UNIVERSIDAD DE CANTABRIA

Departamento de Ciencias y Técnicas del Agua y del Medio Ambiente

Numerical modeling of the global wave climate variability and associated environmental and technological risks

---

A Doctoral Thesis by: Borja González Reguero

Advisors: Iñigo J. Losada and Fernando J. Méndez

Santander, December, 2012

Contents

List of Tables

List of Figures

List of Acronyms

AR4 Intergovernmental Panel on Climate Change Assessment Report 4

AMO Atlantic Multidecadal Oscillation

AO Artic Oscillation

BODC British Oceanographic Data Centre

CFSR Climate Forecast System Reanalysis

CIRES Cooperative Institute for Research in Environmental Sciences

COADS Comprehensive Ocean-Atmosphere Data Set

COV Coefficient of Variability

CSIRO Australia's Commonwealth Scientific and Industrial Research Organization (www.csiro.au)

CS01 Global wave reanalysis, [Cox and Swail, 2001]

CRIEPI Central Research Institute of Electric Power Industry - Japan

DCW Digital Chart of the World

DJF Season: December - January - February

DTM Digital Terrain Model

DMI Dipole Mode Index

EA East Atlantic pattern

EA/WR East Atlantic-Western Russian pattern

ECHAM5 Atmospheric general circulation model

ECMWF European Centre for Medium-Range Weather Forecasts

EF Energy Flux

ENOS / ENSO El Niño-Southern Oscillation

EOFs / EOF Empirical Ortogonal Function

EP/NP East Pacific-North Pacific pattern

ERA-Interim European Centre for Medium-Range Weather Forecasts' atmospheric and wave reanalysis (1979 onwards)

ERA-40 European Centre for Medium-Range Weather Forecasts' atmospheric and wave reanalysis (1957-2002)

ERSST Extended Reconstructed Sea Surface Temperature

ETOPO Earth Topography Digital Dataset. A global relief model of Earth's surface that integrates land topography and ocean bathymetry.

FHBA Floating heave-buoy array

FOWC Floating oscillating water column

GEOS Goddard Earth Observing System

F2HB Floating two-body heaving converter

F3OF Floating three-body oscillating flap device

GCM General Circulation Model

GEBCO General Bathymetric Chart of the Oceans

GEV Generalized extreme Value

GFO Geosat Follow-On satellite mission

GISS Goddard Institute for Space Studies

GIA Glacial Isostatic Adjustment

GLCNMO Global Land Cover of the National Mapping Organizations

GMSL Global Mean Sea Level

GRUMP Gridded Population of the World and the Global Rural-Urban Mapping Project

GODAS Global Ocean Data Assimilation System

GOS Global Ocean Surges reanalysis

GOW Global Ocean Waves reanalysis, [Reguero et al., 2012b]

IH-AMEVA Instituto de Hidraúlica de Cantabria's software for Statistical Analysis of Environmental Variables

IOD Indian Ocean Dipole

IPCC Intergovernmental Panel on Climate Change

JCDAS JMA Climate Data Assimilation System

JJA Season: June - July - August

JMA Japan Meteorological Agency

JRA Japanese Reanalysis Project

LAC Latin America and the Caribbean

MAM Season: March - April - May

MERRA NASA Modern Era Reanalysis for Research and Applications

MRI Meteorological Research Institute (Japan)

NAO North Atlantic Oscillation

NASA National Aeronautics and Space Administration

NCAR National Center for Atmospheric Research

NCEP National Center for Environmental Prediction

NC-GOW Non Calibrated Global Ocean Waves reanalysis data, i.e. hindcast data

NH Northern Hemisphere

NOAA National Oceanic and Atmospheric Administration

NPI North Pacific Index

ODGP2 Third generation wave model

PCs Principal Components

PDO Pacific Multidecadal Oscillation

PNA Pacific North American Index

POT Peak Over Threshold

PSMSL Permanent Service for Mean Sea Level

PWA-R Pacific Weather Analysis, [Graham and Diaz, 2001]

QBO Quasi-biennial Oscillation

RMSE Root Mean Square Error

SAM Southern Annular Mode

SH Southern Hemisphere

SI Scatter Index

SLP Sea level pressure

SLR Sea Level Rise

SOI Southern Oscillation Index

SOM Self-organizing maps

SON Season: September - October - November

SPM Shore Protection Manual

SS Storm Surge

SST Sea Surface Temperature

STRM Shuttle Radar Topography Mission

SWAN Simulating WAves Nearshore, third generation wave model

TOPEX/Poseidon Oceanographic satellite mission (1997-2003)

TPXO Global model of ocean tides based on altimetric data from the TOPEX/POSEIDON mission

TNA Tropical North Atlantic index

TSA Tropical South Atlantic index

UHSLC University of Hawaii Sea Level Center

USA United States of America

USACE United States Army Corps of Engineers

WAM WAve Model, third generation wave model

WEC Wave Energy Converter

WP Western Pacific index or Wave Power

WRD Wave Reanalysis Databases

WW3 Wavewatch III, third generation wave model

20CRv2 20th Century Reanalysis version 2

Chapter 1

Introduction

"When you can measure what you are speaking about and express it in numbers, you know something about it; when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the state of science, whatever the matter may be"

Lord Kelvin

1.1 Introduction

1.1.1 The wind waves in the Global Ocean

The word *climate* etymologically comes from the Greek *climatos* meaning tilt or declination. This concept refers to the influence of the angle of incidence of incoming solar radiation on the weather mean conditions, as evidenced by climate research (e.g., [Gray et al., 2010]). In more recent times, the term *climate* has been associated with near surface climate parameters with significant influence on human well-being, such as atmospheric pressure, surface temperature or precipitation. In the last few decades the concept has broadened. So, the American Meteorological Society (AMS) defines climate as "the slowly varying aspects of the atmosphere-hydrosphere-land surface system"([Glickman, 2000]). The Intergovernmental Panel on Climate Change (IPCC) introduces the description of *climate* as the average weather or more rigorously, "the statistical description in terms of the mean and variability of relevant quantities over a period of time ".

When speaking about marine climate it usually refers to ocean and marine variables such as salinity, sea water temperature or oceans atmospheric conditions. However, the statistical description of wind-waves in the sea is generally known as *wave climate*.

The circulation of wind in the atmosphere is driven by the rotation of the Earth and the incoming energy from the Sun. Wind circulates in each hemisphere in three distinct cells (Figure

1.1), which are from Equator to the poles: the Hadley cell, Ferrel cell, and Polar cell. These cells help transporting energy and heat from the Equator to the poles. The winds are driven by the energy from the Sun at the surface as warm air rises and colder air sinks. To understand the generation of waves, a brief description of this global wind circulation follows.

The circulation cell closest to the equator is called the Hadley cell, where warm air converges on the Equator and rises, forming a belt of low pressure and producing clouds and instability in the atmosphere. This instability causes thunderstorms and releases large amounts of latent heat, which essentially provides the energy to drive the Hadley cell. This air then moves north or south towards the tropics. The Hadley Cell encompasses latitudes from the Equator to about 30°, where the high pressure causes the air near the surface to diverge. As a result, the air is forced to descend, warming in the process and decreasing its relative humidity with lower elevation, finally resulting in clear skies over the tropics. As a consequence, a strong temperature gradient appears between the two different air masses and a jet stream results. At the 30° latitudes, this jet is known as the subtropical jet stream which flows from west to east in both the North and South Hemispheres, creating the *trade winds* (i.e. northeast and southwest trades, respectively). The Coriolis force impacts the direction of the wind flow. In the Northern Hemisphere, the Coriolis force turns the winds to the right, while in the Southern Hemisphere, they are shifted to the left.

From 30° latitude to 60° latitude, a new cell known as the Ferrel Cell takes over. This cell produces prevailing *westerly winds* at the surface within these latitudes. This is because some of the air sinking at 30° latitude continues traveling northward towards the poles, and the Coriolis force bends it to the right (in the Northern Hemisphere). This air is still warm and at roughly 60° latitude approaches cold air moving down from the poles. With the converging air masses at the surface, the low surface pressure at 60° latitude causes air to rise and forms clouds. Some of the rising warm air returns to 30° latitude to complete the Ferrel Cell.

The two air masses at 60° latitude do not mix well and form the *polar front* which separates the warm air from the cold air. The polar jet stream aloft is located above the polar front and flows generally from west to east. The polar jet is stronger in the winter because of the greater temperature contrasts than during the summer.

Above 60° latitude, the polar cell circulates cold, polar air equatorward. The air from the poles rises at 60° latitude where the Polar and Ferrel cells meet, and some of this air returns to the poles completing the Polar cell. Since the wind flows from high to low pressures and taking into account the effects of the Coriolis force, the winds above 60° latitude are prevailing *easterlies*.

In accordance with the latter global circulation pattern, winds on the sea surface generate wind-waves throughout the global ocean. Wind blowing over an initially calm sea surface establishes a very shallow surface current whose velocity is proportional to the wind speed. At some point turbulent fluctuations develop within the air above the sea surface, disturbing the water surface and forming a characteristic pattern of capillary waves. At this stage, surface tension and not gravity represents the major restoring force due to the small scale of the waves. When the wind speed raises further, the wavelength, period, and height of capillary ripples increase. When the wind speed is sufficiently strong to generate waves high enough to perturb the air

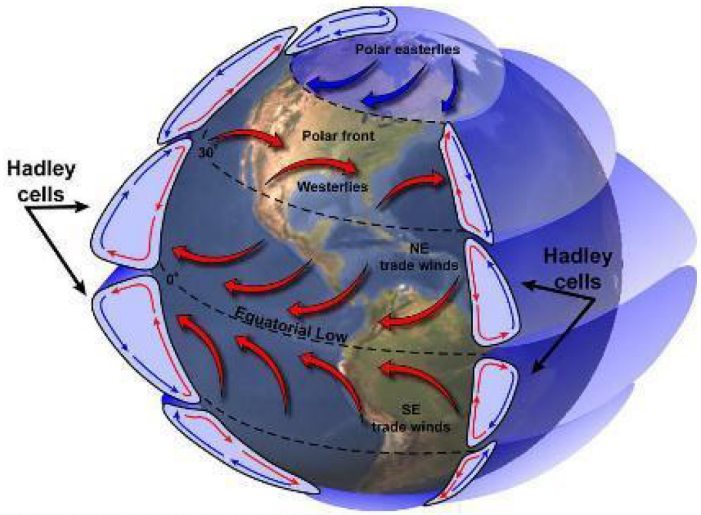

Figure 1.1: General climate circulation. Source: NASA

flow, pressure fluctuations increase the wave height leading to a point where gravity becomes the dominant restoring force (i.e. for wavelengths larger than approximately 10 cm). The process is directionally selective in the direction of the wind.

For deep water, waves at the sea surface are generally determined by three factors: wind speed, duration (the period of time over which the wind is blowing), and fetch (how much water surface is available for the wind to blow). Waves that are still forced by the local wind field and that, as a consequence, are still growing are referred to as *sea* waves, which mainly travels parallel to the wind direction. In their travel, waves suffer of directional (i.e. waves traveling in different directions) and frequency dispersion (i.e. waves of different wavelengths traveling at different phase speeds) which results in the generation of *swell* waves. The energetic budget of each sea-state (a set of waves with similar statistical behavior) is aggregated in a so-called wave spectrum. From this energy characterization of wind-waves, several statistics can be obtained for the description of wave climate, like: the significant wave height, mean direction or the peak period. The description of the wave climatology is usually based on this sort of aggregated statistics of a sea-state.

1.1.2 A varying wave climate and the potential risks

The state of the climate varies and is not constant in time thus presenting external and internal variations. Externally forced variations refers to changes caused by factors not affected by climatic variables themselves like changes in terrestrial forcings. Internally driven variability is due to interactions, non-linear processes and feedbacks between short-term fluctuations with much longer climate responses.

Therefore, the components of climate system are characterized by different time scales. From seasons to years and decades, marine climate conditions are influenced at all scales and it may

have resounding effects for the human system. The common analysis of these time scales will be hereafter referred as *variability* of wave climate.

Considerable interactions between climate at different time scales exist, being the regional climate governed to some extent by large scale climate. A meaningful example is the climate anomalies associated to the El-Niño Southern Oscillation (ENSO) phenomenon that greatly influences other climate variables, such as sea-levels, wave climate, temperature or precipitation. These variations take place in the scale of several years due to general climate patterns anomaly. Such changes will be hereafter referred as *inter-annual variability*.

Besides, human actions have increased the atmospheric concentrations of several gases in the atmosphere, the so-called greenhouse gases, which are closely related with global temperature. Current research is dealing with the relationship between higher concentrations of those gases and the global climate. A warmer climate is likely to mean changes to weather on the whole world. Changes over several decades are usually referred to as *long-term trends*.

Either due to internal or external induced variability, the global wave climate, like many other dynamics in the sea, is being modified. Some of the changes identified so far greatly explain coastal impacts and problems at many regions of the world.

Taking for example the beach nourishment planning-programs at El-Niño affected areas (e.g. California) or coastal flooding events under high NAO values (e.g. western coast of Europe), the effects may be beyond those evidences. Considering the goods exposed in marine areas such as port and offshore infrastructures; fisheries; an increasing socioeconomic development and concentration of population in coastal areas; or the incipient wave energy industry, many are the issues to be considered.

There is an increasing need to consider not only the different scales affecting the coastal and marine system but also to adopt a framework for design and decision making which considers in a holistic view the probability of occurrence and the consequences of the potential affections. This is generally known as risk assessment and it is considered as the first action towards adaptation or mitigation. In many application areas from natural disasters (e.g., tsunamis or earthquakes) to climate change impacts, a risk framework is being widely introduced in studies and guidelines worldwide.

1.2 State of the art

1.2.1 Wave data sources

Four wave data sources can be used for the description of wave climate: (1) buoy measurements, (2) visual observations from ships, (3) altimeter data and (4) numerical modeling. Each one presents specific strengths and weaknesses. The best choice is usually a combination of all of them, although it also depends on the scope of the study.

For offshore waters around the world, long-term buoy wave measurement networks are still relatively scarce and scattered distributed. Networks with directional measurements are even

less numerous, even though directional information is essential for a number of applications (e.g. forecasts, costal defence, etc).

In addition to the scattered distribution of buoys worldwide, mainly located in the Northern Hemisphere, they provide high quality but generally discontinuous data, due to installation and maintenance reasons, spanning only some decades in the past. These characteristics make them a perfect complement for other sources of data and ideal for certain site-specific studies, but inadequate to describe wave climatology at a planetary scale.

Another data source for wave data measurements are the visual observations made on board ships. Some analysis have relied on these data for wave climate analysis. However, because of the accuracy of the observations depends upon the ability, experience of each observer, the quality of these data has been generally considered dubious. An additional drawback is that they are restricted to ship routes and, therefore, vary in space and are discontinuous in time.

However, since the end of the 1980s, satellite altimetry has became a very useful source of information worldwide. The back-scattered signal from satellite altimeters can provide wave height measurements close to the accuracy of a buoy. Measurements are made every second, whilst the satellite flies over a repeated net of ground tracks at about 6 km/s, which as a rule of thumb means that one altimeter registers wave data at one point each 10 to 15 days. This provides enormous amount of wave data worldwide, and with a steady flow of new data from three or more operational satellites at present, millions of new observations are becoming available each month. Notwithstanding these features, the time span covered by these data is not long enough for all purposes. Furthermore, they provide global coverage but measurements are too disperse in time for certain purposes like wave extreme analysis variability over decades.

Global satellite altimeter measurements have been performed during 1985-1989 by the US Navy's Geosat and the follow-on missions from 2000 by ESA's ERS-1 (from 1991 to 1996), ERS-2 (1996 to 2003), EnviSat (launched in 2002), and most importantly due to its longevity, the US/French Topex/Poseidon mission from 1992 to 2005. The Topex-follow-on mission (Jason) has provided data since 2002. Full resolution altimeter significant wave height and wind speed data for most of these satellites can be found on the global network.

Over the last decades, numerical modeling has solved the equations underlying wave generation and propagation processes (i.e. action balance equation). Some of the most well-known models for wave generation and propagation are WaveWatch III, WAM, SWAN or CREST (see [Cavaleri et al., 2007] for a comprehensive review on wave modeling state of the art). Many meteorological centers run today wave models regionally and globally and dedicated long-term hindcasts have also been performed from diverse atmospheric reanalysis projects. Note that the term *reanalysis* is usually used for numerical simulations which embed observational data in the computation, whereas *hindcast* refers to the raw numerical data.

Wave modeling provides long span time series (time coverage depending on the wind forcing data) combined with high spatial resolution and a wide variety of outputs, from wave energy spectra to sea-state statistics. It is not surprising then that wind-wave reanalyses have therefore become a valuable source of information for wave climate research, ocean wave energy and coastal studies over the last decade.

Recently, a series of wave reanalysis generated with third generation models have been carried out to obtain wave climate information to complement the instrumental measurements (buoys and altimeter observations) in time and space. A comprehensive review of climate reanalysis, which are use as forcings for wave modeling, is provided in the following section. Wave reanalysis state of the art is specificality addressed in Chapter 2.

1.2.2 Climate and numerical modeling: the atmospheric reanalyses

Being nowadays the anthropogenic influence on climate widely accepted (e.g., [Solomon et al., 2007]), research has turned to seek for past changes and simulate possible future behavior of climatic variables. For the latter, Global Circulation Models (GCMs) are used to compute projections on the basis of different emission scenarios. However, for the identification of long-term trends in climatic variables, long time series are required. Generally, measurements do not cover such a long span and numerical modeling serves to expand the data into the past which in turn is analyzed to detect long-term trends.

Reanalysis is a scientific method for developing a comprehensive record of how weather and climate are changing over time. Observations and a numerical model that simulates one or more aspects of the Earth system are combined objectively to generate a synthesized estimate of the state of the system. A reanalysis typically extends over several decades or longer and are, therefore, used extensively to identify the causes of climate variations and change, and preparing climate predictions ([Weisse and Von Storch, 2010]). Information derived from reanalyses is also used increasingly in commercial and business applications in sectors such as energy, agriculture, water resources, and insurances.

In the field of wave climate, the meteorological reanalyses have allowed the forcing of wave models to generate numerical wave datasets. Nowadays, it can be stated that wave reanalyses complemented with altimeter and buoy measurements are the best information for global studies on wave climate.

With respect to the different atmospheric reanalysis available, http://reanalyses.org/ can be visited for detail information. In the following a comprehensive list of the global datasets is provided:

ECMWF 40 Year Reanalysis (ERA-40): 1957-2002, [Uppala et al., 2005]

ERA-40 is a global atmospheric reanalysis which was produced using a June 2001 version of the ECMWF Integrated Forecast Model (IFS Cy28r3). The spectral resolution is T159 (about 125 km) and there are 60 vertical levels, with the model top at 0.1 hPa (about 64 km). Observations were assimilated using a 6-hourly 3D variational analysis (3D-Var). Satellite data used include Vertical Temperature Profile Radiometer radiances starting in 1972, followed by TOVS, SSM/I, ERS and ATOVS data. Cloud Motion Winds are used from 1979 onwards. Various data from past field experiments were used, such as the 1974 Atlantic Tropical Experiment of the Global Atmospheric Research Program, GATE, 1979 FGGE, 1982 Alpine Experiment, ALPEX and 1992-1993 TOGA-COARE.

ECMWF Interim Reanalysis (ERA-Interim): 1979-present, [Dee et al., 2011]

ERA-Interim was originally planned as an 'interim' reanalysis in preparation for the next-generation extended reanalysis to replace ERA-40. It uses a December 2006 version of the ECMWF Integrated Forecast Model (IFS Cy31r2). It originally covered dates from 1 Jan 1989 but an additional decade, from 1 January 1979, was added later. ERA-Interim is being continued in real time. The spectral resolution is T255 (about 80 km) and there are 60 vertical levels, with the model top at 0.1 hPa (about 64 km). The data assimilation is based on a 12-hourly four-dimensional variational analysis (4D-Var) with adaptive estimation of biases in satellite radiance data (VarBC). With some exceptions, ERA-Interim uses input observations prepared for ERA-40 until 2002, and data from ECMWF's operational archive thereafter.

Japanese 25-year Reanalysis (JRA-25): 1979-2004/(JCDAS) 2005-present,

[Onogi et al., 2007]

The Japanese 25-year Reanalysis (JRA-25) represents the first long-term global atmospheric reanalysis undertaken in Asia. Covering the period 1979-2004, it was completed using the Japan Meteorological Agency (JMA) numerical assimilation and forecast system and specially collected and prepared observational and satellite data from many sources including the European Center for Medium-Range Weather Forecasts (ECMWF), the National Climatic Data Center (NCDC), and the Meteorological Research Institute (MRI) of JMA. A primary goal of JRA-25 is to provide a consistent and high-quality reanalysis dataset for climate research, monitoring, and operational forecasts, especially by improving the coverage and quality of analysis in the Asian region. JRA-25 was conducted by JMA and CRIEPI (Central Research Institute of Electric Power Industry). It has been continued as JCDAS (JMA Climate Data Assimilation System) operated by JMA on real time basis. The data assimilation systems of JRA-25 and JCDAS are the same. Users can use JRA-25 and JCDAS as one continuous reanalysis dataset. The second Japanese reanalysis JRA-55 (1958-2012) is ongoing.

