Alma Mater Studiorum Universitàdi Bologna

DOTTORATO DI RICERCA IN IL FUTURO DELLA TERRA, CAMBIAMENTI CLIMATICI E SFIDE SOCIALI

Cicolo 35

Settore Concorsuale: 04/A4 - GEOFISICA

Settore Scientifico Disciplinare: GEO/12 - OCEANOGRAFIA E FISICA DELL'ATMOSFERA

TOWARD AN UNDERSTANDING OF THE ATMOSPHERIC FORCING UNCERTAINTIES IN THE MEDITERRANEAN SEA

Presentata da: Mahmud Hasan Ghani

Coordinatore Dottorato

Silvana Di Sabatino

Supervisore

Professor Nadia Pinardi

Co-supervisors

Dr. Francesco Trotta

Dr Giovanni Liguori

Dr Silvia Bianconcini

Esame finale anno 2023

Abstract

The study of the probability distribution of atmospheric surface variables is important for the understanding of uncertainty in numerical ocean forecasting systems due to the direct forcing that the atmosphere exerts on the ocean variability. The thesis has extensively investigated for the first time the statistical distributions of atmospheric surface variables and heat fluxes for the Mediterranean Sea. After retrieving a 30-year atmospheric analysis dataset, we have captured the spatial patterns of the probability distribution of the relevant atmospheric variables for ocean atmospheric forcing: wind components (U, V), wind amplitude, air temperature (T2M), dewpoint temperature (D2M) and mean sea-level pressure (MSL-P). The study reveals that a two-parameter PDF is not a good fit for T2M, D2M, MSL-P and wind components (U,V) and a three parameter skew-normal PDF is better suited. Such distribution captures properly the data asymmetric tails (skewness). After removing the large seasonal cycle, we show the quality of the fit and the geographic structure of the PDF parameters. It is found that the PDF parameters vary between different regions, in particular the shape (connected to the asymmetric tails) and the scale (connected to the spread of the distribution) parameters cluster around two or more values, probably connected to the different dynamics that produces the surface atmospheric fields in the Mediterranean basin. Moreover, using the atmospheric variables, we have computed the air-sea heat fluxes for a 20-years period and estimated the net heat budget over the Mediterranean Sea. Interestingly, the higher resolution analysis dataset provides a negative heat budget of – 3 W/m2 which is within the acceptable range for the Mediterranean Sea heat budget closure which requires the entering heat from Gibraltar to be balanced by a net heat loss at the air-sea interface. The lower resolution atmospheric reanalysis dataset (ERA5) does not satisfy the heat budget closure problem pointing out that a minimal resolution of the atmospheric forcing is crucial for the Mediterranean Sea dynamics. The PDF framework developed in this thesis will be the basis for a future ensemble forecasting system that will use the statistical distributions to create perturbations of the atmospheric ocean forcing.

Contents

List of Figures

List of Tables

Chapter 1

Background

1 Introduction

This thesis is the first comprehensive study of the statistical distribution of surface atmospheric variables and air-sea fluxes in the Mediterranean Sea. The ocean is a mechanical engine that derives its energy from atmospheric forcing variables that compose momentum, heat and water fluxes at the air-sea interface. The study of the distribution of atmospheric variables is then an essential step in understanding the ocean circulation response to these forcings.

To characterize the air-sea fluxes we need to consider several atmospheric surface variables: surface wind components, air and dew point temperature, and mean sea level pressure among others. This thesis offers for the first time the analysis of all of them and tries to map the differences in the PDFs in different regions of the Mediterranean Sea.

To study the climatological probability of atmospheric surface variables and the associated heat fluxes, we use long-term archives of analyses and reanalyses that are the best estimates of the state of the atmosphere deriving from the optimal blending of observations and model outputs. The data sets are available for about the past 30 years (1991-2020) from ECMWF archives and this time series length is considered to be sufficient in computation climatological standard normals. In accordance with WMO Guidelines on the Calculation of Climate Normals (Organization, 2017), 30 years is recommended as standard averaging period, but that period is likely to be low in observing more extreme statistics.

This work is preliminary to the study of optimal perturbations of atmospheric forcing for the ocean general circulation of the Mediterranean Sea and potentially to develop an ensemble prediction system based upon them. Nowadays, weather forecasting systems are not only deterministic with a day-to-day evolution of the weather patterns, but they also present probability distributions of extreme events and long-term statistics using ensemble methods. A reliable statistical interpretation of the atmospheric forcing variables can provide a baseline to perturb the atmospheric forcing in ocean forecasting and analyse the uncertainty in the ocean response, thus providing the basis for probabilistic forecasting in the ocean. The feedback of those atmospheric forces may vary at certain degrees and time scales due to the no-linear state of the atmosphere (Jayaraman et al., 2010). Due to the unpredictable nature of atmospheric variability and numerical models are systematically biased and produce errors that

can incline the forecast errors in time scale and affect the quality of the forecast (Monahan et al., 2011). Therefore, it's a significant theoretical approach to analyse the atmospheric forcing variables. In the arenas of climate and numerical weather or ocean predictability studies, a probability distribution drawn from a long-term observation of an atmospheric variable can support researchers and scientists to generate more reliable forecasts.

1.1 Probability distributions of atmospheric variables

We know that the atmosphere is characterized by a non-linear dynamical state, where it is common to apply a Probability Distribution Function (PDF) analysis to extract basic information about its climatological state and its variability. Furthermore, it is increasingly evident that the probability distributions of large-scale atmospheric and oceanic variables are non-Gaussian (Chu, 2009; Gille, 2005; Kyselỳ, 2002; Moccia et al., 2021; Monahan et al., 2011; Pishgar-Komleh et al., 2015; Tye et al., 2014; Yiou et al., 2008). In other words, PDFs are characterized by asymmetric and fat tails (Sardeshmukh et al., 2015). In general, the atmospheric variables are assumed to be normally distributed in the scientific community (Perron and Sura, 2013). But there is a probability to observe extreme values in the distribution which is usually scarce.

Additionally, atmospheric variables are characterized by a dominant periodic seasonal cycle and the PDF analysis is done using seasonally adjusted. Thus, atmospheric variable PDFs are likely to be fully characterized by 3 parameters, the shape, location and scale, and also defined with positive and negative values of the random variable of interest.

In this thesis, we choose a skew-normal PDF to model the atmospheric variables with skewness and mild non-Gaussian kurtosis. There are not many references for the application of skew-normal PDF in the atmospheric or ocean fields, but this PDF has many characteristics that are considered desirable for our investigation. One of them is that this class of PDF includes the normal one in a natural way, it is mathematically tractable and allows us to consider situations characterised by different levels of skewness and kurtosis (Azzalini, 1985). Furthermore, we use also an Asymmetric Laplace Distribution (Yu and Zhang, 2005) which is better suited to fit distributions of air-sea fluxes, as already noted in the literature (Gulev and Belyaev, 2012).

According to Sardeshmukh and Penland (2015), there are three ways to deal with probability distributions for atmospheric variables ; we can apply a PDF function based on generalized extreme theory (GEV) to sample the extreme values, a second approach can be a probability distribution fit to all observations values, not only the extreme values, with parameters related to mean and variance; and under a third approach, a numerical climate model dataset is used to estimate the tail probabilities using histograms. In this study, we are going to apply the third approach to the time series data derived from a large ensemble of atmospheric model integrations with a possible compromise for unknown errors in model data. A skew-normal PDF can be an alternate PDF to model the atmospheric variables. The most important feature of this skew-normal PDF is that it can cover the normal density and asymmetry in the tails for a distribution. We will discuss and explain the skew-normal PDF in chapter 2. However, there are not many references for the application of skew-normal PDF in the atmospheric or ocean sciences

A considerable number of studies are available for the probability distribution of surface wind amplitude over land and ocean due to its importance for renewable energy converters. In most cases, the wind amplitude distribution has been found to be reasonably well approximated by the Weibull distribution that depends only on two parameters, namely shape and scale parameters (Ali et al., 2018; Monahan, 2006). However, the Weibull random variable is defined on a positive support. Air temperature is another crucial atmospheric variable for climate studies. Yiou et al. (2008) applied the generalized Pareto distribution (GPD) to analyze the extreme value statistics for summer temperature and precipitation. However, we are not aware of general PDF studies of temperature anomalies considering the full climatological spectrum of time scales in the data sets.

1.2 Air Sea dynamics

The ocean interacts with the atmosphere at the air-sea boundary by exchanging heat, momentum, gases and particles that have an influential role in the regional and global climate system (Myhre et al., 2013). The ocean receives heat at the surface from the incoming solar radiation, and seawater evaporates, then condensed moistures return to ocean by releasing heat. With the processes of moistening and warming, the air-sea dynamics provides the net amount of heat received and released by the ocean. The

ocean feels wind stress over the surface and regulates the exchange of heat and energy between the air and sea. Then the heat and momentum fluxes influence in the driving process of surface currents and deep-water circulation (Gille, 2005).

Numerous studies are available that describe air-sea physics using remote sensing, in situ observational data and numerical models. The atmospheric variables composing the air-sea fluxes are the surface winds, surface air temperature, specific humidity at sea surface level levels, cloud cover, surface precipitation and sea-level pressure.

In principle, the ocean is forced by the heat fluxes Q, momentum fluxes τ and fresh-water fluxes F. Even a fraction of the ocean surface is covered by sea-ice, but ignoring the sea-ice part in respect to present study, the general net heat balance is expressed by (Large and Yeager, 2009)-

$$Q_{as} = Q_S + Q_L + Q_E + Q_H \tag{1.1}$$

Where, short wave or solar radiation Q_S are long wave Q_L are the radiative part and, turbulent heat flux component includes wind stress τ, sensible heat Q_H and latent heat Q_E.

Here we present the widely used bulk formula in computing net heat flux budget. In concern to our study area, the Mediterranean Sea, adopted formulation for heat flux components have been used by the previous studies for several decades and those formulation will be described in detail in the chapter 3.

The Mediterranean Sea is a semi-enclosed sea that receives water and heat from the Gibraltar Strait (the Dardanelles volume and heat flux is important only locally). If we assume that the heat averaged over the entire volume of the Mediterranean Sea is in a steady state, we have to assume a balance between heat gained at Gibraltar and the heat lost at the basin surface. Thus, a heat budget closure hypothesis was formulated for the Mediterranean Sea that requires a long-term negative net air-sea flux average over the basin (Béthoux et al., 1998; Castellari et al., 1998). Due to various range of net heat fluxes from the literature, it's necessary to verify the heat budget first and, we are using a high-resolution atmospheric analysis dataset for the first time in the Mediterranean. Moreover, the pdf at each grid point will be sampled to find a realistic perturbation for the specific instantaneous atmospheric forcing variable forcing the ocean predictive model. Then, the resulting perturbed heat fluxes from the ocean

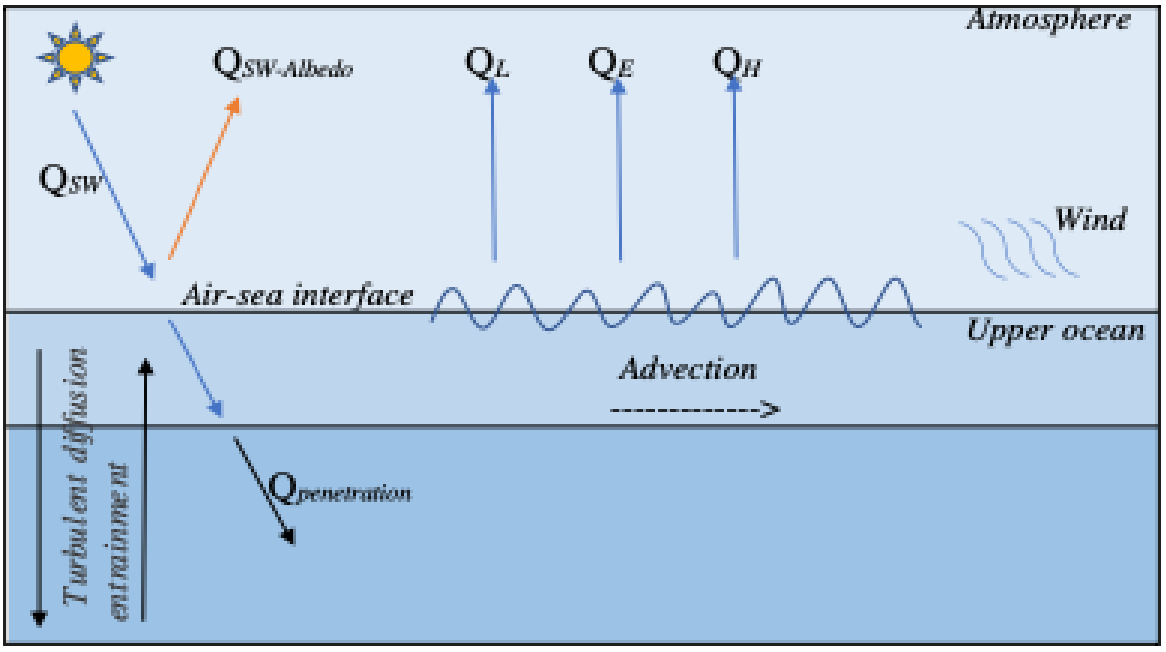

FIGURE 1.1: A schematic view of the air-sea flux dynamics over the ocean (adapted from Marullo et al., 2021)

forecast model will be compared with the estimated heat flux pdf to ascertain their realistic values with respect to the heat flux pdf. We will use this heat budget closure hypothesis to check the consistency of our flux computations and distributions.

1.3 Atmospheric forcing in numerical ocean forecasting system

We are thankful for the tremendous development of conceptual core and coupled ocean–atmosphere numerical models that provides us valuable forecasts in everyday life. The advent of faster computers has been key in processing and analyzing data, and in the development and application of more realistic atmospheric and oceanic models to gain further insight into the nature of the atmosphere and hydrosphere, as well as their interactions. The development of the ocean forecasting system started in the decade of 1940s, and it was initiated from meteorology because it facilitated the ocean prediction systems from long operational forecasting experiences (Pinardi et al., 2017). From the early 1960s, attempts have been started to identify the multi-scale forcing mechanisms in the numerical forecasting systems that can explain the observed characteristics of atmospheric and oceanic variables (Amador et al., 2006).

Based on the Lorenz's seminal work on chaos theory from the early 1960s, probabilistic approaches to prediction have come to dominate the science of weather and climate

forecasting. Lorenz (1963) introduced the concept of prediction errors arising from inadequate knowledge of the forecast initial conditions. Whatever the time frame, any small error that exists in the forecast initial conditions can become large even at the ten days time scale (Toth and Kalnay, 1997). The study of the error growth in an atmospheric forecast system is a central objective for present day weather forecasting. It is now accepted that probabilistic forecasting, based on ensemble methods, can better fit the need for accurate forecasts giving an estimate also of the associated uncertainty. It's a key objective for a numerical forecasting system to provide a reliable forecast. In relation to the perturbation, we are interested on the "first guess" atmospheric analysis that model takes as atmospheric forcing input, which is given based on the statistical analysis results.

Starting from this point, Milliff et al., 2011; Pinardi et al., 2008 developed an ocean ensemble forecast methodology based on perturbations of both initial conditions and atmospheric forcing. According to them, the atmospheric forcing uncertainties are as large as initial condition errors in determining the spread of the forecast. Then, perturbations of atmospheric variables to produce ocean ensemble forecasting were tried in the past by Pinardi et al. (2011) and Lima et al. (2019). The first used a Bayesian Hierarchical Model (BHM) to determine the perturbations in surface winds while the second used ad-hoc perturbations in bathymetry and surface atmospheric variables. The BHM approach cannot be easily generalized to other atmospheric variables such as air temperature or humidity while the second is ad-hoc. Furthermore, Pinardi et al. (2011) shows that the winds coming from the ECMWF ensemble prediction system did not generate the appropriate perturbation in ocean variables. Pinardi et al. (2011) conclude that perturbed winds in the ECMWF ensemble prediction system do not sample properly the uncertainty of atmospheric variables at the surface, being a method developed for the 500 hPa height. In order to progress toward the optimal atmospheric perturbations for ocean ensemble forecasting, the estimation of the basic atmospheric variable pdf is a necessary step, and, in the future, we will develop the pdf sampling methods that will allow to extract realistic perturbations of the surface atmospheric variables.

1.4 Structure of the thesis and objectives

This thesis tries to advance our understanding and estimation of the probability density functions associated with atmospheric surface forcing and associated heat fluxes. The first part of the work is to establish the basic statistics and the pdf for the main five atmospheric surface variables of interest, trying to justify the choice of specific fitting function used. In a second moment the experience gained for the atmospheric variables is used for the air-sea heat fluxes thus discovering and explaining that for some of the fluxes there are specific pdf to be used (skew-normal versus asymmetric Laplace distribution). The last section tries to understand the importance of extreme events in the distributions for the negative net heat budget of the Mediterranean Sea

In short, in Chapter 2 we analyse the PDFs of selected atmospheric surface variables of interest and qualitatively validate the PDF fit results. We then discuss the statistical moments and in particular the skewness of the different state variables and the spatial variability of the PDF parameters. This is interesting because the PDF parameters are connected to the skewness and spread of the distributions that differ across the Mediterranean Sea regions.

Furthermore, in Section 3 we diagnose the heat fluxes and compute the net heat flux to close the Mediterranean Sea heat budget hypothesis. The analysis shows that some of the flux components have distinctive PDF distributions. In chapter 4, we will present an assessment of the computed PDF parameters for atmospheric variables and heat fluxes, and how they can be significant in evaluation of average distribution values and probable extreme ranges for the atmospheric variables and heat fluxes.

Section 5 concludes and provides an outlook for future works.

Chapter 2

The statistical analysis of atmospheric variables in the Mediterranean Sea

1 Introduction

The probability distribution of any continuous variable of natural phenomena provides information on the likelihood of occurrence of the associated events. A time series of an atmospheric variable can help in estimating the uncertainties associated with an ocean forecasting system forced by such atmospheric variable (Pinardi et al., 2011). An ocean forecasting system, similar to a numerical weather prediction system (Pinardi et al., 2017) is 'forced' by atmospheric fluxes of water, momentum and heat that are derived from atmospheric state variables. Such fluxes will be studied in the 2nd Chapter of this thesis.

The ocean state is dynamic and is continuously forced by the complex air-sea interactions at the air-sea interface. Surface wind stress plays a key role in driving ocean currents and is a major source of their energy. Heat and water fluxes, together with the wind stress, generate the specific ocean stratification, which in turn is a very important characteristic of the ocean circulation. Short term predictions deal with uncertainties in the local space-time uncertainties of the atmospheric forcing. In the long term, the ocean climate is determined by how much mechanical energy is extracted from the winds and how much water/heat is extracted at the air-sea interface building from single sporadic events in space/time. Air-sea momentum, heat and water fluxes are composed by near surface atmospheric state variables distributions; thus we are interested in understanding the probability distribution of these variables.

There is a good amount of literature on statistical analysis and Probability Distribution Functions (PDF) applied to surface atmospheric winds. From the literature, several PDFs have been identified particularly for the surface wind speed. The focus of these studies is mainly on the energy potential related to the offshore wind industry (Ali et al., 2018; Kang et al., 2018; Morgan et al., 2011; Pishgar-Komleh et al., 2015). Among the commonly applied PDFs, the Weibull is a very popular one introduced by Weibull in 1950): it has a wide application in industrial fields, but over the years it has also been used in the natural sciences. In connection with the atmospheric variables, the Weibull PDF is the most common PDF used for surface wind speed (Ali et al., 2018; Mudholkar and Srivastava, 1993; Pishgar-Komleh et al., 2015). Monahan (2006) analyzed the probability distribution of sea surface wind amplitude obtained from scatterometer observations on a global scale and showed a good fit using the Weibull PDF. In a continuous distribution, there is a probability to observe some

extreme values which might be rare many cases. Extreme value theory for a distribution is a useful statistical tool to quantify the statistical properties of extreme events (Yiou et al., 2008). Gumbel Gumbel (1958) introduced the idea of Extreme values theory so the tail of the distributions can be parametrized containing information on the extremes. Based on the Extreme Values Theory (EVT), a few studies have analyzed extreme air temperatures, precipitation and rainfall data from reanalysis and observations to extract the statistics of extreme values on land (Kyselỳ, 2002; Moccia et al., 2021; Yiou et al., 2008).

In this thesis, we will seek to characterize the wind components (U, V) in detail in addition to wind speed. In recent times, a few studies have also started to analyze the statistical distribution of oceanographic currents and relate it to meteorological PDFs. Chu (2008, 2009)) analyzed the surface current amplitude using a Weibull PDF and related it to the gaussian distributions of the wind components. This is probably the first paper that differentiates between the distribution of the amplitude and that of the ocean current components.

The Weibull PDF for a continuous random variable is defined on a positive support, so it is not appropriate for wind components that have both positive and negative values. To fit a PDF to data in the $[-\infty, \infty]$ range we need to use a three-parameter PDF as opposed to the two-parameter PDF such as Weibull. For positive and negative defined variables, a skew-normal PDF (Marchenko and Genton, 2010 could be used, and this is the choice that we have taken in this thesis.

In nature, the PDFs of atmospheric and oceanic variables are skewed with heavy tails (Sardeshmukh et al., 2015; Sardeshmukh and Penland, 2015). According to Sardeshmukh and Penland (2015) the observed non-Gaussianity can be modelled with a stochastically generated skewed function. A skew-normal distribution can represent non-gaussian features such as asymmetric tails. This is an extension of the normal distribution while covering the skewness and containing the general characteristics of a Gaussian distribution (Flecher et al., 2010). It represents low probability values with a thin tail which can be a limitation since some of the atmospheric variables instead have a heavy tails (Yiou et al., 2008). We limit our investigation to reproducing a few PDF moments, such as mean, variance and skewness, leaving aside the kurtosis which requires much longer datasets than currently available. WMO established that 30 years is the adequate time series length to extract climatological values such as mean and std (Organization, 2017). However, to estimate other PDF moments this

length might not be enough, and we will show that Kurtosis cannot be estimated by fitting the data with a chosen PDF.

This study is probably the first one that analyses statistical distributions of winds and other surface atmospheric variables over the entire Mediterranean Sea. The atmospheric fields are retrieved from the analyses of the European Centre for Medium Range Weather Forecasts (ECMWF) operational system.

There are two main objectives of the present study. The first one is to evaluate the optimal probability distribution for the selected surface atmospheric variables. The second is the derivation of the moments (expected value, variance, skewness and kurtosis) of the estimated densities to characterize for the first time the surface atmospheric variables for the Mediterranean Sea.

2 Datasets and methodology

2.1 Atmospheric dataset

The dataset is retrieved for a 30-year period covering January 1991 to December 2020 from the ECMWF analysis dataset. The ocean forecasting system is forced by the atmospheric analysis dataset not by the reanalysis. Therefore, first we analyse the pdfs to use the distributions for the ensemble prediction system. Here, we present a short overview of the ECMWF IFS system and analysis dataset used for this study. The analysis dataset is an output of the Integrated Forecasting System (IFS) which works by the core Earth System Model developed at ECMWF. This atmospheric analysis is generated by the IFS in everyday cycle and used as an initial condition for a forecast. During the last decades, the systems, operational modules and spatial resolutions have undergone several changes. ECWMF produces three major forecasting products on the time scale, but this investigation is related to the atmospheric forcing data from the medium range forecast ones. Normally, the IFS provides forecast output where time resolution is six hours, but the spatial resolution is variable depending on the updates of the system. All the technical information and details of the updates are available on the ECMWF website. The ECMWF analysis system has changed noticeably from 1991 to today, especially in terms of horizontal nominal resolution and the number of datasets assimilated. This might affect the quality of the 30 years

TABLE 2.1: ECMWF model evolution used in the analysis system along the years and nominal lat/lon resolution. (https://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model)

Year of changes	**Spectral model truncation (T and TL)**	**Lat/Lon nominal resolution**
1991	T213	0.5625
1998	TL319	0.5625
2000	TL511	0.351
2006	TL799	0.225
2010	TL1279	0.141
2016	01280	0.1

dataset in the initial years. We will discuss the dataset consistency during the analysis but at this time we have chosen to use the entire period. However, we have chosen to consider all the data at the same resolution using the software suggested for automatic downloading at ECMWF for the years from 1991 to 2020. A brief overview on the horizontal resolution of the ECMWF analysis system is given in the Table (2.1)

We have retrieved a dataset of spatial resolution with 0.125* 0.125 degrees in latitude and longitude. The atmospheric variables considered are 10 m wind components (U for the zonal direction and V for the meridional direction), mean sea level pressure, dew point temperature, and 2 m air temperature. These fields are the major atmospheric variables composing the heat, momentum and water fluxes at the air-sea interface. Both precipitation and cloud cover in analyses and reanalyses are still affected by large systematic errors in the Mediterranean Sea, thus our choice was not to study the distributions of these variables at this moment.

