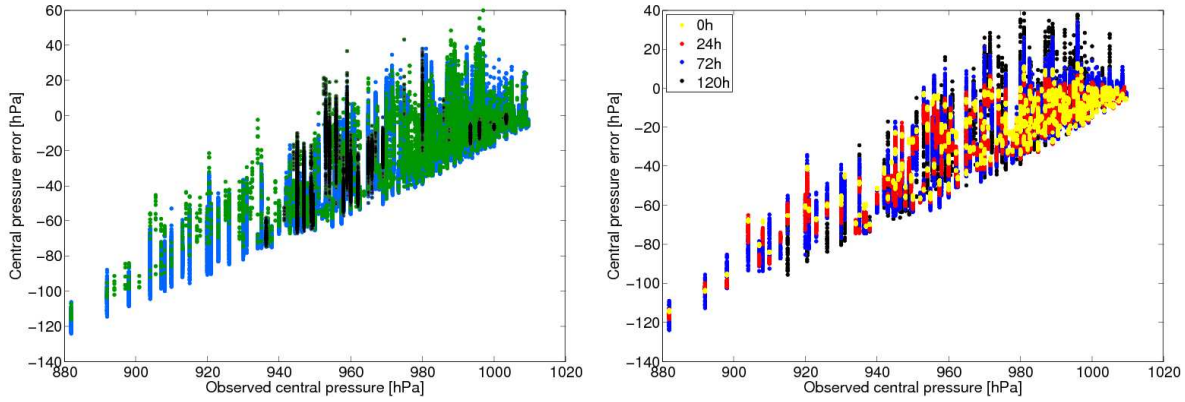
Further investigations on the central pressure bias were conducted in order to find the best possible correction for applications like insurance loss predictions. In figure A.11 the $CPE$ is shown as a function of observed and forecast pressure. Figure A.12 shows $CPE$ as a function of observed and forecast central pressure change in 12 h. The vertical lines in figure A.11 (a) and (b) as well as in figure A.12 (a) and (b) come from the fact that for 1 observed pressure there are always 52 forecasts and thus forecast errors. The parallel lines in the corresponding graphs below ((c) and (d) of the same figures) relating forecast pressure and pressure changes in 12 h to $CPE$ are induced by the tendency of 1 ensemble forecast to produce pressure values spread around a certain median forecast value. The distance from the observed forecast is then always similar taking into account the latter spread.

Only one good correspondence was found, namely between the observed pressure and the $CPE$. Correlations of 0.86 for the analysis down to 0.77 for 120 h lead time can be found. The lower the observed pressure, the higher the $CPE$ for all lead times. Ideally, the $CPE$ should be decoupled from the observed pressure (no systematic error) and the intensity error low (no random error). High correlation, however indicates that the errors can be explained, for examples by the suppositions made in section 6.1.1. Thus, the correlation between observed pressure and $CPE$ is interesting in terms of understanding the high pressure bias, but does not help in the application of a case dependent, targeted bias correction.

No dependency, however is found between $CPE$ and observed pressure change (see figure A.12). This means the hypothesised inability of the model to capture rapid strengthening and weakening (see section 6.1.4), its inertia is clearly not a function of the specific situation (observed pressure change). Correlations of maximum 0.23 were found for the analysis time, reaching approximately 0 for 120 h lead time.

There is generally only low correlation between the forecast central pressure change or central pressure change and $CPE$. Such a dependence would have been useful for a case specific bias correction. The correlation between forecast central pressure and $CPE$ is 0 for the analysis time and decreases to -0.56 for lead time 120 h. For longer forecast ranges, $CPE$ shows a weak correlation with central pressure, but probably not strong enough to deduce a forecast pressure change dependent bias correction.

For the analysis time, the correlation of forecast central pressure change with CPE is strong (-0.97). The detail is shown in figure A.4. The forecast pressure change in the analysis case is the difference between the last observation and the present analysed pressure. Strong forecast pressure changes are linked to high $CPE$. The high dependency comes from the fact that the change is measured from the last observation, which is clearly also expected to correlate well with the present observation. No correlation, however is found between the analysis and the future $CPE$ in the forecasts. Thus this dependency is of no use for a targeted pressure bias correction either.



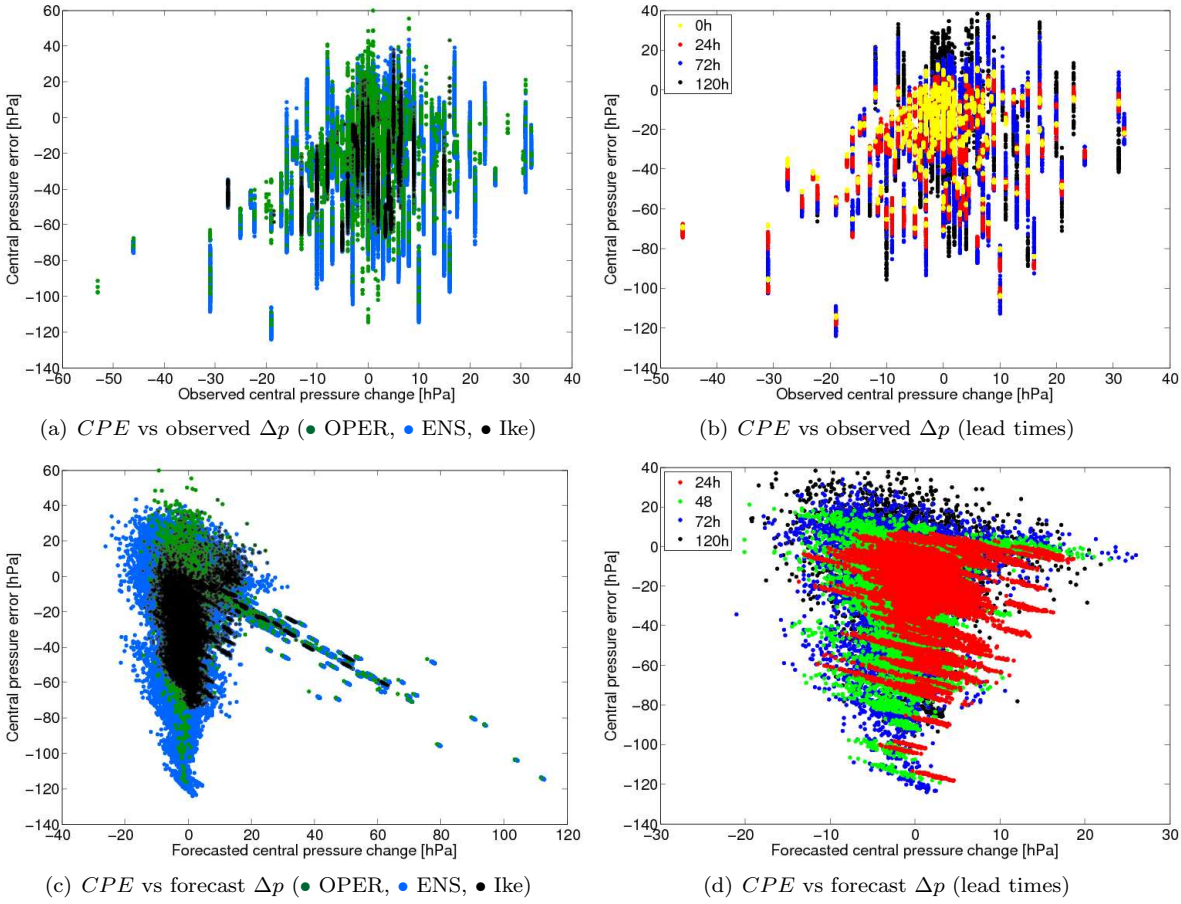

(a) $CPE$ vs observed $p$ (● OPER, ● ENS, ● Ike)

(b) $CPE$ vs observed $p$ (lead times)

(c) $CPE$ vs forecast $p$ (● OPER, ● ENS, ● Ike)

(d) $CPE$ vs forecast $p$ (lead times)

**Figure A.11:** *CPE dependence on forecast and observed pressure. Green: ECMWF operational forecast, Blue: ECMWF ensemble forecast, Black in the plots on the left: hurricane Ike in 2008.*




(a) $CPE$ vs observed $\Delta p$ (● OPER, ● ENS, ● Ike)

(b) $CPE$ vs observed $\Delta p$ (lead times)

(c) $CPE$ vs forecast $\Delta p$ (● OPER, ● ENS, ● Ike)

(d) $CPE$ vs forecast $\Delta p$ (lead times)

**Figure A.12:** *CPE dependence on forecast and observed pressure change. The change in pressure is measured between the present and the previous time step. Green: ECMWF operational forecast, Blue: ECMWF ensemble forecast, Black in the plots on the left: hurricane Ike in 2008.*

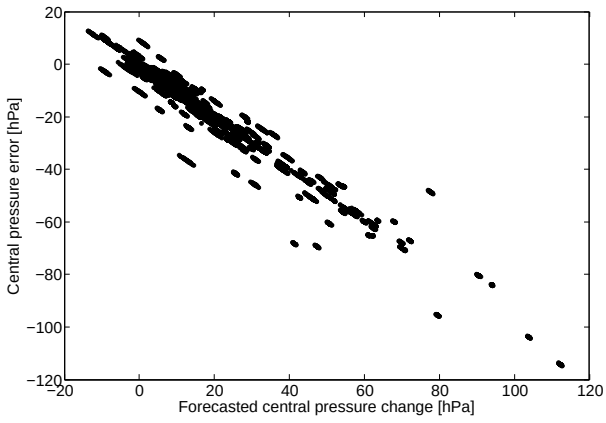



**Figure A.13:** *Dependence of initial central pressure error on pressure change between the previous and the present observation.*