NASA Modern Era Reanalysis for Research and Applications (MERRA): 1979-present, [Rienecker et al., 2011]

MERRA is a NASA reanalysis for the satellite era using a major new version of the Goddard Earth Observing System Data Assimilation System Version 5 (GEOS-5) produced by the NASA Global Modeling and Assimilation Office (GMAO). The Project focuses on historical analysis of the hydrological cycle on a broad range of weather and climate time scales and places the NASA EOS suite of observations in a climate context.

NCEP Climate Forecast System Reanalysis (CFSR): 1979-Jan 2010,

[Saha and Coauthors, 2010]

The National Centers for Environmental Prediction (NCEP) Climate Forecast System Reanalysis (CFSR) was completed over the 31-year period of 1979 to 2009 in January 2010. The CFSR was designed and executed as a global, high resolution, coupled atmosphere-ocean-land surface-sea ice system to provide the best estimate of the state of these coupled domains over this period. Presumingly, the current CFSR will be extended as an operational, real time product into the future.

NCEP/NCAR Reanalysis I: 1948-present, [Kalnay et al., 1996]

This reanalysis was the first of it's kind for NOAA. NCEP used the same climate model that were initialized with a wide variety of weather observations: ships, planes, station data, satellite observations and many more. By using the same model, scientists can examine climate/weather statistics and dynamic processes without the complication that model changes can cause. The dataset is kept updated using near real-time observations.

NCEP/DOE Reanalysis II: 1979-2009, [Kanamitsu et al., 2002]

NCEP produced a second version of their first reanalysis starting from the beginning of the major satellite era. More observations were added, assimilation errors were corrected and a better version of the model was used.

NOAA-CIRES 20th Century Reanalysis v2 (20CR2): 1871-2008, [Compo et al., 2011]

The 20th Century Reanalysis version 2 (20CRv2) dataset contains global weather conditions and their uncertainty in 6-hour intervals from the year 1871 to 2008. Surface and sea level pressure observations are combined with a short-term forecast from an ensemble of integrations of an NCEP numerical weather prediction model using the Ensemble Kalman Filter technique to produce an estimate of the complete state of the atmosphere, and the uncertainty in that estimate. The uncertainty is approximately inversely proportional to the density of observations. Additional observations and a newer version of the NCEP model that includes time-varying CO2 concentrations, solar variability, and volcanic aerosols are used in version 2. The long time range of this dataset allows scientists to examine better long time scale climate processes such as the Pacific Decadal Oscillation and the Atlantic Multidecadal Oscillation as well as looking at the dynamics of historical climate and weather events. Verification tests have shown that using only pressure creates reasonable atmospheric fields up to the tropopause. Additional tests suggest some correspondence with observed variations in the lower stratosphere.

It is worth noting that certain critical factors should be considered for time variability detection. Some numerical databases present inhomogeneities within the computation period, mainly related to different amount of observations embedded in the calculation process. Advances in knowledge and computer science also imply differences within the various climate reanalyses available. This is clearly noted in the satellite era (i.e. decade of 80s) because altimetry data were incorporated to climate computations, much improving the quality as a result. This is not a simple issue at all. For instance, there is an ongoing debate on the effect of altimetry registers of sea-level rise and the detected acceleration in the dataset during the last three decades. The same question remains open for some climate reanalysis for the period prior to the satellite era (e.g. [Bengtsson et al., 2004, Saha and Coauthors, 2010]).

The amount of assimilated data in the reanalyses is also variable in space. Since the Southern Hemisphere has traditionally received less attention, the assimilation of data varies regionally. Consequently, inhomogeneities are much more pronounced in the Southern Hemisphere and results must be taken with care there.

To avoid inhomogeneity problems in the computation, recent climate reanalysis like ERA-Interim or the CFSR, have increased the resolution and adjust the simulation period to improve

the data assimilation process and data quality. However, it comes at the cost of shorter temporal coverage which is a critical feature for long-term trends detection. Hence, a pragmatic balance has to be found when considering the study of long-term variability. On the one hand, the question of homogeneity in the computation period and on the other, the sufficient temporal coverage for variability analysis. Up to now, the NCEP/NCAR Global Reanalysis and the ECMWF ERA-40 reanalysis have been the widely used datasets for wave fields computation, although they are being replaced by new high resolution products, respectively. This is an ongoing process in which in some years time the wave reanalyses should adapt to advances in modeling alongside with new climate data of growing quality and time span.

The state of the art of wave modeling and wave reanalyses is addressed in detail in Chapter 2. The state of knowledge corrections applied to the numerical results are also detailed there.

1.2.3 The wave climate and its variability

Climate *variability* refers to variations in the mean state and other statistics of the climate on spatial and temporal scales beyond individual weather events ([Solomon et al., 2007]). Variations in more than a year time scale are usually known as *inter-annual variability*. In a warming world, many *long-term changes* have been identified in many variables ([Solomon et al., 2007]). Note that in different scientific areas *long-term trends* refers to different time scales. For example, geologists may refer to *long-term changes* when variations on time scales of millennia are considered. In the context of wave climate, the use of the concept refers to changes that occur over several decades, that is, the time horizon accessible to human experience.

1.2.3.1 Current knowledge

Many studies have described wave climatology during the last years, some of them at a global scale: [Barstow, 1996, Young, 1999, Cox and Swail, 2001, Caires et al., 2004b, Chen et al., 2004, Caires and Sterl, 2005a, Semedo et al., 2011b]; and others focusing on extreme wave height values: [Alves and Young, 2004, Chen et al., 2004, Sterl and Caires, 2005, Caires and Sterl, 2005a, Izaguirre et al., 2011].

Ocean wave heights have been often correlated with inter-annual climate variations such as the North Atlantic Oscillation ([Woolf and Challenor, 2002, Dodet et al., 2010]) or the Southern Annular Mode ([Hemer et al., 2010]). [Gulev and Grigorieva, 2004] found that the mean wave height in the Pacific is correlated with the El Niño Southern Oscillation (ENSO). The Pacific North-America Index (PNA) is also largely linked to the eastern Pacific cyclone activity over the Gulf region and the North American coast [Gulev et al., 2001]. Some studies have also even addressed the directional analysis ([Hemer et al., 2010]), although they are scarcer.

Besides, numerous evidences suggest long-term changes in the intensity and frequency of storms characterized by a decrease in activity in the mid-latitudes and an increase in the high latitudes on the Northern Hemisphere (e.g. [McCabe et al., 2001, Gulev et al., 2001, Wang et al., 2006, Ulbrich et al., 2009, Alexandersson et al., 2000, Bengtsson et al., 2006, Leckebusch et al., 2006, Pinto et al., 2007, Wang et al., 2011]).

Since waves are primarily derived from cyclonic and anti-cyclonic conditions on the global ocean, storm activity changes may have a reflection on ocean wave heights. Such changes in wave climate are important to be evaluated for understanding of potential impacts. Long-term variations of storm waves are crucial for infrastructures in the coastal zones (e.g. ports, wave energy farms, coastal development, etc.). Coastal erosion is another issue for concern for the years to come, with most of the world's sandy shorelines subjected to erosion during the past century ([**?**]).

Investigations on trends in oceanic wave heights have used ship observations [Gulev and Grigorieva, 2004, Gulev and Grigorieva, 2006], buoy measurements, numerical modeling ([Group, 1998, Cox and Swail, 2001, Wang and Swail, 2001, Wang and Swail, 2002, Sterl and Caires, 2005, Hemer et al., 2010, Dodet et al., 2010, Semedo et al., 2011b]) or satellite altimeter observations ([Young et al., 2011]). Some efforts have also been made in defining the future wave climate from climate change scenarios (e.g., [Semedo et al., 2011a, Mori et al., 2010]).

1.2.3.2 Methods to study time variability

The study of wave climate (i.e. scale of one year: from seasons to hours) is usually addressed through different statistical techniques. The description of wave climate is generally based on certain statistics such as the means and standard deviations of different scalar and directional variables. Extreme analysis models, both stationary and non-stationary, are specifically used for the study of extreme values.

The inter-annual variability (i.e. scale of years), however, is often address through correlation patterns and regression models between standardized variables and climate indices time series or other techniques to identify spatial patterns and temporal evolution (e.g., EOF).

The long-term changes (i.e. scale of decades) are identified through specific statistical methods, which present the main advantage of including the uncertainty into account. Some of them are common regression techniques but others focus on identifying spatial patterns of change. For directional variables, the statistical approach is more difficult and the available models far less numerous.

However, detection of long-term trends is affected by a number of factors, including the size of the trend to be detected; the time span of available data; and the magnitude of variability and autocorrelation of the noise in the data ([Weatherhead et al., 1998]). Therefore, the number of years of data needed strongly depends on the type of data and its variance. For example, wave data, similarly to other climatic variables like rainfall, show a greater variance than other meteo-oceanographic variables such as sea-level or sea surface temperature. The nature of the variables to study will decisively influence the techniques to use.

The details on the latter methods are explained in detail in an accompanying annex to the main document.

1.2.4 Wave energy

Within the renewable sources of energy, wave energy has experimented a great development over the recent years. This form of energy is currently under development and at its pre-commercial phase, but it may become competitive and play an important role towards sustainability ([Falnes and Lovseth, 1991, Thorpe, 1992, Thorpe, 1999, Jacobson, 2009]).

Several studies on the evaluation of the global wave energy resource are currently available: [Krogstad and Barstow, 1999, Pontes et al., 2002, Barstow et al., 2008, Cornett, 2008, Mork et al., 2010]. They are based on satellite altimeter observations or wave reanalysis data, but in all cases presenting limited time span (approximately a decade).

Besides, there is still a big controversy on the actual figures of available resources. The global offshore wave power was originally estimated to be between 1 to 10 Twh ([Isaacs and Seymour, 1973, IEA, 2009, Lewis et al., 2011]), although recent research indicate that the lower bound would be more feasible. Also, the combination of current technologies with the available resources is very scarce, and limited to academic purposes or site-specific studies.

To date, design and operation rely on past history events to be optimized. An obvious limitation comes from assuming that the time history used in the design is statistically representative of the future period in the life cycle. However, this assumption is only valid where conditions do not change or they do in a negligible degree compared with design conditions. Nevertheless, this variability is usually not contrasted in practice. In the case of wave energy installations, for the design of operations under mean conditions and the structural reliability under extreme conditions, contrasting the former assumption is of paramount importance.

Despite the fact that climate variability affects wave climate in various forms ([Harrison and Wallace, 2005]), to our knowledge, there is no planetary study on the possible affection of the most relevant climate patterns on the wave energy resources. On a regional basis, current research is trying to cover the gap between seasonal and decadal oscillations. See for example [Ching-Piao et al., 2012] and the relationship between El-Niño phenomenon and the wave energy for northeast coast of Taiwan. Indeed, the variability of resources may condition the operability and potential outcome and, furthermore, it may suppose a critical factor for the site installation decision. Maintenance programmes should account for variations over the scale of months and years.

The question of whether the wave energy resource is changing at a long-term scale (i.e. decades) is currently under research. This issue may present different consequences for different regions depending on the availability and intensity of wave energy and the magnitude of the long-term change in such a case. The relative weight of these changes should also receive some consideration.

Finally, survivability is paramount within the sets of requirements a wave energy converter must comply with: very efficient conversion in small to moderate seas together with to survival under storm seas ([Cruz, 2008]). However, few research has focused on this regard.

1.2.5 Coastal impacts

Coasts are of great ecological and socioeconomic importance. They sustain economies and provide livelihoods through fisheries, ports, tourism, and other industries. Coastal ecosystems are amongst the most productive. These areas have also been centers of human settlement since perhaps the dawn of civilization, and have cultural and aesthetic value as well. In fact, coastlines are among the most populated regions, with nearly half the world's major cities located within 50 km landward. Coastal population densities are 2.6 times greater than those of inland areas.

The Intergovernmental Panel on Climate Change (IPCC) in their fourth assessment report (AR4; [IPCC, 2007a]) states that coasts are expected to be exposed to increasing risks, including coastal erosion, due to climate change and sea-level rise.

Little doubt exists that the effect will be exacerbated by increasing human-induced pressures on coastal areas and will have serious affection to tourism, coastal urban areas, ecosystems and sea infrastructures in many nations. Of special relevance are the islands nations, whose particular characteristics make them especially vulnerable to the coastal hazards.

A lot of research has been carried out on these topics, a great part focused on sea-level rise and the economic loss of potentially inundated areas (e.g., [Yohe et al., 1999, West et al., 2001, Nicholls, 2004, Tompkins et al., 2005, Michael, 2007, IPCC, 2007a]). [IPCC, 2007a] estimated that the global mean sea level has risen at an average rate of 1.7 ± 0.5 mm/yr during the twentieth century. Subsequent to the IPCC-AR4 report in 2007, there have been several new estimates ranging from 1.9 ± 0.4 mm/yr ([Church and White, 2011]) to 3.2 ± 0.4 mm/yr ([Merrifield, 2009]). However, new semi-empirical approaches defend that global SLR is accelerating in a way strongly correlated with global temperature ([Rahmstorf, 2007, Vermeer and Rahmstorf, 2009]). Latest findings confirm the rise. Whether an acceleration is occurring or not is an issue under current research (see for example: [Cazenave, 2008, Church et al., 2010, Rignott, 2011, De Santis et al., 2012]). For the coming future, several studies have proposed estimates of SLR for the end of the present century: [IPCC, 2007a, Rahmstorf, 2007, Pfeffer et al., 2008, Horton et al., 2008, Vermeer and Rahmstorf, 2009, Jevrejeva et al., 2010]. Although the authors use different physical bases, a rise greater than 1 m seems to be not feasible by the end of the century and its implications worth considering.

The physical impacts of rising water levels are well known and diverse ([Nicholls et al., 2007, Nicholls and Cazenave, 2010]). For an early but clarifying study on the engineering impacts of SLR, [CEI-RMSL, 1987] can be consulted. It provides a useful basis for design calculations and policy decisions that must take changes in water levels into account. Submergence, erosion, an increasing coastal flooding as well as saltwater intrusion of surface water are some of the most prominent effects.

However, sea-level rise is not the unique factor to be considered in coastal risks. Coastal flooding and exacerbated erosion may be produced by changes in the storm surge levels and in the wave climate intensity or/and direction. Far less attention has been focused on these issues.

It is also noted that these impacts occur in addition to other human-induced changes and pressures in the coastal areas such as coastal defenses, destruction of wetlands, port and harbor works, and reduced sediment supply due to dams (see [USACE, 2002] for further description of

these problems).

A great number of coastal problems are also due to natural induced variability in coastal dynamics. Take for example beach nourishment planning-programs at El-Niño affected areas (e.g. California, USA) or coastal flooding events under high North Atlantic Oscillation phases (e.g. western coast of Europe). Not only wave climate, but also coastal flooding and erosion have been related with large climate anomalies (e.g. [Solomon and Forbes, 1999, Andrews et al., 2006, Harley et al., 2010]). For instance, during the winter of 1997-98, wind-driven waves and abnormally high sea levels significantly contributed to hundreds of millions of dollars in flood and storm damage in the San Francisco Bay region. Evidences of erosion in Australia with ENSO events has also been reported (e.g. [Ryan et al., 1999, Ranasinghe et al., 2004]).

From another perspective, impacts only result in consequences depending on the environmental or human activities on the coast such as fishery areas, urban settlements or port infrastructure, which sometimes renders impossible a natural adaptation of margins to the new hydrodynamics conditions. Since consequences stem from the nature of human development, *vulnerability* and *exposure* of the coast is starting to receive an increasing attention.

It is difficult to estimate accurately how much population is at risk from coastal hazards due to changes in frequency and intensity of events. Coastal areas with elevations of less than 10 m account for 2% of the earth's surface but contain approximately 10% of the population (i.e. more than 600 million people), while 13% are urban zones. Virtually two thirds of the world's cities with populations of more than 5 million are located in these regions. Population trends in coastal zones are particularly important in developing countries, where the proportion more than doubles that in developed countries ([McGranahan and Anderson, 2007, Satterthwaite et al., 2009]). One estimate suggests that 10 million people in the world are today affected annually by coastal flooding, and this number will increase in the various climate change scenarios ([Nicholls, 2004]).

Coastal variability should not be a cause for alarm or complacency but considered by those involved in planning and design, by being sensitized to the probabilities of and quantitative uncertainties of the different changes and risks. Other regions of the world are already incorporating these aspects in their planning and design phases. See for example US cases for US Army corps of Engineers projects or storm surge studies of the Federal Emergency Management Agency: [CEI-RMSL, 1987, USACE, 2011b, USACE, 2011a]. They mainly concern with sea-level rise and not other coastal dynamics changes, like wave climate, presumably due to lack of information and greater uncertainties, but may provide a useful guideline.

For example, USACE has recently developed a guidance ([USACE, 2011b]) for incorporating the direct and indirect physical effects of projected sea-level rise across their projects life cycle. Similarly, their approach for planning and risk analysis in projects is one of the most clear and objective intended examples (e.g., [USACE, 2011a, USC, 1996]).

1.2.6 Risk evaluation

Risk is a measure of the probability (or likelihood) of occurrence of a future hazard (i.e. with a certain reach) and its consequences ([USACE, 2011a]). There is comprehensive terminology to refer to the terms risk, hazard, exposure and vulnerability at each discipline: [Coburn et al., 1991,

Carter et al., 1994, McCarthy et al., 2001, Schneiderbauer and Ehrlich, 2004, Birkmann, 2006, Adger, 2006, Solomon et al., 2007, Füssel, 2007] are some examples.

The hazard term is usually defined by a combination of the occurrence of a certain agent which configures an impact and its associated uncertainty. Coastal flooding or infrastructure failure are meaningful examples. This term can usually be defined from the analysis of past data and the understanding of the process which generates the impact.

Vulnerability is usually defined by a certain set of indicators and measures through an intensive geospatial assimilation of information. Variables like land uses, population, infrastructure, environmental resources or gross domestic product distribution can configure the big picture of vulnerability in the coastal zones. Each risk requires, however, specific vulnerability definition depending on the particular features of the elements to be affected. Other approaches understand the vulnerability as the capacity of zones, regions, entities or bodies to anticipate, confront, resist and recover from the impact produced by an agent of a given magnitude. Many also consider that vulnerability is made up of various terms like susceptibility or sensitivity, resilience and adaptive capacity ([Schneiderbauer and Ehrlich, 2004, Magrin et al., 2007]).

Physical features of the coast, usually referred as *exposure* (i.e. physical exposure), link the coastal dynamical changes and the uses in the coastal system. The integration of consequences and impacts, or in other words, vulnerability, exposure and hazards, configure a framework for the environmental risk analysis of coastal zones.

A more precise review of the risk framework can be seen in the section devoted to the methodology, where the risk approach adopted is described in detail.

1.3 Objectives and scope

1.3.1 General objective

The global aim of this work is to analyze the global wave climatology and its spatial and temporal variability to determine the range of possible coastal and technological implications, under a risk-based framework. For this to be accomplished, several prior steps have to be taken.

First of all, a new robust wave database must be developed to overcome deficiencies in existing alternatives. Risk analysis based on literature on disaster reduction will be carried out providing a more suitable framework to assess risks in a changing climate. To study the risks on wave energy technology and coastal environmental impacts it is first necessary to address the study of the global wave climatology, its energy and the changes in different time scales. Many secondary objectives will arise in the process. They are all explained below.

1.3.2 Specific goals

After revising the state of the art and the current knowledge on wave climatology and potential impacts of its variability, several specific objectives are programmed. They are organized

by thematic areas below.

With respect to the ***methods***, data and approach, two main objectives are identified:

- The analysis intends to address **different time scales**, varying from the annual and seasonal climatology description to inter-annual (i.e. scale of years) and long-term (i.e. scale of decades) variations. Due to the characteristics of the available data, its quality and spatial-temporal resolution, it is necessary to develop a wave reanalysis which can provide the basic data with sufficient spatial coverage (i.e. global) and time span (i.e. as long as possible in order to analyze possible changes in the different scales). This database needs to present a good quality along the continental margins so that it allows the analysis of offshore wave energy technology risks.
- To adopt and apply a **risk-based approach** for the study of the impacts in a changing wave climate. Such approach should consider changes in nature dynamics and in the human system, thus defining a general informed decision-making framework.

With regard to the study of the ***global wave climatology***, the main aim is to provide further insight to the state of knowledge. The specific aims of this part are:

- Address the description of the scalar and directional behavior of the mean climatology of waves at time scales of seasons and years, along with quantification of its variability range over the last decades.
- Identify the correlation and influence of the most prominent climate patterns worldwide on the wave climate.
- Seek for long-term changes in the scalar wave climate. Identify the areas where the potential changes could be of concern.
- Develop a suitable directional regression model for the identification of long-term trends and inter-annual variability study of directional variables. Definition of long-term trends in the directional wave climate and identifying the areas of concern.
- Study the wave extremes at offshore depths along the coastal margins at a planetary scale and based on a non-stationary approach which should model seasonality.
- Infer potential impacts from wave climatology long-term changes for the continental margins of the global ocean, such as potential sediment transport. Provide a quantified identification of regions for concern.

Based on the global wave climatology, a similar approach is carried out for the assessment of the ***wave energy resources and associate technology risks***. In view of the lack of knowledge in the field, it is proposed to:

- Study the wave energy resources climatology by determining the mean annual and seasonal resources, its global spatial variability, and the ranges of seasonal and monthly variations.

- Identify the potential wave energy resources at the continental margins for offshore depths and compare the global estimate with other works.

- Study the long-term changes in wave energy resources over the past six decades, exploring the possibility of a global signal of climate change in the wave energy that may be transferred from the previously reported changing storm activity.

- Define and quantify the effect of climate patterns on the wave energy resources and determine whether they could imply a factor for consideration in the design and operation of wave farms.