2.2 Climatology and seasonal adjustment

The dataset is reduced to daily means since the daily cycle is not resolved by the six hours data and for our first study, we are interested in synoptic variability. All the atmospheric variables were analyzed to extract the seasonal signal. All days of the year are taken into account to compute a daily long-term mean for the entire 30-year period. This climatological value represents the seasonal cycle. Lets' take the atmospheric variable daily time series to be indicated by where "t" indicates the day

of the year and "j" the year. The seasonal climatological cycle is defined (Pezzulli et al., 2005) as:

$$C_t = \frac{1}{N} \sum_{j=1}^{N} X_{tj} \tag{2.1}$$

Then this long-term climatology C_t is subtracted from the original observed time series the anomalies $\tilde{X}_{tj}$ are computed as:

$$\tilde{X}_{tj} = X_{tj} - C_t \tag{2.2}$$

As an example, Figure 2.1 shows the seasonal climatology for February and August. The wind amplitude extent is large in February in the Gulf of Lion region, while this extent is larger in the Aegean Sea in August. The wind vectors are overlaid on the seasonal climatology of the wind components. The winter and summer climatology are characterized by two local wind regimes in the Mediterranean Sea, the Mistral and Etesian winds in the western and eastern Mediterranean respectively. The Mistral wind is a dry and cold northwesterly wind, influencing the marine climate in the Western Mediterranean Sea. The Etesian wind on the other hand is a strong summer wind blowing from the north to the Aegean Sea and around the Greek Islands where it plays important role in the heat flux distribution in that region. In the summer period, the zonal component (U) shows large values over the whole Mediterranean, while the meridional component (V) is large in the Gulf of Lion and the Aegean Sea. The climatological monthly mean values of the wind components vary in the interval [-4.5, 6.5] m/s both in February and August. The dewpoint temperature and air temperature have shown large seasonal differences between observed February and August in their minimum and maximum values. From the seasonal differences (Fig. 2.1), we find the minimum value difference for dew point temperature is 27.78 (K-kelvin) and maximum value difference is 12.85 K between February and August. For the air temperature, these minimum and maximum value differences are 26.89 K and 19.37 K respectively. In the Mediterranean region, normally the month of August is the hottest month during the summer period, and this 19.37 K difference in maximum air temperature indicates the phenomena. The MSL-P shows low pressure values in August over the Levantine Sea, while high pressure prevails over the western Mediterranean Sea, under the influences of the Atlantic high-pressure system.

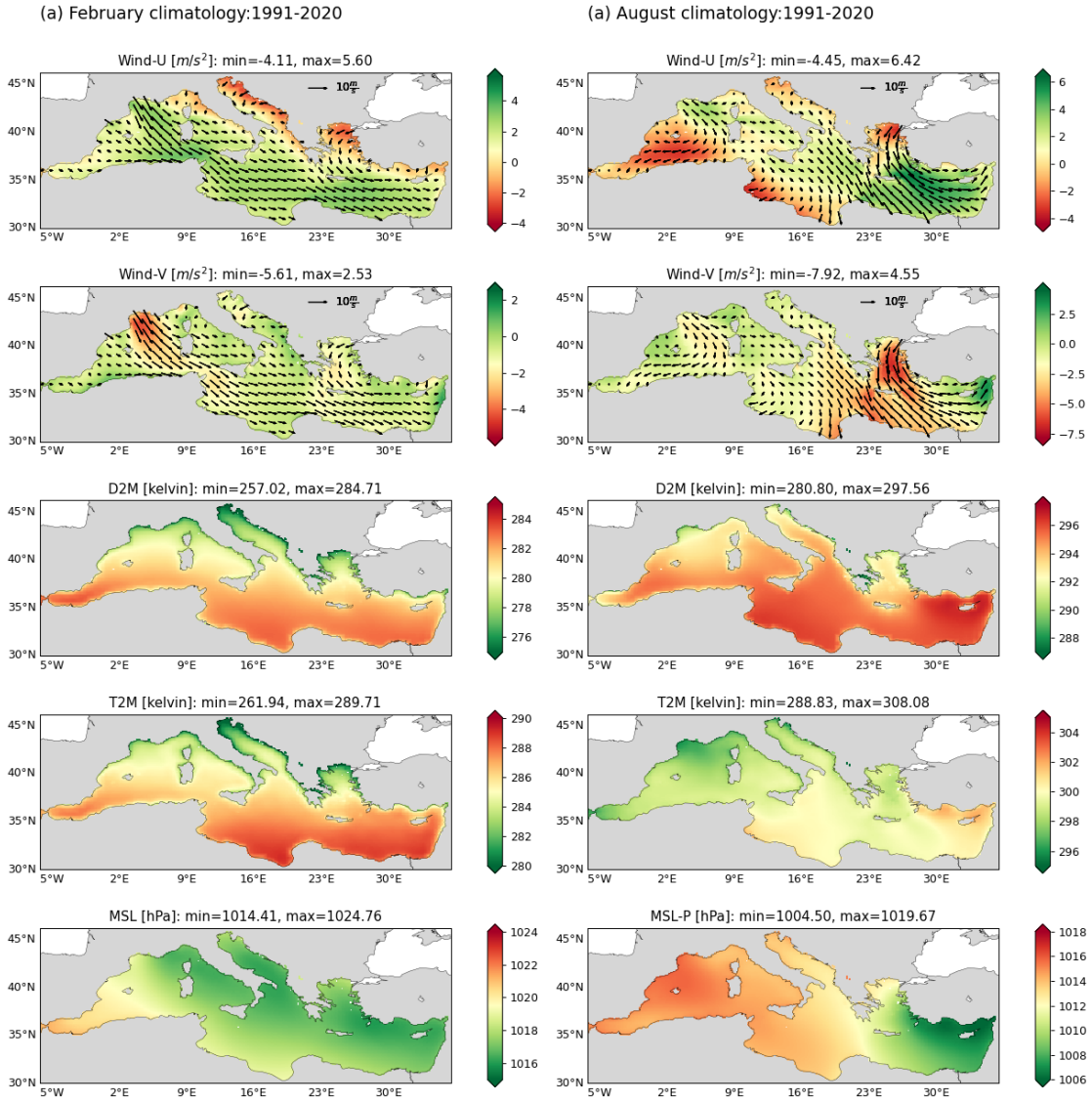

FIGURE 2.1: 30-years monthly mean climatology of atmospheric variables. From top bottom -wind components (U, V), dewpoint temperature, air temperature and mean sea-level pressure for the winter month [February (left panel) and summer month-August (right panel)]

To further characterize the climatology, we have computed the basin averaged monthly mean for all atmospheric variables. In Figure 2.2, the mean is close to 0.5 m/s for the zonal component and about -0.5 m/s for the meridional component and in some years a strong seasonal cycle is visible in each component. However, interpretation of the area average component values is difficult. Using the wind components (U, V), the 30-year monthly basin averaged wind amplitude is computed, where now the seasonal signal is noticeable. There is a maximum wind amplitude value during the winter of 2015. In the casees of temperature and dew point temperature, the monthly mean

values show the prevailing seasonal cycle for the whole period. It shows a normal summer and winter mean distribution except for the year 2012, when the monthly mean temperature dropped to the lowest point in the observation period. The mean sea level pressure shows high frequency components superimposed on the seasonal variability in a similar way to the wind components since they are partially connected by the geostrophic relationship.

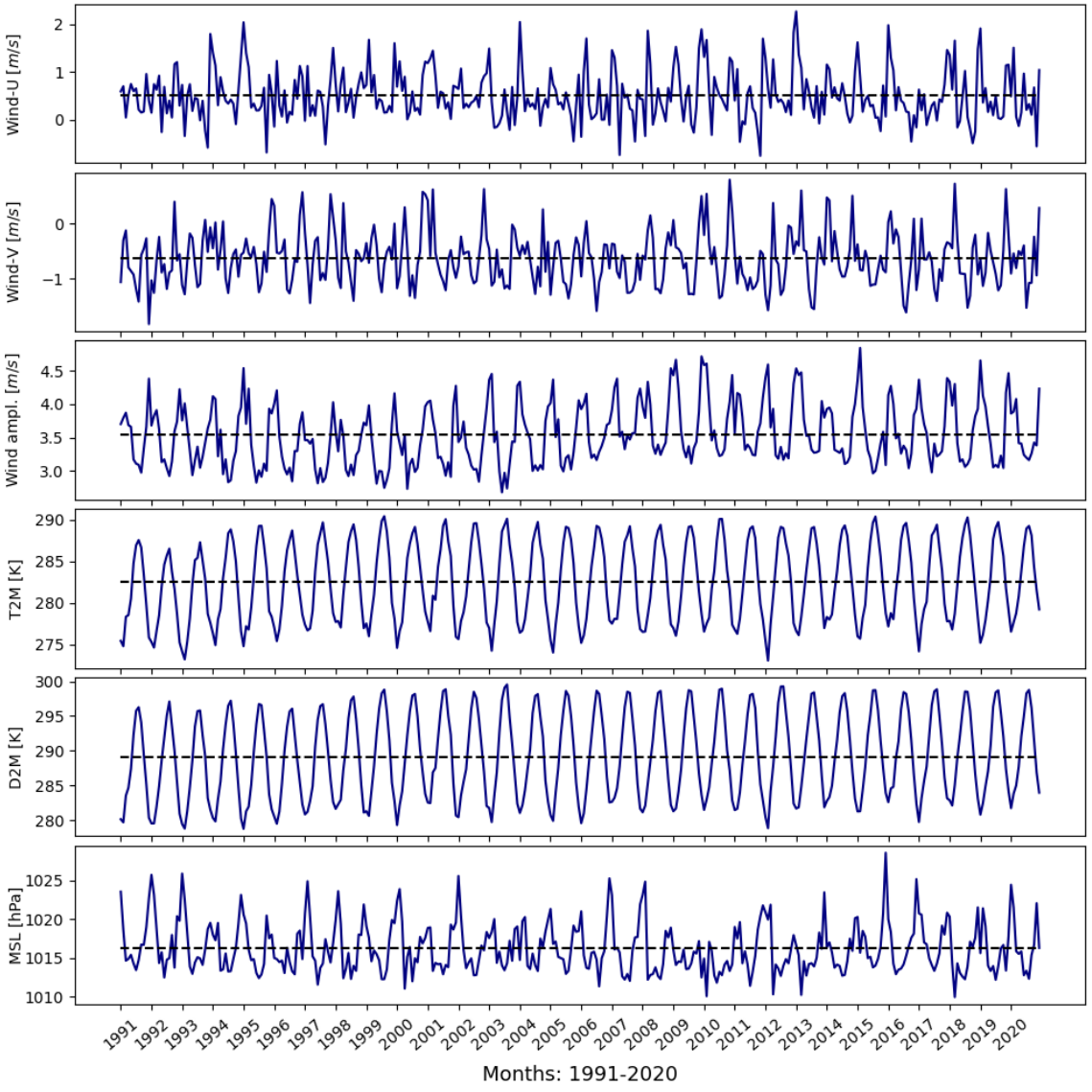

FIGURE 2.2: 30 years basin averaged monthly mean time series of atmospheric variables. From top to bottom: wind components (U, V), dewpoint temperature (D2M), air temperature(T2M) and mean sea-level pressure (MSL-P). The net mean is presented by the dashed line

Figure 2.3 presents the basin averaged monthly anomaly time series during the observed period. The atmospheric analysis dataset is composed of 30-year six hourly

analyses that we have transformed into daily averaged series. From this daily data set we compute the long-term climatology for each day. For each variable we subtract the climatology time series from the atmospheric analyses time series generating our final anomaly time series. This means that the mean of the anomaly time series is null. The anomaly range for both zonal and meridional wind components lies between -1 to +1 m/s. Monthly zonal anomalies lower than -1 m/s are present for the years 1993, 2006, 2008, and 2012 as outliers in the time series, while for meridional anomalies only for the year 2002. In the case of wind amplitude anomalies, we find a mixed trend of incline and decline for the first 15 years in the observation time series. The maximum positive anomaly for wind amplitude is observed for the year 2015 and maximum negative anomaly is found for the year 2016. Otherwise, the wind amplitude anomaly range is distributed between -0.5 to 0.5 m/s. The range of anomalies in the mean sea-level pressure time series is between -5.5 to 5.5 hPa. There is a specific anomaly in MSLP observed during the 30-years period. Low MSLP values are dominant in the period from 2008 to 2015. This MSLP negative anomaly period correspond to high wind amplitude anomalies as expected by a geostrophic balance.

The air temperature and dewpoint temperature time series show a different behavior before the year 2000, very likely due to the lower resolution of the ECMWF analysis in the period 1991-2006. In a future analysis, the first 10 years of data should be removed from the PDF analysis but at the time of writing the thesis, this was not possible.

2.3 The probability distribution functions

The probability distribution is characterized by location and scale parameters, where location parameter is the most frequent value of the distribution and the scale parameter is proportional to the standard deviation. In addition, the shape parameter controls the gradients of the distribution. Following our principal study goal, we apply the PDF functions to the daily anomaly time series of the atmospheric variables. So, in this section, we apply the fits for known PDFs on the anomaly data set described in section 2.2.

The first PDF used is Weibull which is defined (Chu, 2009) as:

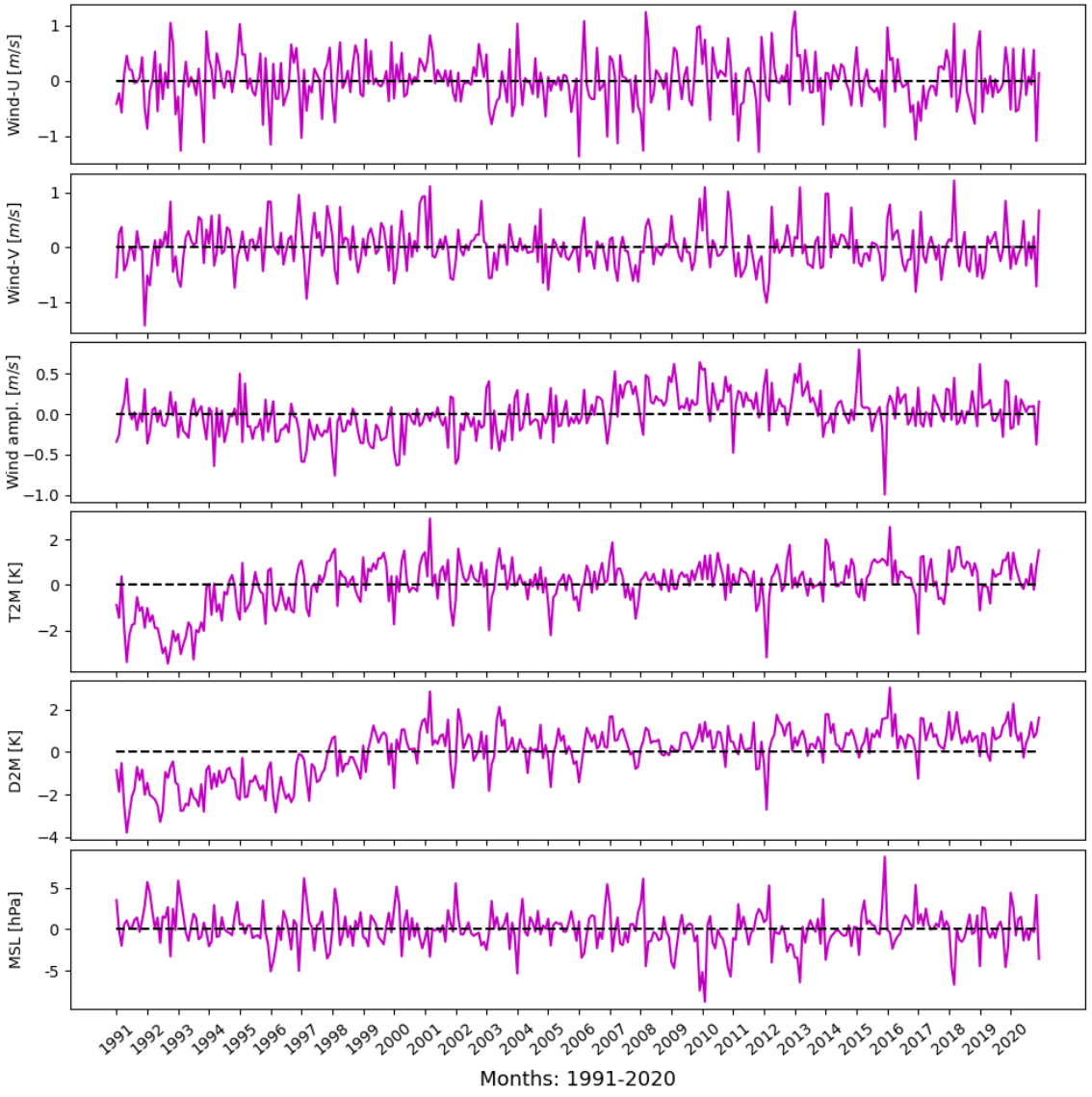

FIGURE 2.3: 30 years basin averaged monthly anomaly time series of atmospheric variables. From top to bottom: wind components (U, V), dewpoint temperature (D2M), air temperature(T2M) and mean sea-level pressure (MSL-P). The mean is presented by the dashed line.

$$f(x, \kappa, \lambda) = \frac{\kappa}{\lambda}(\frac{x}{\lambda})^{(\kappa-1)}\, exp^{-(\frac{x}{\lambda})^{\kappa}} \tag{2.3}$$

where κ is the shape and λ the scale parameter. The Weibull is defined only for x>0 and both the shape and scale parameters are positive, $\kappa > 0$ and $\lambda > 0$.

The second PDF used is the skew-normal with α as the shape parameter, μ as the location parameter and λ the scale parameter (Azzalini, 1985; Marchenko and Genton,

2010) and defined as:

$$f(x, \alpha, \mu, \lambda) = \frac{2}{\lambda}\, \phi\,(\frac{x-\mu}{\lambda})\, \Phi\,(\frac{\alpha(x-\mu)}{\lambda}) \tag{2.4}$$

where,

$$\phi(\frac{x-\mu}{\lambda}) = \frac{1}{\sqrt{2\pi}}\, exp^{-\frac{1(x-\mu)^2}{2\lambda^2}} \tag{2.5}$$

$$\Phi\,(\frac{\alpha(x-\mu)}{\lambda}) = \int_{-\infty}^{\frac{\alpha(x-\mu)}{\lambda}} \Phi(t)dt \tag{2.6}$$

Here x $\in (-\infty, \infty)$. The skew-normal PDF becomes a normal distribution when α =0 (Azzalini, 1985). The distribution is right skewed if $\alpha > 0$ and left skewed, if α <0

Skew-normal distributions are extensions of the Gaussian distributions also describing the skewness (Flecher et al., 2010). The difference between the skew-normal and the Weibull PDF is illustrated in Figure. 2.4

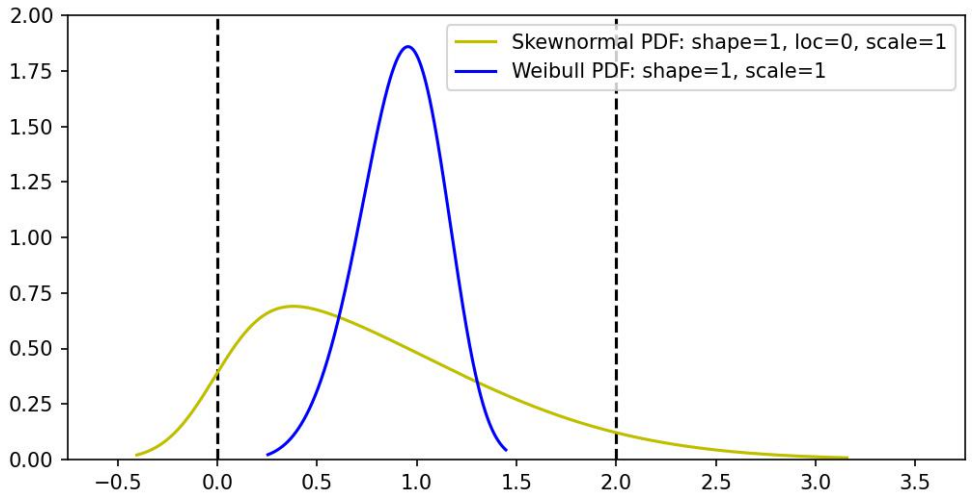

FIGURE 2.4: The probability distributions of skew-normal and Weibull PDFs for the same scale and shape parameters.

2.4 PDF parameter estimation

A complexity arises in using the MLE method due to the presence of asymmetry in data where normality assumption is not valid. The problem is addressed by Dey (2010) for the ratio of normal density and distribution function in the likelihood equation with the presence of skewness parameter. We have detected a similar issue in the estimation of the skew-normal PDF parameters, particularly with the skew or shape parameter. The goal of MLE is to find the values of the model parameters that maximize the likelihood function over the parameter space. It selects the parameter values that make the model data most probable, under the assumed model distribution. We have used the open-source Python library called 'SciPy'. The procedures used allow finding the best-fit parameters by minimizing the negative of the log-likelihood function. To minimize this function, the Nelder-Mead (NM) method (Gao and Han, 2012) has been used. It is an efficient direct search method commonly used to find the minimum of a scalar function of one or more variables using only function values, without any derivative information.

The NM method requires an initial starting point in the parameter space to start searching and through an iterative procedure it converges to the minimum. However, our (negative) likelihood function may have many local minima. To find the unique global minimum, this minimization procedure has been coupled with a multi-start strategy.

The multi-start procedure we have implemented consists of: - Generate uniformly distributed random starting parameters values - Call the iterative (local search) algorithm, determining a set of n local minimum solutions (More details in appendix) - Select the minimum solutions corresponding to the minimum value of the (-log)-likelihood function. We have used a set of 3 to 5 starting parameter values.

To qualitatively validate the parameter fit, the statistical moments are computed from the data themselves and compared with the moments computed with the fitted PDF. Mean, variance, skewness, and kurtosis can be calculated from the data as follows:

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{2.7}$$

$$S^2 = \frac{\sum (x - \bar{x})^2}{N - 1} \tag{2.8}$$

$$\mu_3 = \frac{mean\{[x - \bar{x}]^3\}}{\mu^3} \tag{2.9}$$

$$\mu_4 = \frac{mean\{[x - \bar{x}]^4\}}{\mu^4} \tag{2.10}$$

The first four moments (mean, variance, skewness and kurtosis) for the Weibull distribution are given by –

$$Mean(x) = \lambda\, \Gamma\, (1 + \frac{1}{\kappa}) \tag{2.11}$$

$$Variance(x) = \lambda^2\, [\Gamma\, (1 + \frac{2}{\kappa}) - \Gamma^2\, (1 + \frac{1}{\kappa})] \tag{2.12}$$

$$Skewness(x) = \frac{\Gamma(1 + \frac{3}{\kappa}) - \, 3\Gamma\, (1 + \frac{1}{\kappa})\, \Gamma\, (1 + \frac{2}{\kappa}) + 2\Gamma^3\, (1 + \frac{1}{\kappa})}{[\Gamma\, (1 + \frac{1}{\kappa}) - \Gamma^2(1 + \frac{1}{\kappa})]} \tag{2.13}$$

$$Kurtosis(x) = \frac{\Gamma(1 + \frac{4}{\kappa}) - \, 4\Gamma\, (1 + \frac{1}{\kappa})\Gamma(1 + \frac{3}{\kappa}) + 6\Gamma^2\, (1 + \frac{1}{\kappa})\Gamma(1 + \frac{2}{\kappa}) - 3\Gamma^4\, (1 + \frac{1}{\kappa})}{[\Gamma\, (1 + \frac{1}{\kappa}) - \Gamma^2(1 + \frac{1}{\kappa})]^2} \tag{2.14}$$

where Γ is the gamma function.

The statistical moments for the skew-normal PDF are given by :

$$Mean(x) = \mu + \lambda\delta\sqrt{2}\pi \tag{2.15}$$

$Where\ \delta = \frac{\alpha}{\sqrt{(1+\alpha^2)}}$

$$Variance(x) = \lambda^2(1 - \frac{2\delta^2}{\pi}) \tag{2.16}$$

$$Skewness(x) = \frac{(4-\pi)(\delta\sqrt{2/\pi})^3}{2(1-2\delta^2/\pi)^{\frac{3}{2}}} \tag{2.17}$$

$$Kurtosis(x) = \frac{2(\pi-3)(\delta\sqrt{2/\pi})^4}{(1-2\delta^2/\pi)^2} \tag{2.18}$$

3 Probability distributions of the atmospheric variables

In brief, section three describes the PDF of the 5 selected atmospheric variable anomalies by giving for each of them a description of the PDF fit and the validation of the statistical moments. The wind amplitude is added as a derived variable to compare with the literature.

The fit is shown initially as the comparison between the data histogram and the fitted PDF at for 8 grid points illustrated in Fig. 2.5.

3.1 Wind amplitude

The anomaly of wind amplitude, W, is defined as:

$$W_{an} = \sqrt{(\tilde{U}^2 + \tilde{V}^2)} \tag{2.19}$$

Here the fitting distribution is Weibull since W is positive definite. Figure 2.6 shows the histograms of wind anomaly amplitude and the PDF fit for the 8 sea locations shown in Fig. 2.5 . We have also added a climatological time series for each point fitted with PDF to show an example of climatology computation. At each sampling location, the Weibull PDF fits qualitatively well the anomaly time series. It is evident that the amplitude of the seasonal cycle is comparable with the anomalies and that there is a long tail of extremes. We have observed a normal and acceptable range of wind amplitude in the Mediterranean Sea that can fit with the Weibull PDF well.