## A.5 Object-Based Verification for Hurricane Ike

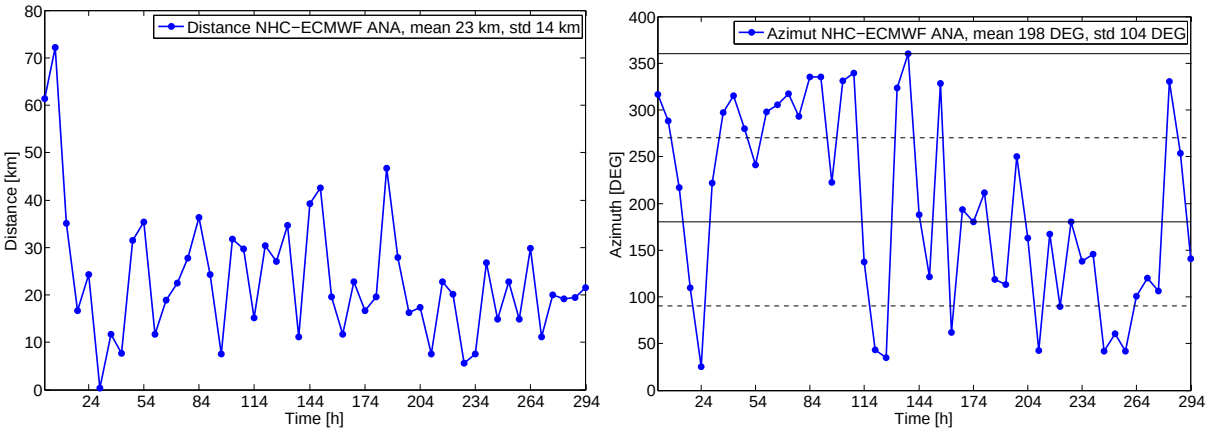



**Figure A.14:** *Difference in position of the minimum central pressure between the ECMWF analysis field and NHC best track data. Left: Distance between NHC best track position and ECMWF analysis. Right: Azimuthal error between NHC best track position and ECMWF analysis.*

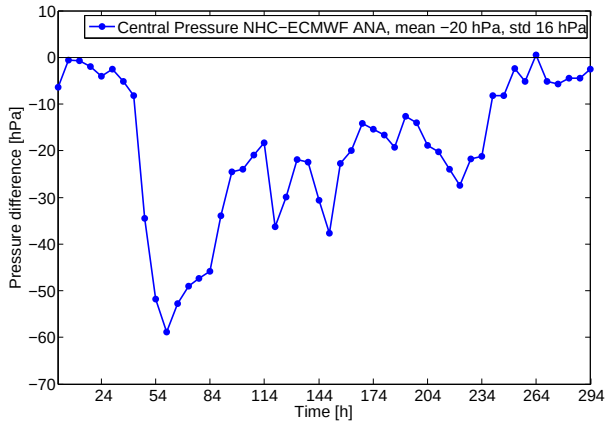



**Figure A.15:** *Central Pressure difference between NHC best track data and ECMWF analysis.*

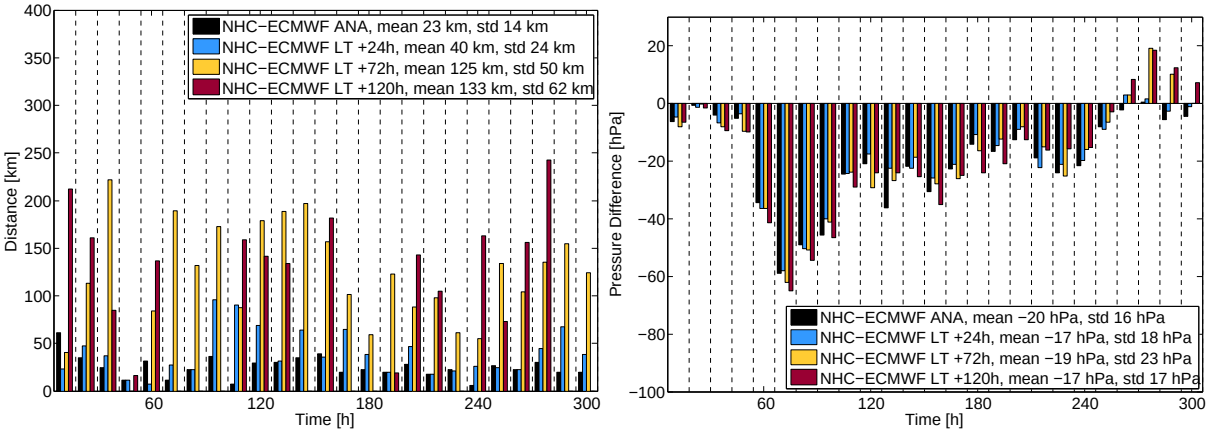
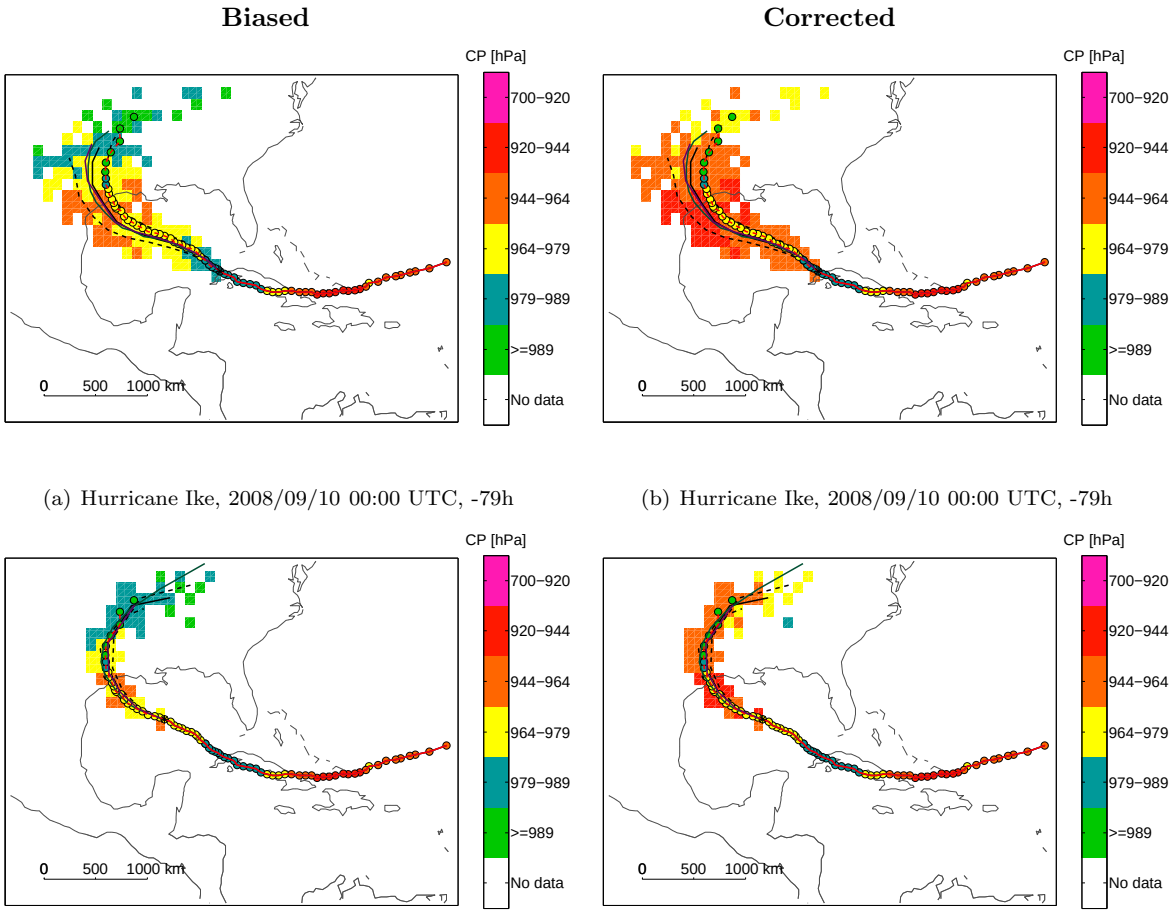



**Figure A.16:** *Position and central pressure difference between NHC best track data and ECMWF analysis and forecasts. Left: Distance between NHC best track position and ECMWF analysis as well as +24 h, +72 h and +120 h lead time forecasts. Right: Central Pressure difference between NHC best track data and ECMWF analysis as well as +24 h, +72 h and +120 h lead time forecasts.*