- Study the risks for operations and outcome from wave farms from expected changes of wave energy resources in the coming future, adapting the general risk framework to do so.

- Study the risks for potential installations from changes in wave extremes.

Due to the interest of assessing potential impacts of a changing wave climate in coastal areas, it is considered to assess the ***environmental risks*** of flooding and erosion for the region of Latin America and the Caribbean (LAC), as a particular case. Specifically, it is aimed to:

- Define the wave climatology in higher detail in the region of study, along with other coastal dynamics that may be changing and may configure drivers and hazards in coastal areas.

- Study of long-term erosion and flooding events, both in the long-term and due to extreme events, as outstanding impacts in the coastal zones.

- Establish which hazards are of greater concern.

- Analyze vulnerability and exposure in the LAC region to provide the information for the application of the risk framework.

- Assess the erosion and flooding risks for the region of study, identifying potential areas of concern, and develop a comprehensive diagnosis of the current situation, intended to further analysis, adaptation and sustainable development policies.

All in all, this Thesis aims at a wider scope than global wave climate analysis and their related risks since it intends to constitute an initial step towards the definition of a global framework for coastal impacts and risks assessment for more informed decision making.

Note also that regarding the advances in modeling and extended measurements of different climate dynamics, we are in an unique position to assemble efforts for contributing to a sustainable development through a more efficient coastal planning and engineering solutions. The information, results, arguments and methods herein included intend to contribute to achieve such a noble goal.

1.4 Methodology

1.4.1 General methodology

From the review of current knowledge, it is clear that wave climate variability may influence human development in diverse ways and time scales. In parallel, risk analysis is being increasingly extended from design to policy integration. Settlements in high risk areas make environments and population more sensitive to climate variations. The application of a risk analysis framework to the problems of a varying wave climate is one specific objective of the present work. The methodology adopted for the risk assessment is described in the next point. The global framework is applied at two particular large scale studies: (1) erosion and flooding risks in Latin America and the Caribbean, and (2) technological risks for four wave energy converters based on the assessment of global wave energy resources variability during their life-cycle.

Figure 1.2 shows a flowchart that represents the process and the main study areas of the present work: description of the global wave climatology, its inter-annual (range of years) and longer-term (range of decades) variability. For the wave climatology description a numerical wave reanalysis is first developed (Global Ocean Waves database, GOW) including satellite information and reanalyzed meteorological data (winds and ice coverage).

To study implications of a varying wave climate, the initial step is to develop a global wave dataset that fulfills the requirements of long time span, representativeness and spatial and temporal homogeneity. Owing the scarceness of data with sufficient spatial and temporal definition (i.e. buoys and altimeter data), this new global wave reanalysis is developed using the wave model WavewatchIII ([Tolman, 2002a]) and the NCEP/NCAR-RI meteorological reanalysis ([Kalnay et al., 1996]). Particular features of the resulting dataset to highlight are: (1) the correction of wave data with altimeter observations with a prior identification of outliers due to tropical cyclones, (2) extensive validation of numerical results with buoy and altimeter data, (3) long time span covered and (4)intention to be up-dated.

As said, marine climate dynamics vary in the range of hours, days, months, seasons, years and decades. Variations over the seasonal scale are far from being negligible and recently have revealed an important influence on many coastal problems and impacts. The natural inter-annual behavior of the wave climate is in this work explored. Climate change seems to be affecting marine climate in the long-term as has been reported during the last decades. Sea-level rise is globally accepted but wave climate variations and implications have been less attended. This Thesis aims to contribute to the knowledge in that particular area. The study of variability is based on advanced statistical analysis of time series (correlation, trend detection and other regression modeling).

From the global wave dataset, a description of the global wave energy resource and the performance of four offshore wave energy converters is tested. Additionally, based on long-term trends in the wave climate and the established risk framework, operations and survivability in a global context is analyzed for the selected technologies.

Analogously, the work intends to address the issue of coastal impacts derived from changes in the wave climate. However, it is necessary to limit the analysis to a regional domain, in this

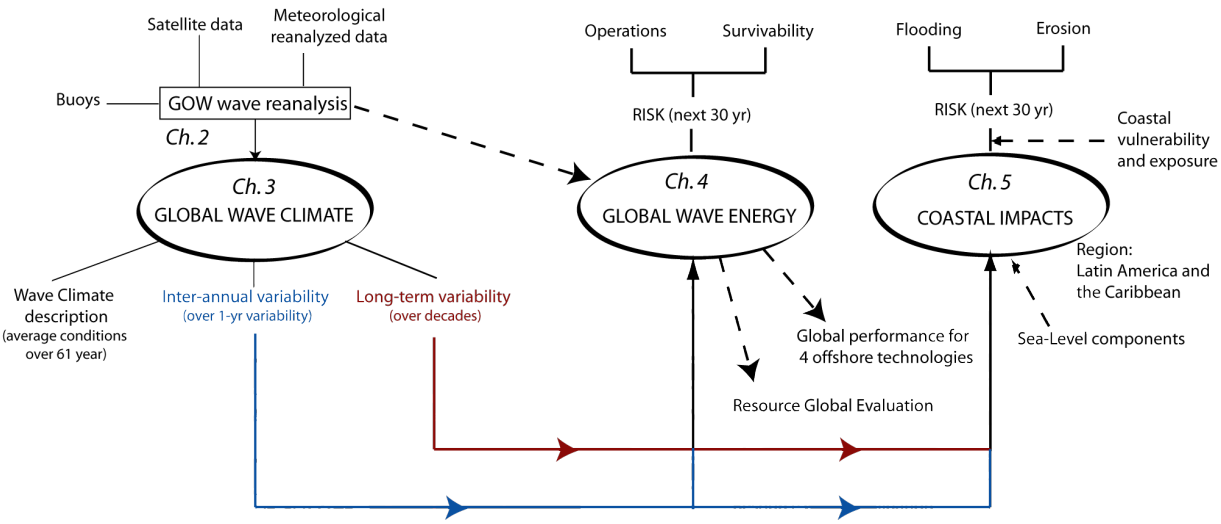

Figure 1.2: Main areas of analysis of the Thesis and relationships between them.

case the region of Latin America and the Caribbean, since data on coastal vulnerability and exposure are needed to derive the consequences upon a risk perspective. The coastal impacts addressed in this research are flooding and long-term erosion. In this regard, the different sea-level components (sea-level and tides) contributing to such impacts are also necessary to be taken into account.

1.4.2 The risk framework and the adopted approach

Despite the diverse terminology used at each discipline (e.g., [Adger, 2006]), it is commonly acknowledged in the literature that risk depends mainly on a number of components: hazard, exposure, and vulnerability. These three terms constitute the "risk triangle", primarily used in natural disaster terminology ([Schneiderbauer and Ehrlich, 2004]). This approach is hereafter adopted since it corresponds well with the coastal problems.

However, before describing the approach adopted for risk assessment, a clear description of the terms used is crucial, considering the vast number of definitions and methods at each discipline.

1.4.2.1 Definitions and concepts

There is comprehensive terminology to refer to the terms risk, hazard, exposure and vulnerability: [Coburn et al., 1991, Carter et al., 1994, McCarthy et al., 2001, Schneiderbauer and Ehrlich, 2004, Birkmann, 2006, Adger, 2006, Solomon et al., 2007, Füssel, 2007] are some examples. Definitions of each term used in the methodology follow.

The concepts used in this work correspond mainly with European Commission terminology

([Schneiderbauer and Ehrlich, 2004]) which defines ***risk*** (R) as "the probability of harmful consequences or expected losses resulting from a given hazard to a given element at danger or peril, over a specified time period". To understand a risk, compare risks or determine the evolution under the effects of climate change, an attempt is usually made to quantify it in some way. This requires data on the effects of the various hazards causing the risk and forecasting the probability of future events. A further critical aspect for the evaluation of future risks is the identification of the causes and effects, and a clear understanding of the processes which occur in the face of disastrous events or phenomena induced by or associated with climate change.

Depending on the way an element at risk is defined, it can be quantified in terms of the expected economic loss, the number of people affected or the extent of physical damage to property. Precision in quantifying a risk depends to a large extent on the amount of information available. Databases for actions or events which may generate negative effects or losses are important in determining the likelihood of occurrence, in terms of both the quality and the quantity of data. These factors influence the determination of the risk, which will be defined by an estimated mean and an interval of confidence. Such intervals represent confidence in relation to any estimate of future risk.

The probability of a coastal storm, associated flooding, wave height intensity, erosion, reliability of structures or functionality of wave farms, as examples of possible impacts, are all factors that result in a certain risk of failure or disfunction with associated consequences. These probabilities can be calculated by engineers based generally on historical information. The consequences are obtained from the analysis of a wide variety of data.

A further important factor is the agent and/or ***hazard*** (H), defined in the context of this work as any event, whether a physical phenomenon or arising from human activity, which may cause damage in terms of losses of human lives, property, socioeconomic losses or environmental degradation in a coastal environment. It must be noted that it is not only necessary to be familiar with the agents or hazards which may affect the coast, but also to quantify their frequency of occurrence in time and their magnitude. Thus the concept of uncertainty is introduced as the probability that an agent will arise with a given magnitude in a region and in a specific period of time. The whole concept is broadly named as hazard.

The agent and its uncertainty in terms of frequency of occurrence are independent both of the coastal zone and of human activities where are likely to affect. Such is not the case with intensity, which may be mitigated or amplified by orientation, geography or other coastal features. Waves are a clear example, their propagation affected as they approach the coast, so altering the intensity and direction, but whose frequency of occurrence is independent of the type of coast. Defining the statistical behavior is a useful tool in this regard.

When a given coastal zone is affected by an agent of a certain intensity, the population, ecosystems, infrastructures, services and/or economic activities are affected. These damages are grouped under the concept of ***vulnerability*** (V), understood as the extent to which changes could harm a system, or to which the community can be affected by the impact of a hazard. Thereby, vulnerability comprises multiple institutional, economic and socio-cultural factors.

The European Commission ([Schneiderbauer and Ehrlich, 2004]) or the Intergovernmental Panel on Climate Change ([Magrin et al., 2007]) divide the vulnerability into several terms, but

this work does not differentiate this degree of fragmentation and it rather takes the vulnerability as an indicator of potential socioeconomic damage, in an aggregated form.

The term ***exposure*** (E) links the agent with the risk at a specific zone. It incorporates information on the coastal relief and remaining physical characteristics such as slope or geographic location. It thereby defines the degree to which a coast will be affected by a particular agent (e.g. land surface affected by coastal flooding). Exposure is defined as the zone, region, entity or body affected by the occurrence of an event associated with the agent. The term is normally employed in the natural disaster literature to define the population living in a geographical area affected by an agent. However, the exposure can also be associated to the physical features that make that population exposed, for example, the topography.

To further illustrate these concepts, consider a coastal zone with a urban area (i.e. population or land uses configuring the vulnerability term) in a coast subjected to changes in the wave dominant direction in the long term. If the settlement is located at a cliff (i.e. exposure), in principle it will not be affected by a shift in wave direction (i.e. hazard), but it could be exposed to changes in the wave intensity which may induce cliff erosion. Meanwhile, if the urban area is protected from sea dynamics by a sandy beach (i.e. exposure), both changes in wave intensity and directional shifts (i.e. hazards) will affect the urban front.

1.4.2.2 Risk assessment approach

Since risk is a measure of the probability (or likelihood) of occurrence of a future hazard (i.e. with a certain reach) and its consequences ([USACE, 2011a]), and although the expression depends on the method and application field, it may be generally expressed as:

$$Risk = f(Probability, Consequences) \tag{1.1}$$

It is acknowledged in the literature that risk depends mainly on a number of components: hazard (H), exposure (E), and vulnerability (V), which make up the "risk triangle", primarily used in natural disaster terminology ([Schneiderbauer and Ehrlich, 2004]). Thus risk, R, can be expressed as:

$$R_{ah} = H_{ah} \cdot E_a \cdot V_{ah} \tag{1.2}$$

where subindexes a and h refer, respectively, to a geographical location (a) and a hazard (h), with their associated magnitude.

In probabilistic terms, a hazard which ultimately generates an impact can be expressed by a probability density function, $f_z(z)$ (i.e. combines frequency of occurrence and intensity). A generic expression, integrating the various factors for a given agent, and associated with a particular geographical location, is defined as follows:

$$R = \int_{-\infty}^{\infty} f_z(z)E(z)V(z)dz \tag{1.3}$$

where E is the exposure of the area of study for the value of agent z. Analogously, V is the vulnerability associated with the value of agent z, and R is the expected cost of the damage (units given by the vulnerability term).

Note that from the standpoint of this methodology, unlike many other studies, vulnerability is a mathematical function defined according to the degree of loss of the element affected, and which depends on the intensity of the agent.

Figure 1.3 represents the latter expression showing the probability density function (hazard) of the random variable which defines the impact (agent, z) and the probability between two thresholds which would cause a certain degree of damage. Note that, for a specific agent, the probability between two values corresponds to the integral of the curve.

There is in general no information to construct the functions $E(z)$ and $V(z)$ in a continuous form so that it must be done discontinuously, and the risk integration is expressed in discrete form:

$$R = \sum_{i=1}^{n} P_i \cdot E_i \cdot V_i \tag{1.4}$$

where n is the number of subintervals into which the magnitude of the agent is discretized; P_i is the probability that agent z falls within the values of the level of the $i-nth$ interval; E_i is the exposure produced for the mean value of the agent $(z_{i-1}+z_i)/2$; and V_i is the vulnerability produced for the value of the agent $(z_{i-1}+z_i)/2$. The probability for the agent to occur with a given intensity (P_i) can be determined in various ways depending on the nature of the variable. Those used here are named for each case, but a comprehensive description of each of them can be found in the appendix A devoted to the methods to study the time variability.

Figure 1.4 shows a diagram of the methodological process proposed in this work. Based on an evaluation of the coastal dynamics by their existing statistical behavior (with numeric data or/and registered measurements), and inferring the future statistics (by simulation or statistical techniques), the impacts can be defined. This, together with an analysis of the vulnerability and the exposure of the socioeconomic and physical resources, generally based on a large assimilation of geospatial data, makes it possible to perform a risk assessment, its subsequent analysis and the resulting decision-making.

This methodology thus takes account of the following: (1) that vulnerability changes with the type of agent and its severity and (2) the risk requires knowledge of the spatial distribution of (a) the occurrence of the events associated with the agents and (b) the elements at risk.

However, when dealing with structural design, in a basic approach (level I), the value that configures the failure of a certain infrastructure is usually defined by the exceedance over a

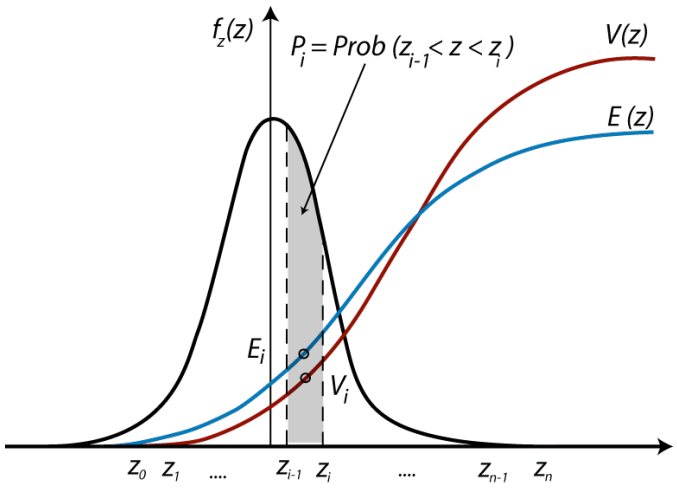

Figure 1.3: Curves for the risk assessment where z represents an agent generating an impact on the coast (hazard; z), $f(z)$ represents the hazard probability density function, $V(z)$ the function of vulnerability, and $E(z)$ the function of exposure for each intensity of agent z.

certain extreme value, which is defined by a certain return period (i.e. mean number of years for the occurrence of the storm value) and a life expectation. The equation for assessing the risk along the installation expected life (i.e. the likelihood of at least one event that exceed the design value) can be expressed as:

$$R = 1 - [1 - Prob(X >= x_t)]^n = 1 - [1 - \frac{1}{T}]^n \tag{1.5}$$

where $Prob(X >= x_t) = 1/T$ is the probability of the occurrence for the failure event in question and n is the expected life of the structure. This expression does not incorporate a term for consequences because it is assuming that total failure occurs if the extreme value is reached. However, that is not completely true as failure is normally a progressive process. This is why actual designs generally follow a more accurate probabilistic approach.

Depending on the state of the infrastructure, the survival risk for a certain year would be:

$$R(t) = [1 - \frac{1}{T(t)}]^n \cdot C(T, t) \tag{1.6}$$

where $C(t)$ would take into consideration the repair costs depending on the return period as a function of time. The total risk for the life cycle could be obtained from the integration in time for all the years of the life cycle, were the the evolution in time of T(t) and C(T,t) known.

Alternatively, the proportion of safety (Security Index, SI) could also be represented trough the ratio between the initial probability of failure at the design age, associated to a certain return period (T), and at the end of the life cycle ($T(t = LC)$):

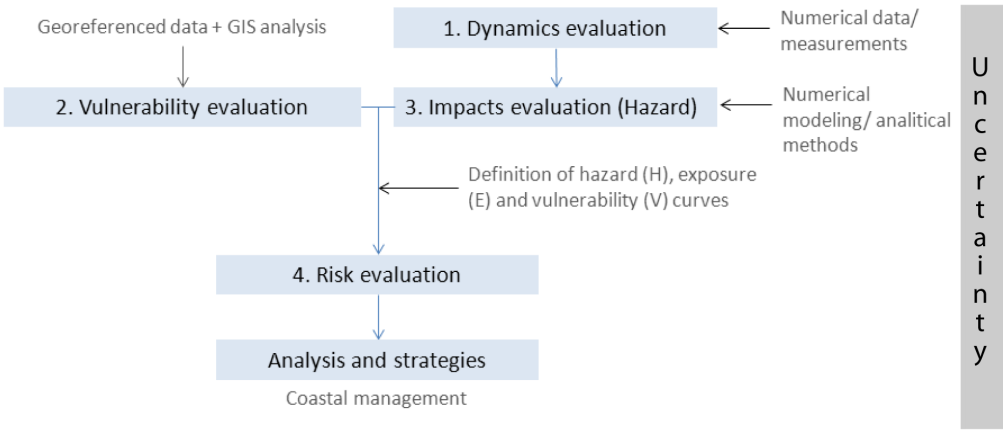

Figure 1.4: Methodology and general framework for the risk assessment.

$$SI = (1 - \frac{T(t = LC)}{T(t = 0)}) \cdot 100 \tag{1.7}$$

1.5 Structure of the document

The document is structured in six chapters, these lines corresponding to the first of them, in form of a general introduction of aims and general context. Chapter 2 describes the development of the global wave reanalysis performed as an initial step for the achievement of the subsequent goals. For introductory purposes, each chapter includes a brief review of the specific state of the art on each subject.

Chapter 3 addresses the study of the global wave climatology considering different statistics parameters (wave heights, periods and directions) and past variations in inter-annual and long-term time scales. Methods for time variability are described in the annex A. A discussion on anthropogenically-forced changes in climate variability is also held to reveal the need of studying both time scales. The relationship between global wave power and climate change is also addressed at the end of the chapter, suggesting a possible interaction between sea surface temperature and the global energy transferred to the ocean waves.

Having seen and understood past changes in wave climate, Chapters 4 and 5 deal with specific derived risks. Technological risks for wave energy converters harvesting and structural design are faced in Chapter 4. A comparison of four different technologies is provided, at present and through a life-cycle. Alongside with the survival risk from changes in wave extremes, a general overview for consideration of wave variability and non-ergodic behavior into wave energy studies follows.

Chapter 5 specifically deals with the risks of erosion and flooding in the region of Latin America and the Caribbean, broadly named as environmental risks. Wave climate is not the

unique factor considered in this section, sea-level components are also studied. A risk analysis is finally applied for these two particular coastal impacts. Such section shows a second application of the risk analysis assessment adopted in the work.

It is necessary to caution the reader, however, that each chapter is structured in a way that intends to be self-explanatory. Each chapter at least includes an introduction, a main body of content and final conclusions. In this way, each chapter deals with a specific thematic area that despite being part of a unique whole, may be read individually.

Chapter 2

A new global wave reanalysis

"Anyone who thinks there's safety in numbers hasn't looked at the stock market pages"

Irene Peter, American writer

2.1 Introduction

Ocean surface gravity waves are the result of an important exchange of energy and momentum at the ocean-atmosphere interface. Waves propagate through the ocean basins transporting the accumulated energy obtained from the wind. During wave propagation, some energy is dissipated through different processes. The knowledge about how energy from winds transfers into the seas and how this energy propagates and dissipates is of great importance for the scientific community, since it allows understanding and modeling of wave fields. This modeling is of utmost importance for design purposes in offshore and coastal engineering. Furthermore, current research topics in ocean studies require long time series of wave climate with high spatial resolution at a global scale. Some examples of these research topics are the evaluation and study of wave energy resources, ocean dynamics variability, definition of operable conditions in shipping routes, maintenance and repair strategies for offshore constructions, extreme wave analysis, etc. Besides engineering, climate change also demands tools and data to define long-term variability of wave climate within different scenarios. Note that for all these research trends, global wave fields containing long time series of wave climate parameters are required.