After the PDF fitting on the eight selected single grid points, the method is applied to the whole study domain. By using the Weibull PDF, shape parameter (κ) and scale parameter (λ) values are computed for all the grid points available within the

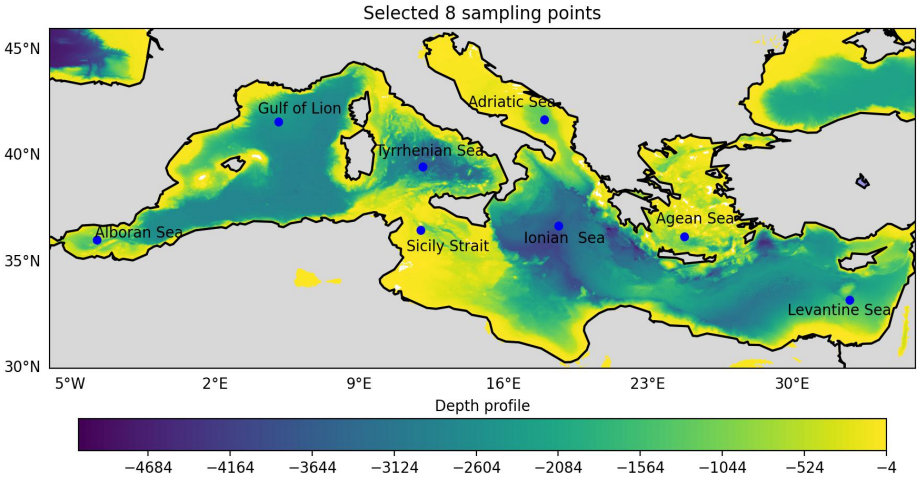

FIGURE 2.5: Selected 8 locations over the Mediterranean Sea [Alboran Sea point (-1, 36.6), Gulf of lion point (4, 42.5), Tyrrhenenian Sea point (12.1, 39.5), Sicily Strait= (12, 36.5), Adriatic Sea= (16.5, 42, Ionian Sea= (17.5, 37), Aegean Sea= (24.5, 37), Levantine Sea= (31.2, 32.3)]

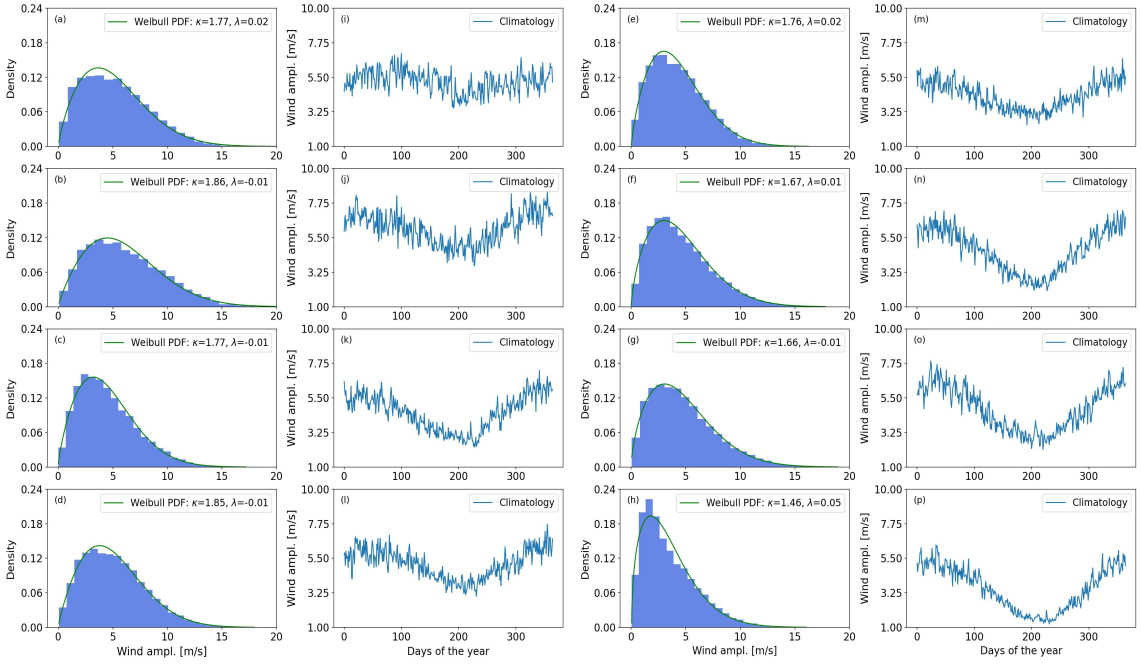

FIGURE 2.6: The single grid point histograms for wind amplitude anomalies (1st and 3rd columns) from the eight sampling locations mentioned in figure 2.5 for the period of 1991-2020, where 2nd and 4th columns represent the climatology (in m/s) of respective sea points [(a) Alboran Sea (b) Gulf of Lion (c) Tyrrhenian Sea (d) Sicily Strait (d) Adriatic Sea (e) Ionian Sea (f) Aegean Sea (g) Levantine Sea]

studied domain. The Weibull shape parameter (κ) and scale parameter (λ) values are computed over the whole of the Mediterranean Sea.

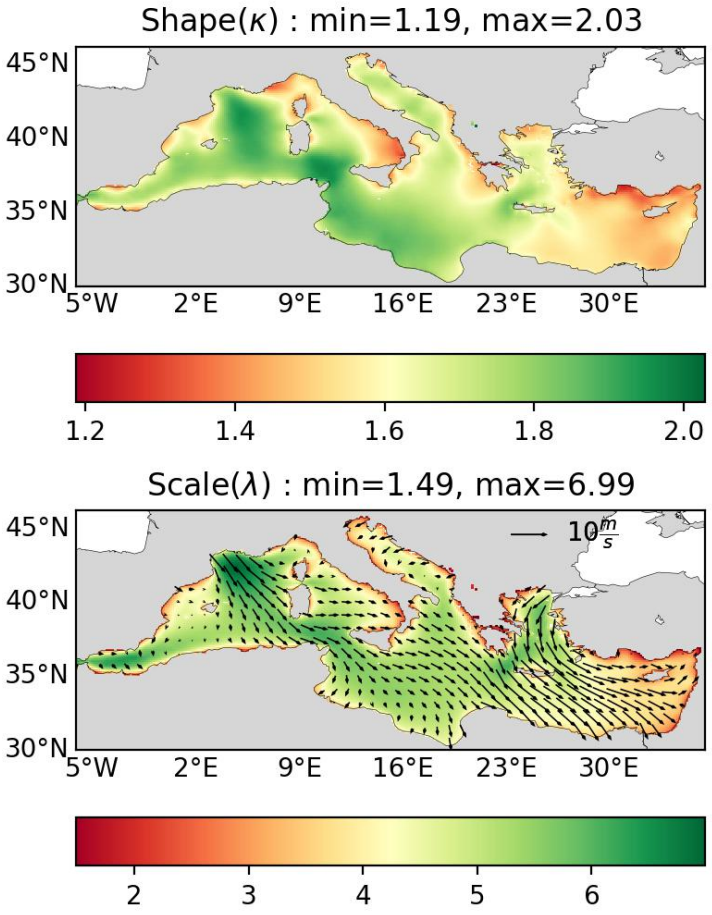

FIGURE 2.7: The Weibull PDF (κ, λ) parameters for wind amplitude anomalies for the 1991-2020 period. The mean wind amplitude vectors (m/s) are overlayed on the scale parameter.

Figure 2.7 shows the spatial variability of the two Weibull PDF parameters. The shape parameter (κ) values fall in the range between 1.19 to 2.03, but it can be seen from the spatial distribution that this range is variable between the western and eastern Mediterranean Sea. We observe the higher value range lies between 1.8 to 2 in the western Mediterranean. However, in some parts of the western Mediterranean, values fall below 1.7, while in the Eastern Mediterranean part, values are lower than 1.6. On the other hand, scale parameter (λ) values are distributed with an identical pattern in the whole of the Mediterranean Sea with a range of 1.49 to 6.99 (2.7). Weibull has only two parameters since it is only for positive definite time series. Thus, there is a correlation between shape and scale depending on the geographic

location. Theoretically, if shape parameter (κ>1) increases, then the wind amplitude (λ) increase proportionally.

3.1.1 Qualitative validation of the PDF fit-wind amplitude

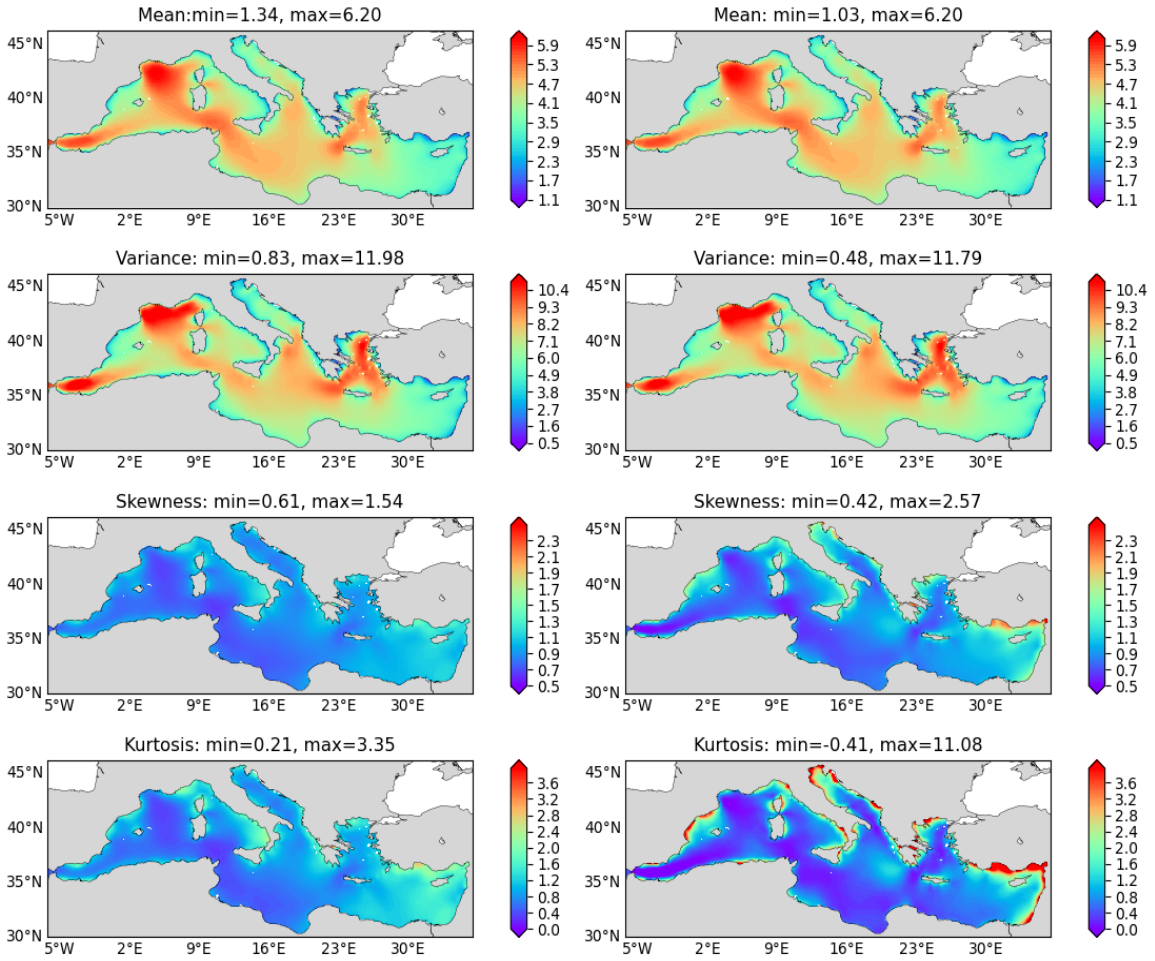

FIGURE 2.8: Distributions of the statistical moments (mean, variance, skewness, kurtosis) for observed wind amplitude anomaly and Weibull PDF fitting parameters, 1991-2020

Figure 2.8 shows the distributions of statistical moments for observed wind amplitude data (left) along with the moments computed from Weibull PDF using equations (2.11) to (2.14). If we compare the moments, where the distribution patterns are identical and value ranges are close enough to show a good quality fit. We can see that the first two mode of moments, mean and variance, have a very good agreement in the pattern and value range, while there is little variation for the skewness and kurtosis in their range. The fitted PDF asymmetry (skewness) is quite similar to the observed values while there is a range discrepancy for the kurtosis in the coastal

regions, probably due to extrapolation of values in these regions from the relatively coarse resolution ECMWF time series in the period 1990-2000. In addition, we assume that wind regimes in the edge of coastal area is slightly different form the open ocean area. We have noticed a difference in kurtosis that cannot be covered by theoretical Weibull PDF only near the edge of coastal sea areas. The kurtosis distribution is observed with high value regions, which are more than 3 in those coastal areas.

3.2 Wind components

Since the wind anomaly components can assume both positive and negative values, we cannot fit any distribution function that supports only a positive integer range and therefore, for the wind anomaly components ($\tilde{U}, \tilde{V}$) the skew-normal distribution is used. From the single grid point histogram of the zonal component ($\tilde{U}$), the Alboran sea point shows a larger discrepancy between the values and the fitted PDF (Fig. 2.9). The rest of the sea points have shown a good fit with the skew-normal fit. Interestingly, for the meridional vector, this uncertainty is found in the Gulf of Lion sea point and other sea points have demonstrated a good fitting with skew-normal PDF. On average, the meridional component is distributed within the range of -10 to 10 m/s. For the Gulf of Lion (location-b), the spread is more visible than other sea locations with a range of -15 to 15 m/s2 and the Levantine Sea point (location -h) shows the least spread with a range of -6 to 10 m/s2 (Fig: 2.9).

For the case fo Alboran Sea, the irregular pattern is probably connected to sub-seasonal periodic frequencies in the anomalies, which might be connected to local wind conditions. Interestingly, this double peak is also shown for the meridional ($\tilde{V}$) in the Gulf of Lion point (Fig: 2.10). In the future analysis, we will try to extract the subseasonal frequencies from the time series before carrying out the PDF fit.

The scale parameters (λ) are overlaid with the mean wind amplitude (m/s) anomaly of the observed period.

Figure 2.11 shows the parameter values for the wind anomaly components ($\tilde{U}, \tilde{V}$) for the skew-normal PDF. The scale parameter for zonal component displays stronger patterns in the Alboran Sea, and some parts of the Gulf of Lion and south of the Sardinia Island have shown the existence of a slightly stronger zonal component. By contrast, the scale parameter for the meridional component is weaker in the Alboran

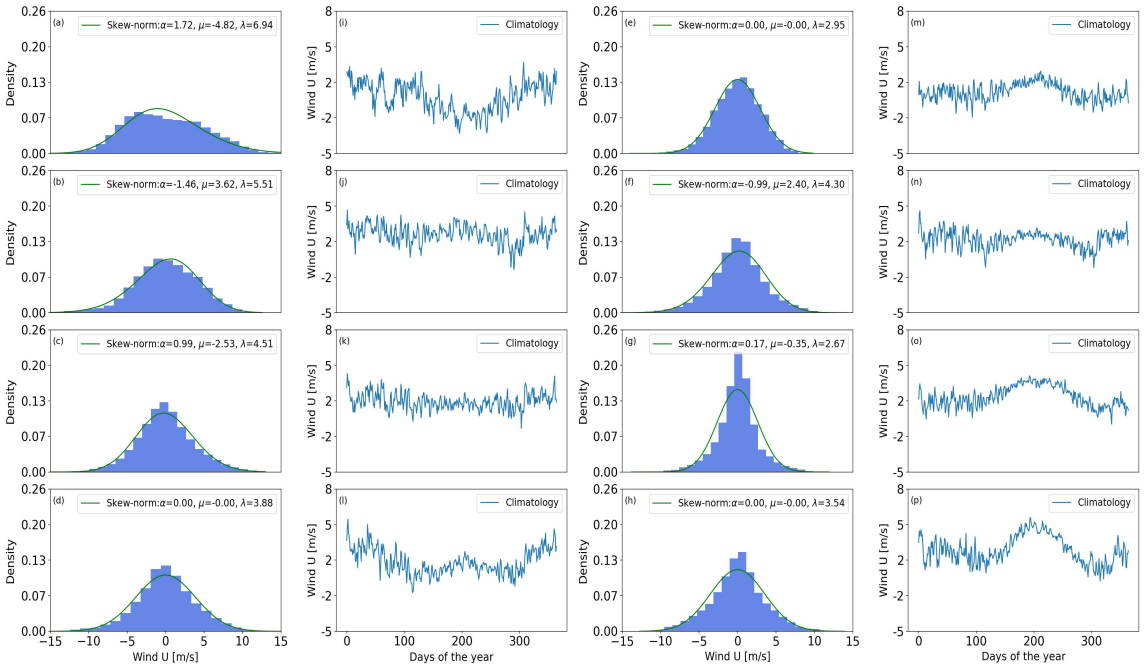

FIGURE 2.9: Histograms for the wind anomaly component $\tilde{U}$ (1st & 3rd columns) for the eight sampling points (fig: 2.5) (2nd and 4th columns represent the climatology for respective points ((a) Alboran Sea (b) Gulf of Lion (c) Tyrrhenian Sea (d) Sicily Strait (d) Adriatic Sea (e) Ioniann Sea (f) Aegean Sea (g) Levantine Sea)

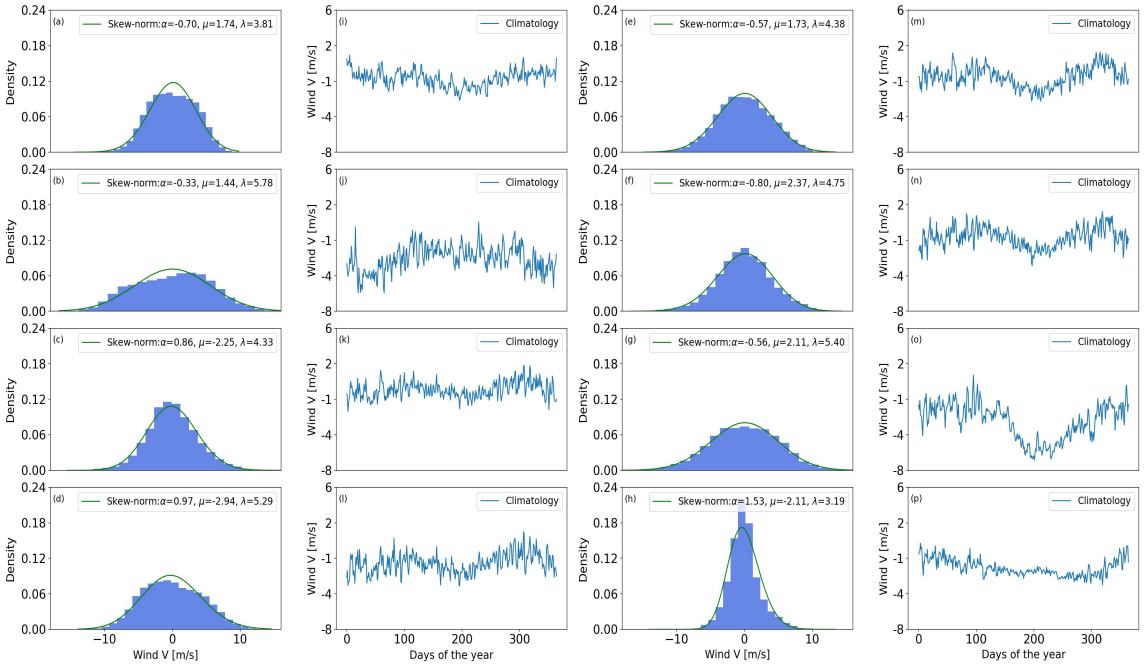

FIGURE 2.10: Histograms for the wind anomaly component $\tilde{V}$ (1st & 3rd columns) for the eight sampling points (fig: 2.5) (2nd and 4th columns represent the climatology for respective points. [(a) Alboran Sea (b) Gulf of Lion (c) Tyrrhenian Sea (d) Sicily Strait (d) Adriatic Sea (e) Ionian Sea (f) Aegean Sea (g) Levantine Sea]

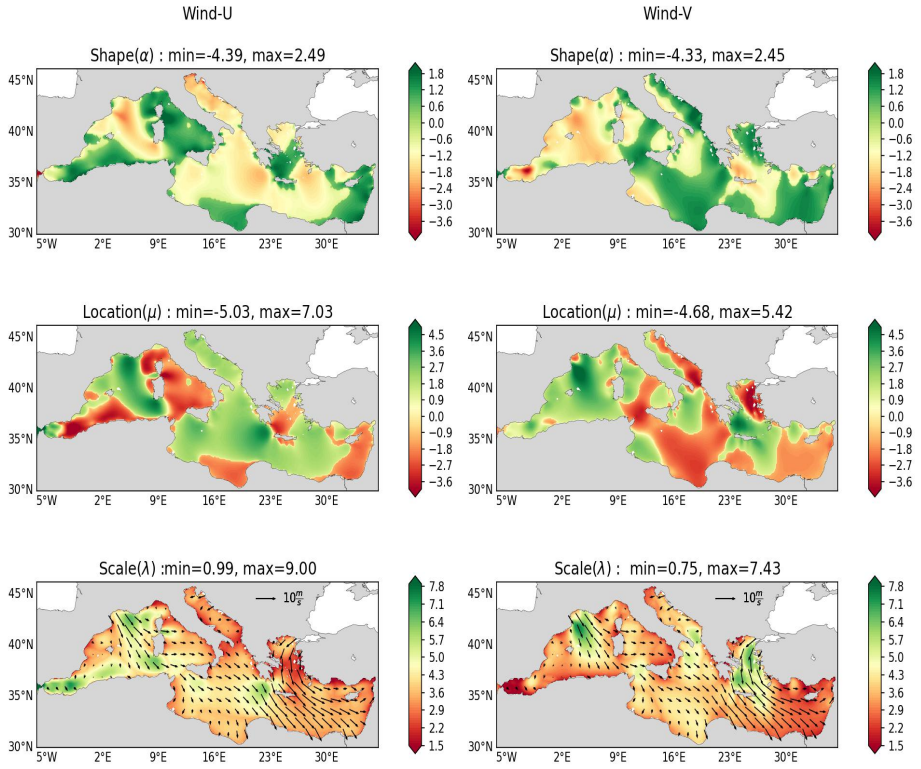

FIGURE 2.11: The skew-normal PDF parameter distributions for the zonal wind anomaly $\tilde{U}$ (left panel) and meridional wind anomaly (right panel) components during the period 1991-2020. The scale parameters (lambda) are overlayed with the mean wind amplitude (m/s) anomaly of the observed period.

Sea but shows strong circle with a maximum value distribution greater than 7. If we consider the absolute range, the shape parameter is distributed in the range of -3$<\alpha>$2 which mostly transforms the distribution into a normal one having almost equal positive and negative symmetry in the tails (see also from single grid point histograms). But this symmetry is not true for wind $\tilde{U}$ in the Alboran Sea point and wind $\tilde{V}$ in the Tyrrhenian Sea and Levantine Sea points. The scale parameter is 1-2 m/s with larger values in the Western Mediterranean Mistral area and the Aegean Sea Etesian area.

3.2.1 Qualitative validation of the PDF moments for wind components

To qualitatively validate the fitting of the skew-normal PDF on the wind components we have compared the moments between the observed and PDF fitted distribution. The statistical moments of the anomalies of $\tilde{U}$ and $\tilde{V}$ are computed with eq. 2.7 to eq. 2.10. Then, the moments of the skew-normal PDF using the fitted parameters are computed using equations mentioned in section 2.4 .

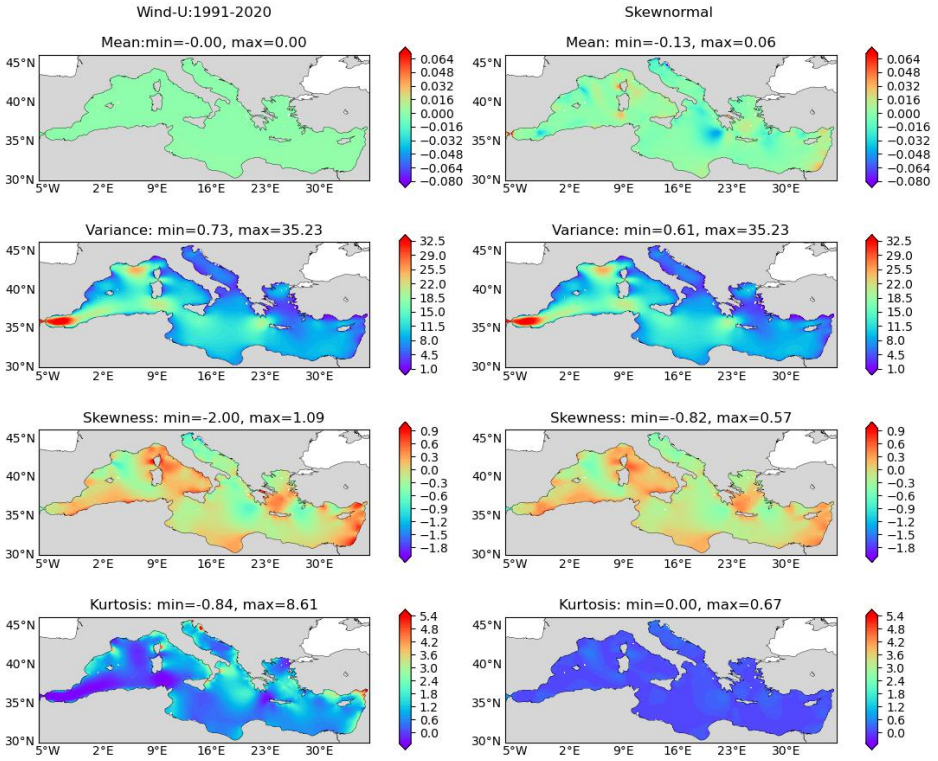

FIGURE 2.12: Distribution of the statistical moments (mean, variance, skewness, kurtosis) computed for observed wind $\tilde{U}$ anomaly component and the skew-normal PDF parameters.