## A.6 Meteorological Forecast Maps of Selected Case Studies for Insurance Loss Predictions



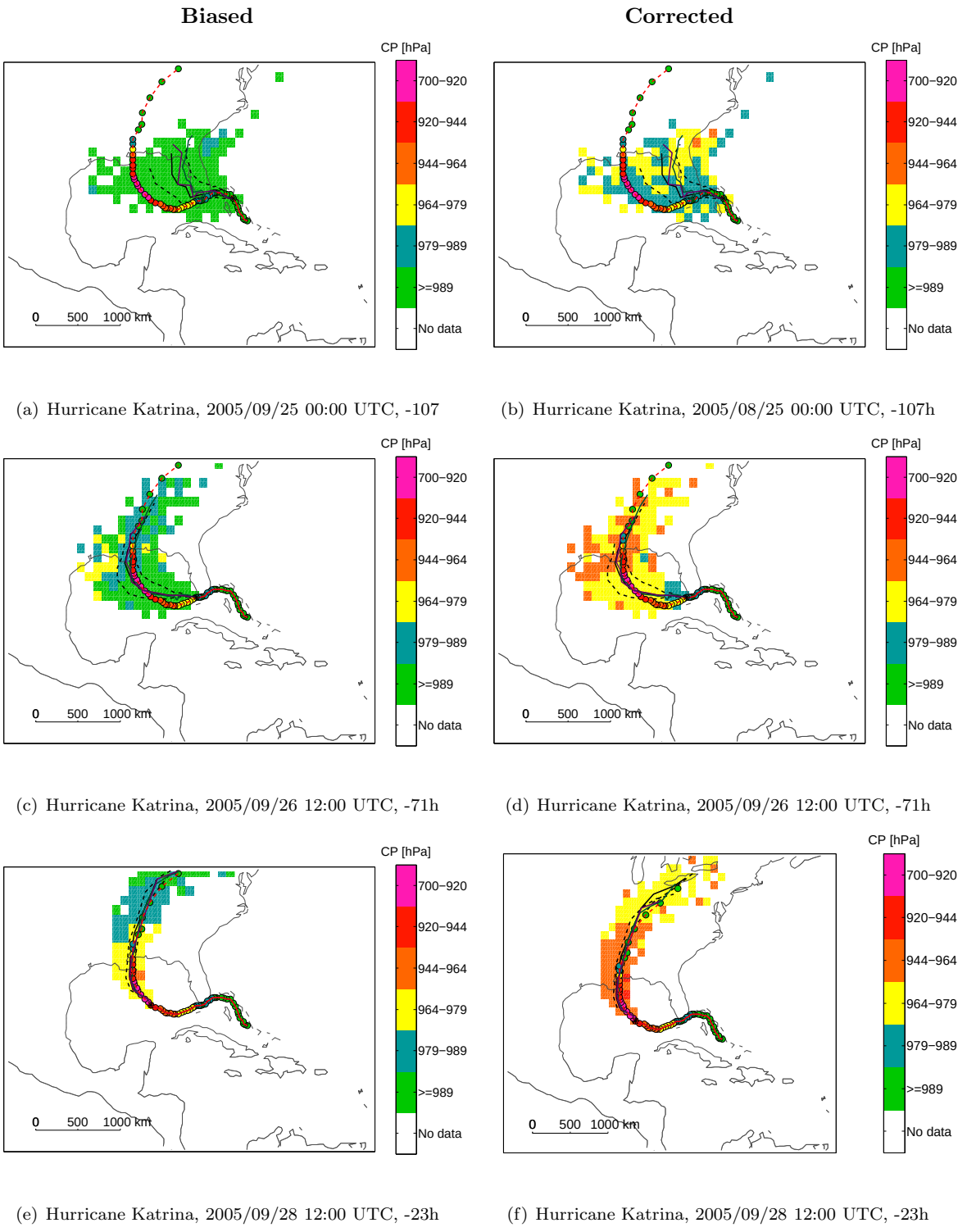

(a) Hurricane Ike, 2008/09/10 00:00 UTC, -79h

(b) Hurricane Ike, 2008/09/10 00:00 UTC, -79h

(c) Hurricane Ike, 2008/09/12 00:00 UTC, -39h

(d) Hurricane Ike, 2008/09/12 00:00 UTC -39h

**Figure A.17:** *ECMWF ensemble track and central pressure (CP) in hPa forecasts for hurricane Ike for the indicated dates and times. On the left the original, biased pressure maps, on the right, the bias corrected pressure maps. — Past observations, * Analysed position, – – – Future observations, — Operational forecast, — Control forecast, — Ensemble median, – – – Ensemble quartiles.*




(a) Hurricane Katrina, 2005/09/25 00:00 UTC, -107

(b) Hurricane Katrina, 2005/08/25 00:00 UTC, -107h

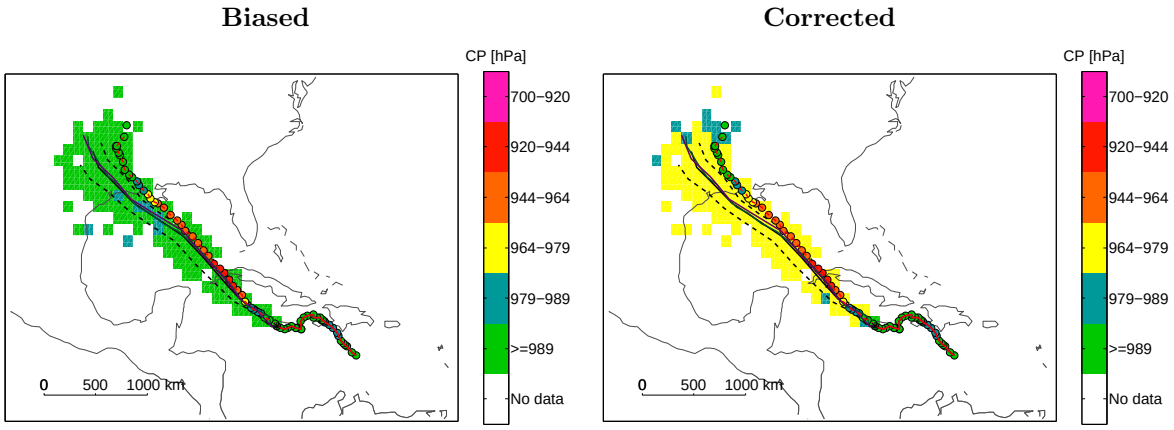
(c) Hurricane Katrina, 2005/09/26 12:00 UTC, -71h

(d) Hurricane Katrina, 2005/09/26 12:00 UTC, -71h

(e) Hurricane Katrina, 2005/09/28 12:00 UTC, -23h

(f) Hurricane Katrina, 2005/09/28 12:00 UTC, -23h

**Figure A.18:** *ECMWF ensemble track and central pressure (CP) in hPa forecasts for hurricane Katrina for the indicated dates and times. On the left the original, biased pressure maps, on the right, the bias corrected pressure maps. Lead times correspond to the landfall in New Orleans. — Past observations, * Analysed position, – – – Future observations, — Operational forecast, — Control forecast, — Ensemble median, – – – Ensemble quartiles.*



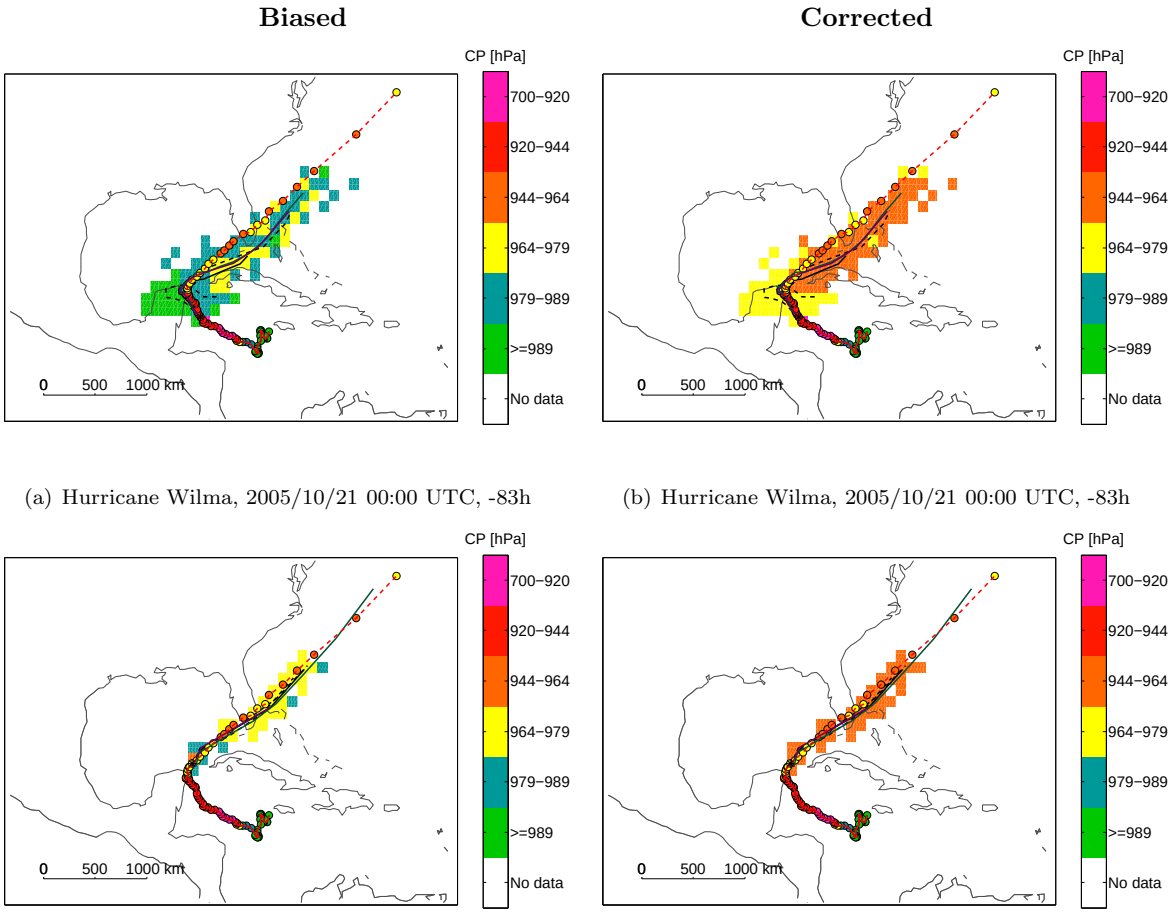

(a) Hurricane Gustav, 2008/08/29 12:00 UTC, -72h

(b) Hurricane Gustav, 2008/08/29 12:00 UTC, -72h

(c) Hurricane Gustav, 2008/08/31 12:00 UTC, -24h

(d) Hurricane Gustav, 2008/08/31 12:00 UTC, -24h

**Figure A.19:** *ECMWF ensemble track and central pressure (CP) in hPa forecasts for hurricane Gustav for the indicated dates and times. On the left the original, biased pressure maps, on the right, the bias corrected pressure maps. — Past observations, * Analysed position, – – – Future observations, — Operational forecast, — Control forecast, — Ensemble median, – – – Ensemble quartiles.*