Over the last decades, there has been an increasing interest in collecting wave climate information through instrumental devices such as buoys and satellite altimetry. Buoy measurements provide very accurate time series records but they are relatively short and are sparsely located in space, most of them in the Northern Hemisphere. In addition, they usually present interruptions due to disruptions on the normal use by buoy failure and maintenance activities. In contrast, satellite observations present a global coverage and also provide information with a high level of precision (± 3 cm, [Krogstad and Barstow, 1999]). However, this source of data is only available since 1992 and with a non-regular time resolution.

Both sources of information, buoys and altimetry, do not configure a temporal and spatial homogeneous record of ocean wave climate variables for most of the purposes mentioned above. This issue has motivated an increasing interest in wind wave models, which allow obtaining spatially homogenous long-time series of wave climate parameters, i.e. Wave Reanalyses Databases (WRD). However, as it has been pointed out by several authors ([Cavaleri and Sclavo, 2006, Caires et al., 2004a]), WRD are not quantitatively perfect, presenting several deficiencies with respect to instrumental data. Despite those shortcomings, WRD constitute an optimal way to accurately interpolate data both in time and space, even for those locations where no instrumental measurements exist ([Weisse and Von Storch, 2010]). Results are accurate enough to make them suitable, if carefully applied, to be used for coastal engineering purposes as well to assess long-term changes and trends.

Wind wave models are driven by wind fields and constrained by ocean sea/ocean cover. The quality of any WRD depends upon the quality of wind forcing ([Feng et al., 2006]), and based on them, many efforts have been made in the last decades to generate consistent sets of data to define the wave climatology. However, changes over time in data sources, advances in data analysis techniques and evolution of the wind wave models have conducted to inhomogeneities between the wave results of the different reanalyses described above. The Global Ocean Waves (GOW) reanalysis, to be here presented, like its predecessors, also shows slight inhomogeneities caused by changes in the amount of assimilated observations within NCEP/NCAR reanalysis. They are particularly relevant in the Southern Hemisphere and before 1980 ([Sterl and Caires, 2005]). Potential users must be aware of this fact, specially if dealing with long-term trends in the Southern Hemisphere.

Therefore, available numerical data of wave climate vary both in time range and in quality. The aim of this chapter is to provide a comprehensive explanation of a new WRD with the following characteristics:

1. It intends to be continuously updated, constituting a valuable dataset of wave climate parameters for wave climate and engineering applications.
2. Global coverage.
3. Long length of the simulated records (time series of different wave statistical parameters and energy spectra from 1948 onwards).
4. High temporal resolution of the outputs (on a hourly basis).
5. Exhaustive validation using instrumental measurements from buoys and satellite altimetry.
6. Post-process using altimetry observations consisting of: (a) identification of possible outliers (from 1992) related to hurricanes and typhoons not appropriately reproduced by the numerical modeling, and (b) calibration of the model hindcast results to obtain a more accurate description of the wave statistical distribution according to instrumental data.

The rest of the chapter is organized as follows. Section 2.2 provides a review of the most important wave datasets with global coverage. Section 2.3 explains the methodology to obtain the Global Ocean Waves (GOW) reanalysis. The model description, the forcing fields and

boundary conditions, and model set-up are described in points 2.4, 2.5 and 2.6, respectively. A validation of numerical results (hindcast data) is developed in section 2.7. Correction procedures (outliers removal and calibration of wave heights), final results and verification of the correction procedure are presented in section 2.8. Homogeneity of data is discussed in section 2.9. Finally, concluding remarks are outlined in 2.10.

2.2 State of the art of wave modeling and wave reanalyses

The most advanced state-of-the-art wind wave models are the third generation wave models ([Komen et al., 1994]). Two of the most relevant and widely used within this group are the wave models WAM ([Hasselman et al., 1998]) and Wavewatch III ([Tolman, 2002b, Tolman et al., 2002, Tolman, 2009]), denoted as WW3 in the following. Both models are recommended to be used for open ocean wave simulation because non-linear and wave-bottom interactions are not appropriately addressed. For this reason, wave climate in coastal and shallow waters is poorly described with these models. For details about models focused on coastal and shallow waters conditions see [Schneggenburger et al., 1997, Booij et al., 1999] or [Camus et al., 2011].

Several WRD have been developed in the recent decades, only some conceived with a planetary scope. The most prominent are:

ERA-40: [Sterl et al., 1998] produced the first global wave reanalysis fields by forcing WAM on a 1.5° latitude x 1.5° longitude grid covering the whole globe with the ERA-15 winds from 1979 to 1993. In their study they analyzed the significant wave height (Hs) climatology in terms of annual cycles and trends. Following ERA-15, the ECMWF conducted ERA-40 ([Uppala et al., 2005]) for the period of 1957–2002. This is a reanalysis of, among other things, global ocean wind and waves, modeled with WAM ([Hasselman et al., 1998]). [Caires and Sterl, 2005a] produced a corrected version of the dataset, named as C-ERA-40, with a significant improvement in the diagnostic statistics ([Caires and Sterl, 2005b]).

ERA-40/ODGP2: In an independent study, the ERA-40 wind fields for 1988 were also used to force the ODGP2 spectral ocean wave model on a 1.25° latitude x 2.5° longitude grid covering the whole globe.

NCEP–NCAR winds and derived wave reanalyses: The first long global reanalysis of the surface winds was produced by NCEP–NCAR, initially covering from 1958 to 1997 and later extended. This reanalysis also benefited from the assimilation of data from COADS, but no satellite wind measurements were used. These wind fields are available on a Gaussian global grid of 1.875°. [Cox and Swail, 2001]performed a global wave reanalysis (CS01 hereafter), on a 1.25° latitude x 2.5° longitude grid for the period 1958-1997 with the spectral model ODGP2, becoming the first 40-yr wave reanalysis covering the whole globe. Their results were later used by [Wang and Swail, 2002] to study seasonal extremes of wave height.

On the other hand, there are data available from NOAA's operational hindcast system since 2005, which uses the version 2.22 of WW3 and these winds. Many other hindcast

data have used the NCEP/NCAR reanalysis but they are at a regional scale (i.e. ocean basins).

ERA-Interim: ERA-Interim ([Dee et al., 2011]) is an 'interim' reanalysis from the ECMWF for the period 1989-present in preparation for the next-generation extended reanalysis to replace ERA-40. Several significant advances in data assimilitation and observations were included with respect to ERA-40.

Non-public reanalyses: Several companies provide wave data based on wave reanalysis. One of the most prominent datasets are "MetOcean Solutions". They have developed 32 years of gridded wave hindcast data (1979-2011) at 0.5° for the globe, with the model WW3 and the CFSR wind fields (NCEP Climate Forecast System Reanalysis), which could be considered as the equivalent to ERA-Interim dataset.

There is a vast number of regional reanalysis performed in the last decade. For example, motivated by deficiencies in the NCEP/NCAR input wind fields, [Swail and Cox, 2000] carried out an intensive kinematic reanalysis of these winds in the North Atlantic using a finer wave model grid (0.62°° x 0.833°). Pacific Weather Analysis ([Graham and Diaz, 2001]) produced a 50-yr wave reanalysis (PWA-R) using NCEP/NCAR winds and the model WW3 on a 1°x1.5° grid for the North Pacific Ocean during the winter season. One of the first but prominent attempts to reconstruct past wave climate in the North-East Atlantic was carried out by the WASA group ([Group, 1998]). Later on, within the project HIPOCAS, a high resolution wave (and sea level) reanalysis ([Pilar et al., 2008]) was developed using the wave model WAM. More recently [Dodet et al., 2010] computed a reanalysis with WW3 for the last six decades to analyze wave climate variability in the North-East Atlantic. It is worth noting that the temporal resolution of results vary between reanalyses, from one to six hours. [Caires et al., 2004a] present a comparative study between some of the most relevant global WRD.

2.3 Methodology

This study presents the development of a wave reanalysis based on calibration of model hindcast data with satellite wave height observations and with verification against historical buoy data. The development of the GOW (Global Ocean Waves) database encompasses several stages, which are summarized in the flow chart of Figure 2.1.

Firstly, the wave generation is obtained by using the WW3 model and the NCEP/NCAR global wind and ice cover datasets. In order to check the performance of the wave generation model and the quality of the forcing fields, a preliminary validation is done using both buoy and satellite altimetry data as benchmarks. Next stage consists of the calibration of the numerical results (i.e. significant wave height) using satellite altimetry data. This process aims to make several systematic corrections to reduce deviations between the probability distribution function of corrected hindcast and instrumental data. However, previous to these corrections, an outlier detection process over the data pairs, both numerical and instrumental, is addressed. This filtering process is important to eliminate data related to hurricanes or typhoons episodes, which may be captured by the altimetry but are not appropriately reproduced in the forcing wind fields

because of insufficient resolution. Finally, an additional validation using buoy data is carried out to check the performance of the calibration process and the quality of the final database. Hereafter, we will refer to the hindcast data as the Non-calibrated GOW (NC-GOW) results and we will use GOW for the results after applying the corrections.

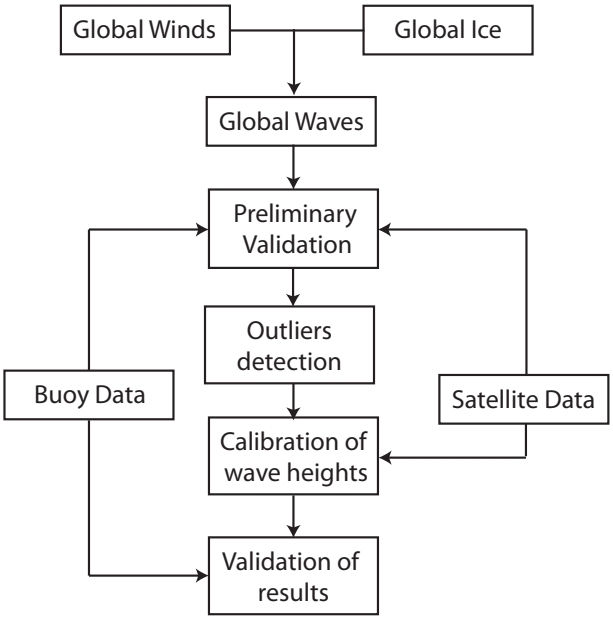

Figure 2.1: Methodology diagram for the global wave reanalysis development

2.4 Numerical model and set-up

WAVEWATCH III ([Tolman, 2002a, Tolman, 2002a, Tolman, 2009]), is a third generation wave model developed at NOAA/NCEP in the spirit of the WAM model ([Komen et al., 1994]). It is a further development of the model WAVEWATCH, as developed at Delft University of Technology ([Tolman, 1989]) and WAVEWATCH II, developed at NASA, Goddard Space Flight Center (e.g., [Tolman, 1992]). WW3, however, differs from its predecessors in many important points such as the governing equations, the model structure, the numerical methods and the physical parameterizations. Furthermore, with model version 3.14, WW3 is evolving from a wave model into a wave modeling framework, which allows for easy development of additional physical and numerical approaches to wave modeling.

WW3 solves the random phase spectral action density balance equation for wavenumber-direction spectra. The implicit assumption of this equation is that properties of medium (water depth and current) as well as the wave field itself vary on time and space scales that are much larger than the variation scales of a single wave. With last version 3.14 some source term options for extremely shallow water (surf zone) have been included, as well as wetting and drying of grid points. Whereas the surf-zone physics implemented so far is still fairly rudimentary, it does imply

that the wave model can now be applied to arbitrary shallow water. Despite the latter capacity, version 2.22 was used for this work since shallow water should be addressed through other models like SWAN and other strategies (e.g., see [Cavaleri et al., 2007] and [Camus et al., 2011]) and the last version was not yet available when the study started. However, at offshore depths, the numerical modeling has not changed.

Some physical features of the model are:

- The governing equations of WW3 include refraction and straining of the wave field due to temporal and spatial variations of the mean water depth and of the mean current (tides, surges etc.), when applicable.
- Parameterizations of physical processes (source terms) include wave growth and decay due to the actions of wind, nonlinear resonant interactions, dissipation ('whitecapping'), bottom friction, surf-breaking (i.e., depth-induced breaking) and scattering due to wave-bottom interactions. The model is prepared for triad interactions and is prepared for other, as of yet undefined source terms, but the latter have not been implemented yet.
- Wave propagation is considered to be linear. Relevant nonlinear effects such as resonant interactions are, therefore, included in the source terms (physics).
- The model includes options for choosing various term packages, some intended for operations, others for research. The source term packages are selected at the compile level.
- The model includes several alleviation methods for the Garden Sprinkler Effect ([Tolman, 2002a])
- The model includes sub-grid representation of unresolved islands
- The model includes dynamically updated ice coverage.
- The model is prepared for data assimilation, but no data assimilation package is provided with the distribution of the source code.
- Spectral partitioning is available (new in model version 3.14) for post-processing of point output, or for the entire wave model grid using the [Vincent and Soille, 1991] algorithm ([Hanson et al., 2009].

Numerical features, output options, user manuals and further information can be consulted on: http://polar.ncep.noaa.gov/waves/wavewatch/.

2.5 Forcing and boundary conditions

Wave model output is very sensitive to the wind field input choice. Therefore, different studies have been developed to judge the quality of the wind fields. [Tolman, 2002b] determines that the NCEP/NCAR winds provide the best results in terms of significant wave height estimates for the model WW3. More recently, [Feng et al., 2006] analyze four different wind forcing fields

with the WW3 model: (1) NCEP/NCAR reanalysis winds, (2) the ECMWF wind fields, (3) the QuickSCAT wind observations blending the NCEP/NCAR reanalysis winds and, finally, (4) an enhanced ECMWF wind field with assimilation of wind speed measurements. Their results indicate that NCEP/NCAR winds as input data produce the best agreement with TOPEX altimetry wave measurements, at both global and regional scales, while the others present a higher spatial variability and are all positively biased.

The long temporal coverage (from 1948), the up-to-date characteristic and the good evaluation obtained in the mentioned works, indicate that the use of NCEP/NCAR wind fields is an adequate choice for wave modeling performance using the model WW3. However, the wave data in some regions of the Southern Hemisphere should be taken with caution due to wind input homogeneity problems ([Sterl and Caires, 2005]).

Some analysis (not shown) was made on the forcing of the model with other atmospheric reanalyses, in particular JRA-25 and ERA-40 (see full list of available atmospheric reanalyses in the state of the art in chapter 1). The trials confirmed the choice of the NCEP/NCAR.

Considering the latter and the fact that NCEP/NCAR reanalysis provides the longest time span available, for the purpose of wave climatology analysis, this reanalysis was used for the development of the new wave dataset. Therefore, WW3 model was forced with 6-hourly wind fields from the NCEP/NCAR Reanalysis project ([Kalnay et al., 1996]), from 1948 onwards. Wind data are defined on a Gaussian grid with a spatial resolution of approximately 1.9° in latitude and 1.875° in longitude.

Simulations also include ice coverage fields from NCEP/NCAR. The boundary conditions were configured by the continental margins, so the internal sea such as the Mediterranean sea was not considered for the global computation. Bathymetry data used for the simulation was obtained and specially treated from the ETOPO dataset ([NOAA, 2006]).

2.6 Model set-up

The generation and propagation of the wind waves were simulated with the model WW3, version 2.22 ([Tolman, 2002a]). Simulations are computed on a global grid with a spatial resolution of 1.5° in longitude and 1° in latitude (a total of 22,945 computational nodes). The resolution chosen resulted from a balance between computation effort (over 4,100 h of CPU), quality of results and analysis capacity (i.e. study of the climatology on a planetary scale). The processing time for approximately 23,000 time series was in the range of months. However, it must be stressed that a considerable improvement will be obtained with resolution of 0.5°x0.5° according to our experience. This will be specially remarkable for the coastal boundaries but it will boost the computation and processing time (i.e. by over a factor of six).

Regarding the numerical configuration, wave growth uses source terms ([Tolman and Chalikov, 1996]) to account for wind input, non-linear wave-wave interactions and whitecapping. Effects of depth-induced refraction are also considered in the propagation model. Wave interactions with currents and island shadowing were not considered in the simulation process. The minimum propagation time-step used for the computation was

60 seconds and the spectral resolution covered 72 regularly spaced directions. Frequencies extended from 0.03679 hz with 25 frequency steps and a frequency increment factor of 1.1.

The output parameters obtained all over the grid are: the significant wave height (H_s), mean wave period (T_m), peak period (T_p), peak direction (θ_p), mean wave direction (θ_m), directional spread and energy spectra in specific locations along the coast to analyze multimodal sea-states, with different swell and wind sea components. This increment of data storage in coastal areas allows the coupling with higher resolution wave propagation models (see [Camus et al., 2011]) for engineering applications (ports, breakwaters, sediment transport, etc.). In this work the different swell and wind sea components are not analyzed separately as in [Semedo et al., 2011b], although they could also have been obtained.

2.7 Wave hindcast results

2.7.1 Validation data

An important aspect within WRD design is the validation process using instrumental information as a benchmark. For this particular issue, wave model results are compared with measurements from deep-water buoys at different locations over the Atlantic and Pacific oceans and the Caribbean Sea. In-situ buoy measured wave data were obtained from three different sources: NOAA National Data Buoy Center, the Environment Canada and Puertos del Estado (Spain). Table 2.1 includes the locations of the buoys used in the validation process, which are also represented in Figure 2.6.

Although buoy observations are considered the most reliable wave measurements, they are scattered in time and space, mainly located in the Northern Hemisphere and are generally only available for the last two decades. In contrast, altimetry wave measurements provide the best possible spatial coverage to evaluate global wave data. Wave data from different satellite missions show very good agreement between each other, so they were combined for comparison with the numerical results, from 1992 to 2008. Satellite missions correspond to Jason 1, Jason 2, TOPEX, ERS-2, Envisat and GFO Data were corrected for each mission according to procedures explained in [Cotton, 1998] and [Woolf and Challenor, 2002], and later updated by [Hemer et al., 2010].

2.7.2 Preliminary validation

In order to assess the quality of the numerical simulated results, these are validated against buoy and altimetry data. Different wave parameters (H_s, T_p and θ_m) are compared in a total of 21 buoy stations, some of them directional gauges. The following diagnostic statistics parameters are calculated for comparing model performance (y) with respect to instrumental data (x):

- The systematic deviation between two random variables (BIAS; usually model minus observations):

$$\text{BIAS} = \bar{x} - \bar{y}. \tag{2.1}$$

- The root mean square error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{n_d}\sum_{i=1}^{n_d}(x_i - y_i)^2}. \tag{2.2}$$

- The residual scatter index (SI), which measures dispersion with respect to the line $x = y$:

$$\text{SI} = \frac{\text{RMSE}}{\bar{x}}. \tag{2.3}$$

- The Pearson's correlation coefficient (ρ):

$$\rho = \frac{\text{cov(x,y)}}{\sigma x \cdot \sigma y}. \tag{2.4}$$

  where *cov* represents the covariance between the two variables and ρ varies between -1 and 1.

- Sample distribution moments: mean (μ), standard deviations (σ), skewness (γ), and kurtosis (ξ).

The mean μ, of the discrete random variable X, that is a 'weighted average' of all the values the variable may be expressed as,

$$\mu = E(X) = \sum_{x \in S(X)} xP(x), \tag{2.5}$$

$$\text{or} = \int_{x \in S(X)} x\, f(x)dx. \tag{2.6}$$

Thus, the mean, μ, is the first moment of X with respect to the origin.

The second central moment, better known as the *variance*, σ^2, of the discrete random variable X, is

$$\sigma^2 = E[(X-\mu)^2] = \sum_{x \in S(X)} (x-\mu)^2 P(x), \tag{2.7}$$

$$\text{or} = \int_{x \in S(X)} (x-\mu)^2 f(x)dx, \tag{2.8}$$

The *standard deviation*, σ, of the random variable X is the positive square root of its variance.

The third central moment, better known as the *skewness*, or lack of symmetry, of the discrete random variable X, is given by,

$$E[(X-\mu)^3] = \sum_{x \in S(X)} (x-\mu)^3 P(x), \tag{2.9}$$

$$\text{or} = \int_{x \in S(X)} (x-\mu)^3 f(x) dx, \tag{2.10}$$

$E[(X-\mu)^3]$ will be positive if there is greater dispersion of values $X > \mu$ than for values $X < \mu$; the sign and magnitude of $E[(X-\mu)^3]$ governs the sign and degree of 'skewness'

$$\gamma_1 = \frac{E[(X-\mu)^3]}{\sigma^3}. \tag{2.11}$$

Positive skewness is indicated by the longer 'tail' of the distribution in the positive direction.

The fourth central moment, better known as the *kurtosis*, of the discrete random variable X, is,

$$E[(X-\mu)^4] = \sum_{x \in S(X)} (x-\mu)^4 P(x), \tag{2.12}$$

$$\text{or} = \int_{x \in S(X)} (x-\mu)^4 f(x), dx, \tag{2.13}$$

which measures the 'flatness' of the distribution. The greater the moment, the flatter (less peaked) is the distribution. The kurtosis is defined as

$$\gamma_2 = \frac{E[(X-\mu)^4]}{\sigma^4}, \tag{2.14}$$

with $\gamma_2 = 3$ for the Normal distribution. This measure is used only for large samples.

Higher order moments can be developed. In general the set of all moments of a probability function describes the function exactly. Thus, for some probability functions, a limited set of moments is sufficient to describe the function completely.

These statistics are used to measure the quality of the results at the two validation stages:

1. Comparing numerical results (NC-GOW) with buoy data at a first stage.
2. Comparing calibrated numerical results (GOW) with buoy data and altimetry after the calibration process.