For wind $\tilde{U}$, the fitted distributions approximate well the second and the third moments (i.e., variance and skewness), while the mean and the kurtosis have evident discrepancies from the observed values (Fig: 2.12).. The error in the mean is in the order of 0.1 in absolute value, that is, about 10% of the moderate anomaly values. We assume this is due to the dependency on the three parameters to compute the theoretical mean of skew-normal PDF.

The results are similar for wind anomaly component $\tilde{V}$ (Fig: 2.13). The skewness deserves some additional consideration because the major difference between the two

estimates is in the values. The observed skewness values are slightly larger for both $\tilde{U}$ and $\tilde{V}$ than the one obtained from the fitted parameters, indicating an underestimation of tails by the skew-normal distribution fit. This might be a limitation of the chosen PDF that does not approximate well fat tails datasets. Also, we assume this higher asymmetry range in the observed datasets is probably linked with the dynamics of the model and different parameters. This is because our observed atmospheric analysis dataset retrieved from a numerical atmospheric model output which itself is simulated by using many theoretical equations and approximations

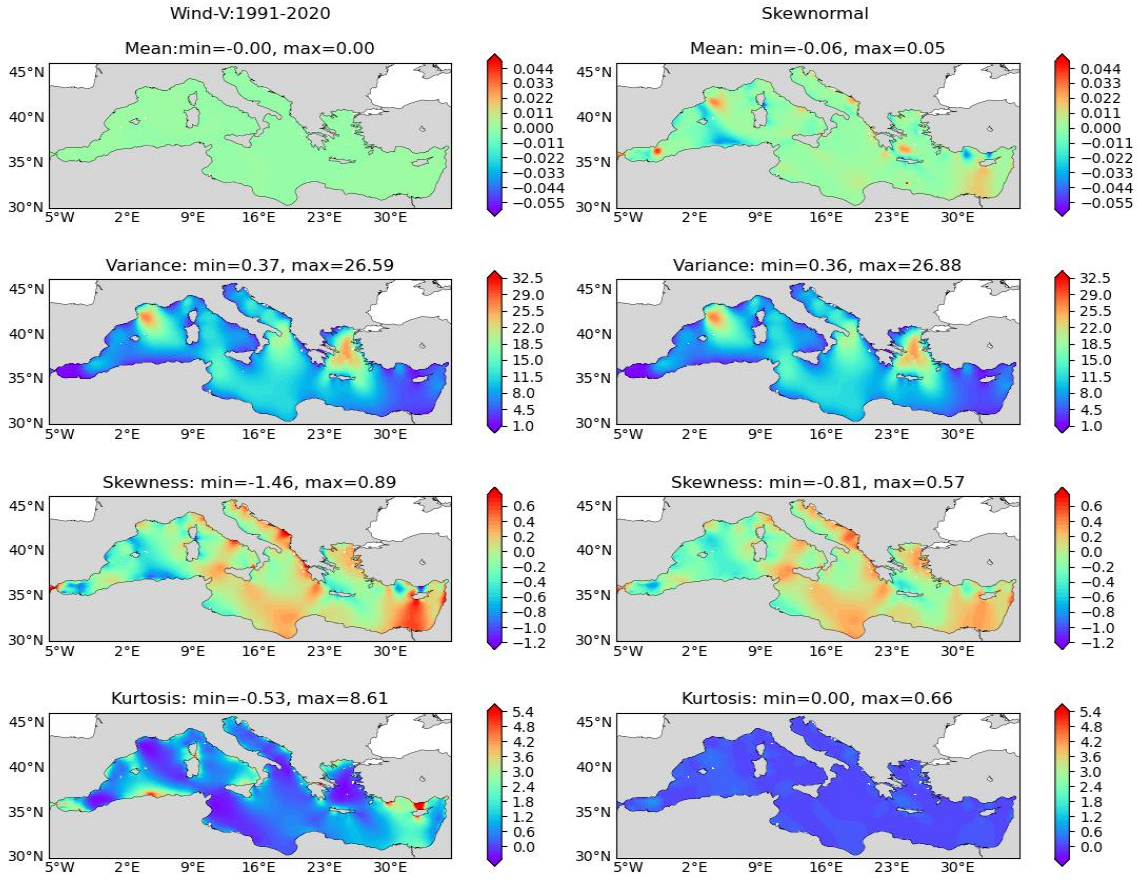

FIGURE 2.13: Distribution of the statistical moments (mean, variance, skewness, kurtosis) computed for observed wind $\tilde{V}$ anomaly component and the skew-normal PDF parameters.

3.3 Air temperature

The quality of any atmospheric variable assessment depends on the chosen probability density function (PDF) to describe the statistical characteristics. The air temperature

is a vital surface atmospheric field in the atmospheric forecasting model as well as a major forcing parameter to perturb the ocean forecasting simulations. Also air temperature at 2 m is an important surface atmospheric input that composes the heat fluxes that force numerical ocean models. Thus, its distribution and connected uncertainties are as important to know as the wind field distributions and moments. In the literature, the statistical analysis of air temperature is carried out mainly to extract extreme events and their distribution. Likewise, Yiou et al. (2008) applied the extreme value analysis to temperature and precipitation, showing that they do not follow a Gaussian distribution. In another study of extreme temperature events, like daily maximum values and extreme heat waves, Kyselỳ (2002) examined the probability distribution of extreme temperatures with the Gumbel distribution.

In our case, we wish to analyze the full distribution of temperature values, not only extremes and we will use the skew-normal distribution. In the case of air temperature, it is even more important to subtract the large seasonal cycle present at the mid-latitudes before performing the analysis. The air temperature dataset is also retrieved from the same observation period 1991-2020 and examined with the three-parameter skew-normal PDF that shows a good fit on single point distributions.

Figure 2.14 shows the distribution of skew-normal PDF fitting for temperature anomalies at the 8 sampling locations. The temperature anomaly range for each grid point is between – 5 to 5 Kelvin. In this initial fitting, we observe the distributions are not spread in the Alboran Sea and Levantine Sea points, when the rest of the points have a similar spread range.

Figure 2.15 shows the parameter distribution for air temperature where a latitudinal trend from the northern to the southern Mediterranean shores is evident. From the maps, a clear inverse relationship between the shape and location parameters can be observed. But a significant difference can be noticed in the southern coast of Mediterranean Sea, where the distribution values near the coastline point to the relation between shape and location parameters for surface air temperature. In a skew-normal distribution, a shape parameter - also called skew parameter - that characterises the degree of asymmetry in the distribution and location parameters, defines the mean of the distribution. For this study, we are interested only the atmospheric variables over the ocean that can be used to force the ocean forecasting simulations, so we skip the parameter distribution over the land.

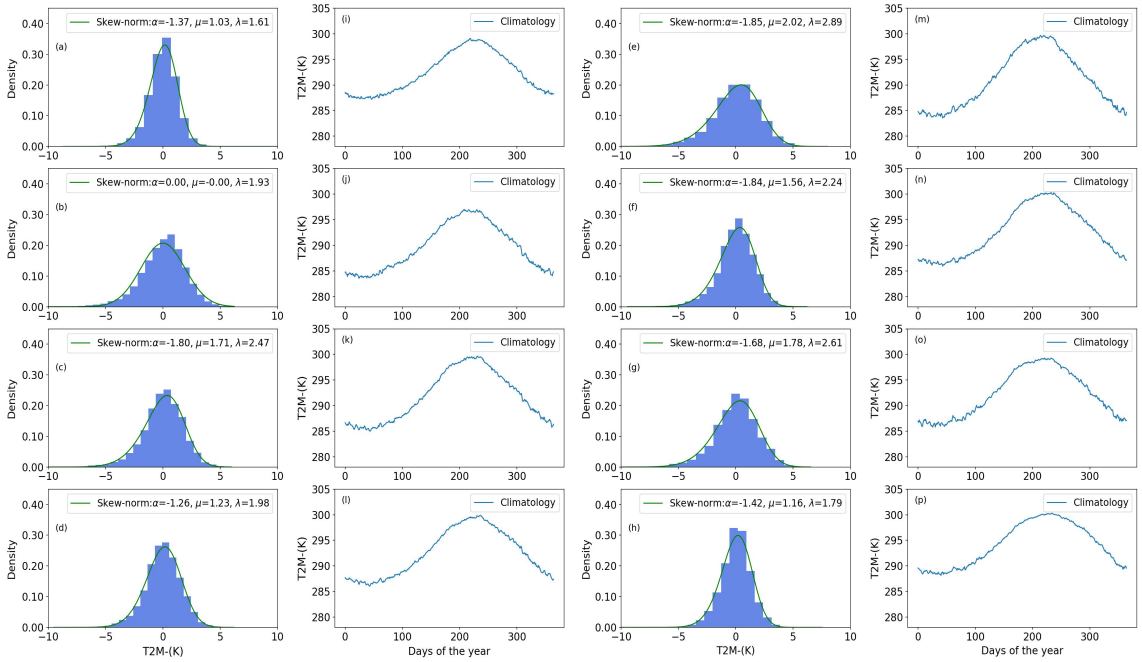

FIGURE 2.14: The single grid point histograms (1st & 3rd columns) for air temperature anomalies for the eight sampling points (fig: 2.5) (2nd and 4th columns represent the climatology for respective points). [(a) Alboran Sea (b) Gulf of Lion (c) Tyrrhenian Sea (d) Sicily Strait (d) Adriatic Sea (e) Ionian Sea (f) Aegean Sea (g) Levantine Sea]

3.3.1 Qualitative validation of the PDF fit for temperature anomaly

In this section, we carry out the qualitative validation of the skew-normal PDF fitted parameters for the surface air temperature anomaly. Figure 2.16 shows the statistical moments of air temperature anomalies from the data and the skew-normal fitting parameters side by side. The fitted moments approximate well the data values for the variance and skewness, while they present differences for the mean and the kurtosis. The error in the mean value (in the range of [-0.04, 0.06] K) is smaller than the one for the wind component if compared to the value of the temperature anomalies that are mostly in the range of plus or minus 5 K. For the variance, the distribution patterns are almost identical with the same minimum and maximum. The air temperature anomaly skewness values are again slightly smaller than the observed data values (minimum -1.07 to maximum 0.86 for the data and -0.77 and maximum 0.51 for the fitted distributions). From the skewness distribution of air temperature (left panel), there are some highly positive value areas present in between the 15° to 30° longitude area. These patches of high positive asymmetrical zone indicate the existence of heavy tails in the observed dataset distribution. The kurtosis is as usual very different

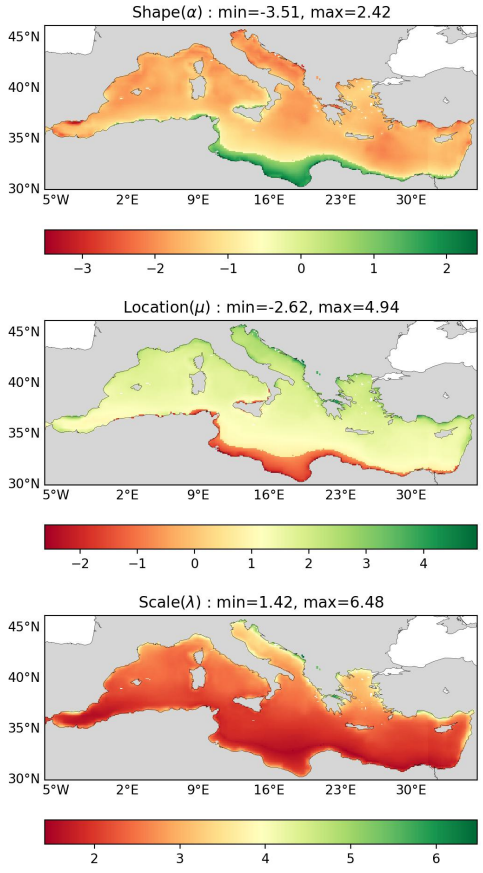

FIGURE 2.15: The skew-normal parameter distribution values for air temperature anomalies for the period of 1991-2020.

between the data and the fitted distributions, indicating large fat tails in the data that are not well fitted by the skew-normal PDF.

3.4 Dewpoint temperature

The dewpoint temperature is the temperature that brings the air water vapor to saturate. The dewpoint temperature data plays an important role in various hydrological, climatological and agronomical related studies (Mohammadi et al., 2015). It is not only a significant climatic parameter for long-term climate change research but also an input forcing parameter in ocean forecasting simulations. In an ocean forecasting

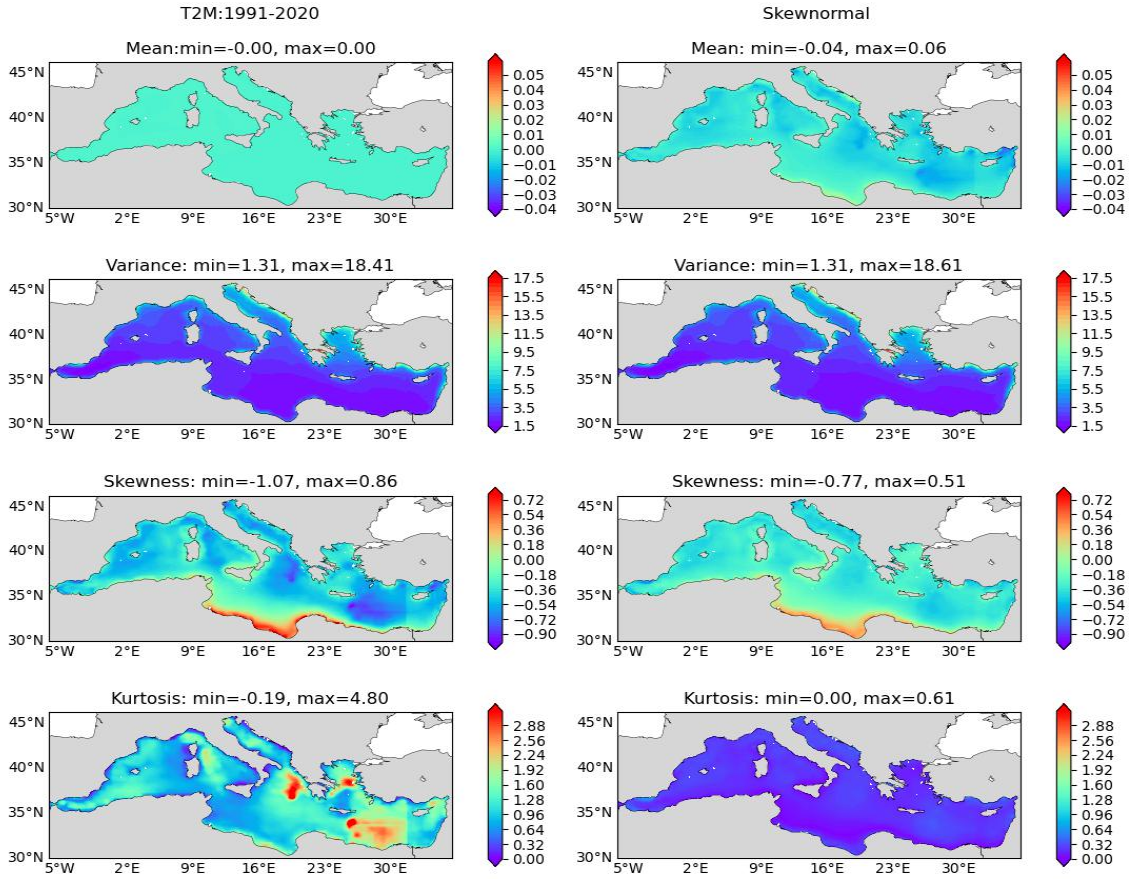

FIGURE 2.16: Distributions of the moments (mean, variance, skewness, kurtosis) for air temperature anomalies and from skew-normal fitting parameters for the observed period 1991-2020

system, model simulations are perturbed with atmospheric fields retrieved from other atmospheric forecasting models. In oceanography, the dewpoint temperature enters the formulation of latent heat fluxes, as explained in the next chapter. No previous study is found that deals with the PDF for surface dew point temperature, even if we do not expect they would be very different from the 2 m air temperature. This section investigates the statistical characteristics of the dewpoint temperature using the skew-normal PDF.

The dataset for dewpoint temperature covers the 30-year period with 6-hour intervals for a total of 10958 days of observation points. In the next stage, the process of subtraction of long-term climatology from the observed time series removes the influence of seasonal cycles on dew point temperature variation and only the interannual signal

remains in the observation.

The skew-normal PDF function captures well the skewness on single grid point distributions (Fig. 2.17). In comparison with the air temperature , the dewpoint anomalies are more skewed, showing longer tails, particularly in the western Mediterranean.

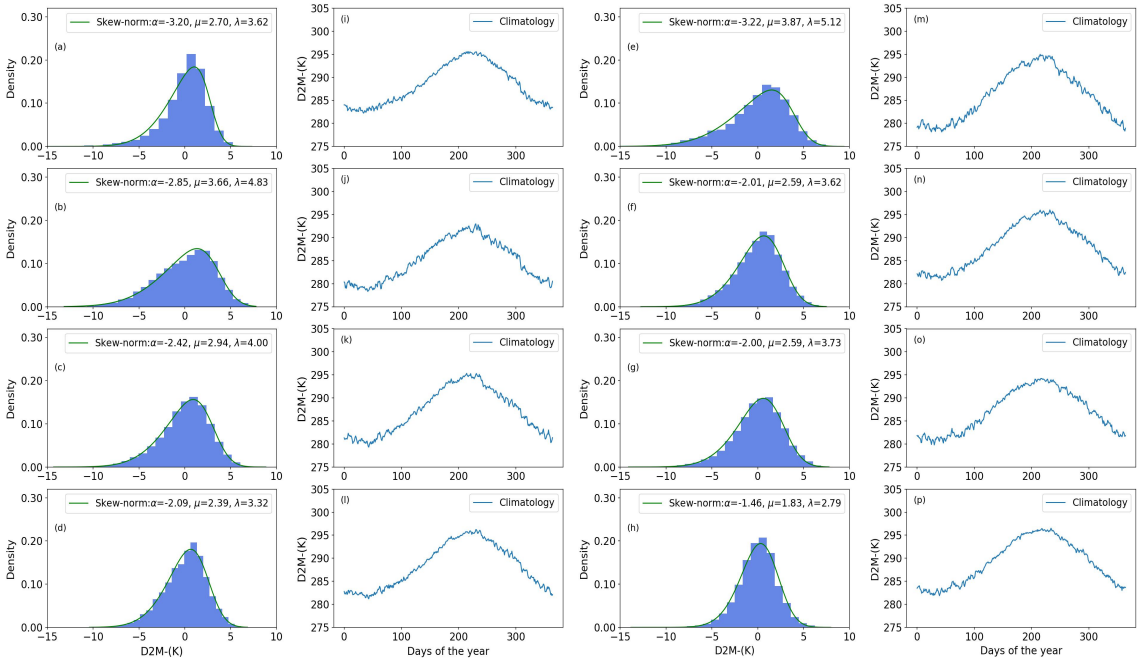

FIGURE 2.17: The single grid point histograms (1st & 3rd columns) for dewpoint temperature anomalies from the eight sampling points (Fig: 2.5) during the observed period of 1991-2020 (first and third columns represent the climatology for respective points). [(a) Alboran Sea (b) Gulf of Lion (c) Tyrrhenian Sea (d) Sicily Strait (d) Adriatic Sea (e) Ionian Sea (f) Aegean Sea (g) Levantine Sea]

Figure 2.18 shows the skew-normal fitted parameter structures which are at large scales, similar to the air temperature parameters. The shape parameter is always negative indicating the prevalence of negative value tails.. Location and scale parameters lie in the positive range of 1.33 to 5.89 and 2.32 to 7.59 respectively. It is also visible that most of the location parameter values are within 1.33 to 3, namely an opposite range to the shape parameter.

3.4.1 Qualitative validation of the PDF fit: dewpoint temperature

In Figure 2.19 we show the comparison between the moments computed from data and the fitted PDF parameters. Once again, variances and skewness show a remarkably good match. The variance distributions are within the range 3. to 23 K with the

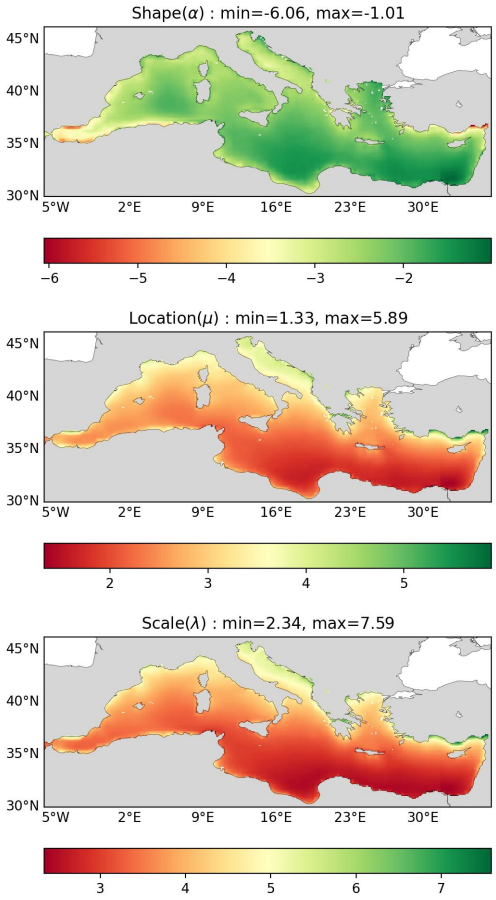

FIGURE 2.18: The skew-normal PDF parameter value distribution for dewpoint temperature anomalies for the observation period of 1991-2020.

largest values near the northern shores of the Eastern Mediterranean. The mean is again different from zero but, as for the air temperature, the error amounts to a small percentage of the anomaly signal. The variance distributions are within the range 3.01 to 22.88, with just a little difference in minimum and maximum values. This validation figure provides a clear justification that the fitted theoretical PDF has covered the asymmetry of the distribution well. With the identical distribution maps verify the skew-normal PDF can be a good fit for dewpoint temperature, For the kurtosis, again we see that the fit cannot reproduce the values, as for the previous cases.

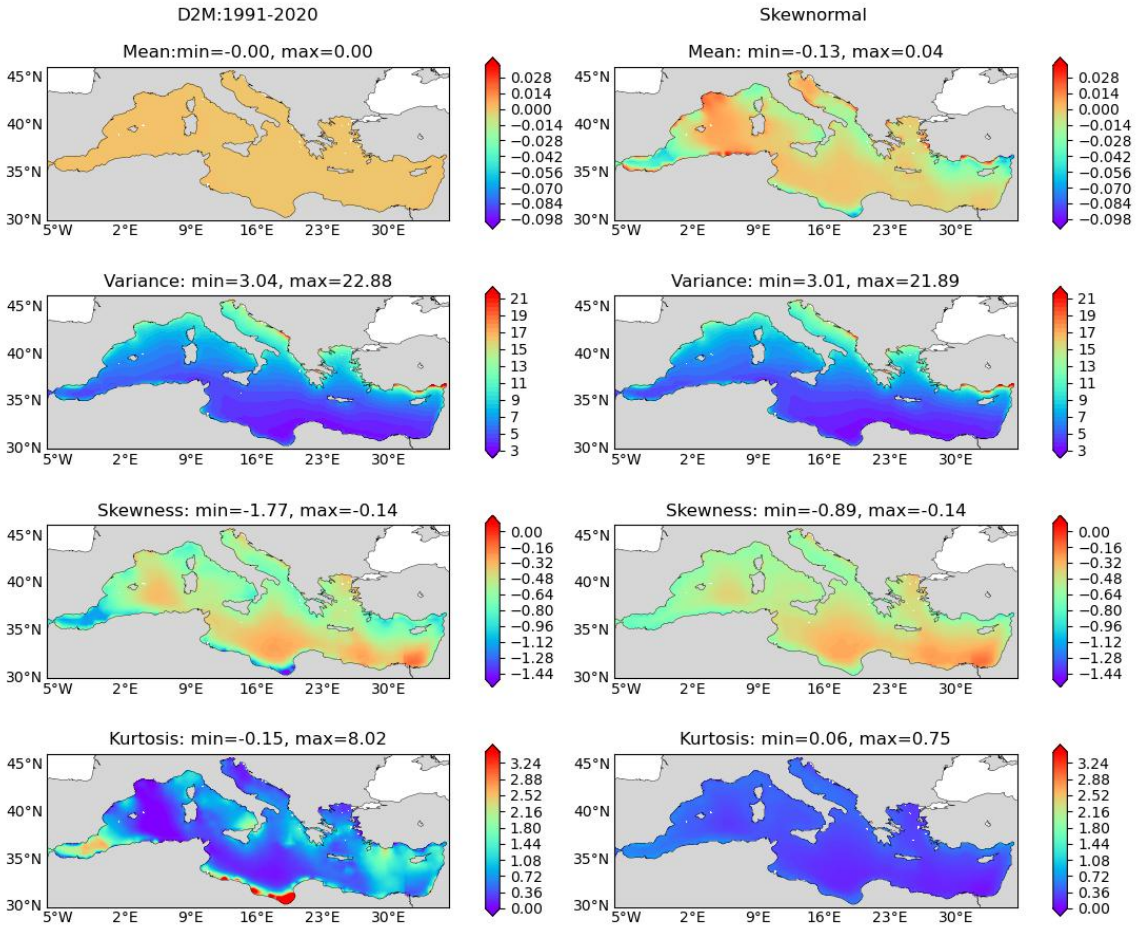

FIGURE 2.19: Distributions of the moments (mean, variance, skewness, kurtosis) for dew-point temperature anomalies and Skew normal parameters for period 1991-2020

3.5 Mean sea-level pressure

The mean sea-level pressure (MSL-P) is the atmospheric pressure that forces at the mean sea-level (MSL) and MSL is the measured average surface level of water bodies measured against a reference geoid. The study of probability distribution of MSL-P data is a related one when we deal with uncertainty analysis of atmospheric or ocean forecast. The study of the probability distribution of MSL-P is important since MSL-P directly forces the ocean models; it is a variable composing the heat fluxes and its gradients are related via the geostrophic relationship, to surface winds.