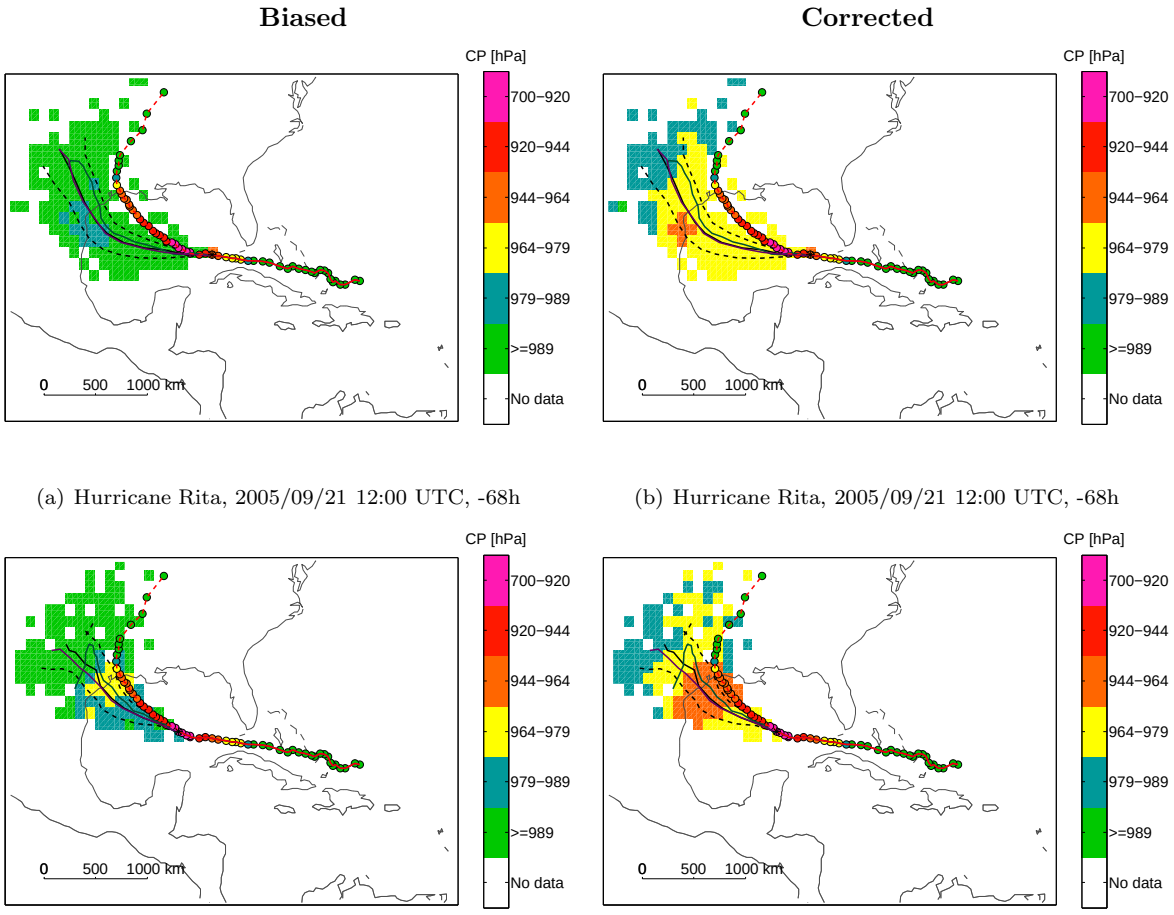

(a) Hurricane Wilma, 2005/10/21 00:00 UTC, -83h

(b) Hurricane Wilma, 2005/10/21 00:00 UTC, -83h

(c) Hurricane Wilma, 2005/10/23 12:00 UTC, -23h

(d) Hurricane Wilma, 2005/10/23 12:00 UTC, -23h

**Figure A.20:** *ECMWF ensemble track and central pressure (CP) in hPa forecasts for hurricane Wilma for the indicated dates and times. On the left the original, biased pressure maps, on the right, the bias corrected pressure maps. Lead times corresponds to the landfall in Florida. — Past observations, * Analysed position, – – – Future observations, — Operational forecast, — Control forecast, — Ensemble median, – – – Ensemble quartiles.*



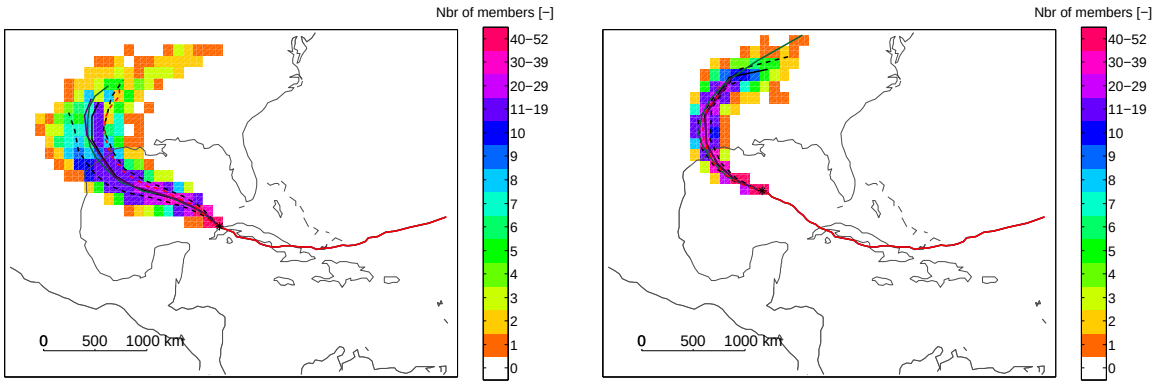

(a) Hurricane Rita, 2005/09/21 12:00 UTC, -68h

(b) Hurricane Rita, 2005/09/21 12:00 UTC, -68h

(c) Hurricane Rita, 2005/09/22 12:00 UTC, -44h

(d) Hurricane Rita, 2005/09/22 12:00 UTC, -44h

**Figure A.21:** *ECMWF ensemble track and central pressure (CP) in hPa forecasts for hurricane Rita for the indicated dates and times. On the left the original, biased pressure maps, on the right, the bias corrected pressure maps. — Past observations, * Analysed position, – – – Future observations, — Operational forecast, — Control forecast, — Ensemble median, – – – Ensemble quartiles.*




(a) Hurricane Ike, 2008/09/10 00:00 UTC, -79h

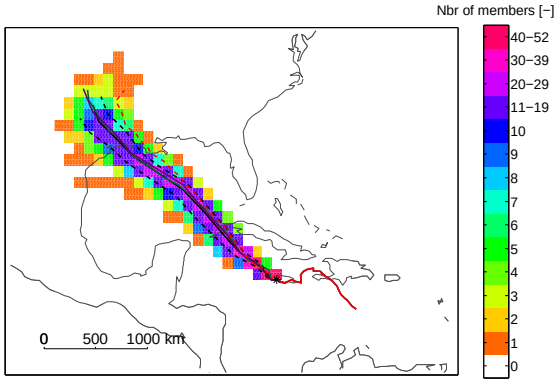
(b) Hurricane Ike, 2008/09/12 12:00 UTC, -31h

(c) Hurricane Katrina, 2005/08/25 00:00 UTC, -107h

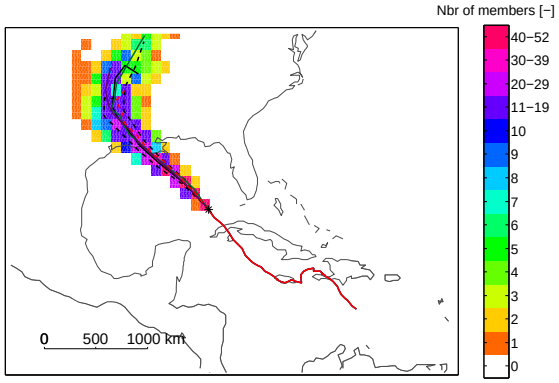
(d) Hurricane Katrina, 2005/08/26 12:00 UTC, -71h

(e) Hurricane Katrina, 2005/08/28 12:00 UTC, -23h

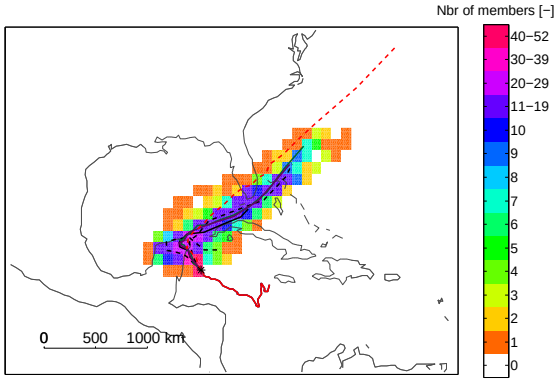
**Figure A.22:** *ECMWF ensemble track forecasts for hurricane Ike and Katrina for the indicated dates and times. The number of ensemble members passing through a specific grid point is shown. The lead times before landfall for Hurricane Katrina correspond to the landfall in New Orleans. — Past observations, * Analysed position, – – – Future observations, — Operational forecast, — Control forecast, — Ensemble median, – – – Ensemble quartiles.*



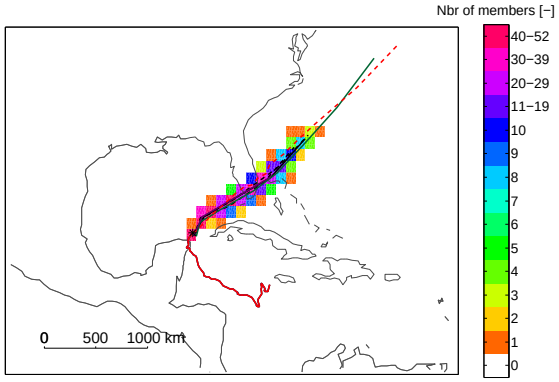

(a) Hurricane Gustav, 2008/08/29 12:00 UTC, 72h



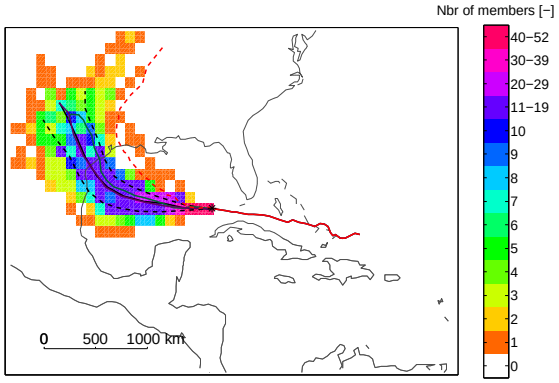

(b) Hurricane Gustav, 2008/08/31 12:00 UTC, -24h



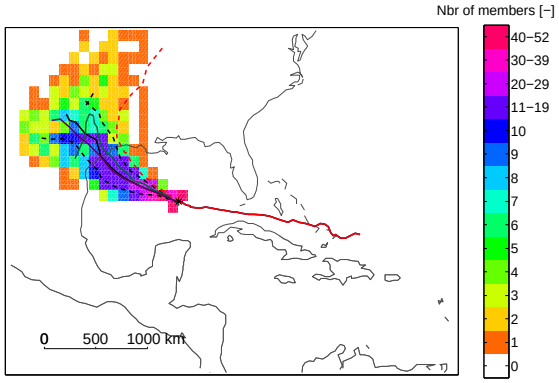

(c) Hurricane Wilma, 2005/10/21 00:00 UTC, -83h




(d) Hurricane Wilma, 2005/10/23 12:00 UTC, -23h




(e) Hurricane Rita, 2005/09/21 12:00 UTC, -68h




(f) Hurricane Rita, 2005/09/22 12:00 UTC, -44h

**Figure A.23:** *ECMWF ensemble track forecasts for hurricane Gustav, Wilma and Rita for the indicated dates and times. The number of ensemble members passing trough a specific grid point is shown. For Hurricane Wilma the lead times corresponds to the landfall in Florida. — Past observations, * Analysed position, – – – Future observations, — Operational forecast, — Control forecast, — Ensemble median, – – – Ensemble quartiles.*