Due to the scattered distribution of buoy locations in the oceans, the only way to validate numerical results on a global scale is by comparing them with altimetry observations. To make

meaningful comparisons, reanalysis data are interpolated to the instants and positions of the instrumental observations (denoted as GOW_i data). For every node of the simulation grid, all altimetry observations within cells of equal dimension as the model resolution (1° x 1.5°) are selected.

Figure 2.2 shows scatter and quantile-quantile (25 equally distributed quantiles on a Gumbel scale) plots comparing buoy data and wave model results at different locations. In all cases, NC-GOW shows good agreement. Note that in the case of the buoy CAR-41041, a tropical cyclone appears in the buoy record with a maximum significant wave height around 7 m, which is not appropriately reproduced by the wave model (H_s around 2 m height). This is due to the poor resolution of the input wind fields. Note also that for high quantiles, wave heights appear to be under-estimated in some buoys while in others they are slightly over-estimated. This result supports the need of a correction, especially for the highest quantiles, which are the most relevant for engineering applications.

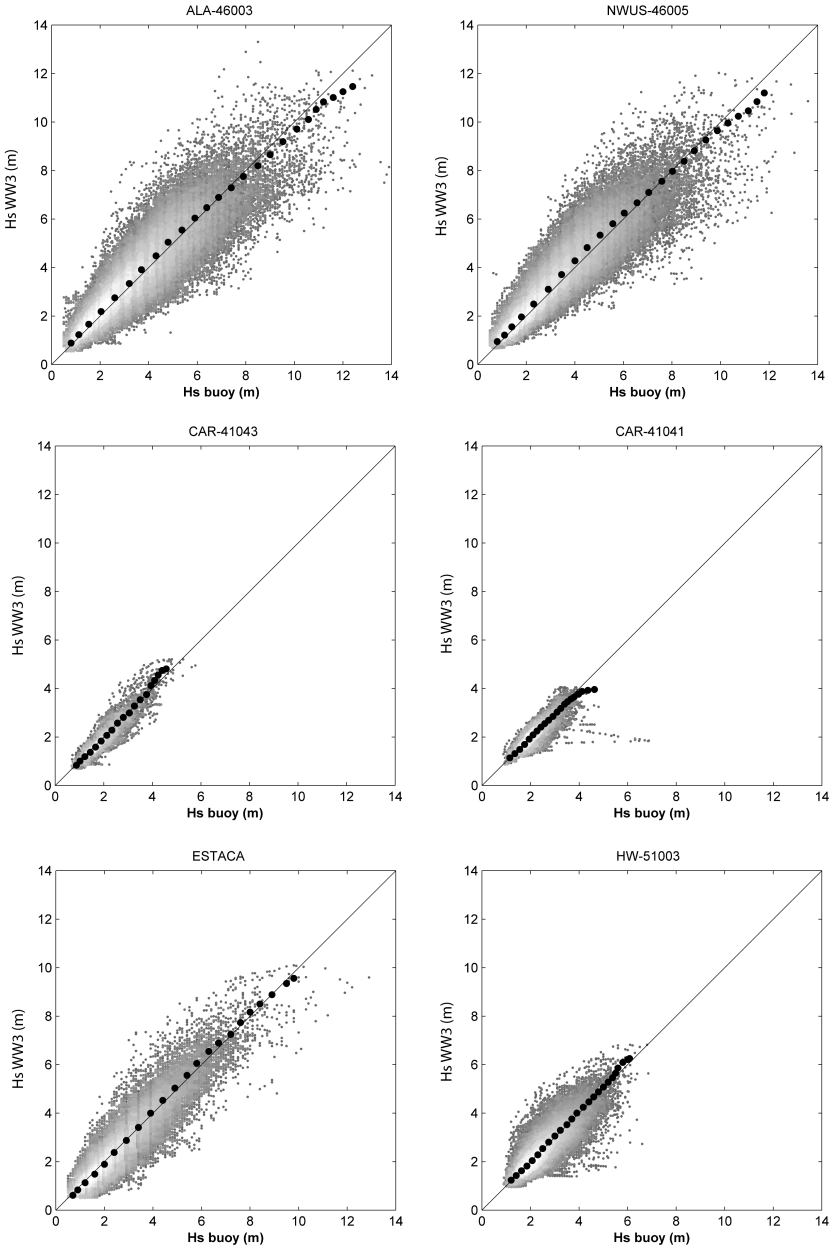

Figure 2.2: Scatter and quantile-quantile plots of buoy measurements (horizontal axis) and wave model results (vertical axis). Large black dots represent the quantile values (plotted equally spaced in a Gumbel scale), small dots correspond to data pairs of significant wave heights (buoy versus model) and the color intensity represents the density of data. Data for each buoy correspond to the time span given in table 2.1.

In Figure 2.3, hindcast and instrumental H_s time series at six different buoy locations, covering different years, are shown. For all cases, model data reproduce appropriately the magnitude and temporal evolution of the instrumental H_s records. Note that the highest differences correspond to peak events, where some of them are accurately reproduced in magnitude while others are not. For instance, the peak events of NWUS-46006 buoy show little discrepancies during the first months of the year whereas differences up to 2 m occur for the last months of the year.

Figure 2.4 shows the model performance on i) significant wave height (H_s), ii) peak period (T_p) and iii) mean wave direction (θ_m) at Silleiro buoy station during year 2006. Note that besides H_s, both peak period and mean wave direction present very good agreement with respect to instrumental data. In general terms, good agreement has been found between model results and observations, being the higher discrepancies associated with the highest wave events. This result is also observed, from a statistical point of view, in the quantile-quantile plots (see Figure 2.2).

Table 2.1 provides for different buoy locations the following information: the name of the buoy, longitude and latitude, length of records ($\boldsymbol{n}$), and several diagnostic statistics related to H_s and T_p, respectively, comparing NC-GOW data versus buoy observations. From this validation the following remarks are pertinent:

1. The biases related to wave heights are relatively low. The highest absolute values correspond to negative biases, which means that the model overestimates wave heights on average.

2. The buoys located in areas with frequent storms (BER-46035, ALA-46003, NWUS-46006, CAN-46004, NWUS-46005) tend to show a poorer performance in wave heights and are also associated with higher dispersion ranges (see scatter plots in Figure 2.2).

3. The scatter indexes and correlation coefficients of H_s are below 0.3 and above 0.85 respectively, which are appropriate diagnostic values for these kind of comparisons. Note that lower scatter indices and higher correlation coefficients correspond to higher reanalysis quality. Diagnostic statistics for CADIZ and CH-32301 (Chile) buoys are comparatively worse than for the rest of locations. The former is due to the coarse spatial grid resolution in semi-enclosed areas such as small gulfs and the latter may be due to the doubtful quality and short length (1984-1986) of the record.

4. Correlations related to T_p are lower than those associated with H_s. This result is consistent with results from other reanalyses existing in the literature.

5. The bias associated with T_p is negative for all cases (overestimation of the model) except for the three buoys in the tropical Atlantic. This result is probably induced by (i) not enough swell dissipation in the model and (ii) the discrete interaction approximation (DIA) for nonlinear wave interaction, which would also be consistent with the observed overestimate of the wave heights.

[Caires et al., 2004a] made a comparison of several global wind wave reanalysis, contrasting results during four different years (1978, 1988, 1994 and 1997) with several buoy records over the globe. Table 2.2 provides those results at the Peruvian coasts (buoy CH-32302), including also

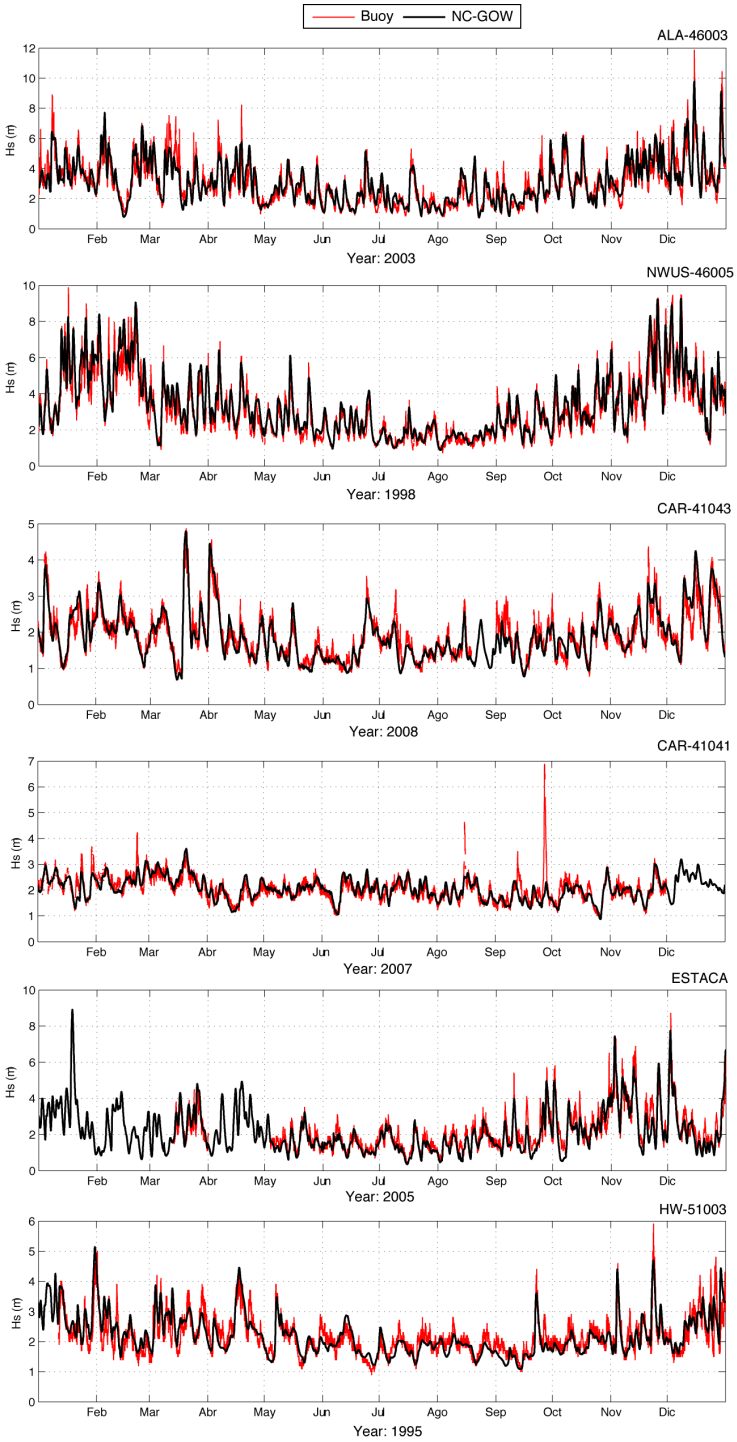

Figure 2.3: Reanalysis (black) and instrumental (red) significant wave height (H_s) time series at several buoy locations. Data for each buoy correspond to the time span given in table 2.1.

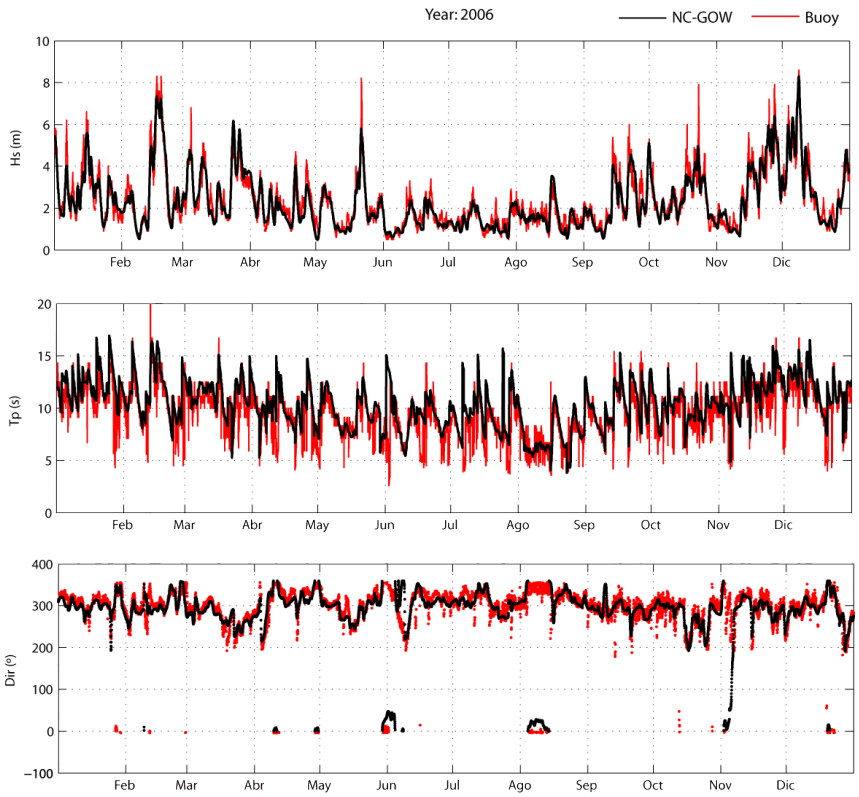

Figure 2.4: Time series of significant wave height (H_s), peak period (T_p) and mean wave direction (θ_m) at SILLEIRO buoy. Red: buoy measurement, Black: wave model.

NC-GOW results for comparison purposes. Analogously, tables 2.3 and 2.4 provide the same information as table 2.2 but related to four islands in Hawaii (buoys HW-51001, HW-51002, HW-51003 and HW-51004) and three buoys close to the coast of Alaska (ALA-46001, ALA-46003 and CAN-46004), respectively. Note that their work did not include the comparison with more modern and better global reanalyses, like C-ERA-40 or ERA-Interim ([Dee et al., 2011]) datasets; this reason prevent us from making this comparison. It is important to note that expression (2.3) used for the Scatter Index in the present paper differs slightly from the one used in [Caires et al., 2004a]:

$$\mathrm{SI}_C = \frac{\sqrt{\frac{1}{n_d}\sum_{i=1}^{n_d}\left[(y_i-\bar{y})-(x_i-\bar{x})\right]^2}}{\bar{x}} = \frac{\sqrt{\frac{1}{n_d}\sum_{i=1}^{n_d}\left[(y_i-x_i)-\mathrm{BIAS}\right]^2}}{\bar{x}}, \tag{2.15}$$

However, they are equal ($\mathrm{SI}_C = \mathrm{SI}$) in case the BIAS is null. The comparison hereafter has been made with the SI_C index.

In addition, there are some differences about how the data are pre-processed for comparison purposes. [Caires et al., 2004a] process the time series using the procedure described in [Caires and Sterl, 2003], and compare reanalyses with a 6-h average from buoy observations.

				Hs (m)					Tp (s)				
Buoy Name	**Time span**	**Lon.**	**Lat.**	n	**BIAS (m)**	**RMSE (m)**	ρ	**SI**	n	**BIAS (s)**	**RMSE (s)**	ρ	**SI**
BILBAO	1990-08	356.95	44.00	49132	0.00	0.51	0.92	0.27	51517	-0.72	2.13	0.79	0.22
PENYAS	1997-08	353.83	44.00	67365	-0.24	0.66	0.85	0.34	68035	-0.66	2.02	0.79	0.21
ESTACA	1996-08	352.38	44.06	62277	0.05	0.52	0.92	0.22	62471	-0.69	2.13	0.78	0.22
VILLANO	1998-08	350.08	43.50	60848	0.03	0.55	0.91	0.23	60899	-0.74	2.09	0.77	0.23
SILLEIRO	1998-08	350.61	42.13	39043	0.06	0.44	0.94	0.20	39184	-0.74	1.92	0.78	0.20
CADIZ	1996-08	352.50	36.48	30755	-0.08	0.60	0.71	0.49	35682	-3.09	4.17	0.71	0.58
CAR-41043	2007-08	294.99	20.99	14363	0.04	0.25	0.93	0.14	14363	0.33	1.43	0.79	0.16
CAR-41040	2005-08	306.96	14.48	23060	0.07	0.26	0.90	0.13	23061	0.38	1.60	0.76	0.18
CAR-41041	2005-08	313.99	14.36	28581	0.04	0.27	0.88	0.13	28581	0.30	1.75	0.78	0.20
CH-32301	1984-86	254.80	-9.90	3359	0.00	0.33	0.76	0.15	3359	-0.98	3.46	0.71	0.29
CH-32302	1986-95	274.90	-18.00	68944	0.05	0.32	0.88	0.15	68939	-0.20	3.07	0.71	0.24
BER-46035	1985-07	182.42	57.05	158843	-0.16	0.67	0.92	0.25	159523	-0.37	2.73	0.67	0.30
ALA-46003	1976-07	205.02	52.70	191687	-0.15	0.66	0.92	0.22	185827	-0.24	2.67	0.71	0.26
ALA-46001	1972-07	211.83	56.30	234247	0.03	0.65	0.91	0.24	222623	-0.50	3.20	0.65	0.32
CAN-46004	1988-07	226.10	48.35	109988	-0.22	0.77	0.89	0.27	109988	-0.49	3.06	0.69	0.28
NWUS-46005	1976-07	228.98	46.05	210187	-0.19	0.59	0.93	0.21	201655	-0.86	3.42	0.69	0.32
NWUS-46006	1977-07	222.52	40.80	187124	-0.14	0.57	0.94	0.20	179605	-0.60	3.23	0.70	0.29
HW-51001	1981-07	197.79	23.43	189167	-0.02	0.40	0.91	0.17	189126	-0.42	3.08	0.70	0.29
HW-51002	1984-07	202.22	17.19	174107	0.07	0.43	0.84	0.18	174122	-1.58	3.67	0.66	0.37
HW-51003	1984-07	199.18	19.22	168201	-0.01	0.37	0.87	0.17	168196	-1.17	3.46	0.68	0.33
HW-51004	1984-07	207.52	17.52	171378	0.02	0.38	0.85	0.16	171375	-1.61	3.76	0.65	0.37

Table 2.1: Correlation statistics for significant wave height and peak period between NC-GOW and buoy observations.

We compare 6 hour average data interpolated to the position and time (hourly) when the buoy records were registered, considering that the reanalysis winds are only available on a six hourly data. In this manner, the number of data for comparison is of the same order than the reference. From results given in tables 2.2, 2.3 and 2.4, the following observations are pertinent:

1. The mean values corresponding to ERA-40 and NC-GOW coincide for buoys on the Peruvian coast and Hawaii, and they are very close with respect to results in Alaska.

2. NC-GOW data preserves the quality of the correlation coefficients ($\boldsymbol{\rho}$) and scatter indexes (SI).

3. NC-GOW gives lower BIAS results with respect to compared reanalyses for all locations.

Although results obtained from different reanalyses are not directly comparable due to preprocessing, our results are consistent with respect to analogous reanalyses and instrumental data, increasing the confidence on the NC-GOW performance. In addition, validation using buoy observations confirms the quality of the hindcast data related to: i) the time series evolution, and ii) the quantile statistical distribution.

Regarding the validation of NC-GOW on a global scale using altimetry data from 1992 up to 2008, Figure 2.5 shows color plots of the mean and the 95^{th} percentile H_s for both altimetry and NC-GOW data. The storm tracks regions can be clearly identified, both in the Northern and Southern Hemisphere. Contour plots related to altimetry data present the same patterns as those associated with the NC-GOW data, which indicates that the model is properly describing

Year	**REAN**	**n**	**Mean (m)**	**BIAS (m)**	**RMSE (m)**	$\boldsymbol{\rho}$	SI_C
1988	ERA-40	1461	2.21	-0.03	0.33	0.84	0.15
	ERA-40/ODGP2			-0.30	0.41	0.87	0.13
	CS01			-0.24	0.40	0.83	0.15
	PWA-R			-0.14	0.38	0.84	0.16
	NC-GOW	1410	2.22	-0.03	0.28	0.90	0.12
1994	ERA-40	1457	2.18	-0.14	0.30	0.92	0.12
	CS01			-0.43	0.56	0.82	0.17
	PWA-R			-0.33	0.48	0.83	0.16
	NC-GOW	1412	2.18	0.13	0.32	0.89	0.14

Table 2.2: Comparison in Peruvian coast of several wave reanalysis (modified from [Caires et al., 2004a]).

Year	**REAN**	**n**	**Mean (m)**	**BIAS (m)**	**RMSE (m)**	$\boldsymbol{\rho}$	SI_C
1988	ERA-40	3399	2.20	-0.23	0.42	0.87	0.16
	ERA-40/ODGP2			-0.31	0.45	0.87	0.15
	CS01			-0.16	0.40	0.83	0.17
	PWA-R			-0.45	0.62	0.81	0.19
	NC-GOW	4013	2.20	0.04	0.37	0.86	0.17
1994	ERA-40	4570	2.55	-0.38	0.51	0.90	0.13
	CS01			-0.46	0.62	0.81	0.17
	PWA-R			-0.59	0.73	0.83	0.17
	NC-GOW	4342	2.55	0.18	0.46	0.85	0.17
1997	ERA-40	5569	2.37	-0.16	0.35	0.90	0.13
	CS01			-0.31	2.48	0.85	0.15
	PWA-R			-0.37	0.58	0.84	0.19
	NC-GOW	6820	2.34	-0.03	0.39	0.86	0.17

Table 2.3: Comparison in Hawaii of several wave reanalysis (modified from [Caires et al., 2004a]).