The mean sea-level pressure dataset is retrieved in the pascal unit for the 30-year period, and it was converted to hPa during the pre-processing of the dataset. Under

the initial trail of skew-normal PDF fitting for eight sampling points, we have a good fit for each point. Noticeably, we do not observe a big difference in the observed and anomaly time series, since there is less variation in the MSL-P distribution over the sea area. From Figure 2.20, we note that the general anomaly distribution is between -15 to 15 hPa for all the examined sea points except for the Levantine Sea where the distribution shows a range between -10 to 10 hPa.

Figure 2.21 shows the maps of the 3 PDF parameters over the Mediterranean Sea. The shape (α) parameter range is always negative with values between -2 and -0.7. The scale parameter (l) shows a clear difference between the eastern and western Mediterranean with values higher in the Eastern Mediterranean area. The location parameter lies in the range of 1.5 to 6 hPa. The location parameter shows a latitudinal gradient, similar to the scale parameter, while the shape parameter a longitudinal gradient. Given that the shape parameter is connected to the skewness we deduce that Western and Eastern Mediterranean have very different tail structures, the former with larger tails or anomalous events.

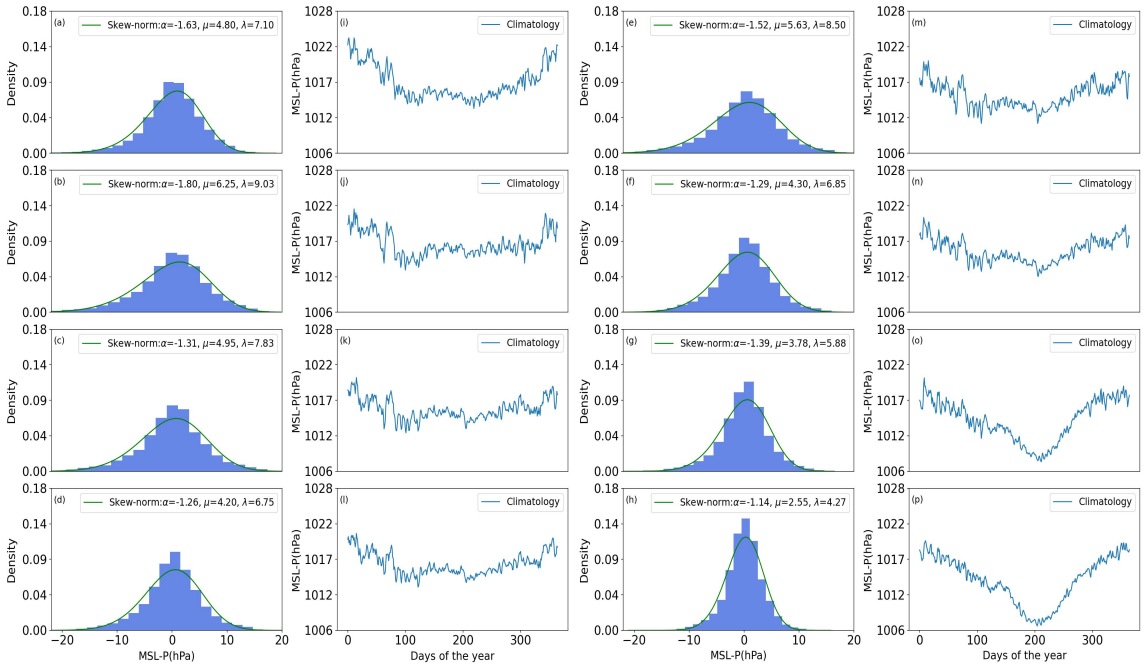

FIGURE 2.20: The single grid point histograms (1st & 3rd columns) for mean sea-level pressure (MSL-P) anomalies from the eight sampling points (Fig: 2.5) during the observed period of 1991-2020 (first and third columns represent the climatology with seasonal variations). [(a) Alboran Sea (b) Gulf of Lion (c) Tyrrhenian Sea (d) Sicily Strait (d) Adriatic Sea (e) Ionian Sea (f) Aegean Sea (g) Levantine Sea]

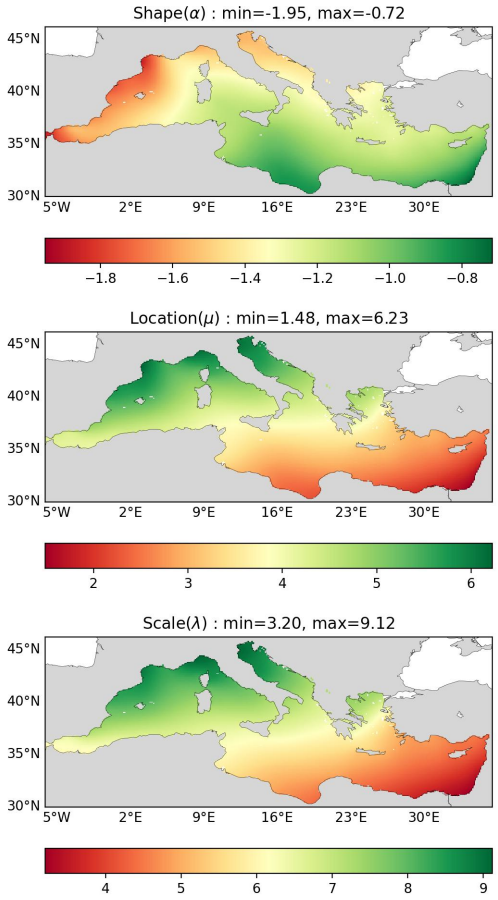

FIGURE 2.21: The skew-normal PDF parameter distributions of MSL-P for the observed period of 1991-2020.

3.5.1 Qualitative validation of the PDF fit: MSL-P

Here again, the fitted distribution moments agree with the data for the variance and skewness, while the estimated mean is in partial disagreement. The mean shows small values if compared to zero; and kurtosis is completely missing the structure (Fig. 2.22). Here, the theoretical distribution approximates well the second and the third modes (i.e., variance and skewness), while it presents some deficiencies for the 4th mode (i.e., Kurtosis). The slight difference in the mean pattern is marked for all atmospheric variables already as the similar difference is observed in the theoretical PDF mean

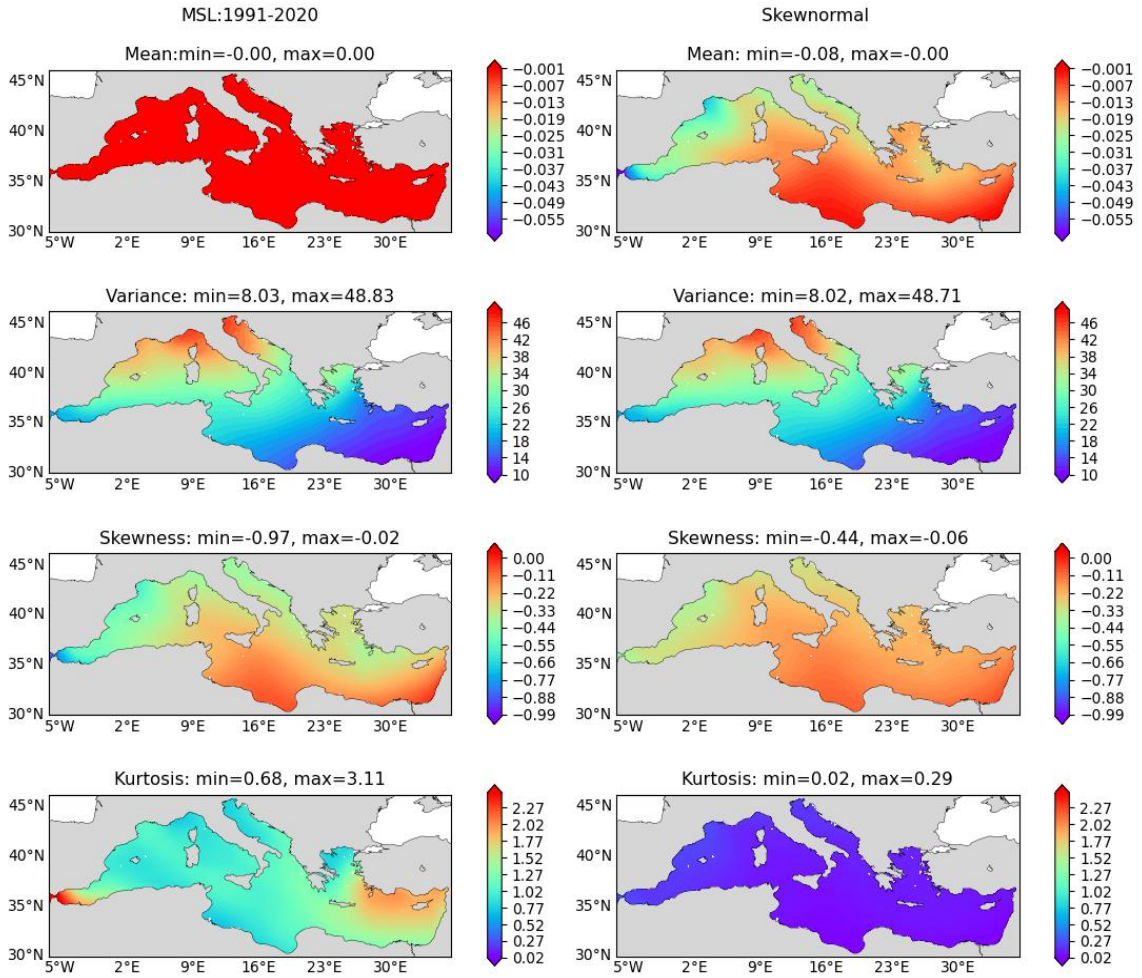

FIGURE 2.22: Distributions of the moments (mean, variance, skewness, kurtosis) for mean sea-level pressure and skew-normal PDF parameters 1991-2020

which is dependent on the PDF distribution parameters. For a theoretical PDF, matching in the variance is important. In this case, both variances have displayed an identical distribution range over the entire Mediterranean Sea. The case is similar for skewness distributions, the difference being between minimum and maximum values only, otherwise the distribution range is almost identical for the whole domain.

4 Goodness of the fit

In this section, the applied PDF fits are examined with a statistical hypothetical test which can explain how our statistical model fits well with the set of observations. Goodness-of-fit tests are used to measure the deviation between the estimated and the observed PDF. The chi-squared test is one of the most accepted and widely used tests to measure the differences in observed and expected values. In this section, the fitted PDFs are evaluated with the classic chi-squared (χ^2) test. This test determines how well the theoretical distribution fits the given data distribution.

The null hypothesis in the χ^2 test is that there is no significant difference between the observed and the expected distribution. To decide whether or not we can reject the null hypothesis, the χ^2 value needs to be computed, which is given by:

$$X^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)}{E_i} \tag{2.20}$$

N is the number of bins used to construct the histogram, O_i is the observed probability distribution (from model results), Ei is the expected probability distribution value computed from the theoretical distribution.

If the chi-squared value is lower than a critical χ^2 value, we retain the null hypothesis and conclude that there is no significant difference between the observed and the expected distributions. The decision rule for the χ^2 test depends on the level of significance (in our case it is set to 0.05) and the degrees of freedom, defined as df $= N - np$, where N is the number of data bins (25) and np is the number of distribution parameters (i.e. 2 for Weibull and 3 for Skew normal distribution), so the critical value of χ^2 is 34 for the 3 parameter skew-normal PDF and 35 for the 2 parameter Weibull PDF (taken from the χ^2 distribution table), (Elderton, 1902).

Figure 2.23 shows the result of the chi-squared test for all the studied variables. The best fit is on the wind amplitude, but also for the other atmospheric variables the values are acceptable over most of the Mediterranean Sea regions. The chi-squared results of both wind components, $(\tilde{U}, \tilde{V})$ have shown mismatch areas when compared to the observed values.

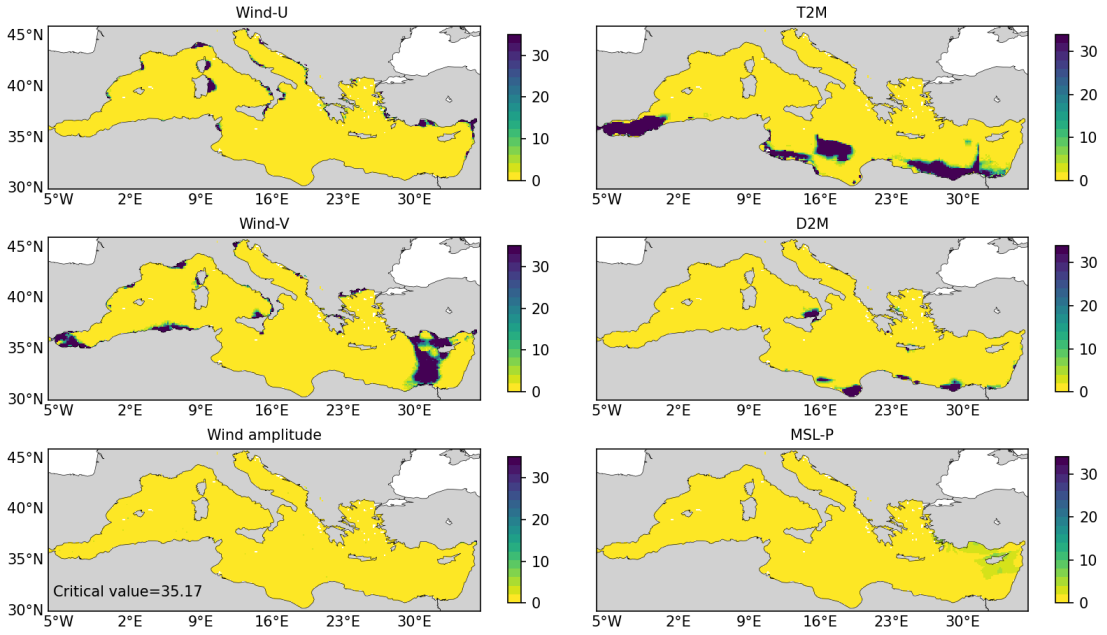

FIGURE 2.23: The distribution of chi-squared test result for atmospheric variables, wind amplitude, wind components $(\tilde{U}, \tilde{V})$ air temperature (T2M), dew point temperature (D2M), mean sea-level pressure (MSL-P). The color bar shows the critical value range of acceptance

For the zonal component of the wind vector, the test fails in few coastal areas of the northern shores, while for the meridional component the test fails in the Alboran and Levantine Sea.

For surface air temperature, the chi-squared test also fails in a large open ocean area including the southern coastal areas and the Alboran Sea, while on the contrary the test is passed for dewpoint temperature almost everywhere except for some small patches near the and at the edge of the southern coastline. At the moment it is not clear why the test fails in some specific regions. Normally, temperature ranges vary near the coast over the land. Here, for both temperatures, most of the misfit areas are located near the coastal regions, particularly along the south-east coastline of the Mediterranean Sea.

5 Conclusions

This study has evaluated the probability distributions of surface atmospheric variables from the ECMWF analysis dataset over the Mediterranean Sea for a period of 30 years (1991-2010). The study aims to quantify the structure and variability of the

atmospheric fields that are used to force ocean models and that produce large scale ocean currents, the stratification of the general circulation and its variability.

The surface atmospheric variables of interest are zonal and meridional components of the 10 m winds, 2 m air and dewpoint temperature, mean sea level pressure. Other atmospheric variables are used to force the ocean models, such as cloud cover and precipitation, but at this moment their uncertainty is too large and their statistics will be of no use in a few years to come.

A two parameter PDF is widely used in numerous works testifying that a Weibull family is good fit to the surface wind speed and surface ocean current. But a three parameter PDF is also an important PDF to analyse other atmospheric variables with the presence of seasonal fluctuations and noises in the time series. A 2- and 3-parameter PDF have been used to analyze the surface atmospheric variables that have a large seasonal cycle in the Mediterranean Sea. For positive definite variables, such as wind anomaly amplitude, a Weibull PDF is used while for all the other variables a skew normal PDF has been considered. This difference is also not well explained in the literature which does not appear to justify the choice of a 2- or 3-parameter distribution. While the 2-parameter distribution can be applied only to positive definite data, the 3-parameter distribution is needed for positive and negative data time series.

The statistical PDFs are normally used to describe the basic structure of wind variability that is particularly relevant for ocean models forced by stochastic wind forcing using analyzed statistical characteristics (Gille, 2005; Weijer, 2005). This study differs by covering other important atmospheric forcing variables that could potentially be used to perturb the air-sea fluxes of numerical ocean forecasting systems. So far, to our best knowledge, we have not found any published article dealing with major atmospheric variables together for an extended statistical study. As we have mentioned before, studies on the characterization of surface wind PDF are abundant, while other important variables, such as air temperature, dewpoint temperature and mean sea-level pressures, are missing. Through this extended approach along with surface wind, we seek to fill the gap by focusing on the complete forcing of numerical ocean modeling.

It is well-known that atmospheric variables are not Gaussian (Sardeshmukh et al.,

2015); in fact, their distribution is asymmetric with tails indicating a consistent presence of extreme, low probability events. In some cases, atmospheric variables are heavily skewed. Furthermore, for wind velocity components, temperature and mean sea level pressure in the Mediterranean Sea, the seasonal cycle is large and thus, seasonality should be removed from the time series. All these variables require a 3-parameter PDF to be used and we have chosen the skew-normal distribution which is asymmetric and allows tails. The skew-normal PDF parameters are a shape parameter, a location parameter, and a scale parameter.

There are many reasons that could generate differences between the momenta estimated from the data themselves and the fitted pdf. One of the reasons could be the limited extent of the time series. Another reason is that periodic signals and trends could have not been completely removed. In our future work we will use different temporal extensions and trend removal procedure to test the sensitivity of the pdf fit to the employed methods. At this moment, we have found that the PDF fitted distribution can represent the mean, variance and skewness moments well, while kurtosis is not captured by the skew normal. There are two possible reasons for this discrepancy. We realized too late that the atmospheric variables at the beginning of the time period, between 1991 to 2000, have a very distinct qualitative behaviour that is associated with the different ECMWF analysis system in place at the time. The second reason is that skewnormal PDF cannot fit well extensive flat tail time series but this needs confirmation in the future. Thus, it will be a matter of future work to use a different 3-parameter distribution which will allow a better representation of kurtosis in the data.

To check the quality of the fit, the chi-squared test is applied to the observed variables and skew-normal PDF fitting. The fit is accepted for most of the grid points within the study domain with different exceptions for the different variables. In consideration of the variables, air-temperature distribution has shown some uncertainty in the chi-squared test statistics, while wind amplitude, mean sea-level pressure resulted as having a very good fit in the goodness of PDF test. We have also noticed some uncertainty in the quality of fit for dew point temperature and wind components . Again, it will be a matter of future work to understand the nature of the discrepancies in specific regions. We argue that in general the land-locked nature of the Mediterranean Sea and the coastal areas could contain trends and periodic signals that are specific of

these areas, and they should be removed in each grid point and regions before fitting the PDF.

Chapter 3

Revisited heat budget and probability distribution of heat fluxes in the Mediterranean Sea

1 Introduction

The air-sea interface creates complex air-sea fluxes that exchange momentum, water and heat coupling the atmosphere and the ocean dynamics (Kara et al., 2000). Air-sea fluxes control the surface poleward branch of the North Atlantic meridional circulation; extreme surface fluxes lead to deep water formation in the Labrador and Mediterranean Sea; surface fluxes change the intensity of vertical mixing; and the diurnal cycle caused by solar heat fluxes affect mixed depth layer of the ocean (Beena and Von Storch, 2009; Bernie et al., 2007; Kuhlbrodt and Monahan, 2003; Tian et al., 2017; Walin, 1982).

Air-sea heat fluxes are a function of atmospheric variables and SST (Small et al., 2019). This chapter focuses on the Mediterranean Sea, one of the few semi-enclosed basins where, on a yearly average, heat is lost to the atmosphere and dense waters are formed because of intense air-sea exchanges (Pinardi et al., 2023). The analysis of the heat balance in the Mediterranean Sea has always been a challenge (Bignami et al., 1995; Castellari et al., 1998; Pettenuzzo et al., 2010, whether done by using numerical models or observations. The fundamental problem of in-situ observations is the limited space-time extension of the data sets, while for the numerical modelling the limitations lie in the semi-empirical approach required to write the air-sea fluxes.

Numerous efforts have been made (Large and Yeager, 2009) to compute the air-sea heat fluxes from atmospheric state variables retrieved from in-situ observations, remote sensing data or numerical model outputs. Using atmospheric and ocean variables from numerical models and observations, we have the possibility to produce longer air-sea flux time series estimated from bulk parameterizations. In this study, we consider atmospheric analysis and reanalysis data that are an optimal reconstruction of the atmospheric state variables in the past, helping reduce uncertainties in computing the heat fluxes.

There are well-established bulk transfer formulas used in the literature (Cronin et al., 2019; Fairall et al., 2003) for the computation of air-sea fluxes. The turbulent heat flux components, latent heat and sensible heat fluxes are estimated from surface wind speed, sea surface temperature, near surface air temperature, and near surface humidity (Large and Yeager, 2009). Gulev and Belyaev (2012) reported the differences in existing heat flux products and according to them, these differences were due to the different bulk formula and input variables used in the studies.

The Mediterranean Sea lies between latitudes 30°N and 45°N and is mainly a semi-enclosed basin connected to the Atlantic Ocean through the narrow Gibraltar Strait. The heat and water losses are balanced by the warm and fresher waters from the North Atlantic (Sanchez-Gomez et al., 2011; Song and Yu, 2017). The freshwater deficit is large due to excess evaporation over precipitation, inputs from the local rivers and Black Sea (Sanchez-Gomez et al. (2011)). The Mediterranean Sea estimation of the net heat budget is also referred as a heat budget "closure" problem because in the long-term heat and water lost at the surface is compensated by heat and water gained at the Gibraltar Strait (Castellari et al., 1998; Marullo et al., 2021). The value of net heat flux under this closure hypothesis should be negative, but at this stage it is probably an issue to determine if the Mediterranea has still a negative balance. The net heat budget value has been computed by Béthoux et al. (1998) to be -7 W/m^2 and this is the reference value for many studies. However, this value refers to data acquired in the 1970s and 80s, and it is possible that in a changing climate such a value could be different. Recent papers on the Mediterranean heat budget show that the net heat budget is positive but with coarse resolution data sets: 2±12 w/m2 from Song and Yo (2017) and 0.73 w/m2 from Criado-Aldeanueva et al. (2012). Furthermore, ocean warming due to climate change could change the balance between heat loss over the basin and heat gain at Gibraltar. In this study, however, we will focus on the negative value of the mean net heat flux, leaving aside the connection of the change to the analysis of climate projections .

In an enclosed sea basin, such as the Mediterranean Sea, determining the components of the heat budget is critical for a better understanding of the water budget and the regional climate. To a large extent, this air-sea dynamics regulates the Mediterranean deep water formation processes Béthoux et al., 1998; Liguori et al., 2017; Sanchez-Gomez et al., 2011; Tsimplis et al., 2006 and the Mediterranean Sea Overturning circulation (Pinardi et al., 2019). In another study, Liguori et al. (2017) used an ensemble of 12 regional climate models to reconstruct the heat flux variability over the deep-water formation regions in the Mediterranean Sea, noting the importance of the heat flux variability in the predictability of climate conditions in the Mediterranean Sea.