## A.7 Insurance Loss Predictions for Selected Case Studies

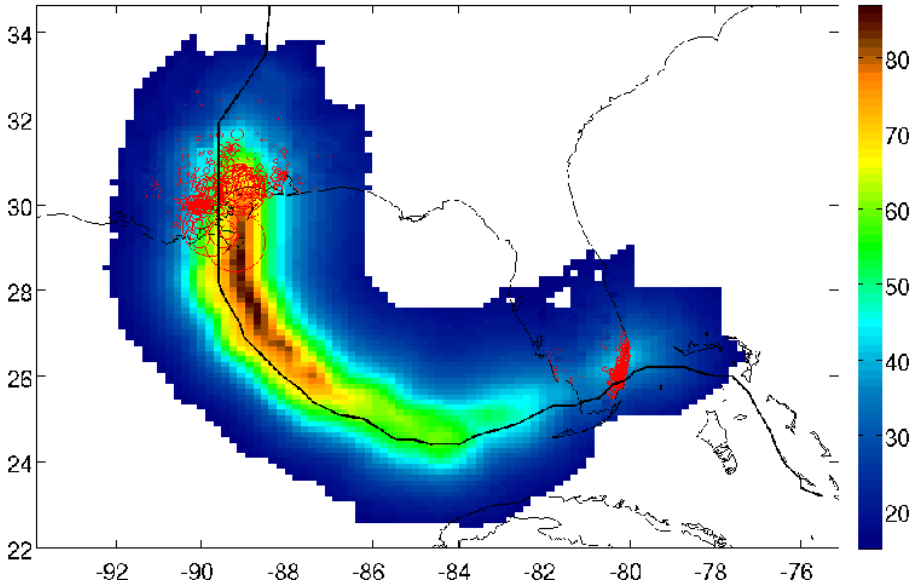
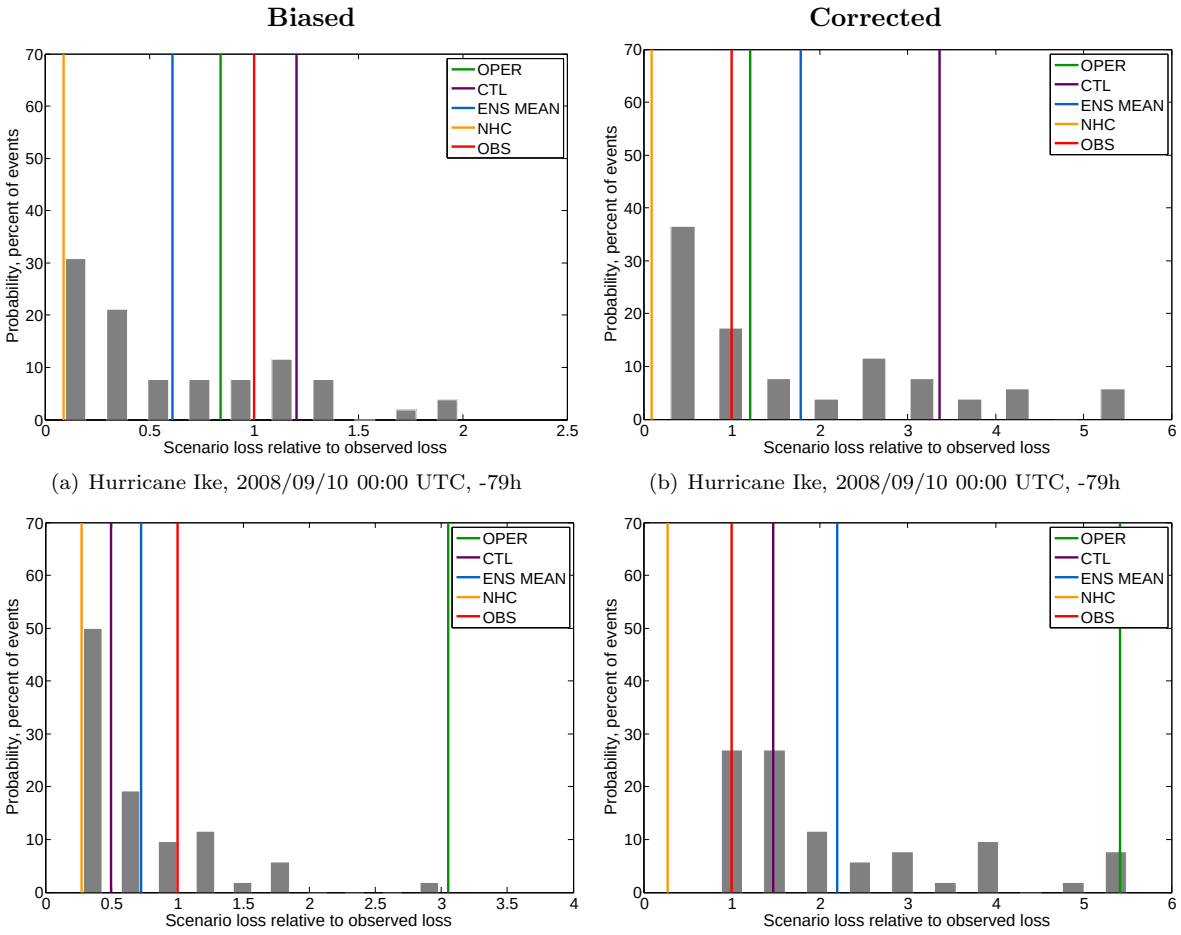



**Figure A.24:** *Observed losses and wind footprint for hurricane Katrina: US Market Portfolio from the 18.06.2008. Importance of losses are indicated by the diameter of the red circles. The colorbar indicates wind speeds in m/s*



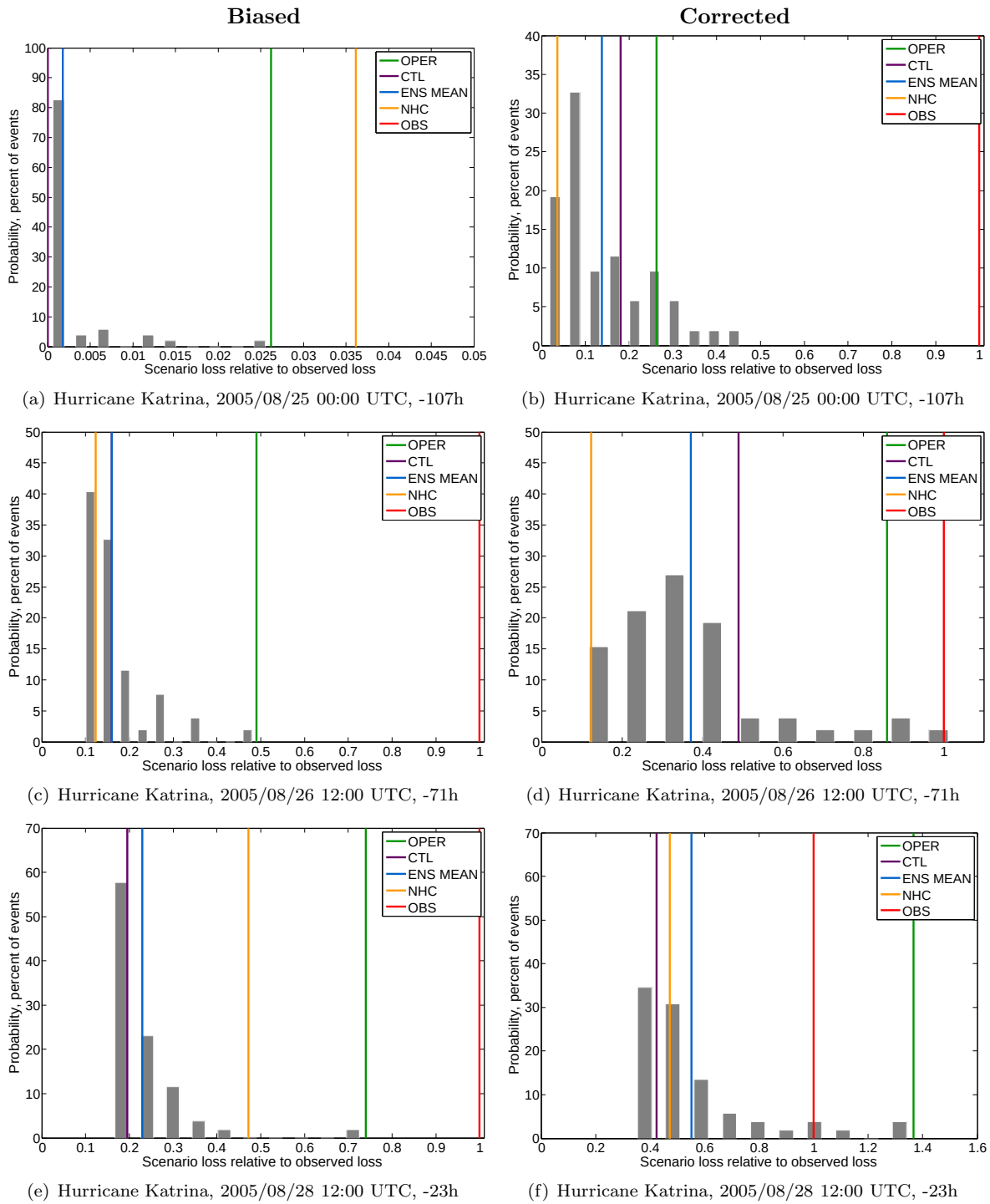

(a) Hurricane Ike, 2008/09/10 00:00 UTC, -79h

(b) Hurricane Ike, 2008/09/10 00:00 UTC, -79h

(c) Hurricane Ike, 2008/09/12 00:00 UTC, -31h

(d) Hurricane Ike, 2008/09/12 00:00 UTC, -31h

**Figure A.25:** *Loss predictions for hurricane Ike for the indicated forecast date and time. On the left, the original, biased loss predictions. On the right, the bias corrected loss predictions. The absolute losses are in the order of magnitude of 10 billion US Dollars.*