Year	**REAN**	**n**	**Mean (m)**	**BIAS (m)**	**RMSE (m)**	$\boldsymbol{\rho}$	SI_C
1978	ERA-40	3313	2.54	-0.24	0.38	0.90	0.29
	CS01			0.35	0.48	0.86	0.31
	NC-GOW	1134	2.49	-0.45	0.70	0.88	0.22
1988	ERA-40	4054	3.18	-0.35	0.68	0.94	0.18
	ERA-40/ODGP2			-0.14	0.57	0.93	0.18
	CS01			0.29	0.71	0.92	0.20
	PWA-R			-0.14	0.80	0.91	0.25
	NC-GOW	3139	3.07	-0.03	0.60	0.93	0.20
1994	ERA-40	3793	2.91	-0.37	0.60	0.97	0.16
	CS01			0.20	0.59	0.94	0.19
	PWA-R			-0.17	0.70	0.93	0.24
	NC-GOW	2700	2.98	-0.07	0.60	0.94	0.20
1997	ERA-40	3788	2.87	-0.21	0.50	0.96	0.16
	CS01			0.20	0.65	0.92	0.22
	PWA-R			-0.22	0.75	0.91	0.26
	NC-GOW	3684	2.81	-0.08	0.61	0.92	0.21

Table 2.4: Comparison in Alaska buoys of several wave reanalysis (modified from [Caires et al., 2004a]).

the wave climate at a global scale. The larger differences are detected on those areas of higher significant wave heights, which are associated with high latitudes on both hemispheres.

Results indicate that Northern Hemisphere presents higher variability in wave climate conditions than the Southern Hemisphere. According to [Izaguirre et al., 2011], by using satellite data, and [Caires and Sterl, 2005a], by numerical modeling, the differences between average wave variations (variance) and extreme wave heights are larger in the Northern than in the Southern Hemisphere.

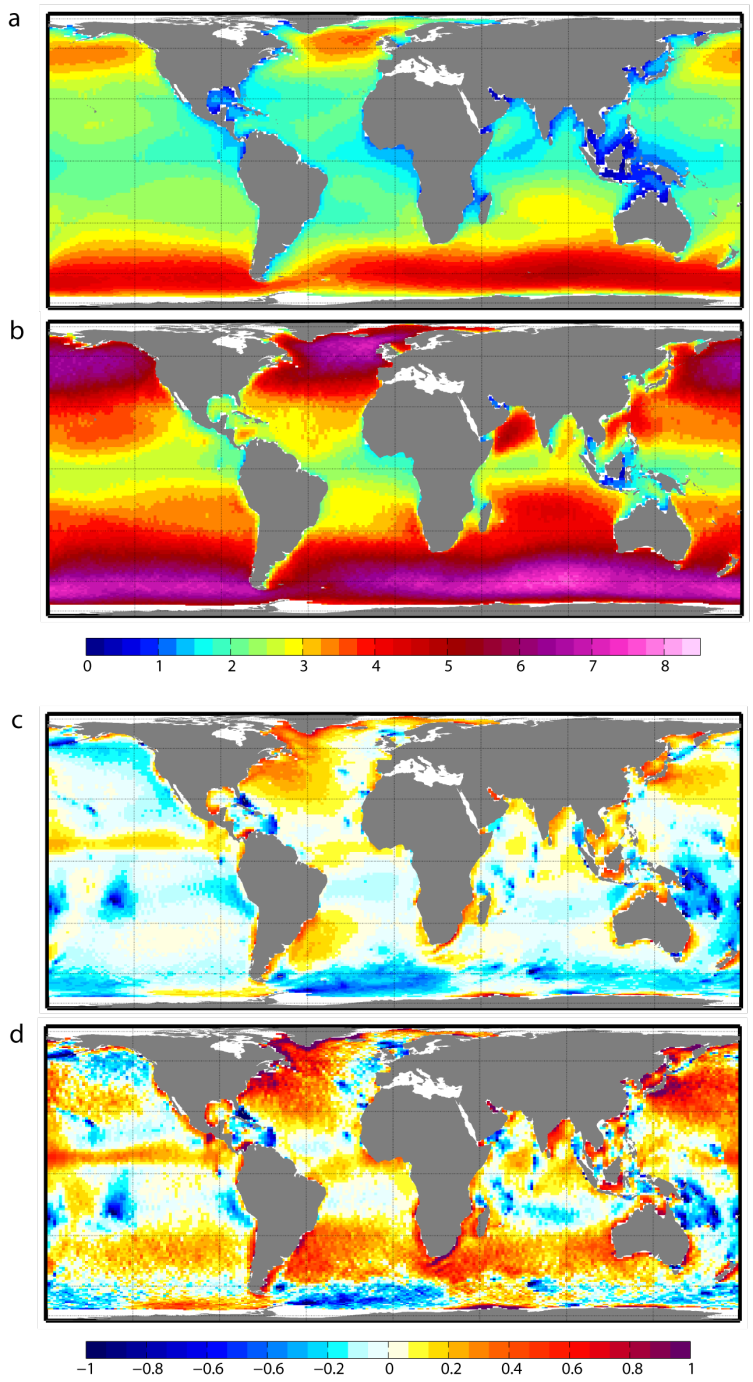

Figure 2.5: Mean and 95$^{\text{th}}$ percentile of significant wave height from NC-GOW hindcast data (panels a and b respectively) and differences with altimeter satellite data ([SAT] - [NC-GOW]) for both statistics from 1992 to 2008 (panels c and d, respectively).

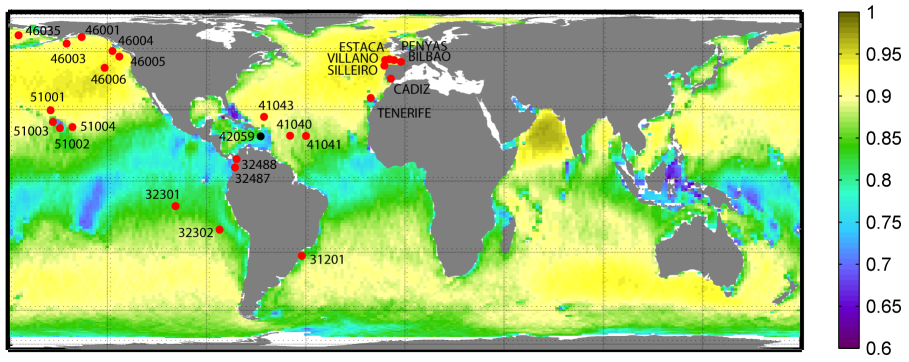

Figure 2.6: Global map of Pearson´s correlation coefficient between satellite observations and hindcast results (NC-GOW) for the full period of available altimetry observations (1992-2008). Red dots represent buoy location.

Visual inspection of Figure 2.5 allows the identification of the same spatial patterns for both mean and 95$^{\text{th}}$ percentile values, which is a qualitative measure of the goodness of the dataset. In contrast, Figure 2.6 shows the Pearson's correlation coefficient ($\boldsymbol{\rho}$) for the hourly H_s from 1992 to 2008. Note that diagnostic statistics are calculated removing outliers, i.e. data related to hurricanes and typhoons, using the method given in [Mínguez et al., 2012], which is briefly described in the next subsection. The higher and lower values for the correlation coefficient and scatter index, respectively, are obtained in areas of large mean significant wave heights (see Figure 2.5). Comparatively worse correlation and scatter index results are obtained in tropical areas, big archipelagos and semi-enclosed basins. The same conclusion is reached if different statistics, such as RMSE and Bias, are used instead (not shown due to space limitations). The discrepancies between numerical and altimetry data sets, especially in those areas where reanalysis is more limited due to the temporal and spatial resolution, justify the application of additional corrections to embed instrumental information.

2.8 The GOW wave reanalysis

2.8.1 Wave field corrections procedures

Validation results given previously show the good performance of the wave reanalysis with respect to: (i) analogous reanalysis existing in the literature, and (ii) instrumental data (buoys and altimetry). These characteristics make this reanalysis a useful design tool for offshore and coastal structures, since it offers long continuous time series and good spatial coverage for the statistical characterization of wave climate with respect to other sources of information. However, several authors (see [Caires and Sterl, 2005b], [Cavaleri and Sclavo, 2006],[Mínguez et al., 2011] or [Mínguez et al., 2012]) point out that there are still discrepancies when comparing WRD with instrumental data. These differences are mainly due to insufficient forcing resolution, and it becomes more evident under the presence of hurricanes and typhoons, which make instrumental data to appear as outliers.

In order to improve robustness of the reanalysis and configure the GOW database, a calibration procedure based on [Mínguez et al., 2011] is performed using altimetry instrumental data. Previous to calibration, an outlier filter (see [Mínguez et al., 2012]) is applied to remove instrumental data related to hurricanes and typhoons. These two procedures are briefly described in the following sections.

2.8.2 Identification and removal of outliers

The bad performance during hurricanes and typhoons, which can be observed in the scatter plot of CAR-41041 in Figure 2.2 and in the corresponding time series in Figure 2.3, is produced because the tropical cyclones are not appropriately resolved using WW3, due to the resolution of the input wind fields. Note that failing to exclude those outlier observations may provoke the distortion of any corrective action. Besides, these data should be treated and analyzed separately for the results of the correction to be fully reliable. For this reason we apply an outlier filter to eliminate these unresolved processes.

[Mínguez et al., 2012] presents different outlier detection regression techniques applied to WRD. The methods are intended for an automatic hurricane and typhoon identification. The advantage of using any of these techniques are: (i) it allows the identification and removal of H_s related to tropical storms, inappropriately reproduced by the reanalysis, (ii) it does not require the availability of a tropical storm database, and (iii) it allows the identification of areas where the influence of tropical storms are relevant, which should be further studied using appropriate models, higher temporal and spatial resolution, etc.

For this particular case, we have selected the method based on a nonlinear heteroscedastic regression model because it is robust and the parametrization is flexible to be applied on different wave climates, which is particularly important as the filtering is applied all over the grid. The model can be expressed in the form:

$$y_i = f_\mu(x_i; \boldsymbol{\beta}) + \varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{2.16}$$

where x_i corresponds to the ith predictor variable (interpolated hindcast data), and y_i is the ith value of the response variable (instrumental data). The model mean and standard deviation are parameterized as follows:

$$f_\mu(x_i, \boldsymbol{\beta}) = \beta_0 x_i^{\beta_1} \tag{2.17}$$

$$\sigma(x_i, \boldsymbol{\gamma}) = \gamma_0 x_i^{\gamma_1}, \tag{2.18}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are parameter vectors related to the model mean (2.17) and standard deviation (2.18), respectively. Note that the standard deviation heteroscedasticity is modeled through the nonlinear function (2.18), and $\varepsilon_i; i = 1, \ldots, n$ are jointly normally distributed $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma_i)$ errors.

Once data pairs (instrumental versus reanalysis) are selected, the outlier identification technique encompasses the following steps:

1. Estimate the parameters $\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}_1$ using the method of maximum likelihood.
2. Calculate the residual vector:

   $$\hat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - f_\mu(\boldsymbol{x};\ \hat{\boldsymbol{\beta}}) \ , \tag{2.19}$$

   where the tilde "$\hat{}$" refers to estimated values.
3. Obtain the residual variance-covariance matrix $\boldsymbol{\Omega}$ using a first-order Taylor series expansion of the regression model at the optimum.
4. Compute the studentized residuals as follows

   $$z_i = \frac{\hat{\varepsilon}_i}{\sqrt{\Omega_{i,i}}} = \frac{y_i - f_\mu(x_i;\ \hat{\boldsymbol{\beta}})}{\sqrt{\Omega_{i,i}}} \quad i = 1, \ldots, n, \tag{2.20}$$

   where $\Omega_{i,i}$ is the ith diagonal element of $\boldsymbol{\Omega}$.
5. Outlier identification: For a given confidence level, i.e. $\alpha = 0.0001$, any case is identified as an outlier if $|z_i| > \Phi^{-1}(1 - \alpha/2)$.

For the purpose of this study, simulation and sensitivity tests performed in [Mínguez et al., 2012], allow us to set the significance level to $\alpha = 0.0001$, for an appropriate removal of data associated with hurricanes and typhoons.

Figure 2.7 presents an example of filtering for a particular location close to the Caribbean sea, which is an area where the presence of hurricanes and tropical storms is highly frequent. The figure shows the scatter plot, empirical and fitted regression model, and information the data removed for different confidence levels. Black dots in the scatter plot indicate points detected as outliers using a significance level of $\alpha = 0.0001$. Note that there are instrumental significant wave heights above 7 m which do not exceed 3 m within the hindcast. Those points are related to high values of standardized residuals and are removed for calibration purposes. Alternation in colours indicate data located between different significance levels.

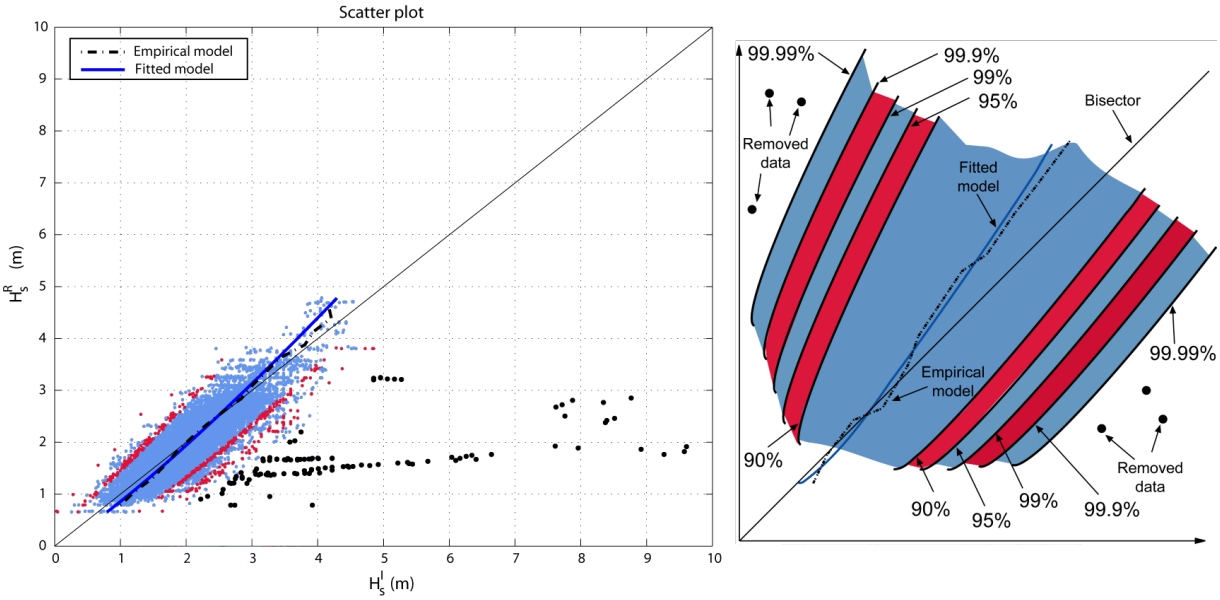

Figure 2.7: Significant wave height outlier identification scatter plot (buoy versus model) in one location near the Caribbean affected by hurricanes. Black dots represent the identified outliers and blue and red dots correspond to data at different confidence levels (see legend). The black discontinuous line depicts the empirical quantile distribution and the blue solid line the quantiles for the fitted model.

2.8.3 Wave heights calibration

Even though there has been an important improvement in numerical wave generation models, validation of results still present discrepancies with respect to instrumental data. There are several reasons, such as a bad descriptions of wind fields and insufficient forcing and model resolution ([Feng et al., 2006]). Additional factors also contribute to poor model performance on shallow waters, such as, inappropriate shallow water physics in wave models, unresolved island blocking, imperfect bathymetry, etc. (see [Cavaleri et al., 2007] for a summary). For this reason, several attempts to correct wave heights with instrumental data has been presented in the literature.

[Caires and Sterl, 2005a] propose a relation between buoy and ERA-40 wave data for the 100-year return values, based on a nonparametric method with "analogs" from a learning dataset. [Tomás et al., 2008] present a spatial calibration method based on empirical orthogonal functions. More recently, [Mínguez et al., 2011] presents a calibration procedure which depends on mean directions. Once the outliers have been identified and removed, we use the method proposed by [Mínguez et al., 2011] to embed satellite information in the GOW database.

The model can be mathematically expressed as:

$$H_s^C = a^R(\theta)\left[H_s^R\right]^{b^R(\theta)} \tag{2.21}$$

where H_s^R is the reanalysis significant wave height, H_s^C is the calibrated or corrected significant

wave height, and $a^R(\theta)$ and $b^R(\theta)$ are the parameters dependent on the mean wave direction θ from reanalysis. Note that for sea states with multiple components this correction does not consider the different directions of each component and its effect should be further explored depending on the relative importance of each energetic component. This deficiency on the calibration process is also acknowledged in [Mínguez et al., 2011].

The parameter values for all possible directions are obtained by interpolation using smoothing cubic spline functions:

$$a_i^R(\theta_i) = a_j + x_j^a(\theta_i - \theta_j) + y_j^a(\theta_i - \theta_j)^2 + z_j^a(\theta_i - \theta_j)^3, \quad (2.22)$$

$$b_i^R(\theta_i) = b_j + x_j^b(\theta_i - \theta_j) + y_j^b(\theta_i - \theta_j)^2 + z_j^b(\theta_i - \theta_j)^3, \quad (2.23)$$

where a_i^R and b_i^R are the interpolated model correction parameters for a given direction θ_i, $a_j, b_j;\ j = 1, \ldots, n_p$ are the parameters to be estimated, i.e. the parameter values associated with directions $\theta_j;\ j = 1, \ldots, n_d$, and x_j^a, y_j^a, z_j^a, x_j^b, y_j^b, $z_j^b;\ j = 1, \ldots, n_d$ are the corresponding cubic spline parameters, which are obtained using zero, first and second order continuity conditions along the circumference ($0 \leq \theta \leq 2\pi$).

Model parameters $a_j, b_j;\ j = 1, \ldots, n_p$ are estimated using the least squares method. Once these parameters are obtained it is possible to correct the complete reanalysis time series using mean wave direction records and expression (2.21). The calibration method makes the directional correction based on quantiles on a Gumbel scale, which gives more importance to the upper tail of the wave heights distribution. Note that for this reason the outlier removal is necessary, because if there exists data related to hurricanes or typhoons they will present discrepancies with instrumental data and the calibration may distort model data misguidedly all over the quantile range in the attempt to correct the discrepancies, as seen in Figure 2.8. For more details about the methodology and its hypothesis see [Mínguez et al., 2011].

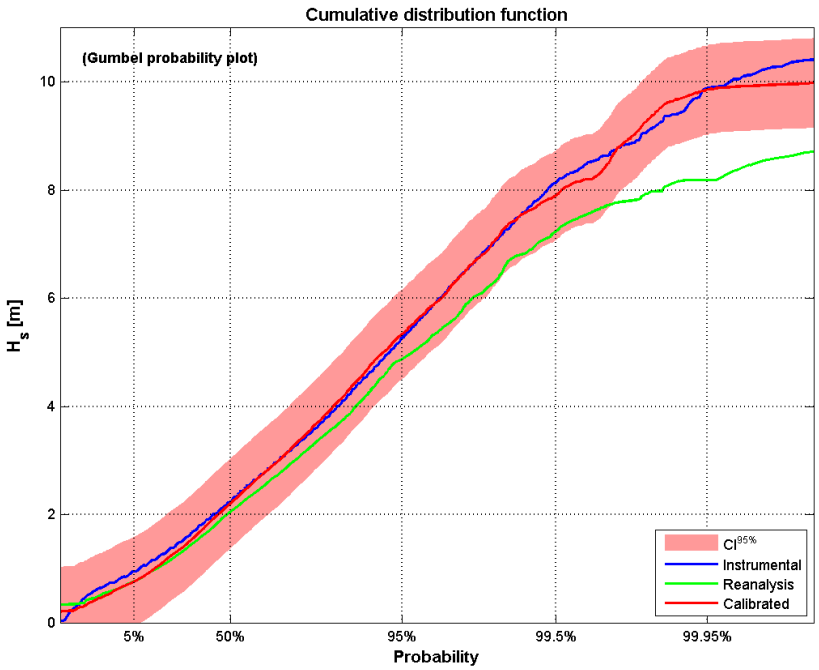

Figure 2.8: Diagnostic plot for the calibration process for one particular location. Blue, green and red lines represent, respectively, the cumulative distribution function of instrumental, model (NC-GOW) and calibrated significant wave heights (GOW). Red shadowing correspond to the 95% confidence calibrated confidence bounds.

Figure 2.8 presents the calibration results for the same example used in Figure 2.7 after outlier removal. The figure shows the cumulative distribution functions of instrumental, NCGOW and GOW data including 95% confidence bands on a Gumbel probability plot. Although the performance for buoy data previously shown did not improve significantly, the empirical cumulative distribution function (ecdf) related to calibrated data is closer to the instrumental ecdf, especially in the upper tail of the distribution. Calibrated results always present better agreement from a statistical viewpoint.