Over the last 40 years, several studies have dealt with the problem of estimating the Mediterranean Sea heat budget, mainly the long-term yearly mean and its seasonal fluctuations. In a model study on air-sea interactions, Castellari et al. (1998) analysed

the heat balance of the Mediterranean basin with a set of bulk formulas and proposing a calibration method for the heat flux bulk formulae calibration. Later, Pettenuzzo et al. (2010) re-analysed the Mediterranean heat and water budget using an atmospheric reanalysis dataset (ECMWF Reanalysis dataset-ERA-40). It was found that the shortwave radiation flux was overestimated by 10% when compared to observed data. An ensemble climatology of surface heat flux was done by Song & Yu and Zhang (2005), who reported a net heat budget of 2±12 W/m2, commenting that this ensemble estimate has still a warm bias. Using an ensemble of high-resolution regional climate models (RCMs), Sanchez-Gomez et al. (2011) presented a result indicating that single Regional Climate Models(RCMs) do not achieve the heat budget closure but resulted with an ensemble mean net heat range of -7 ± 21 W/m2. The latest atmospheric reanalysis efforts by Marullo et al., 2021 have led to result of large net heat flux variability with a range between 1.6 and 40 W/m2 depending on the dataset used. However, Marullo et al. (2021) has referred this variability due to longwave irradiance which needs to be improved.

In addition to the above problems, air-sea fluxes have mainly been studied in the past to establish mean and variance, while we have limited knowledge on their statistical distributions (Korolev et al., 2015; Tian et al., 2017). Knowing the probability density statistics of the air-sea fluxes and their higher moments, could rationalize the uncertainties connected to the air-sea physics. Furthermore, the air-sea fluxes variability has been found in the past to be an important source of uncertainty for ocean forecasting (Lima et al., 2019; Milliff et al., 2011; Pinardi et al., 2023).

In next step, this thesis has started the first in-depth statistical analysis of the air-sea fluxes in the Mediterranean Sea from atmospheric analysis dataset by describing their statistical distribution over the past 30 years. There is no dedicated research work on the statistical distributions of heat fluxes in the Mediterranean Sea, but obviously some authors have stated the importance of statistical inference of the air-sea fluxes in relation with climate and ocean dynamics. To get a detailed assessment of the heat flux characteristics along with the estimation and evaluation of extreme values implies the knowledge on the probability density distribution of heat fluxes over spatial and temporal scales. Moreover, it's an important approach to make an interconnection on the evaluation of reanalysis product like ERA-5. The results of evaluation for the reanalysis products have been published already over the globe, mainly for land areas but a rarer one over the oceans and few studies are available for radiative heat

fluxes. Through this research work, we also try to compute the heat flux components from a low-resolution dataset (ERA-5) beside the high-resolution atmospheric analysis dataset with an objective to compare the net heat budget

The chapter is divided into 5 sections. At the beginning, under this section, here we outline the background of this research approach based on the heat budget closure issue and importance of statistical analysis of heat fluxes in the Mediterranean Sea. In section 2, we present the atmospheric analysis and reanalysis ECMWF datasets together with the oceanic input satellite SST data. In section 3, the results of the computed heat flux components are presented with climatological analyses and net heat budget estimation, and compared with the literature. In section 4, we present the results of applied probability density function (PDFs) of the heat flux components and theoretical PDF validation. In section 5, we compare the heat fluxes computed with analysis and reanalysis input datasets along with a possible uncertainty in the distributions. Finally, in section 6, we summarize the results and highlight some key findings from this investigation.

2 Methodology and datasets

2.1 Air-sea physics in the Mediterranean Sea

In short, we have discussed air-sea dynamics in general by referring the bulk formulas used for heat flux computation in chapter 1.2. However, in our study area, the Mediterranean Sea, some adopted formulations have been established in the past decades through intensive studies. In this section, we present those adopted formulations starting with the net heat flux formula following by the heat flux components used in this study.

The net surface heat flux, Q_{net}, comprises the net shortwave radiation Q_{SW}, net longwave radiation Q_{LW} and surface turbulent flux components that combine latent heat flux of evaporation Q_{LH} and sensible heat flux Q_{SH} (Cronin et al., 2019; Pettenuzzo et al., 2010):

$$Q_{net} = Q_{SW} + Q_{LW} + Q_{LH} + Q_{SH} \tag{3.1}$$

Here, we use the convention that positive heat fluxes denote heat gain from the ocean.

Then this long-term climatology C_t is subtracted from the original time series the anomalies are computed as:

2.1.1 Longwave radiation flux

The longwave surface radiation flux, Q_{LW} is the difference between the infrared radiation (IR) emitted from the ocean surface (LU) and the atmospheric downwelling infrared radiance (LD). The LD component is adapted from Bignami et al. (1995) and the longwave radiation flux is written as:

$$Q_{LW} = Q_{LU} + Q_{LD} \tag{3.2}$$

$$Q_{LU} = -\sigma_{SB} T_s^4 \tag{3.3}$$

$$Q_{LD} = [\sigma_{SB} T_A^4 \, (0.653 + 0.00535 e_A)](1 + 0.1762 C^2) \tag{3.4}$$

where: T_S and T_A indicate the sea surface temperature and air temperature in degrees Kelvin, σ_{SB} is the Stefan-Boltzmann constant and e_A is the atmospheric vapor pressure computed from the mixing ratio of the air W_{air} (Wallace and Hobbs, 2006).

$$W_{air} = \frac{q_a}{(1 - q_{A)}} \tag{3.5}$$

$$e_A = \frac{w_{air}}{(w_{air} + \epsilon)} p \tag{3.6}$$

Where, q_A is the specific humidity of air, p is the surface air pressure and ϵ is a constant (0.622).

2.1.2 Shortwave radiation flux

The shortwave radiation flux, Q_{SW}, is adapted from Rosati and Miyakoda (1988). It is the largest heat flux component and requires a number of variables including cloud coverage and albedo. The shortwave heat flux formula is expressed as –

$$\begin{aligned} Q_{SW} &= Q_{TOT}(1 - 0.62C + 0.0019\beta)(1 - \alpha), \;\; \mathrm{if} C \geq 3 \\ Q_{SW} &= Q_{TOT}(1 - \alpha), \;\; \mathrm{if} C < 3 \end{aligned} \tag{3.7}$$

where Q_{TOT} indicates the clear sky radiation that reaches the ocean surface, C is the fractional cloud coverage, α is the noon solar altitude in degrees and β is the albedo varying month wise. The incoming solar radiation varies on locations with sun zenith angel and total solar radiation that reaches at the ocean surface after diffusion can be represented by the components Q_{DIR} and for direct solar radiation and Q_{DIF} for downward diffused radiation. Then net solar radiation Q_{TOT} can be represented by the summation of components Q_{DIR} and Q_{DIF} –

$$\begin{aligned} Q_{TOT} &= Q_{DIR} + Q_{DIF} \\ &= Q_o \pi^{secz} + [(1 - A_a)Q_o - Q_o \pi^{secz}] * 0.5 \end{aligned} \tag{3.8}$$

Where Q_o is the solar radiation at the top of atmosphere, π is equal to 0.7 and is the atmospheric transmission coefficient, A_a is a constant value (0.09) and z is the sun zenith angle.

2.1.3 Turbulent heat fluxes

The turbulent heat flux is composed of sensible heat Q_{SH} and latent heat Q_{LH} given by the following formula:

$$Q_{SH} = -\rho_A C_P C_H |\vec{V}|(T_S - T_A) \tag{3.9}$$

$$Q_{LH} = -\rho_A L_E C_E |\vec{V}|(q_S - q_A) \tag{3.10}$$

where $\tilde{V}$ is the wind speed, ρ_A is the density of moist air, C_P is the specific heat capacity (1005), C_H and C_E are turbulent exchange coefficients for temperature and humidity, L_E is the latent heat of vaporization, q_A is the specific humidity of air and q_S is the specific humidity of air saturated at the sea surface temperature T_S . Since the average wind speed in the Mediterranean is 5 m/s, Pettenuzzo et al. (2010) suggested using constant turbulent exchange coefficients for C_H=1.3.10$^-$3 and C_E= 1.510$^-$3

2.2 Description of the datasets

Two atmospheric datasets have been chosen for this study: the first is a high-resolution analysis dataset (Rabier et al., 2000) provided by ECWMF (European Center for

Medium-range Weather Forecasts) available from the CMCC archive (personal communication) at six-hour intervals with a constant spatial resolution of 0.125 degrees. The original operational dataset from which the atmospheric fields have been extracted changed in the period from 1991 to 2010 in terms of model resolution and number of observations assimilated. However, we decided to use this dataset with uniform resolution for the study period,

The second dataset is a reanalysis, the so-called ERA5 (Hersbach et al., 2020) with the same time resolution of the previous but only 0.25 degrees horizontal resolution. The ERA5 dataset is produced using a 4D-Var data assimilation as established at the cycle 41 of ECMWF's Integrated Forecast System (IFS), with 137 hybrid sigma/pressure (model) levels (vertical) and top level at 0.01 hPa.

Both datasets cover a 30-year period from January 1991 to December 2020. They are further processed into daily mean values for the whole period to eliminate the unresolved daily cycle. Experiments done with daily, and six-hour daily input fields showed little difference in the distribution, and we chose to filter out the daily variability as much as possible.

For the oceanic SST data, we have used the satellite dataset distributed by the Copernicus Marine Environment Service (CMEMS). The SST is a multi-satellite sensor blended product, so-called L4, at the horizontal resolution of. 0.05* 0.05 degree. The SST dataset is interpolated into the atmospheric analysis and reanalysis dataset grids using the extrapolation method known as the 'sea-over land' method (De Dominicis et al., 2013). The 'sea-over land' method is based on an iteration process to assign the first land value by averaging the surrounding sea points. In our case, using the 'sea-over land' method, the high-resolution SST data set is corrected and merged with the grid points of the atmospheric input ECMWF dataset.

3 Heat budget closure problem revisited

In this section, each heat flux component is computed over a 20-year period using the atmospheric analysis input dataset and analysed for their long-term mean, seasonal cycle, and net heat budget.

3.1 Long term mean and seasonal cycle

The annual mean heat flux components are obtained by averaging the time series over the entire 20-year period of the ECMWF analysis data set. The components are shown in Figure 3.1

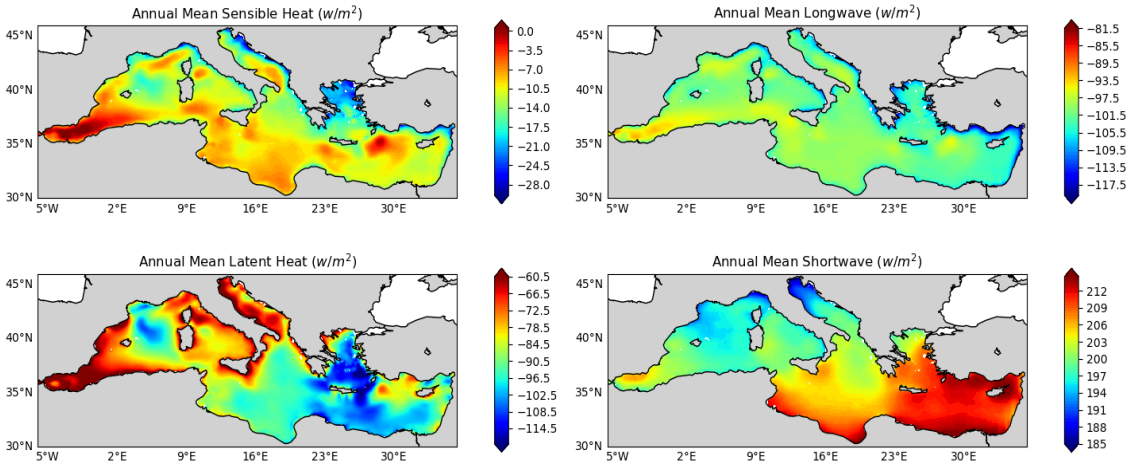

FIGURE 3.1: Mean annual heat flux distribution for the period of 2001-2020 computed from the input of the ECMWF analysis dataset.

The western Mediterranean is marked with more heat loss for turbulent heat fluxes while the eastern part shows less heat loss phenomena. The mean annual solar radiation distribution is high in the eastern part and the evaporation rate is more or less equally distributed over the entire domain. In addition to east-west differences, we observe a north-south gradient, and clearly, the solar radiation flux is lowest in the Adriatic Sea, while clearly the highest values are along the African coasts. The turbulent heat fluxes show clear sub-basin scale structures, different between the eastern and western Mediterranean Sea along with middle Mediterranean region. The largest mean sensible heat loss in absolute value occurs in the coastal regions of the Adriatic and the Aegean Sea where it reaches an absolute maximum of 31 W/m2. We observe largest absolute values of latent heat fluxes in the Gulf of Lion and the Aegean and Levantine Sea, as expected. This is due to the strong and cold winds, called Mistral and Etesian in the north-western and eastern Mediterranean regions respectively. The eastern Mediterranean has the largest evaporation equalling about 120 W/m2 in absolute value. At the edge of north-eastern coastline, we notice a maximum value range in the annual evaporation distribution.

The mean SW shows a gradual decrease from the eastern to western Mediterranean due to the variation of the solar zenith angle with latitude. The northern Adriatic is differentiable in the distribution that indicates this northern part receives comparatively less solar radiation annually. On the other hand, the mean LW distribution is with a closure range of 98-105 W/m2 over the whole domain except the highest LW observed in the Alboran Sea.

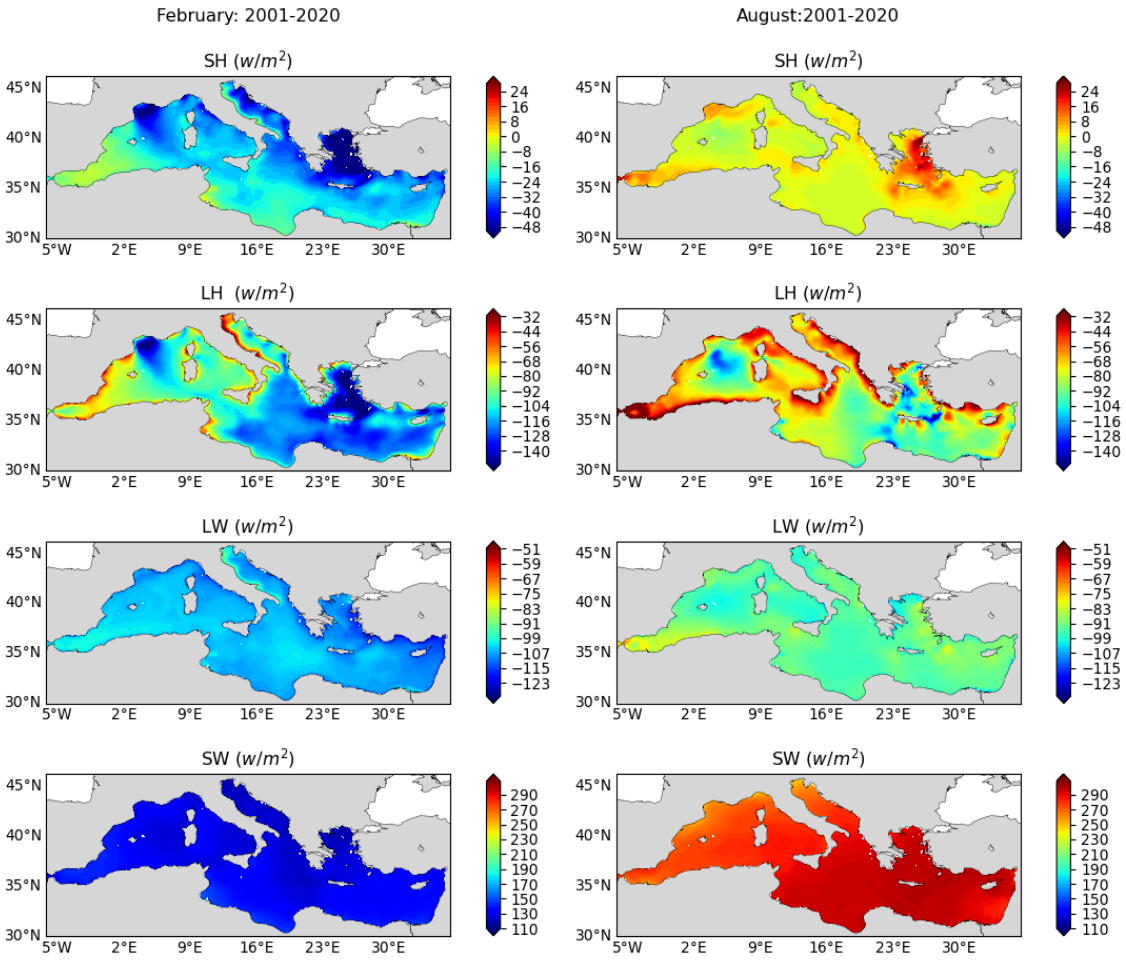

FIGURE 3.2: Seasonal variations of heat flux components: Left column is the monthly average values for February and right column the average for August for the observed period 2001-2020.

The turbulent heat flux distribution is strongly influenced by the surface wind in both winter and summer as the SH and LH distribution is marked by the zones of wind regimes. On the contrary, the radiative fluxes are observed with smoother distribution pattern in the seasonal fluctuation.

Figure 3.2 shows the heat flux component seasonal variations for February and August during the observed period. It is evident that SH flux loss is high during the winter and low in the summer. Both SH and LH fluxes show a higher spatial gradient in February than in August. Such higher gradients are marked in the Gulf of Lion and Aegean Sea particularly for SH distribution in February including some area in the Adriatic Sea region. During the summer, the Aegean Sea receives the most SH flux and some small areas in the Alboran Sea and Gulf of Lion receive high SH flux compared to the interior region. In February, LH loss is higher in the eastern Mediterranean in general, where the gradient is distinct around the Adriatic Sea including a patch in the Gulf of Lion. But for August, the LH vaporization process is high in the lower part of western Mediterranean including the Tyrrhenian Sea, the Sicily strait, the whole of the Adriatic, and some small areas in the eastern Mediterranean. Interestingly, there is a thin gradient of lower LH flux in the Aegean Sea in August. The seasonal variation is observed with a large margin in the distribution range for the radiative heat fluxes, when the LW and SW radiations are low in February and high in August for the entire domain.

The SH flux is the smallest of the four terms (Fig. 3.3). It reaches small positive values during summer and with large negative peaks during autumn and winter. The surface averaged monthly SH fluxes range from a maximum of 2 W/m2 to a minimum of -58 W/m2. We observe a large minima for the year 2012 and a few other relative minima during the years 2002, 2004, 2013.

Taking the LH absolute value (Fig. 3.3), LH shows a summer minimum of 30 W/m2 and a winter maximum of 162 W/m2 . There are three noticeable absolute value maxima during the years 2012, 2013, and 2018, which relate to strong wind regimes anomalies.

The time series of monthly surface averaged SW radiation is distributed with a winter minimum of 120 W/m2 to a summer maximum of 330 W/m2, The seasonal cycle is clearly predominant in the rest of time series. In a similar way, the seasonal cycle is evident in the LW radiation but with a large fluctuation range in summer and winter during the years of 2007 to 2011.

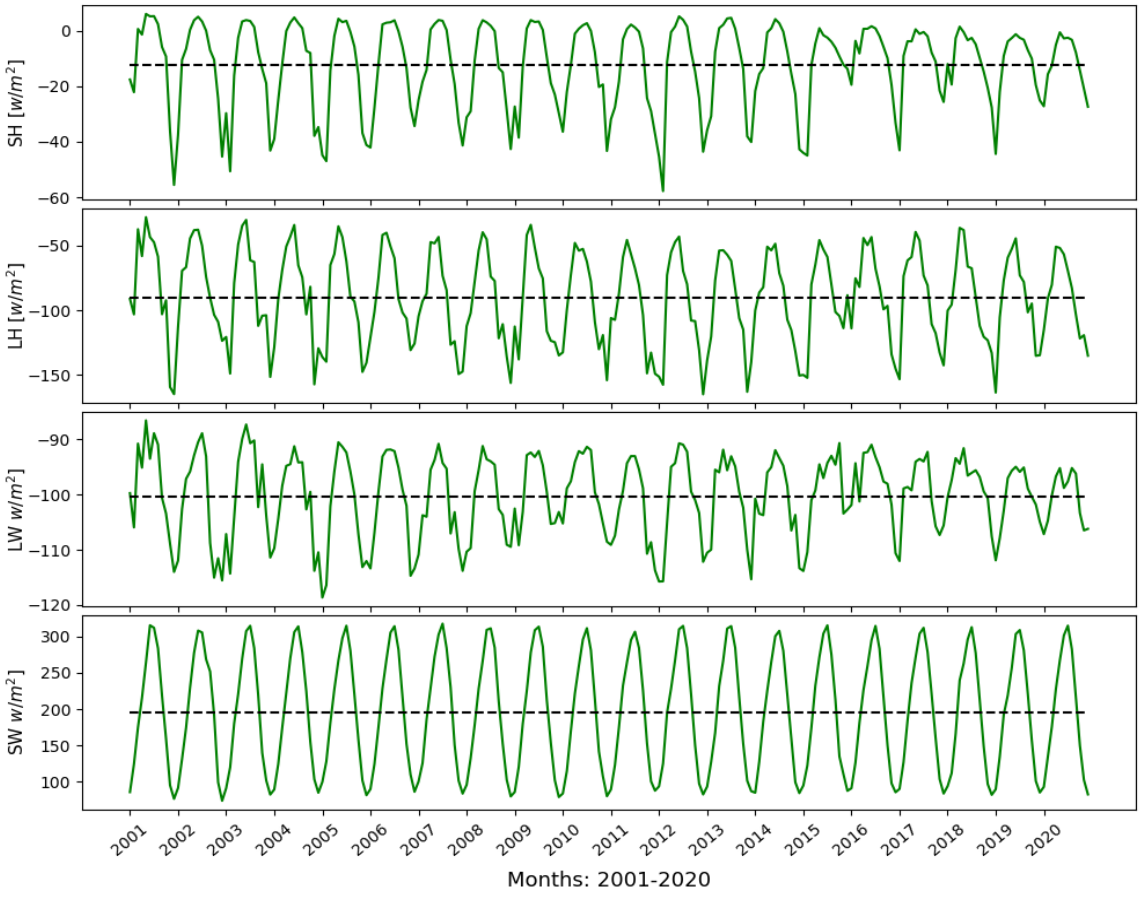

FIGURE 3.3: Monthly time series of surface averaged heat fluxes for the observed period 2001-2020.The mean value with the solid dashed line is also indicated for each flux component.

3.2 Net Heat budget

The spatial distribution of net surface heat flux Q_{net} is computed and presented in Figure 3.4. As we have observed an irregular signal in the surface monthly climatology for the earlier years of the observation period, specifically for the first 6-7 years between the years 1991-2000, we have computed fluxes for the period of 2001-2020. We have suspected a probable resolution change impact over the 90s when changes in the ECMWF system occurred with higher resolution grid. Due to avoid any technical discrepancy, here we have used the 20 years observation, from 2000 to 2020, to compute the net heat flux in this section

The basin average Q_{net} amounts to -3 W/m2 which is consistent with the previous estimates for the Mediterranean Sea domain (Table 3.1). Thus, the ECMWF analysis

TABLE 3.1: Computed fluxes and net fluxes, and comparison with references

Authors	**SH**	**LH**	**LW**	**SW**	**Net flux**
Bethox(1979)	13	120	68	195	-6
Bunker (1982-1)	13	101	68	202	20
Bunker et al.1982-2	11	130	68	202	20
May (1986)	11	130	68	193	2
Garret et al. (1993)	7	99	67	202	29
Petenuzzo et al., (2010)	-14	-92	-79	178	-7
Sanchez-Gomez, et al. (2010)	13±5	13±5	75±6	181±18	-7±21
Criado-Aldeanueva et al. (2012)	-15.1	-93.5	-76.9	186.3	0.73
Song & Yoy (2017)	-13±4	-98±10	-78±13	192±19	2 ±12
ECMWF input (This study)	-11	-91	-100	198	-3
ERA5 input (This study)	-12	-97	-81	208	18

dataset, together with the specific bulk formula used, can accurately reproduce the net heat flux that satisfies the Mediterranean Sea heat budget closure problem.

If we consider the seasonal variation, during the winter the Q_{net} ranges from 0 to -160 W/m2, while in the summer it reaches values between 30 and 150 W/m2.

The Q_{net} shows a complex structure over the Mediterranean Sea, ranging from -40 W/m2 to 50 W/m2 (Fig. 3.4). In general, the Q_{net} shows a heat gain pattern in the mean distribution in the eastern Mediterranean, though a large gradient region is visible near the Aegean Sea where a maximum heat loss area is observed. In addition, some areas in the Adriatic Sea and south-eastern coast of Sicily also show net heat loss. High net heat loss around the south-eastern coast of Sicily and some areas in Aegean Sea are clearly connected to the large winds channelled by the near coastal orographyA large heat gain region during summer is in the Alboran Sea due to the specific wind regime occurring there. The other local winds, the Etesian, force the Aegean Sea to have notable net heat losses in that region.