(a) Hurricane Katrina, 2005/08/25 00:00 UTC, -107h

(b) Hurricane Katrina, 2005/08/25 00:00 UTC, -107h

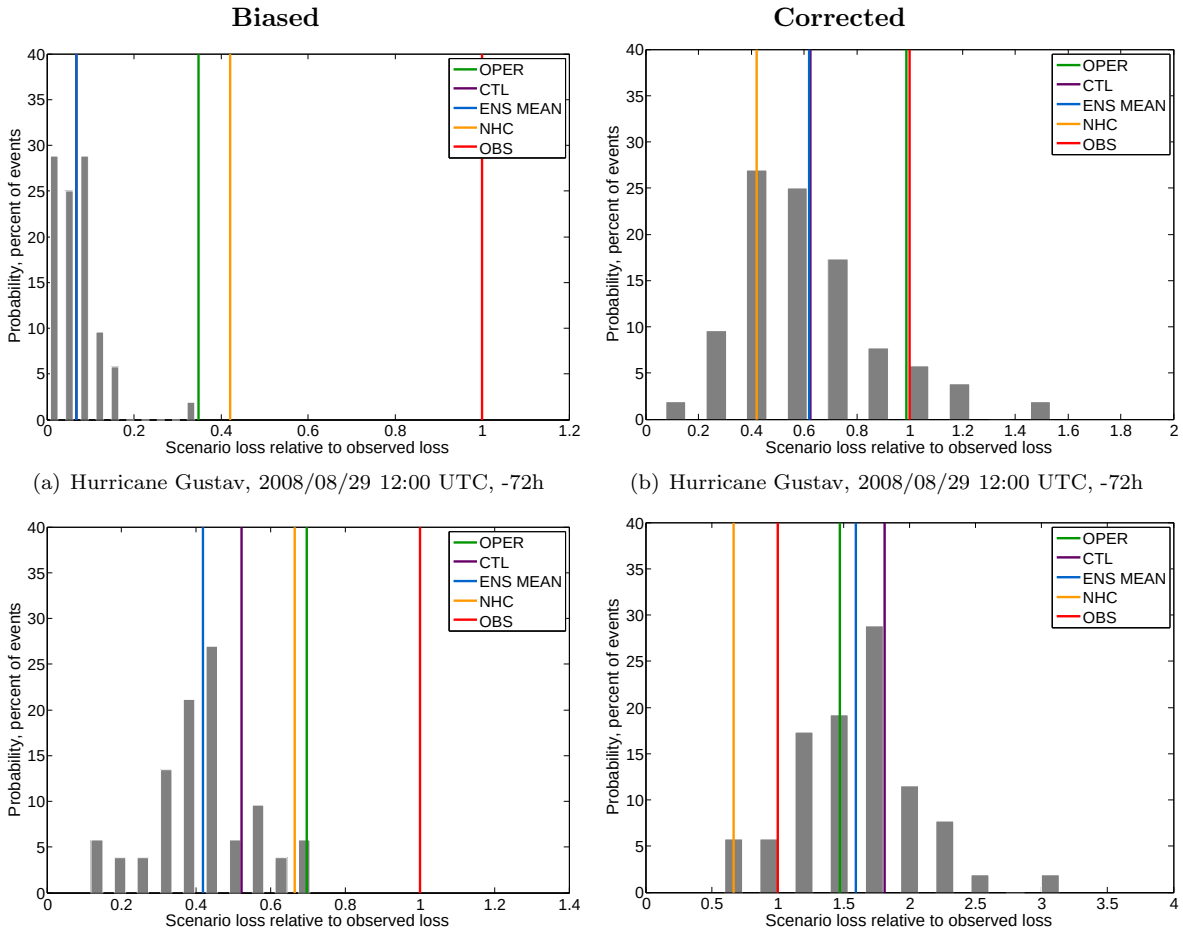
(c) Hurricane Katrina, 2005/08/26 12:00 UTC, -71h

(d) Hurricane Katrina, 2005/08/26 12:00 UTC, -71h

(e) Hurricane Katrina, 2005/08/28 12:00 UTC, -23h

(f) Hurricane Katrina, 2005/08/28 12:00 UTC, -23h

**Figure A.26:** *Loss predictions for hurricane Katrina for the indicated forecast date and time. On the left, the original, biased loss predictions. On the right, the bias corrected loss predictions. The absolute losses are in the order of magnitude of 10 billion US Dollars. Lead times correspond to the landfall in New Orleans.*



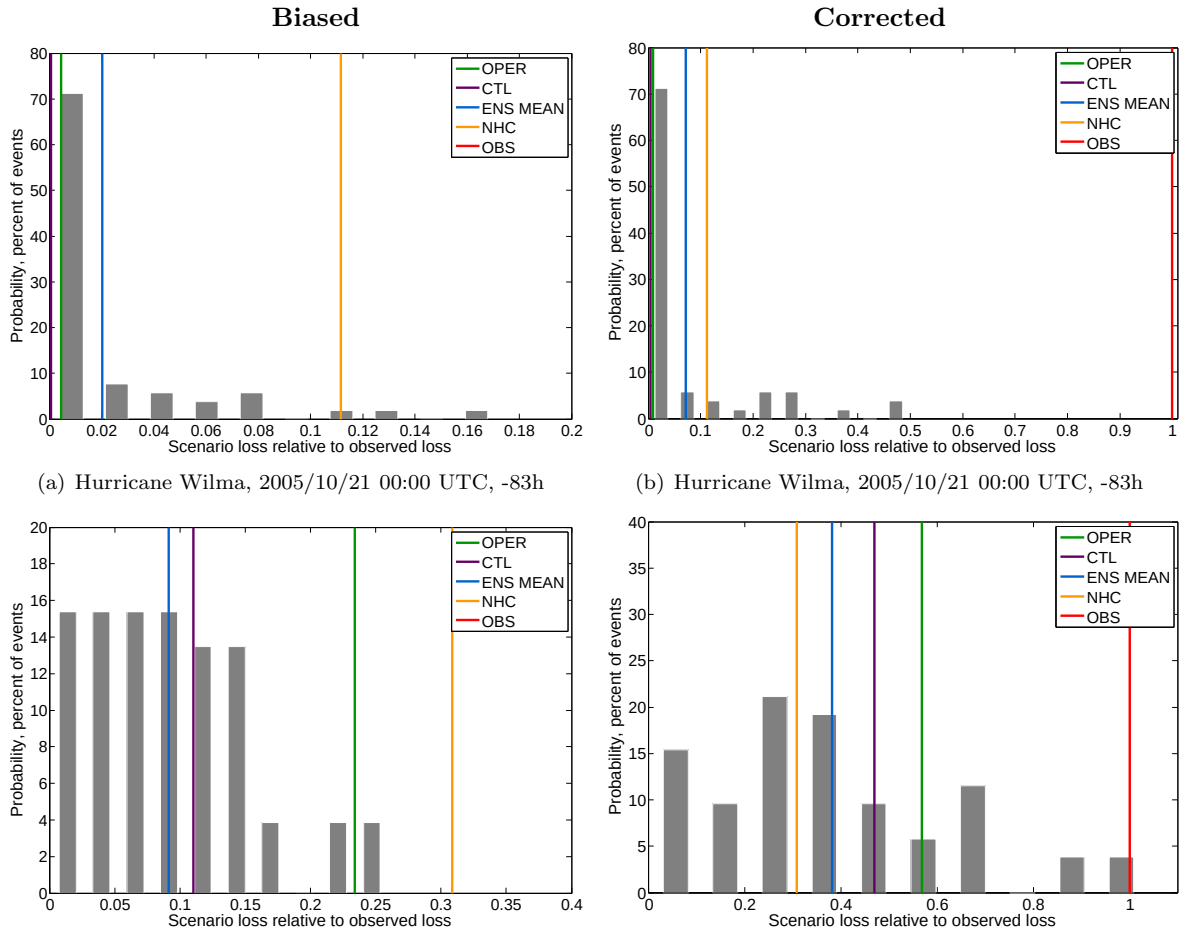

(a) Hurricane Gustav, 2008/08/29 12:00 UTC, -72h

(b) Hurricane Gustav, 2008/08/29 12:00 UTC, -72h

(c) Hurricane Gustav, 2008/08/31 12:00 UTC, -24h

(d) Hurricane Gustav, 2008/08/31 12:00 UTC, -24h

**Figure A.27:** *Loss predictions for hurricane Gustav for the indicated forecast date and time. On the left, the original, biased loss predictions. On the right, the bias corrected loss predictions. The absolute losses are in the order of magnitude of 10 billion US Dollars.*