Figure 2.9 shows the relevance of outliers filter in the calibration procedure for a buoy (NOAA, 42059: 15.054° N, 67.47° W) located in the Caribbean sea. The H_s scatter and quantile-quantile plots of instrumental versus calibrated reanalysis, after (right panels: b, d) and before (left panels: a, c) calibration, are shown without no removal of outliers (upper panels: a, b) and after applying the outlier filter (lower panels: c, d). Hurricanes Dean (year 2007) and Omar (year 2008) have been marked in the upper-left panel (a). Note that for Omar hurricane, instrumental wave heights between 3 and 5 m correspond to model wave heights between 1.2 and 2.5 m, and also instrumental wave heights above 7 m for Dean do not exceed 3 m in the model. These circumstances provoke the highest four quantiles to move away from the bisector. If the calibration procedure is applied without removing those observations which are not properly resolved in the model, results given in the upper-right panel (b) are obtained. Note that, initially, the calibration process results in a lower agreement, obtaining worse diagnostic statistics with respect to reanalysis data, i.e. higher bias, root mean square error and scatter index, and lower correlation coefficient. However, if the outlier filter is applied previously to make the calibration, the scatter and quantile-quantile plots, shown in the lower-left panel (c) of Figure 2.9, are obtained, which after the calibration process transforms into results shown in the lower-right panel (d) of Figure 2.9. Diagnostic statistics after the calibration process improve, i.e. lower bias, root mean square error and scatter index, and higher correlation coefficient.

Figure 2.10 shows the number of data suspicious to be outliers for a significance level $\alpha = 0.0001$ all over the GOW grid domain. Note that the larger values are located in areas where the occurrence of hurricanes, typhoons and tropical cyclones is frequent. Reanalysis data over those locations should be used with care if high values of significant wave heights are analyzed.

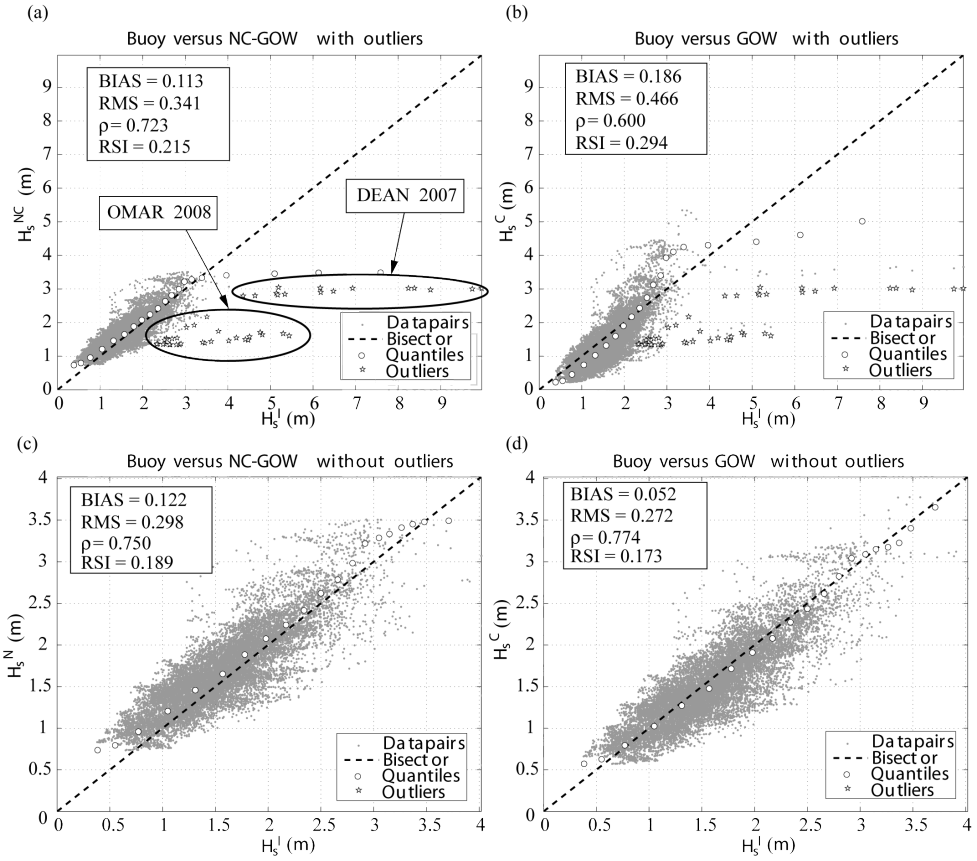

Figure 2.9: Scatter and quantile-quantile plots before (a, c) and after (b, d) the calibration process, without the outlier identification (a, b) and after the removal of the outliers (c, d). Outliers due to Dean and Omar hurricanes are outlined in panel a.

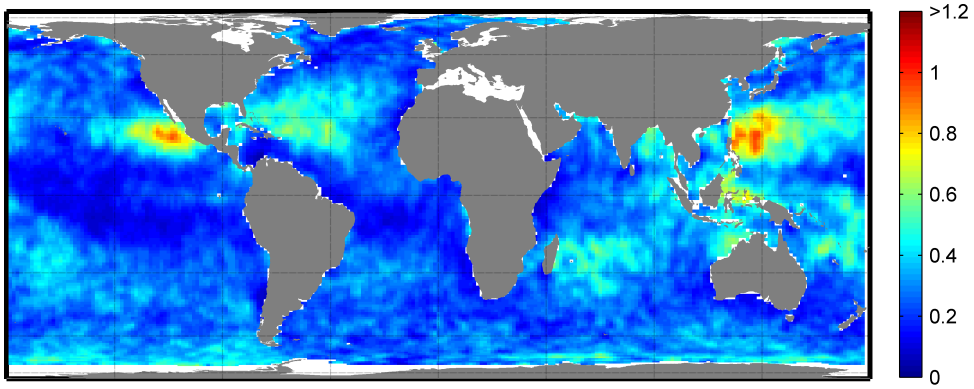

Figure 2.10: Spatial distribution of the percentage of outliers removed at each location within GOW domain, for a given confidence level $\alpha = 0.0001$.

2.8.4 The GOW wave reanalysis dataset

As mentioned in the previous section, the calibration is applied to each node of the simulation grid using the satellite data cells. In order to further compare the effect of the calibration process, we have applied the calibration method using several buoy time series records, as shown in Figure 2.11, where different quantiles before and after the calibration process are provided. Note that although reanalysis quantiles present good agreement with respect to buoy time series quantiles, the calibration process improves results, especially in the upper part of the distribution, with the exception of the extreme tail at ESTACA buoy. The calibration is applied using the estimated parameters of calibration of the closest reanalysis grid node, and using instrumental data. This result reinforces the consistency of the calibration method.

Figure 2.12 shows H_s time series related to (i) buoy (red line), (ii) NC-GOW hindcast (black line), and (iii) GOW (green line), for different buoys over different years (1987, 1995, 1997, and 2007). The corrected record maintains the concordance in the temporal fluctuations and do not imply significant changes in magnitude for the CH-32302 buoy. For the ALA-46001 station, the calibration process improves the storm peak value occurred in March 1995. It is worth noting the higher variability observed in the buoy CAR-41043 record. This effect is probably produced by the variability of winds below the 6 hour temporal resolution of the wind database. Note that despite the correction, the time series differences during October remain unsolved. Related to PENYAS time series record, the improvement of the calibration process implies a decrease of storm peaks.

The influence of the calibrated procedure for the annual mean and the 95^{th} percentile of H_s is shown in Figure 2.13. As can be seen, the GOW wave reanalysis is able to model adequately those parameters is also presented.

Comparison with buoy records does not provide the spatial verification of how the correction performs. For this reason, next section presents the verification of the calibration process, analyzing the spatial distribution of: i) correlation statistics and ii) wave parameters.

2.8.5 Verification of the calibration method

The verification of the calibration method is presented based on an analysis of the performance by using altimetry data available from 1992 to 2005 (training set) for calibration parameter estimation purposes. These estimated parameters allow comparison of calibrated times series from 2006 to 2008 with respect to altimetry observations during the same period (verification set). This verification is needed in order to compare the corrected wave heights in a global domain and not only in scattered locations.

Figure 2.14 represents the difference in the 95^{th} percentile of H_s between altimetry versus NC-GOW hindcast and altimetry versus GOW for the verification set of data. The areas where the differences before calibration are higher correspond to high latitudes, coastal and island areas. Over these areas, with differences up to 1.5 meters in wave height, the calibration process reduces the difference to less than 0.5 meters. The higher differences after corrections are found in the Southern Hemisphere, which on average presents lower discrepancies previous to calibration

with respect to altimetry data. The mean difference in the Northern Hemisphere is about -0.174 meters for the 95^{th} percentile (model underestimation), which is reduced to 0.017 meters after the calibration process. In the Southern Hemisphere, the difference decreases from -0.064 to 0.020 meters. Considering data on a global scale, mean differences change from -0.106 before calibration to 0.019 meters after. The mean significant wave height differences are lower than those obtained for the 95^{th} percentile which decrease from 0.026 to -0.008 m after calibration.

The absolute change in H_s although important may not be completely representative of the effect of the calibration because the wave conditions vary considerably in latitude (see Figure 2.5). In areas with higher waves the differences are expected to be greater, while in relative terms the effect can be not so noticeable. For this reason, Figure 2.15 focuses only on the transformation of the simulated results, before and after the calibration, in terms of difference and relative change. The critical zones, where the effect of the correction is more important, can be clearly identified: high latitudes in the Atlantic and Pacific Oceans and coastal and island areas. Note that it is in theses coastal and island areas where the correlation statistics presents the worst agreement with respect to instrumental data, and for this reason, the percentage of change is higher.

In terms of correlation statistics, Figure 2.16 shows global maps of the RMSE before and after the calibration for the verification period. Again, the areas with higher discrepancies with respect to altimetry observations can be clearly identified, as previously remarked in Figure 2.6 for the full period of the altimetry data. After calibration, the results improve considerably in the areas with deviations and remain the same in most of the domain where the reanalysis data are satisfactory. Summing up in terms of global average values, RMSE decreases from 0.554 to 0.529, SI from 0.219 to 0.206, the BIAS from -0.026 to 0.008 m and the correlation coefficient increases from 0.869 to 0.873. Globally this change is not very significant, but the improvement in coastal regions is remarkable.

From the verification analysis, we can conclude that: (1) after the calibration, the differences with altimetry observations are reduced, (2) larger wave heights are more affected by the correction; (3) critical areas that showed the worst agreement against observations are corrected with the calibration method, (4) coastal regions are considerably improved; and (5) the correction performance is supported by the improvement achieved when comparing with respect to altimetry and buoy data.

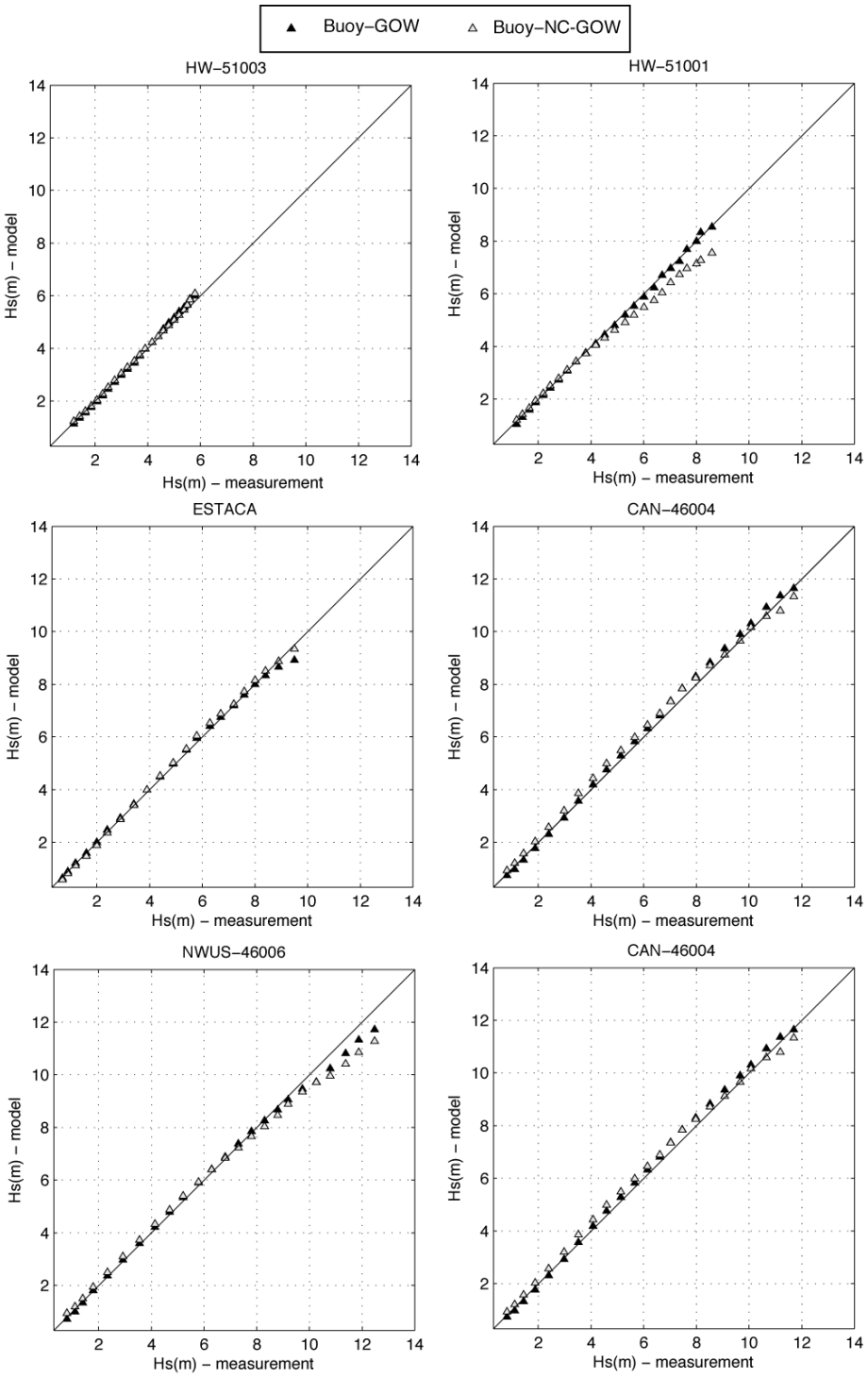

Figure 2.11: Quantile distributions before and after the calibration process at several buoys (plotted equally spaced in a Gumbel scale).

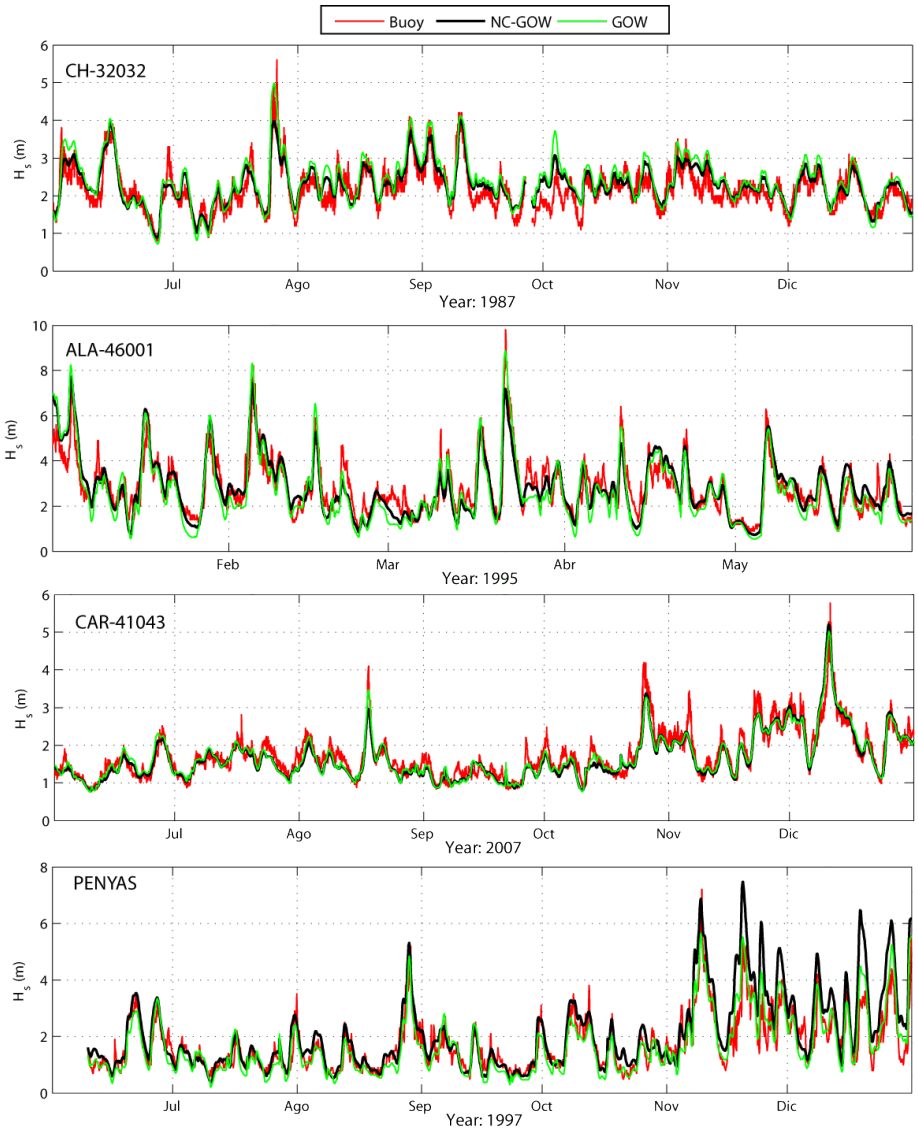

Figure 2.12: Buoy (red), NC-GOW (black) and GOW (green) significant wave height (H_s) time series at several locations.

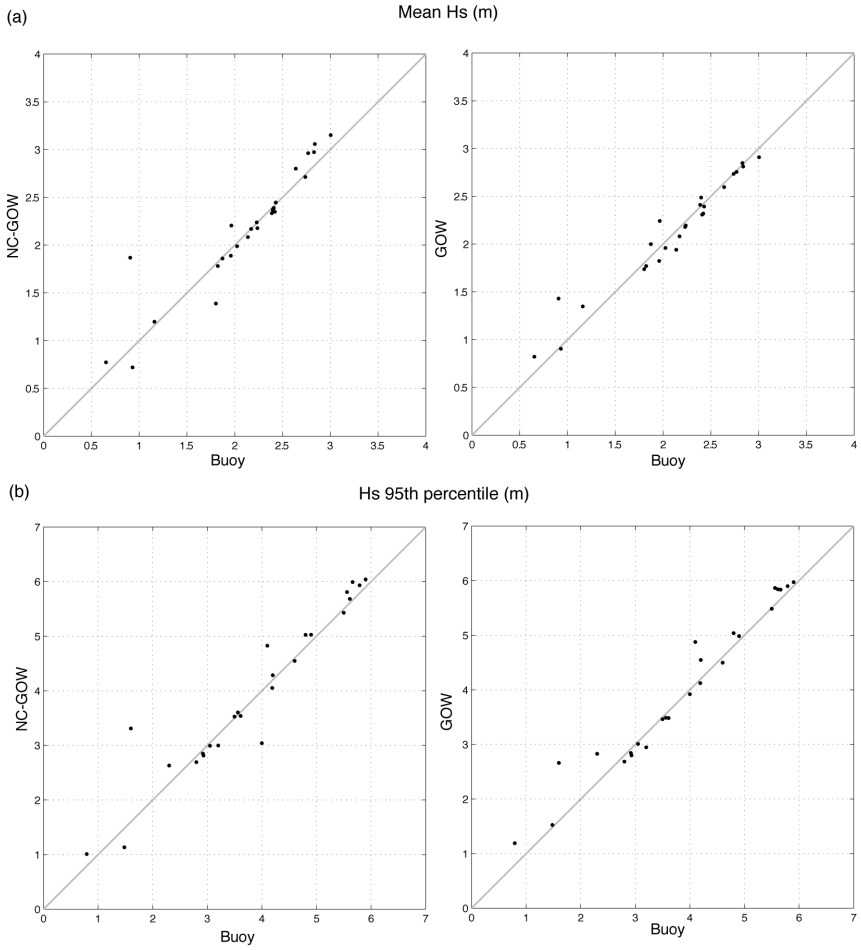

Figure 2.13: Comparison of (a) annual mean significant wave height and (b) 95^{th} percentile, between buoy data from table 2.1 with respect to NC-GOW (left panels) and GOW (right panels) data, respectively.

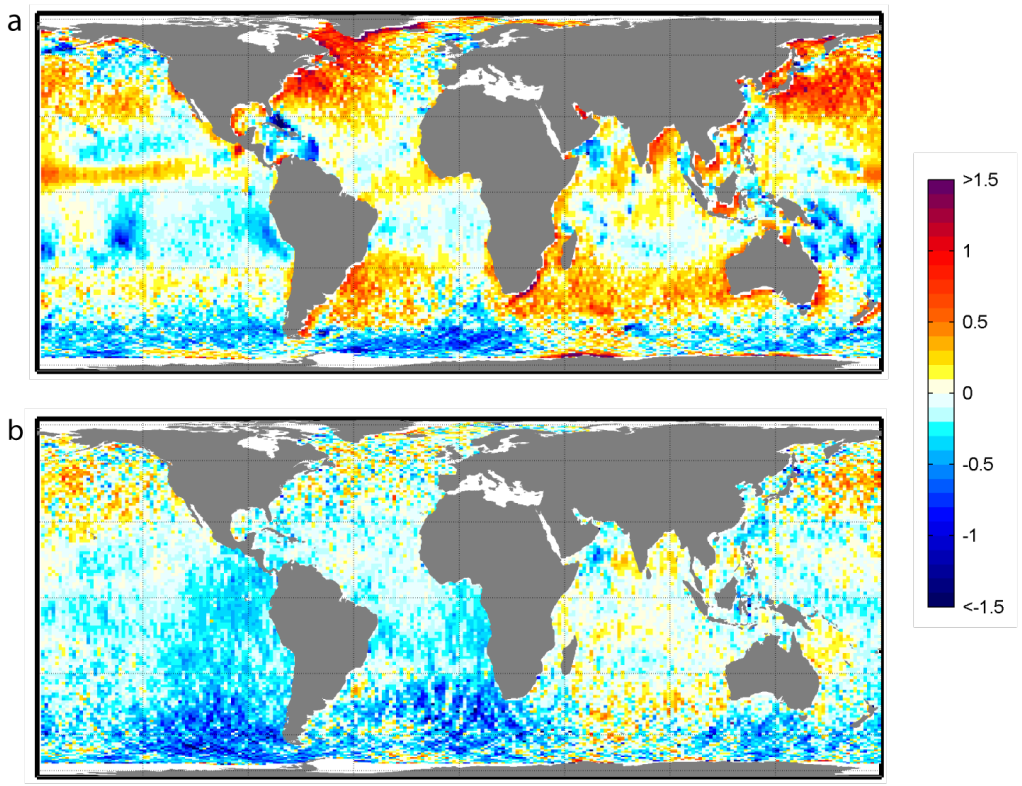

Figure 2.14: Absolute value differences of the significant wave height 95$^{\text{th}}$ percentile from satellite observations (SAT) with respect to: i) NC-GOW results (panel a; SAT - NC-GOW) and ii) GOW data (panel b; SAT - GOW), for the validation period from 2006 to 2008. Calibration is computed with a training set of altimeter data from 1992 to 2005.