From the basin averaged Q_{net} time series, we can notice a smooth monthly climatological signal for the whole observation period (2001-2020) (Fig. 3.5). The seasonal cycle variability is regular, with several anomalous minima over the observed years. In particular, the winter of 2016 is much warmer, and the heat loss is low with respect

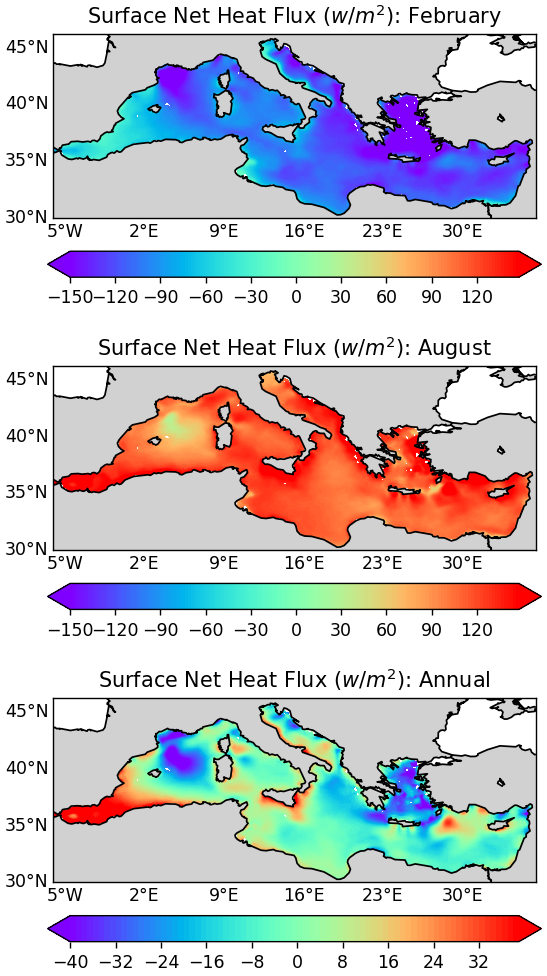

FIGURE 3.4: Surface Net heat flux values (from top to bottom) for the summer (August) and winter (February) months and the annual neat heat flux for the period 2001-2020 from the ECMWF analysis input dataset.

to other winter values. This feature is also connected to the similar lowest fluctuation in the winter period for the same year for flux components (Fig. 3.3). Specifically, we assume this lowest winter net heat flux is marked by the lowest LW flux value for the year 2016, in relation to this, we have observed the lowest LW flux value during the winter cycle in that year in Figure 3.3.

4 Probability distributions of the heat fluxes

In this section, we have analysed the probability distribution of heat fluxes for observation period of 30 years. Similar way to the atmospheric variables, we need to eliminate the seasonal fluctuations from our time series.

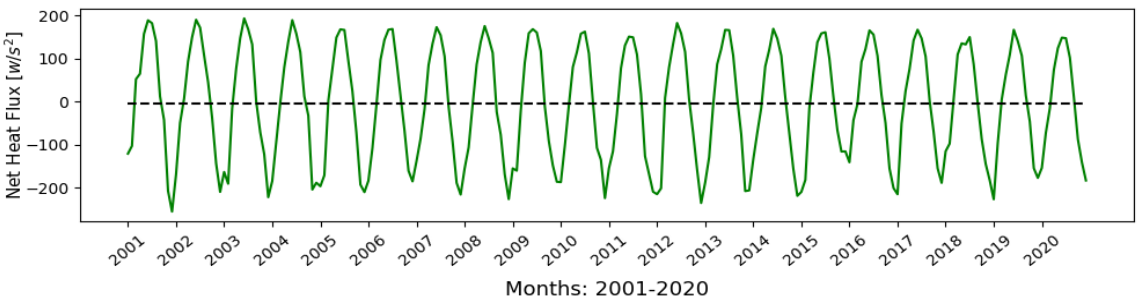

FIGURE 3.5: Monthly time series of surface averaged net heat fluxes Q_{net} for the observed period 2001-2020. The mean value is indicated with the black dashed line.

If we indicate the time series of each component of the heat budget with X_n we can define the heat flux climatology as:

$$Q_t = \frac{1}{N}\sum_{i=1}^{n} X_{tj} \tag{3.11}$$

where 't' indicates the day of the year and ' j' is the number of years. Then the anomaly time series is computed by subtracting the long-term climatology Q_t from the observed heat flux time series X

$$\tilde{X} = X_{tj} - Q_t \tag{3.12}$$

4.1 SH flux distribution

The skew-normal probability distribution function used to fit the atmospheric variables in chapter 2 cannot be used for all the heat flux components as is evident from the histograms at single grid points shown in Figure 3.6 for the SH. From the histograms, it is clear that the values around zero almost indicate a singularity. This result has already been found by Gulev and Belyaev (2012) and we offer an explanation of this in the Appendix.

The most common distribution with such near-discontinuous behaviour at the origin is the three-parameter Asymmetric Laplace Distribution (ALD) (Yu and Zhang, 2005) that we can defined as:

$$f(x,\alpha,\mu,\lambda) = \frac{\lambda}{\alpha+\frac{1}{\alpha}} \begin{cases} exp(\frac{\lambda}{\alpha}(x-\mu)) & if x < \mu \\ exp(-\frac{\lambda}{\alpha}(x-\mu)) & if x \geq \mu \end{cases} \tag{3.13}$$

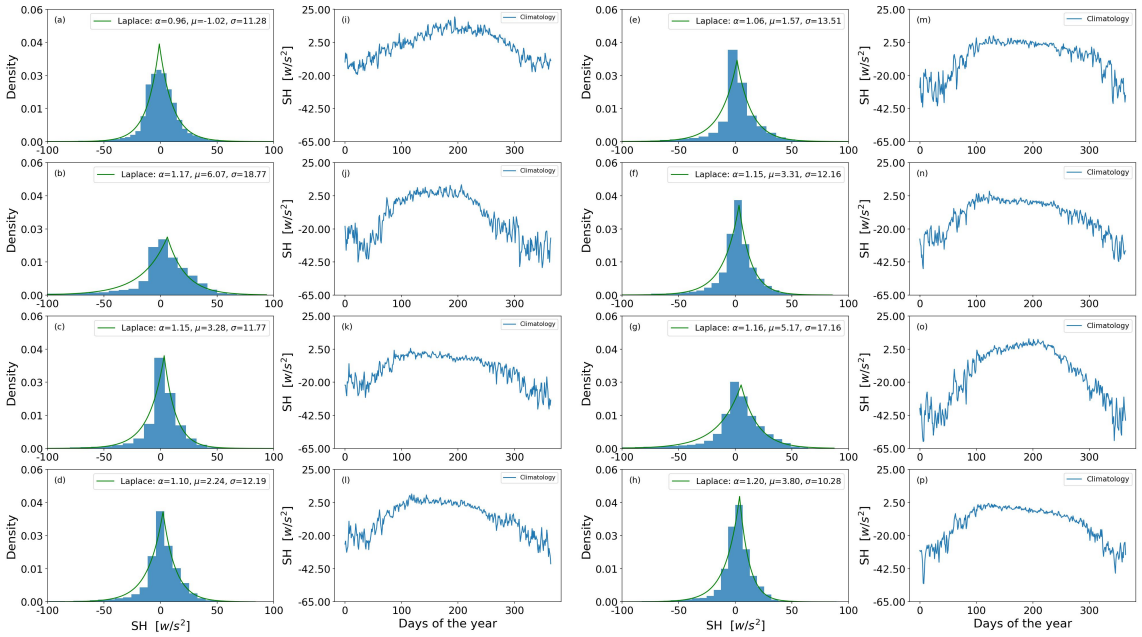

FIGURE 3.6: The single grid point histograms for SH flux anomalies (1st and 3rd columns) from the eight sampling locations for the period of 1991-2020, and 2nd and 4th columns represent the climatology of respective sea points.

where x is the random variable time series, α is the shape parameter, μ is the location and λ the scale.

After fitting the ALD PDF on the transformed SH flux distribution, we have computed the moments of the ALD parameter distributions for qualitative validation. The statistical moments of the ALD PDF are computed using the following equations:

$$Mean(x) = \mu + \frac{1-\alpha^2}{\lambda\alpha} \tag{3.14}$$

$$Variance(x) = \frac{1+\alpha^2}{\lambda^2\alpha^2} \tag{3.15}$$

$$Skewness(x) = \frac{(2-\alpha^6)^{3/2}}{(\alpha^4+1)} \tag{3.16}$$

$$Kurtosis(x) = \frac{6(1+\alpha^3)^2}{1+\alpha^4} \tag{3.17}$$

From the single grid point histogram, we have observed a one or two sharp peaks in the distribution that matches well with the Asymmetric Laplace Distribution (ALD) PDF

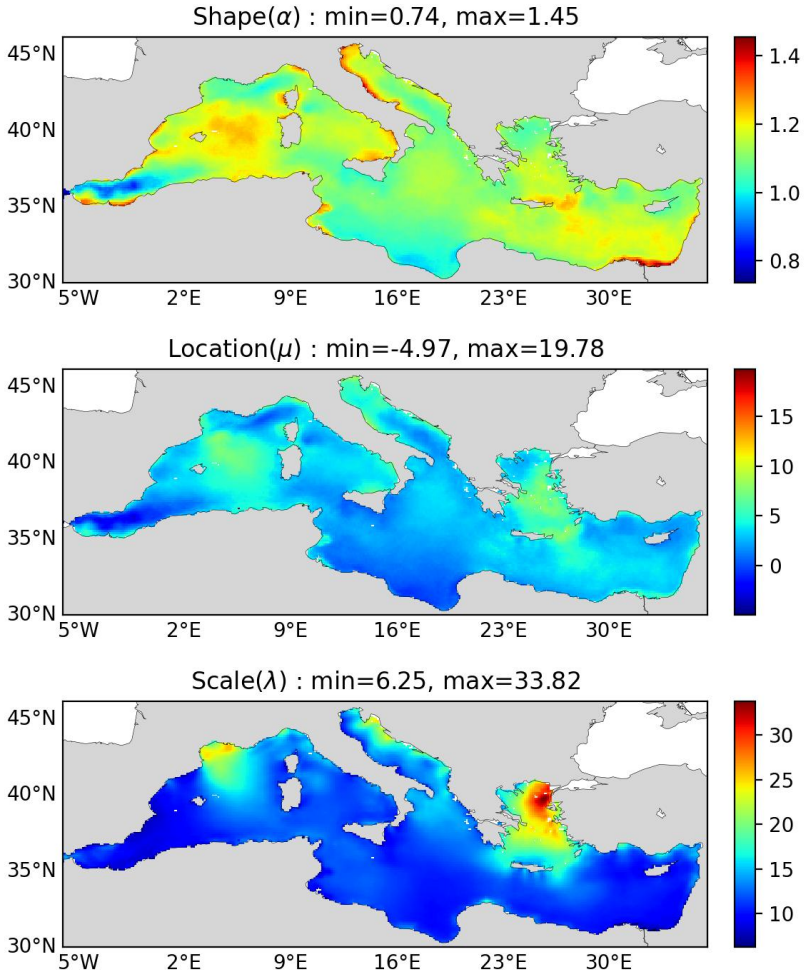

FIGURE 3.7: The Asymmetric Laplace PDF parameter (α, μ, λ) distributions from computed SH flux anomaly for the observation period of 1991-2020.

(Yu and Zhang, 2005) that covers double exponential distribution. At the same time, there are positive and negative skewness with long tails are visible in the histogram.

Figure 3.7 shows the Laplace pdf parameters for the SH flux anomaly time series. The shape parameter α is in the positive range of 0.74 to 1.45, while the location parameter (μ) is both positive and negative. The scale parameter λ has a similar structure to the SH flux climatology that we have already seen in the figure (Fig. 3.1).

To check the quality of the fit, moments of both applied and theoretical PDFare compared in Figure 3.8 and it shows the estimations of the four moments in the left panel for the observed SH flux and right panel for ALD PDF parameters. It can be seen that the mean and variance are similar while skewness and kurtosis differ greatly.

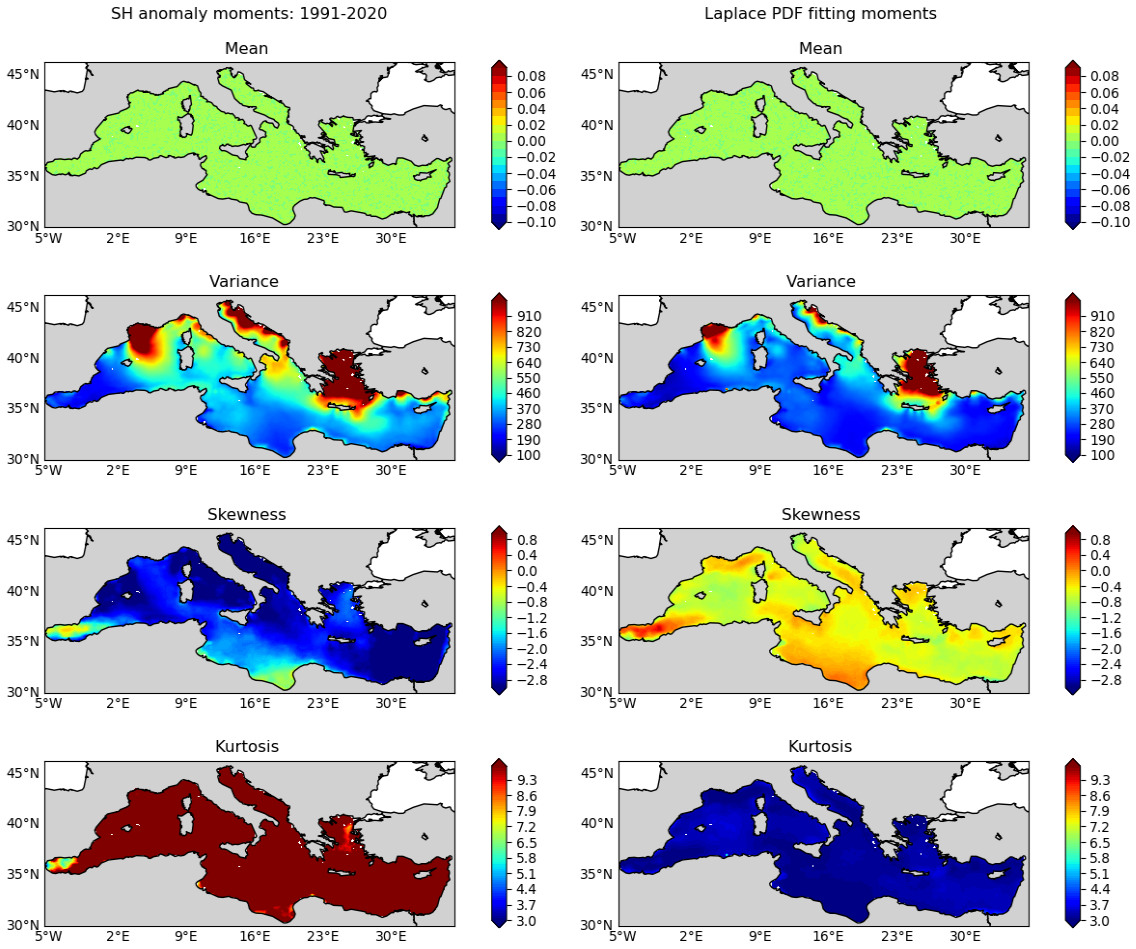

FIGURE 3.8: Comparison of the statistical moments (mean, variance, skewness, kurtosis) computed from the SH flux anomaly values and estimated by the ALD PDF parameters (period 1991-2020).

This is probably due to the fact that kurtosis for the asymmetric Laplace distribution has a kurtosis that does not depend on the scale parameter. In the future, only the 2001-2020 period should be used, and this should improve the agreement between the statistical moments from analysis data and from fitted distribution since the first 10 years do not seem to belong to the same random time series samples.

4.2 LH flux distribution

In case of LH flux, we have not observed any transformation or sharp exponential picks rather long tails are found in the LH flux distributions. Here, the LH flux distribution is analyzed with the skew-normal PDF.

To check which PDF to fit on the LH flux values, 8 sea locations histograms are fitted with skew-normal PDF 3.9 . In this case, we find long tails in the time series and all the points are negatively skewed. The 2nd and 4th columns in figure 4.4 present the yearly climatology of the respective points fitted with the PDF. In the case of LH flux, the skew-normal PDF seems to be adequate

Figure 3.10 shows the parameter distribution computed for the LH flux anomalies. The shape parameter distribution shows negative range values from a minimum $-10.$ to maximum of -1.4. This spatial distribution of α with a negative range supports the single grid point PDF fitting test that shows a negatively skewed pattern. The spatial distribution structure of location and scale parameter show a positive correlation almost everywhere.

The comparison of statistical moments is done in next step to show a qualitative validation of the fit. Both moments of the observed LH flux and skew-normal PDF are computed using the moment equations mentioned in Chapter 2 and presented in Figure 3.11 . There is a good agreement in mean and variance distribution in both observed LH flux anomaly and skew-normal PDF. The skewness distributions are matched in some areas, but there is a difference in the distribution pattern where the skewness for theoretical PDF mainly lies within the range of -0.7 to -1. and for observed skewness, the variation range is large with a range -0.7 to over -2.7. Lastly, the skew-normal PDF kurtosis distribution is different between the observed and estimate PDF. We assume that again the analysis dataset for the period 1991-2000 is the source of the discrepancy and this hypothesis will be tested in future work and further discussed in Section 5.3.

4.3 LW radiation distribution

After trial fitting of the PDF on the 8-point histograms (Fig 3.12), it was evident that the skew-normal PDF shows a reasonable fit. For LW radiation, we do not observe long tails in the time series, but slight negatively skewed patterns. is more visible in the distribution.

The distribution of shape (α) shows negative values, indicating negative skewness where the minimum value is -3.5 and maximum is 0.3 (Fig. 3.13). There is again a relationship between the location (μ) and scale parameter (λ) structures, especially

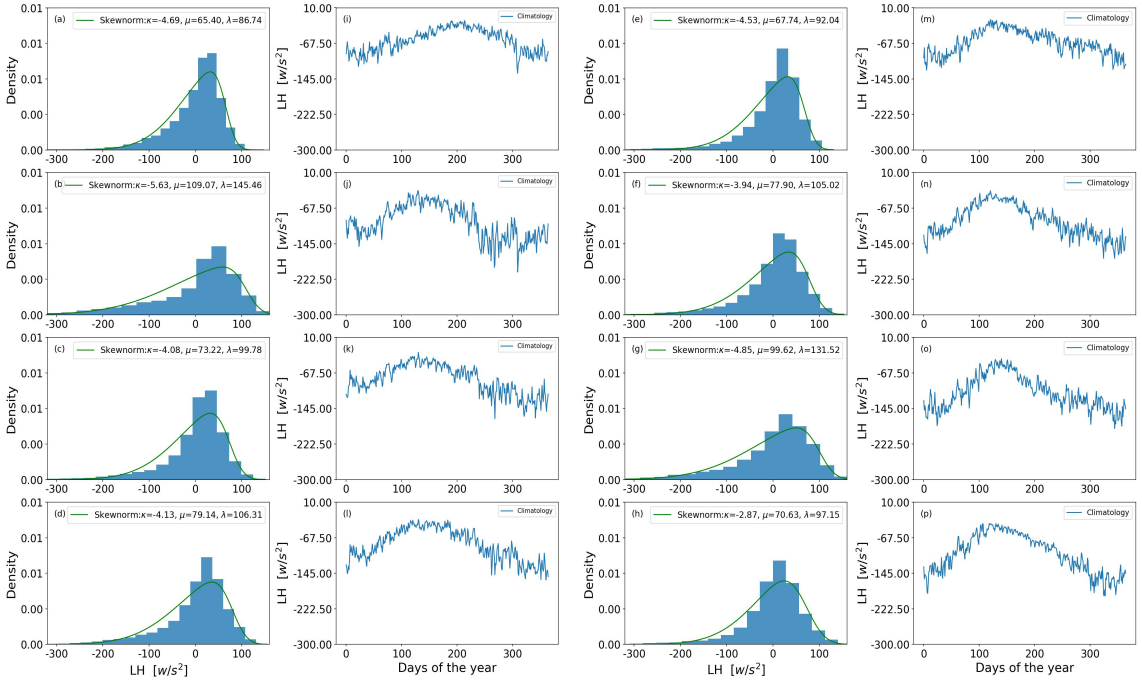

FIGURE 3.9: The single grid point histograms for LH flux anomalies (1st and 3rd columns) at the eight sampling locations for the period of 1991-2020, where 2nd and 4th columns represent the climatology of respective sea points.

in the northern Adriatic Sea region and north-western coastal areas. The statistical moments of the observed LW radiation and skew-normal PDF are computed and presented in Figure 3.14 . There is large difference in mean distribution in between observed flux and theoretical PDF. We assume this variation is due to shape (α) parameter distribution which is distributed from negative to positive and the theoretical skew-normal PDF mean depends on this parameter. For both, SH and LH fluxes, we have observed either positive and negative shape values completely and the means are matched with observed data means. Only the variance distributions are matched with similar pattern. Again, the skewness and kurtosis are not similar as for the other fluxes.

We have found a significant difference in between the theoretical PDF mean and observed long wave mean. We assume this is due to estimation of the parameter values used in the computation of theoretical skew-normal PDF mean. This issue is applicable for the all the fluxes, not only for the LW flux. This is because of the possible error generated during in the estimation of parameter values. The theoretical mean of skew-normal PDF is computed using all three parameters. But we have observed a good agreement in the variance distribution in both observed LW radiation

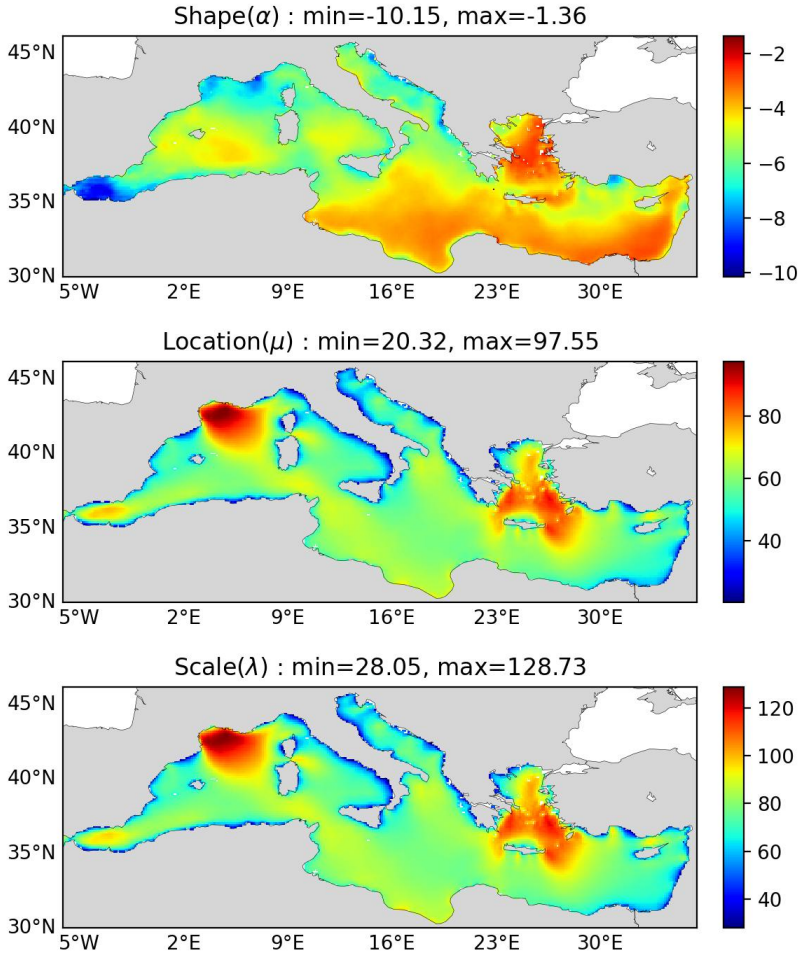

FIGURE 3.10: The skew-normal PDF parameter (α, μ, λ) distributions for computed LH flux anomaly for the observation period of 1991-2020.

anomaly and skew-normal PDF fitting. The skewness distributions are matched at some extent, but not a completely match over the interior basin. Mostly in the upper latitude area, there is a similarity in the skewness distribution with a range of minimum -0.45 to maximum -0.15. The theoretical skew-normal kurtosis does not match with the observed skewness in general. However, we find some areas of the coastal regions of northern Gulf of Lion and Adriatic Sea where both kurtosis distributions show a match in the value range.