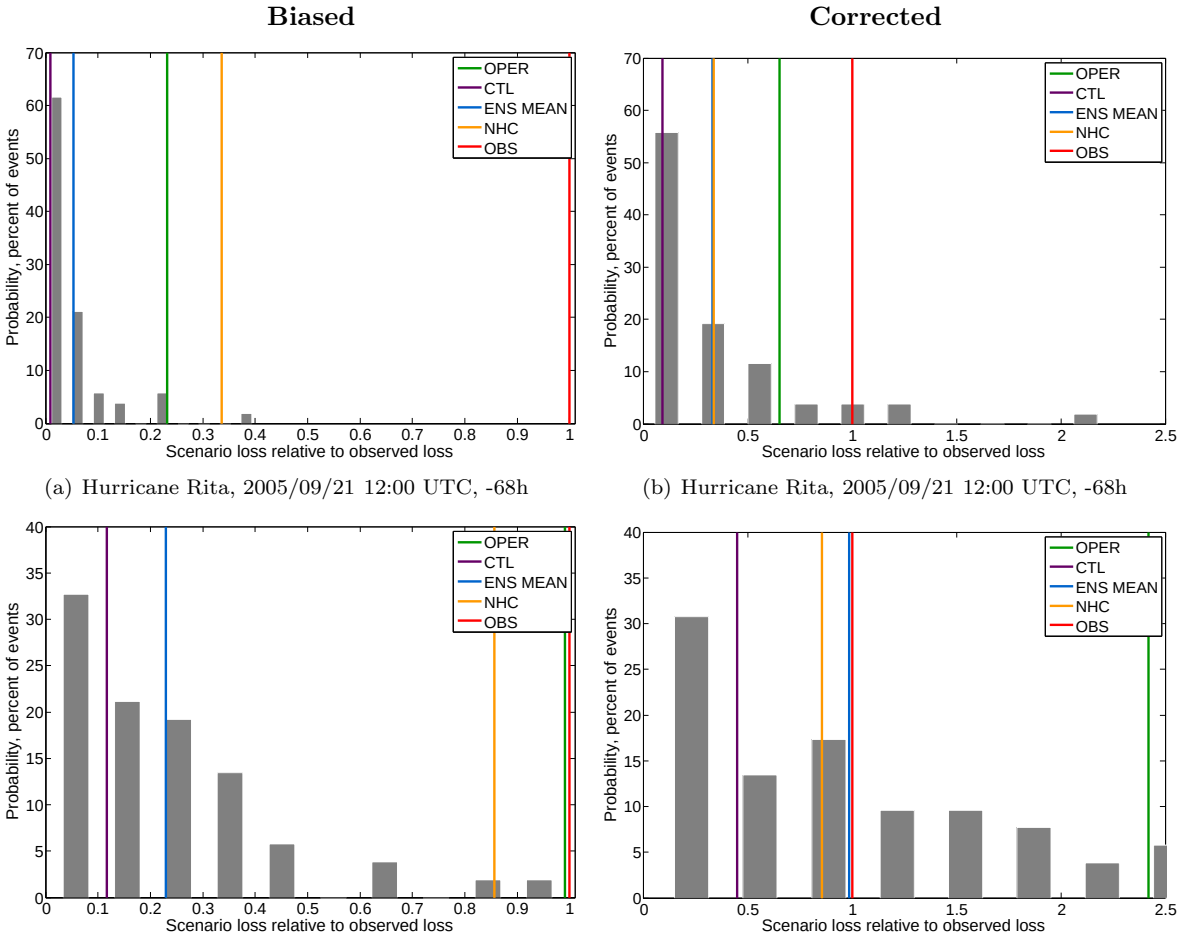

(a) Hurricane Wilma, 2005/10/21 00:00 UTC, -83h

(b) Hurricane Wilma, 2005/10/21 00:00 UTC, -83h

(c) Hurricane Wilma, 2005/10/23 12:00 UTC, -23h

(d) Hurricane Wilma, 2005/10/23 12:00 UTC, -23h

**Figure A.28:** *Loss predictions for hurricane Wilma for the indicated forecast date and time. The lead times correspond to the landfall in Florida. On the left, the original, biased loss predictions. On the right, the bias corrected loss predictions. The absolute losses are in the order of magnitude of 10 billion US Dollars.*




(a) Hurricane Rita, 2005/09/21 12:00 UTC, -68h

(b) Hurricane Rita, 2005/09/21 12:00 UTC, -68h

(c) Hurricane Rita, 2005/09/22 12:00 UTC, -44h

(d) Hurricane Rita, 2005/09/22 12:00 UTC, -44h

**Figure A.29:** *Loss predictions for hurricane Rita for the indicated forecast date and time. On the left, the original, biased loss predictions. On the right, the bias corrected loss predictions. The absolute losses are in the order of magnitude of 10 billion US Dollars.*

# Bibliography


Aberson, S. D., 1998: Five-Day Tropical Cyclone Track Forecasts in the North Atlantic Basin. *Weather and Forecasting*, **13**, 1005–1015.

ASCE, 2005: Minimum Design Loads for Buildings and Other Structures. *American Society of Civil Engineers*.

Avila, L. A., 2002: Best Track Determination at NHC. Technical report, NOAA/NWS/NCEP/National Hurricane Center, Miami, Florida.

Bell, G. D., M. S. Halpert, R. C. Schnell, R. W. Higgins, J. Lawrimore, V. E. Kousky, R. Tinker, W. Thiaw, M. Chelliah, and A. Artusa, 2000: Climate Assessment for 1999. *Bulletin American Meteorological Society*, **81**, s1–s50.

Berg, R., 2009: Tropical Cyclone Report Hurricane Ike. Technical report, NHC.

Berkmeijer, J., R. Buizza, T. N. Palmer, K. Puri, and J. F. Mahouf, 2000: Tropical Singular Vectors Computed with Linearized Diabatic Physics. *Quarterly Journal of the Royal Meteorological Society*, **127**, 685–708.

Beven, J. L., L. A. Avila, E. S. Blake, D. P. Brown, J. L. Franklin, R. D. Knabb, R. J. Pasch, J. R. Rhome, and S. R. Stewart, 2008: Annual Summary, Atlantic Hurricane Season of 2005. *Monthly Weather Review*, **136**, 1109–1173.

Brier, G. W. and R. A. Allen, 1951: *Compendium of Meteorology*, Boston: American Meteorological Society, chapter Verification of Weather Forecasts.

Brown, D. and J. Franklin, 2002: Accuracy of Pressure-Wind Relationships and Dvorak Satellite Intensity Estimates for Tropical Cyclones Determined from Recent Reconnaissance-Based "Best Track" Data. *American Meteorological Society, Preprints, 25th Conference on Hurricanes and Tropical Meteorology, San Diego*.

Buizza, R., 1997: Potential Forecast Skill of Ensemble Prediction and Spread and Skill Distributions of the ECMWF Ensemble Prediction System. *Monthly Weather Review*, **125**, 99–119.

Buizza, R., M. Miller, and T. N. Palmer, 1999: Stochastic Representation of Model Uncertainties in the ECMWF Ensemble Prediction System. *Quarterly Journal of the Royal Meteorological Society*, **125**, 2887–2908.

Cummins, J. D., 2007: CAT Bonds and Other Risk-Linked Securities: State of the Market and Recent Developments. Technical report, Temple University, Fox School of Business and Management.

Davis, C., W. Wang, Y. Chen, J. Dudhia, G. Holland, J. Klemp, J. Michalakes, R. Rotunno, C. Snyder, Q. Xiao, S. S. Chen, K. Corbosiero, M. DeMaria, and H. Reeves, 2008: Prediction of Landfalling Hurricanes with the Advanced Hurricane WRF Model. *Monthly Weather Review*, **136**, 1990–2005, doi:10.1175/2007MWR2085.1.

DeMaria, M., J. A. Knaff, and J. Kaplan, 2006: On the Decay of Tropical Cyclone Winds Crossing Narrow Landmasses. *Journal of Applied Meteorology*, **45**, 491–499.

Dvorak, V. F., 1984: Tropical Cyclone Intensity Analysis Using Satellite Data. Technical report, NOAA.

ECMWF, 2007: *IFS Documentation Cy31r1, Operational Implementation 12 September 2006, Part V: The Ensemble Prediction System*. European Centre for Medium-Range Weather Forecasts.

Emanuel, K., 1991: The Theory of Hurricanes. *Annual Review of Fluid Mechanics*, **37**, 99–128.

— 2005: Increasing Destructiveness of Tropical Cyclones over the Past 30 Years. *Nature*, **436**, 686–688.

— 2009: Tropical Cyclone Energetics and Structure. Programme in Atmospheres, Oceans, and Climate Massachusetts Institute of Technology.

Finley, J., 1884: Tornado Predictions. *American Meteorological Journal*, **1**, 85–88.

Forchaux, S., 2006: Se couvrir contre les catastrophes naturelles, un véhicule à la mode. *L'AGEFI, Quotidien de l'agence économique et financière à Genève*.

Fovell, R. G. and H. Su, 2007: Impact of Cloud Microphysics on Hurricane Track Forecasts. *Geophysical Research Letters*, **34**, L24810.

Franklin, J. L., 2006: 2005 National Hurricane Center Forecast Verification Report. Technical report, NHC.

— 2009: 2008 National Hurricane Center Forecast Verification Report. Technical report, NHC.

Froude, L. S. R., 2009: Regional Differences in the Prediction of Extratropical Cyclones by the ECMWF Ensemble Prediction System. *Monthly Weather Review*, **137**, 893–911.

Froude, L. S. R., L. Bengtsson, and K. I. Hodges, 2007: The Predictability of Extratropical Storm Tracks and the Sensitivity of their Prediction to the Observing System. *Monthly Weather Review*, **135**, 315–333.

Heming, J., 1994: Verification of Tropical Cyclone Forecast Tracks at the UK Met Office. *NWP Gazette*, **1**, 2–8.

Hodge, T., 2006: The Impact of Tropical Cyclones on Gulf of Mexico Crude Oil and Natural Gas Production. Technical report, U.S. Energy Information Administration (EIA).

Hoff, H., L. Bouwer, G. Berz, W. Kron, and T. Loster, 2003: Risk Management in Water and Climate - the Role of Insurance and Other Financial Services. Technical report, Munich Re, Greenwatch GE.

Holland, G. J., 1980: An Analytic Model of the Wind and Pressure Profiles in Hurricanes. *Monthly Weather Review*, **108**, 1212–1218.

Houze, R. A., 1993: *Cloud Dynamics*, volume 53 of *International geophysics series*. Elsevier Science and Technology Books.

IPCC, 2007: Climate Change 2007: Synthesis Report: Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Technical report, IPCC, Geneva, Switzerland.

Jarvinen, B. R. and C. J. Neumann, 1979: Statistical Forecasts of Tropical Cyclone Intensity for the North Atlantic Basin. Noaa technical memorandum, nws, nhc, NOAA.

Jarvinen, B. R., J. C. Neumann, and M. A. S. Davis, 1984: A Tropical Cyclone Data Tape for the North Atlantic Basin 1886-1983. Noaa technical memorandum, nws, nhc, NOAA.

Jolliffe, I. T. and D. B. Stephenson, 2003: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. Wiley edition.

Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press.

Kantha, L., 2006: Time to Replace the Saffir-Simpson Hurricane Scale? *EOS Transactions, American Geophysical Union*, **87**, 3–6.

Krishnamurti, T. N., C. M. Kishtawal, T. E. LaRow, D. R. Bachiochi, Z. Zang, C. E. Williford, S. Gadgil, and S. Surendran, 1999: Improved Weather and Seasonal Climate Forecasts from Multimodel Superensemble. *Science*, **285**, 1548–1550.

Li, C. and J. L. Rego, 2009: On the Importance of the Forward Speed of Hurricanes in Storm Surge Forecasting: A Numerical Study. *Geophysical Research Letters*, **36**, L07609.

Liu, H., L. Xie, L. J. Pietrafesa, and S. Bao, 2007: Sensitivity of Wind Waves to Hurricane Wind Characteristics. *Ocean Modelling*, **18**, 37–52.

Lohmann, U., 2009: Cloud Dynamics. ETHZ lecture course.

Longshore, D., 2008: *Encyclopedia of Hurricanes, Typhoons and Cyclones*. FactsOnFile Inc.

Lorenz, E. N., 1963: Deterministic Nonperiodic Flow. *Journal of Atmospheric Sciences*, 130–141.

Maclay, K. S., M. DeMaria, and T. H. V. Haar, 2008: Tropical Cyclone Inner-Core Kinetic Energy Evolution. *Monthly Weather Review*, **136**, 4882–4898.

Magnusson, L., J. Nycander, and E. Kallén, 2009: Flow-Dependent Versus Flow-Independent Initial Perturbations for Ensemble Prediction. *Tellus*, **61A**, 194–209.