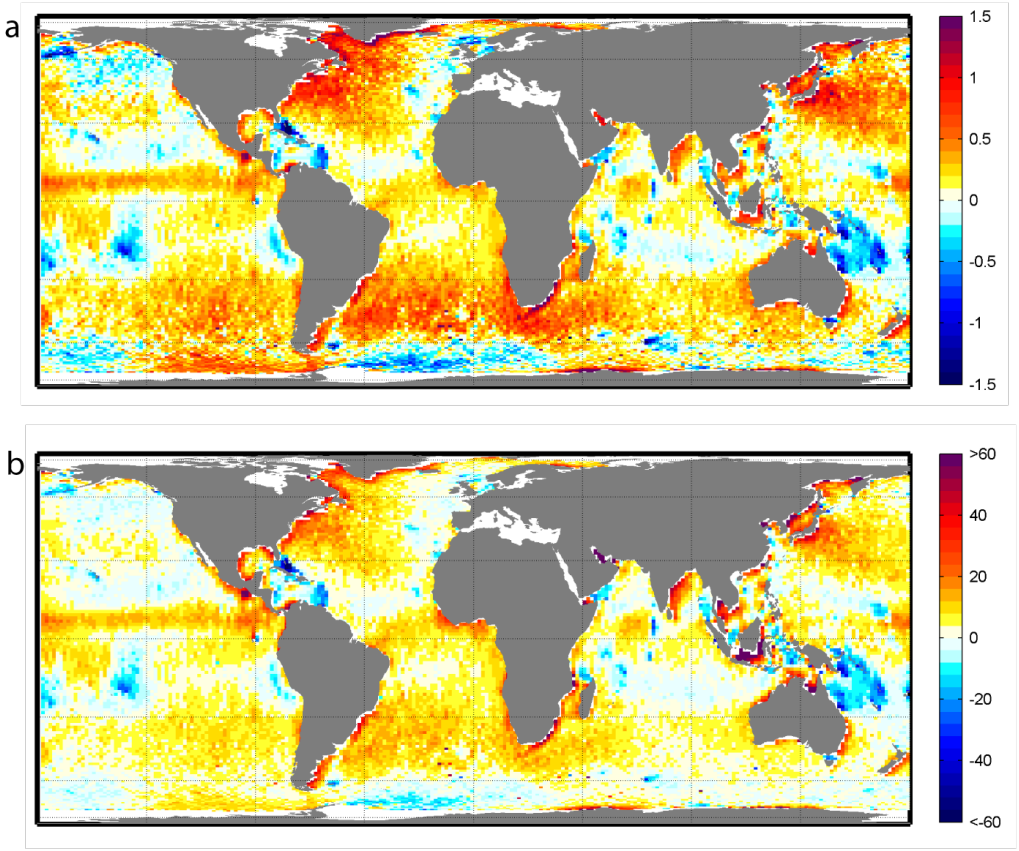

Figure 2.15: Absolute (a; GOW - NC-GOW) and relative (b; [GOW - NC-GOW] / NC-GOW) change in the 95^{th} percentile of significant wave height in the reanalysis data for the period from 2006 to 2008 after the calibration process. Calibration is computed with a training set of altimeter data from 1992 to 2005.

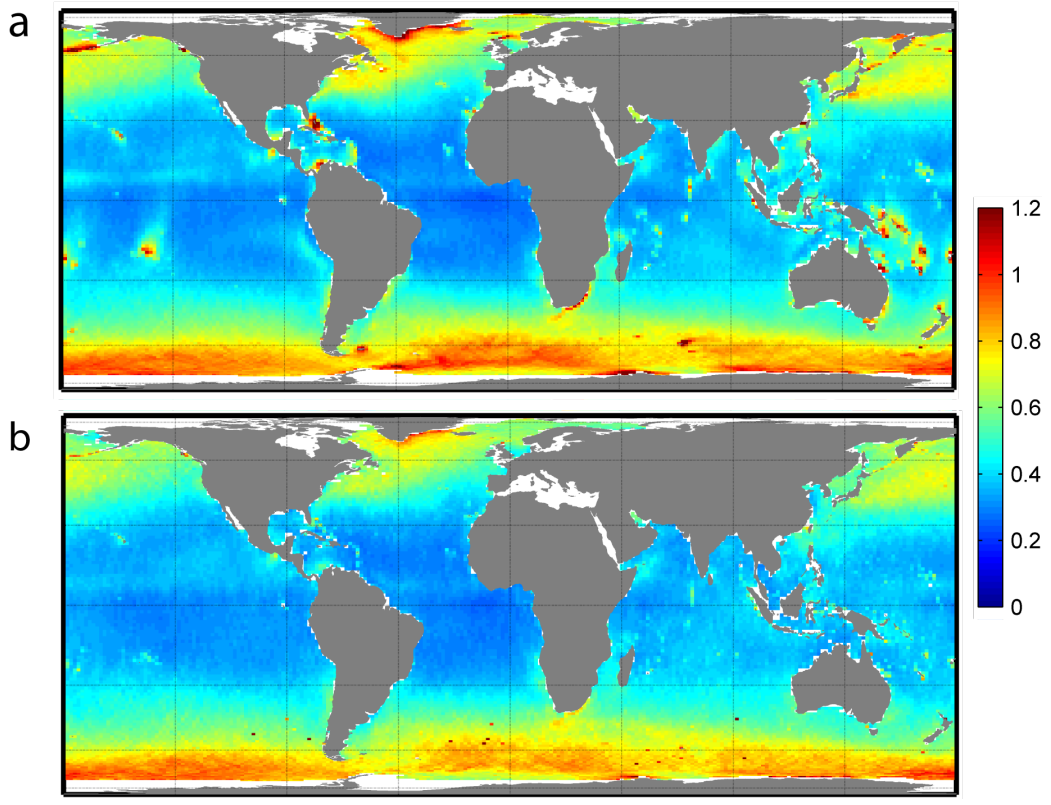

Figure 2.16: RMSE (m) of altimetry (SAT) and the reanalysis significant wave height data, for the period from 2006 to 2008: (a) before (SAT, NC-GOW) and (b) after the calibration process (SAT, GOW). Calibration is computed with a training set of altimeter data from 1992 to 2005.

2.9 Discussion about reanalysis homogeneity and stationarity of corrections

Certainly, the homogeneity of the forcing used to hindcast the historical wave climate and the satellite time span applied in the calibration needs to be discussed before assessing a climate study of the data.

The NCEP climate forecast system reanalysis is intensively analyzed in [Saha and Coauthors, 2010], especially for possible inhomogeneities. A different assimilation process was developed before 1957 and the evolution of observations may affect NCEP/NCAR data in some regions of the southern Hemisphere ([**?**, Sterl, 2004]).

[Kanamitsu et al., 2002] executed a corrected version of the original reanalysis R1 (here used), often called R2, covering only the satellite era from 1979 to the present. In spite of correcting many errors and updating some components of the system, only minor differences were found between R1 and R2 in the primary analysis variables ([Kanamitsu et al., 2002]). The most noticeable difference between Northern Hemisphere (NH) and Southern Hemisphere (SH) scores is the slope of the increase in skill over the period of 1979-2008 compared to those for the NH. With respect to the use of the reanalysis for climate studies, in spite of the constant data assimilation system used in R1, artificial changes may have been introduced by ingesting data from constantly changing observational platforms. This would thereby lead to reservations about the use of R1 in the context of climate change detection ([Chelliah and Ropelewski, 2000]).

Despite these cautions, the global spatial domain, the long-coverage, the up-to-date characteristic and the good evaluation obtained in many works reinforce the use of NCEP/NCAR. Of course, the problem for long-term trends estimation, especially in the SH, remains. However, as said in [Weisse and Von Storch, 2010], it is essential to note that for any study on long-term changes in the marine storm climate, R1 and R2 reanalyses should not be mixed but clearly distinguished. Alongside those reanalysis, other new versions like ERA-Interim ([Dee et al., 2011]) are worthy to considered for more accurate analysis.

Other concerns on the homogeneity arise when considering the homogeneity of the calibration period and the extension to the full reanalysis time span. Special caution must be taken with calibrated data out of the satellite time-span, especially for long-term trends analysis.

To check this, the satellite data (SAT) with the interpolated GOW results for the same time and locations (GOWi) have been compared. The interpolated data do not coincide with the GOW hourly time series due to a lack of temporal resolution in the satellite measurements. The data is corrected in latitude (multiplying the cosine of the latitude of each point) and averaged globally and at each hemisphere. There is an important bias within the results in [Sterl and Caires, 2005], probably because the computation or treatment of the global average differs in some way. Figure 2.17 represents the global H_s time series from interpolated reanalysis and satellite data, while Figure 2.18 does the same but for each hemisphere. Checking with the hourly global time series (rather than the GOWi), the signal is very close to the one shown. The differences are negligible in any case. The data for the year 1992 have not been used because it only covers the last months of the year and they are not representative of the whole year.

The differences in time between both time series vary from positive to negative in the time span considered. It may partly attributed to the number and varying quality of altimeter data with time.

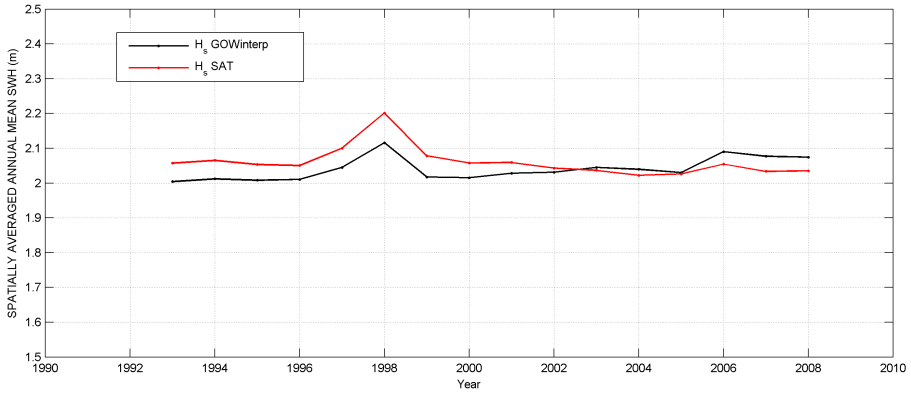

Figure 2.17: Global mean significant wave height time series for GOWi (black) and altimeter data (red), for the available satellite time span (1993-2008).

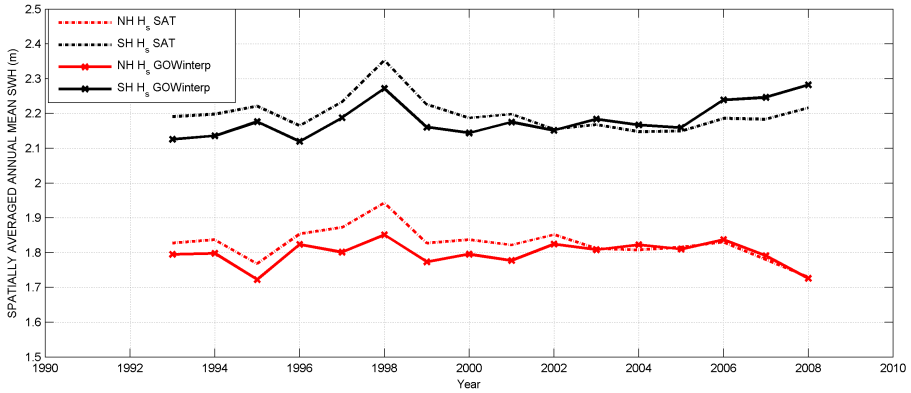

Figure 2.18: Global mean significant wave height time series in the Northern Hemisphere (red) and Southern Hemisphere (black) for GOWi (points) and altimeter data (solid line), for the available satellite time span (1993-2008).

Only the GOWi data present a significant trend in the Southern Hemisphere of 0.625 cm/yr and globally (induced by the relative weight of the water surface in the SH) of 0.39 cm/yr. The rest of the trends are non-significant. This is consistent with the lack of homogeneity of the wind reanalysis in the SH previously commented on.

To further investigate this matter and the stationarity of the correction,the bias has been computed for every year to check if any significant trend shall appear. As can be seen, there is a statistically significant trend of 6 cm/yr in the southern Pacific at high latitudes (both at non-calibrated and calibrated results). This can also be seen in the standard deviation of the bias from 1993 to 2008.

To further investigate this issue, trends over the BIAS in the available altimeter period have been computed. Results are represented in Figures 2.19 and 2.20, for the calibrated and the non-calibrated data respectively. These results show a diagnostic plot of the quality of the results in terms of homogeneity of the wind reanalysis. The stationarity of the correction for most of the globe except for high latitudes of the south-east Pacific may be assured. The standard deviation of the BIAS (Figure 2.21) in the same period further confirms this conclusion.

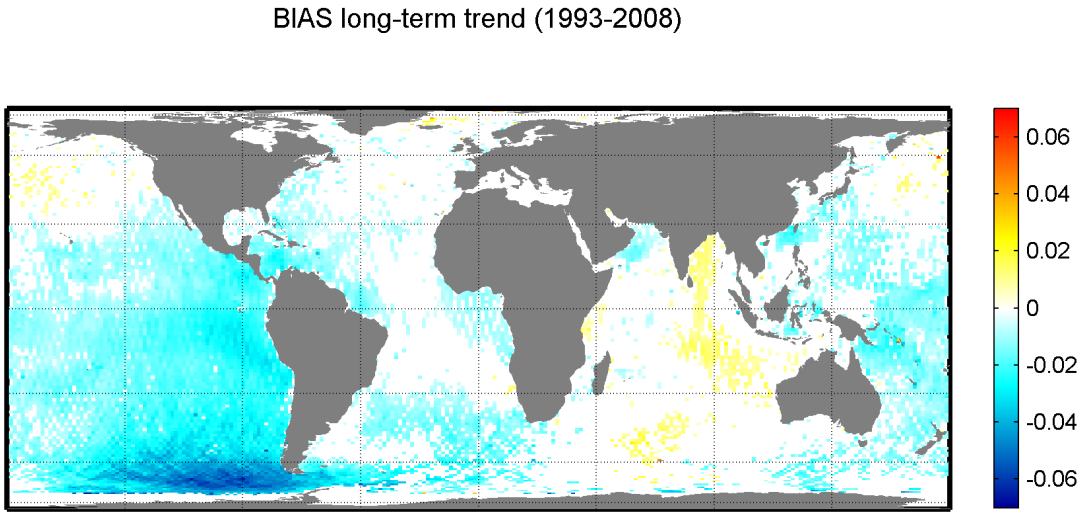

Figure 2.19: Long-term trend of the BIAS between the altimeter data and the GOWi data for the available satellite time span (1993-2008).

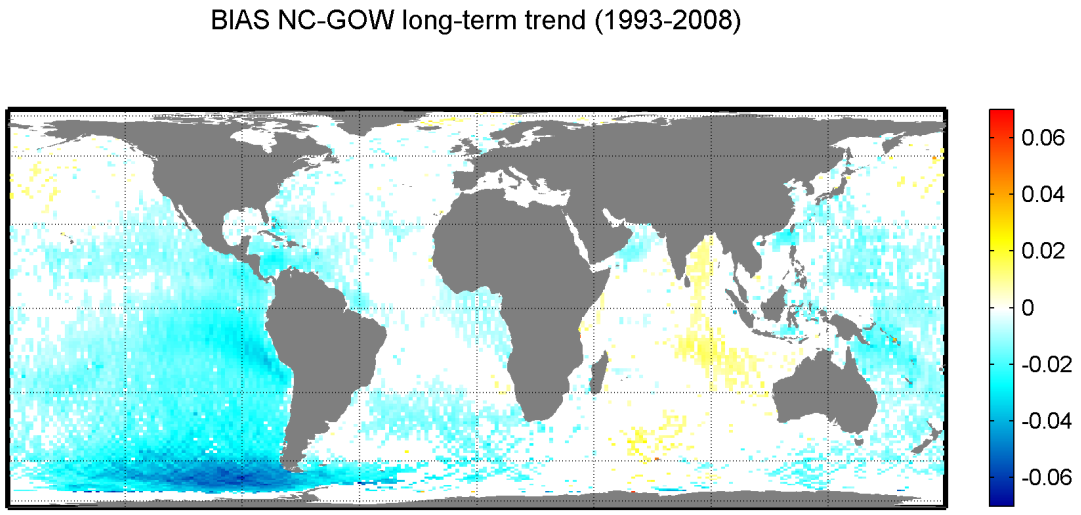

Figure 2.20: Long-term trend of the BIAS between the altimeter data and the NC-GOWi data for the available satellite time span (1993-2008).

2.10 Conclusions

A global wave dataset simulated with the model Wavewatch III and driven by the NCEP/NCAR reanalysis winds and ice fields, with a planetary scope and resolution of 1.5° longitude x 1.0° latitude which covers the period from 1948 to 2008, and aim to be periodically updated has been presented.

Based on the application of a calibration method the dataset has been corrected using altimetry data from the period from 1992 to 2008. The outliers due to tropical cyclones are not appropriately reproduced in the simulation process, due to lack of resolution in the forcing wind fields. For that reason, these data have been identified and removed from the analysis. The quality of the results and the corrections applied have been compared with buoy and satellite altimetry measurements. The results show a satisfactory transformation in the high quantiles distribution when necessary and no changes in areas where the initial simulated data present good agreement with respect to observations.

Additionally, a verification of the calibration method has been performed, obtaining a correction based on the altimetry data from 1992 to 2005 and judging the effect with the remaining observations. A regionally varying correction is confirmed, especially remarkable for high wave heights range and coastal regions.

After incorporating the altimeter data through the calibration process, an exhaustive validation of the results has been performed with altimeter and buoy measurements. The diagnostic statistics shows a fine agreement both in the scatter data and in the statistical distribution of the wave heights indicating that the reanalysis appropriately reflects the wave characteristics identified by the satellites from 1992 to 2008.

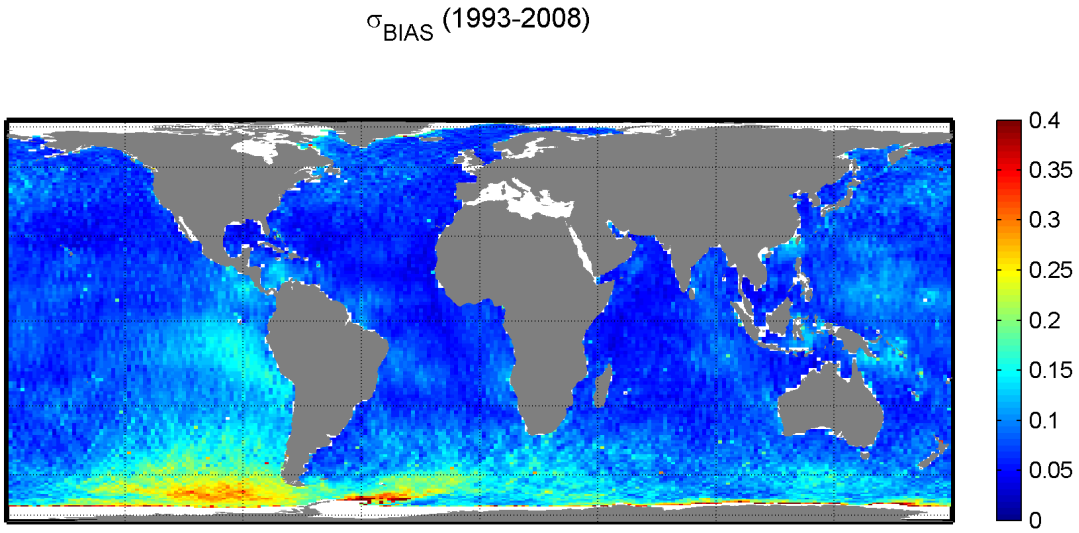

Figure 2.21: Standard deviation of BIAS between the altimeter data and the GOWi data for the available satellite time span (1993-2008)

The spatial and temporal coverage (1948 onwards) of the dataset and the results obtained in the statistical distribution for the full range of wave heights, make the GOW database a long-term and consistent reanalysis, suitable for global applications in ocean wave climate and long-term trends analysis as well as for coastal engineering purposes.