Mahendran, M., 1998: Cyclone Intensity Categories. *Weather and Forecasting*, **13**, 878–883.

Mahouf, J.-F., 1999: Influence of Physical Processes on the Tangent Linear Approximation. *Tellus*, **51A**, 147–166.

McCollor, B. and R. Stull, 2008a: Hydrometeorological Short-Range Ensemble Forecasts in Complex Terrain. Part II: Economic Evaluation. *Weather and Forecasting*, **23**, 557–574, doi:10.1175/2007WAF2007064.1.

McCollor, D. and R. Stull, 2008b: Hydrometeorological Short-Range Ensemble Forecasts in Complex Terrain. Part I: Meteorological Evaluation. *Weather and Forecasting*, **23**, 533–556.

Merill, R. T., 1980: A Statistical Tropical Cyclone Motion Forecasting System for the Gulf of Mexico. Noaa technical memorandum, nws, nhc, NOAA.

— 1984: A Comparison of Large and Small Tropical Cyclones. *Monthly Weather Review*, **112**, 1408–1418.

Meyer, P., M. Bisping, M. Weber, and C. Brauner, 1997: Tropical Cyclones. Technical report, Swiss Reinsurance Company.

Mills, E., 2005: Insurance in a Climate of Change. *Science*, **309**, 1040–1044.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF Ensemble Prediction System: Methodology and Validation. *Quarterly Journal of the Royal Meteorological Society*, **122**, 73–119.

Muller, R., 1944a: Verification of Short-Range Weather Forecasts (a Survey of the Literature), Part I. *Bulletin American Meteorological Society*, **25**, 18–27.

— 1944b: Verification of Short-Range Weather Forecasts (a Survey of the Literature), Part II. *Bulletin American Meteorological Society*, **25**, 47–53.

— 1944c: Verification of Short-Range Weather Forecasts (a Survey of the Literature), Part III. *Bulletin American Meteorological Society*, **25**, 88–95.

Murnane, R. J., 2004: The Importance of Best-Track Data for Understanding the Past, Present, and Future of Hurricanes and Typhoons. *Hurricanes and Typhoons: Past, Present and Future*, 249–266.

Murphy, A., 1973: A New Vector Partition of the Probability Score. *Journal of Applied Meteorology*, **12**, 595–600.

Murphy, A. and R. Winkler, 1987: A General Framework for Forecast Verification. *Monthly Weather Review*, **115**, 1330–1338.

Neumann, J. C., 1972: An Alternate to the Hurran (Hurricane Analog) Tropical Cyclone Forecasting System. Noaa technical memorandum, nws, nhc, NOAA.

Neumann, J. C. and J. M. Pelissier, 1981: An Analysis of Atlantic Tropical Cyclone Forecast Errors, 1970-1979. *Monthly Weather Review*, **109**, 1248–1266.

NHC, 2008: *National Hurricane Center Forecast Verification*. http://www.nhc.noaa.gov/verification.

— 2009: Experimental Saffir-Simpson Hurricane Wind Scale. Note.
URL `http://www.weather.gov/infoservicechanges/sshws.pdf`

Pellerin, G., L. Lefaivre, P. Houtekamer, and C. Girard, 2003: Increasing the Horizontal Resolution of Ensemble Forecasts at CMC. *Nonlinear Processes in Geophysics*, **10**, 463–468.

Persson, A. and F. Grazzini, 2005: *User Guide to ECMWF Forecast Products*. European Center for Medium Range Weather Forecast (ECMWF).

Pielke, R. A. and C. W. Landsea, 1998: Normalized Hurricane Damages in the United States: 1925–95. *Weather and Forecasting*, **13**, 621–631.

Poincaré, H., 1908: *Science et méthode*. Flammarion.

Powell, M. D. and T. A. Reinhold, 2007: Tropical Cyclone Destructive Potential by Integrated Kinetic Energy. *Bulletin of the American Meteorological Society*, **88**, 513–526.

Puri, K. and T. N. Palmer, 2001: Ensemble Prediction of Tropical Cyclones Using Targeted Diabatic Singular Vectors. *Quarterly Journal of the Royal Meteorological Society*, **127**, 709–731.

Rhome, J. R., 2007: Technical Summary of the National Hurricane Center Track and Intensity Models. Technical report, National Oceanic and Atmospheric Administration (NOAA), National Hurricane Center, USA.

Richardson, D., 2000: Skill and Relative Economic Value of the ECMWF Ensemble Prediction System. *Journal of the Royal Meteorological Society*, **126**, 649–667.

RMS, 2008: A Guide to Catastrophe Modelling. *The Review, Worldwide Reinsurance*.

Roulin, E., 2007: Skill and Relative Economic Value of Medium-Range Hydrological Ensemble Predictions. *Hydrology and Earth System Sciences*, **11**, 725–737.

Santos, P., D. Sharp, M. DeMaria, and S. Kiser, 2009: The Determination of Optimal Thresholds of Tropical Cyclone Incremental Wind Speed Probabilities to Support Expressions of Uncertainty in Text Forecasts. *AMS Symposium on Urban High Impact Weather, Phoenix, AZ*.

Schär, C., 2008: Numerical Modelling of Weather and Climate, lecture at ETHZ.

Shapiro, L. J., 1983: Asymmetric Boundary Layer Flow under a Translating Hurricane. *Journal of Atmospheric Sciences*, **40**, 1984–1998.

Sigma, 2006: Securitization - New Opportunities for Insurers and Investors. Technical Report 7, Swiss Re.

Simpson, R. and H. Saffir, 2007: Comment on Powell and Reinhold's 2007 Tropical Cyclone Destructive Potential by Integrated Kinetic Energy article. *Bulletin of the American Meteorological Society*, **88**, 1799–1800.

Simpson, R. H. and H. Riehl, 1981: *The Hurricane and Its Impact*. Louisiana State University Press.

Soille, P., 1999: *Morphological Image Analysis: Principles and Applications*. Springer.

Sommerville, R., 2009: Energy Market Review: Gulf of Mexico Windstorm: Still the Insoluble Risk Management Problem. Technical report, Willis.

Spinner, K., 2009: Hurricane Forecasters Perplexed over Intensity Predictions. *Herald Tribune*.

Stephenson, D. B. and F. J. Doblas-Reyes, 2000: Statistical Methods for Interpreting Monte Carlo Ensemble Forecasts. *Tellus*, 300–322.

Thompson, P. D., 1957: Uncertainty of initial state as a factor in the predictability of large scale atmospheric flow patterns. *Tellus*, **9**, 275–295.

Van der Grijn, G., 2002: Tropical cyclone forecasting at ECMWF: new products and validation. ECMWF Technical Memo 386, European Centre for Medium-Range Weather Forecasts.

Van der Grijn, G., J. E. Paulsen, F. Lalaurette, and M. Leutbecher, 2004: Early medium-range forecasts of tropical cyclones. Technical Report 102, ECMWF Newsletter.

Wallace, J. M. and P. V. Hobbs, 2006: *Atmospheric Science: An Introductory Survey*. Academic Press.

Wells, J., 2006: Natural Gas: Factors Affecting Prices and Potential Impacts on Consumers. Technical report, U.S. Government Accountability Office.

Wernli, H., M. Paulat, M. Hagen, and C. Frei, 2008: SAL—A Novel Quality Measure for the Verification of Quantitative Precipitation Forecasts. *Monthly Weather Review*, **136**, 4470–4487, doi:10.1175/2008MWR2415.1.

Woods, A., 2006: *Medium-Range Weather Prediction*. Springer.

Zimmerli, P., 2003: Natural Catastrophes and Reinsurance. Technical report, Swiss Reinsurance Company.

# Acknowledgements

This master thesis on "Tropical Cyclone Forecast Verification: Three Approaches to the Assessment of the ECMWF Model" has been written in a collaboration framework between the Institute of Environmental Engineering (IFU), the Institute of Atmospheric and Climate Sciences (IAC) at ETHZ and the Swiss reinsurance company Swiss Re.

I would like to express my gratitude to all the people and institutions who helped me with the realisation of this master thesis. My special thanks to

Dr. Olivia Martius vom IAC für deine grosszügige Unterstützung bei allen möglichen Problemen, für die Bereitstellung der ECMWF Daten, für die Einführung ins IACleben, für die interessanten Besprechungen, für deine nützlichen Tipps und konstruktiven Feedbacks, für die Gegenlesung dieses Berichts.

Dr. Marc Wüest von Swiss Re für die Zeit, die du dir so oft für unsere spannenden Besprechungen genommen hast, für den Einblick den du mir in den Swiss Re-Alltag gewährt hast, für die Durchführung der Verlustrechnungen, für die interessante Zusammenarbeit.

Professor Dr. Paolo Burlando, head of the Chair of Hydrology and Water Resources Management at the IFU for having made the project collaboration in this master thesis possible under your supervision and for your interesting questions and feedback.

Professor Dr. Huw Davies, head of the Chair of Dynamical Meteorology at the IAC for giving me the opportunity to do my master thesis in your group.

Professor Dr. Heini Wernli for the interesting discussions and all your valuable comments, inputs and ideas.

The Swiss Federal Office of Meteorology and Climatology MeteoSwiss, in particular Christophe Voisard, for providing the ECMWF tropical cyclone ensemble forecasts.

ECMWF for providing the analysis and forecast data.

Swiss Re for the fascinating topic proposition and the access to the catastrophe modelling software.

Der ganzen IAC Dynamikgruppe, sowie Thierry Corti für euer Feedback.

Meinen Bachelor-/Masterarbeitskollegen, sowie frisch diplomierten IACler vom CHN D und N-Stock für die Unterstützung und all die interessanten Kaffeediskussionen. Insbesondere Chrigi, Lukas, Ruth, Andreas, Ines und Julian. Meinen Mitstudenten aus dem Studiengang Umweltingenieurwissenschaften für die wunderbare Abschlussreise in Kroatien.

Meinen Eltern, die mir all diese Jahre mein geistiges Abenteuer in Zürich grosszügig mitfinanziert haben und mich vor allem auch immer wieder motiviert haben. Olivier pour ta patience, ton humour et ta bonne humeur, ainsi que nos nombreuses discussions sur les analogies entre ouragans et microscopes à émission de champ. Mirjam und Marius, meinen Grosseltern, Kathi, Nuria, Stefan und den Kindern, Bettina, Röbi für eure Unterstützung. Les Praguoises, les romands de Zurich et ceux en Romandie pour tout le reste.