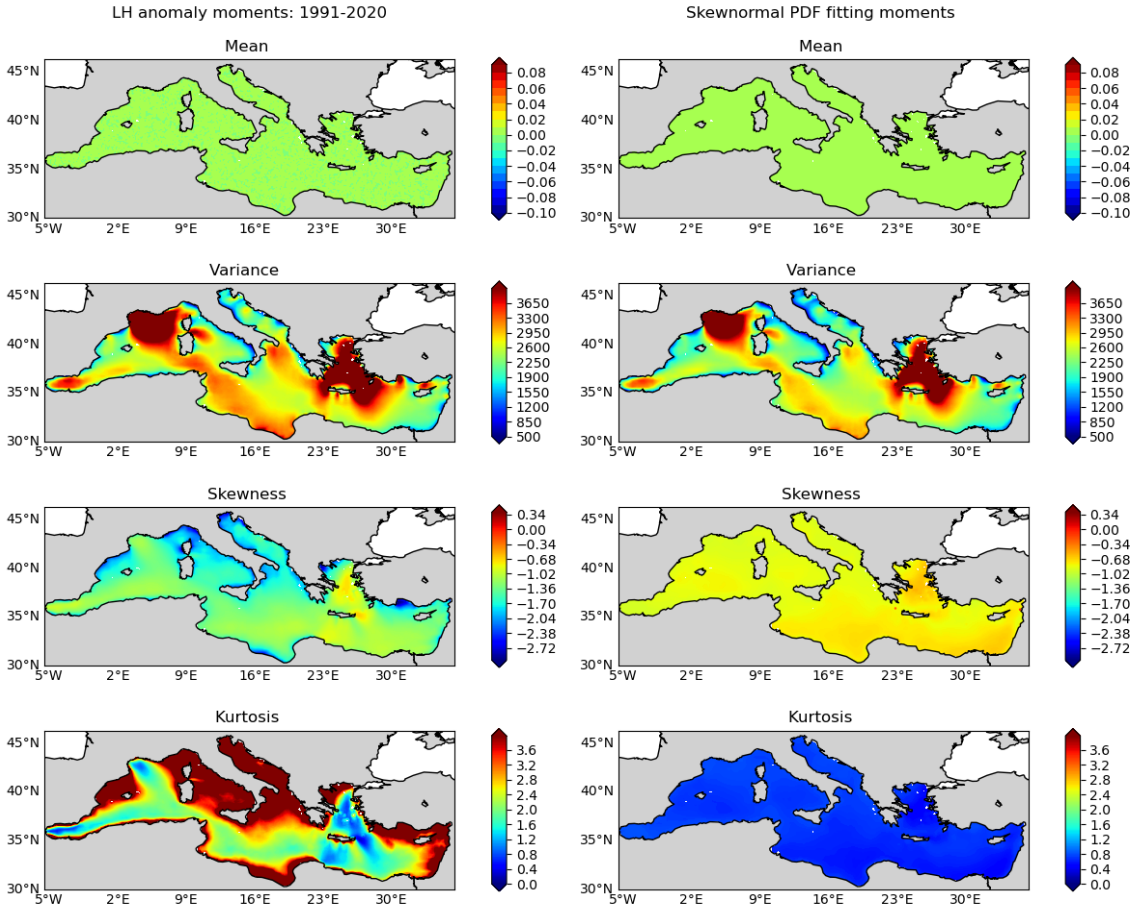

FIGURE 3.11: Comparison of the statistical moments (mean, variance, skewness, kurtosis) for observed LH flux anomaly and skew-normal PDF fitting parameters, 1991-2020.

5 Comparison of high- and low-resolution heat fluxes

5.1 Climatological differences

An attempt has been made to compare the statistical analysis results of heat fluxes obtained from the high-resolution ECMWF dataset with the heat fluxes computed from the lower resolution ERA5 dataset. The ERA5 dataset has been used in many studies and considered reliable given the consistent use of the same model and assimilation system despite its lower resolution.

We recompute the heat fluxes for the 20-year period, 2001-2020 using the ERA5 dataset, and compare with the fluxes computed from ECMWF dataset (Figure 3.15).

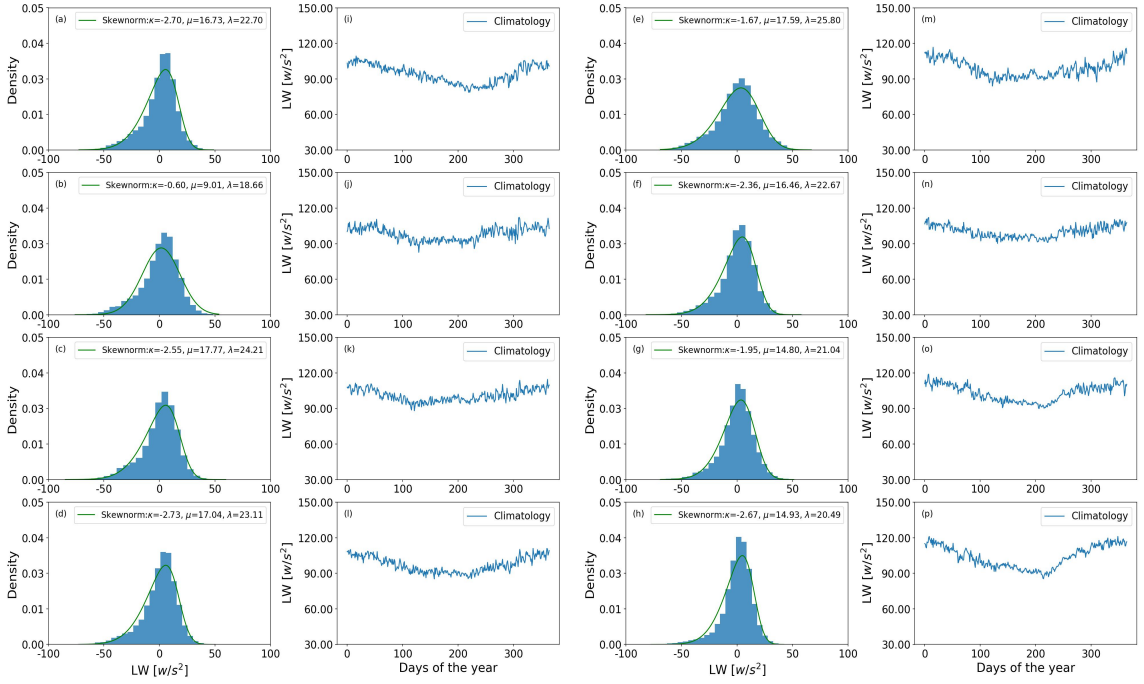

FIGURE 3.12: The single grid point histograms for anomalies of LW flux (1st and 3rd columns) from the eight sampling locations for the period of 1991-2020, where 2nd and 4th columns represent the climatology of respective sea points.

Heat fluxes structure are comparable except for the LW where the differences are notable. The annual SH means are distributed between +2 W/m2 and -32 W/m2 in both datasets. The LH flux shows a smaller heat loss in the western Mediterranean with respect to ERA5 computed fluxes, while on the contrary, the eastern Mediterranean loses more heat with the ERA5 dataset. Comparatively, the eastern Mediterranean gains more heat with ERA-5 dataset as we find the SW radiation range is very high in that area including the central Mediterranean The SW radiation differences are indicative of differences in cloud cover amounting to a larger heat gain in the eastern Mediterranean probably due to an inaccurate estimation of the clouds by ERA5. Overall, The heat fluxes of ERA5 dataset shows a limited difference on the results of turbulent heat fluxes, while the difference is significant for the radiative heat fluxes.

But the advantage of ECMWF analyses over the ERA5 to estimate fluxes is evident from the annual mean Q_{net} difference. The Mediterranean Sea gains comparatively more heat with the ERA5 inputs. The Q_{net} value of ECMWF varies between -30 W/m2 to +55 W/m2, when this range is -15 W/m2 to +55 W/m2 for the Q_{net} of ERA-5 (Fig. 3.16). Besides the difference in surface domain for Q_{net}, for both cases air-sea flux dynamics is strongly visible in the easter Mediterranean for net heat gain,

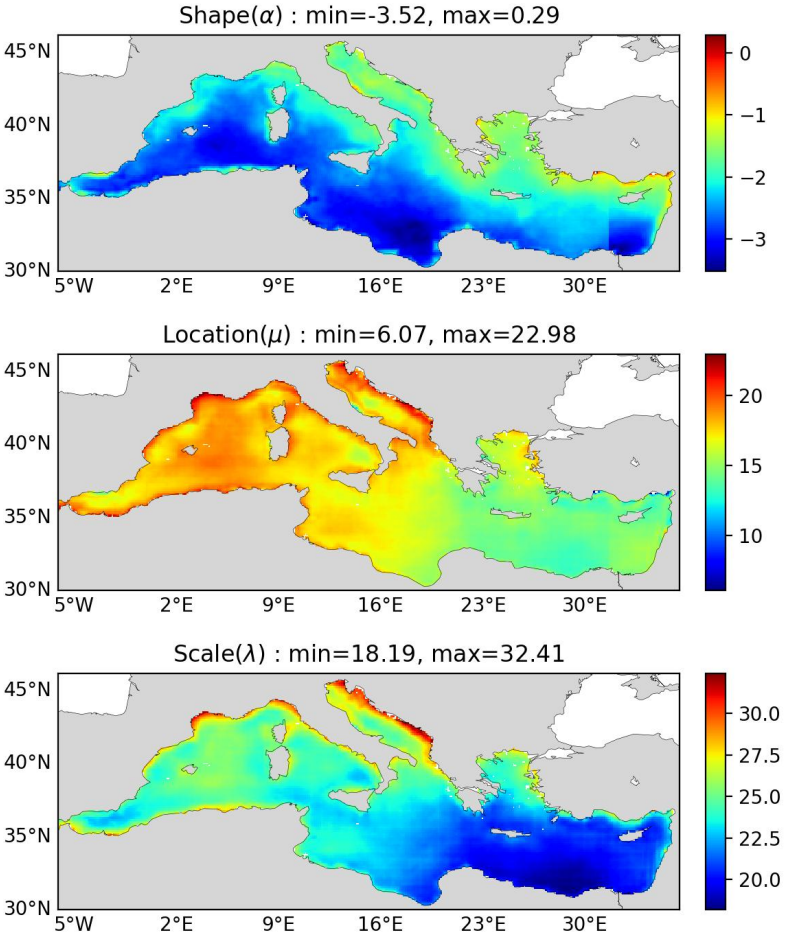

FIGURE 3.13: The skew-normal PDF parameter (α, μ, λ) distributions for computed LW flux anomalies for the observation period of 1991-2020.

in the Gulf of Lion for heat loss due to the continental cold wind (Mistral wind) and in the Aegean Sea due to strong wind (Etesian wind) that blows during the summer period. Calculating the mean basin average value, the ECMWF Q_{net} amounts to -3. W/m2, while for ERA5 it is 18 W/m2.

Thus, it is found that the minimum resolution of atmospheric fields required to have a correct heat budget closure problem for the Mediterranean Sea is given by the analyses at 0.125-degree resolution.

5.2 Evaluation of the PDF fitting and uncertain areas

In this part, an attempt to verify the probability analysis for turbulent surface fluxes has been done with a goodness of the fit test. The objective of this evaluation test

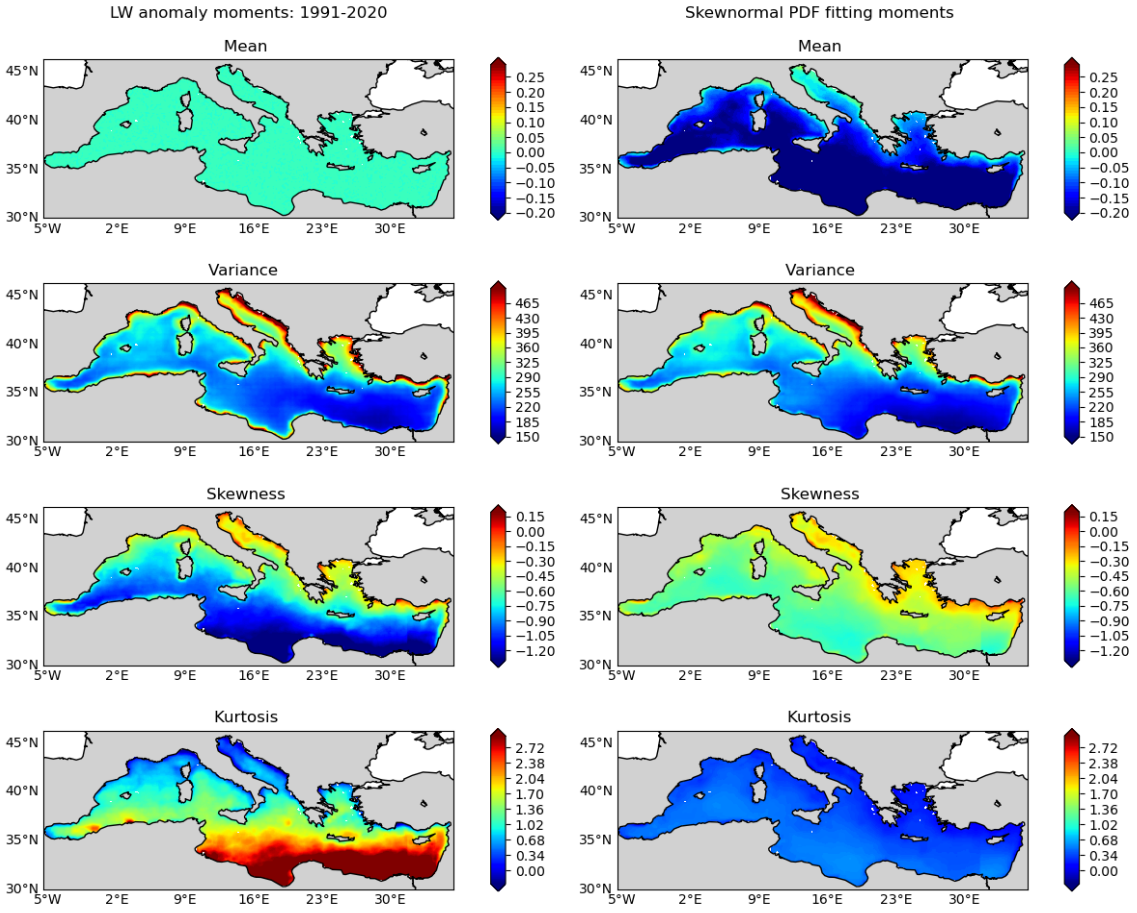

FIGURE 3.14: Comparison of the statistical moments (mean, variance, skewness, kurtosis) with observed LW flux anomalies and skew-normal PDF fitting parameters, 1991-2020.

is to check, how the applied theoretical distribution is reasonably matched with the observed time series. Chi-squared method is a well-accepted test to measure the distance between two independent distributions.

Here, we verify the PDF fit for turbulent surface fluxes with a Chi-squared test method. Figure 3.17 presents the chi-squared test results for the turbulent heat fluxes computed using ECMWF and ERA-5 datasets. The decision rule for the χ^2 test is decided based on the level of significance (in our case it is set at 0.05) and the degrees of freedom, defined as df = N-np, where N =number of bins(25) and np is the number of distribution parameters (i.e. 3 for both ALD and Skew-normal distribution), so the critical value of χ^2 is 34 is set for the test.

For the ECMWF fluxes there are relatively large areas with higher values than P: they are observed near the coastal regions and in the Alboran Sea, Balearic Sea, Tyrrhenian

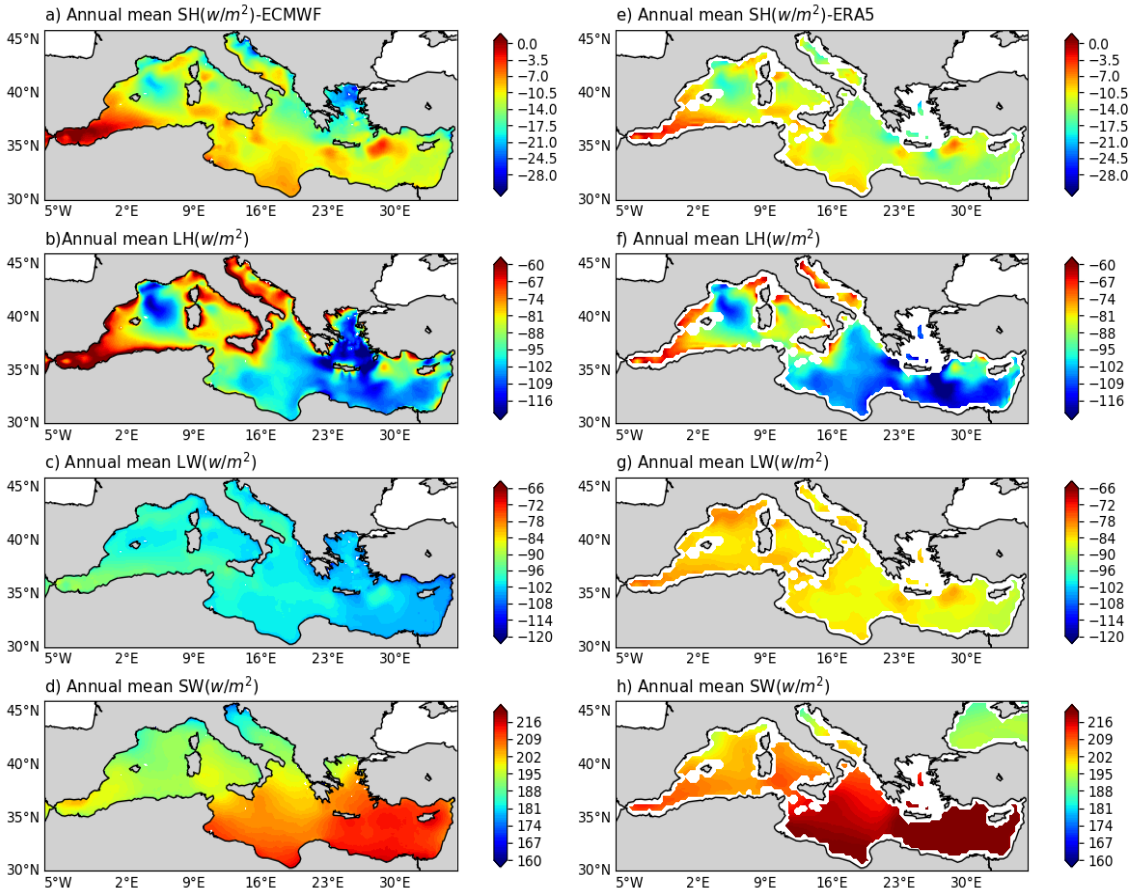

FIGURE 3.15: Comparison of the long-term climatology for heat fluxes computed from ECMFW (left panel) and ERA5 (right) datasets.

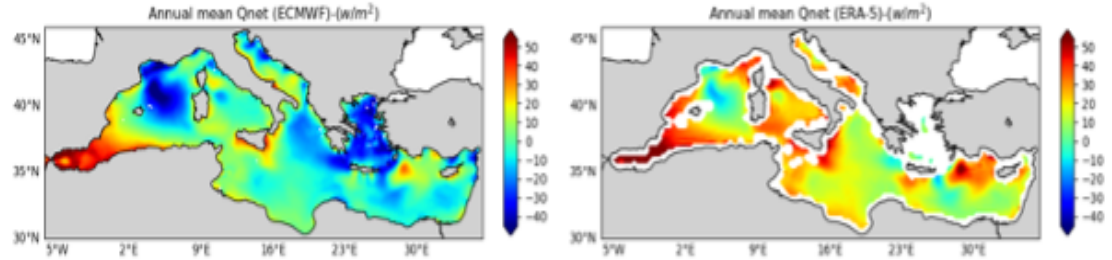

FIGURE 3.16: Comparison of the annual Q_{net} computed from ECMWF (left panel) and ERA-5 (right) input dataset

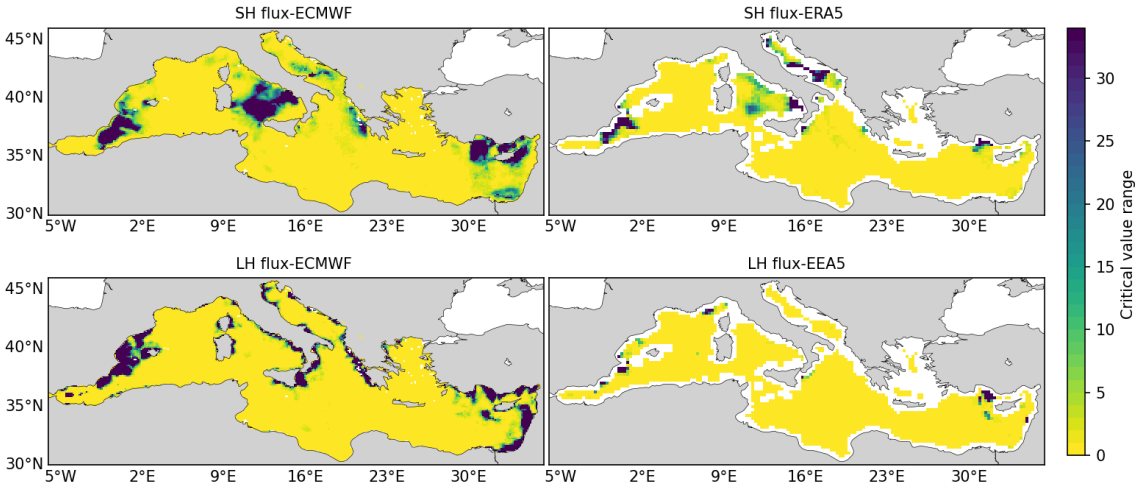

FIGURE 3.17: Comparison of the Chi-squared tests between turbulent heat fluxes for ECMWF (left panel) and ERA-5 (right) datasets.

Sea, Adriatic Sea, and Levantine Sea. A large uncertain region is common for both SH and LH in the Alboran Sea and Balearic Sea areas.

In addition, we compare the chi-squared statistics in the context of high- and low-resolution fluxes. In Figure 5.3, the right panel presents the similar chi-squared test result obtained for the fluxes computed for ERA-5 data against applied PDFs. On comparison, the ERA5 dataset shows a better match between the fitted PDF and the data than ECMWF.

The chi-squared results of SH and LH fluxes (ERA-5 input) show that most of the surface grid points are well fitted with the applied theoretical PDFs (ALD and skew-normal). If we consider the few uncertain region, there are few small areas presented with higher P values than critical value 34 (Elderton, 1902), mostly for SH flux (ERA-5 input) and some small areas for LH flux (ERA-5). It's important to mention that the uncertain regions (ERA-5 flux) are matched with the regions for fluxes with ECMWF inputs.

6 Conclusions

A primary motivation for this investigation is to establish a methodology for the investigation of the probability distribution of heat-fluxes using atmospheric variables that are used to force the ocean models and forecasts. The final aim is to assess the

uncertainty of ocean model atmospheric heat fluxes that could produce errors in the ocean stratification conditions and water mass structure.

This study has analyzed probability distribution analysis of heat fluxes obtained computationally using ECMWF atmospheric analyses input at high horizontal resolution (1/8°*1/8° grid) for the Mediterranean Sea. In addition, we have computed fluxes using lower resolution atmospheric input from a reanalysis dataset (ERA-5) with an objective to compare with the net heat flux of higher resolution (ECMWF) flux PDF. We used the land-sea mask from ERA-5 considering values only on the mask=0 grid points. This gives us confidence that we do not use a value of SST contaminated by land.

Firstly, the heat budget of the Mediterranean Sea is analysed for average annual mean and seasonal variations. The largest heat budget component is the net solar radiation SW, followed by the longwave radiation LW, the latent heat LH third and then the sensible heat SH. All heat flux components display significant seasonality (Fig: 3.2). The basin average, net heat flux, Q_{net} resulted at -3 W/m2 for the 20-year period (2000-2020), excluding the initial 10 years from the time series, which is within an acceptable range with the well-known negative value required to close the Mediterranean Sea heat balance (Table 3.1). However, it is found that this value is not achievable for ERA5 computed fluxes, thus pointing for the first time to the required resolution of atmospheric forcing to close the budget.

Secondly, we have demonstrated that like atmospheric variables (shown in chapter 2), the probability density of surface heat fluxes can also be modelled and fitted with a three parameter PDF composed of a shape, a location, and a scale parameter. For this part, we have covered a 30-year observation period to compute the heat fluxes as well as for the probability density analysis. For LH and LW fluxes, the skew-normal was used, while for SH an Asymmetric Laplace Distribution was required due to presence of double picks in the distribution. There is an encouraging agreement between the first two statistical moments of the fitted PDF and the observed values. Even the asymmetry of the fitting PDFs has shown a good agreement some cases well, but the large difference is observed for all the kurtosis comparisons. However, there is a notable difference in the skewness and kurtosis comparison.

The comparison of radiative heat fluxes between high-resolution and low-resolution atmospheric input has shown very different values, especially for the LW flux. This

result is actually in agreement with the conclusion by Marullo et al. (2021) on the sensitivity of LW estimates from the atmospheric dataset used to calculate fluxes. However, we intended to use the Mediterranean heat budget as a means to evaluate the quality of the data set used to compute the pdf. Heat fluxes pdf that have been estimated in this chapter, will allow in the future to understand the importance of extreme events to compose the net negative heat budget of the Mediterranean Sea. Furthermore, heat fluxes pdf could be used to understand if perturbations in atmospheric variables produce realistic heat flux perturbations

Chapter 4

Assessment of the probability distribution results

1 Introduction

In Chapter 2, we analysed the time series of atmospheric variables over a 30-year period for an atmospheric analysis dataset as the output of a medium-range weather prediction model. Then, in Chapter 3, we calculated the heat flux components using the atmospheric forcing variables and modelled the calculated heat fluxes (SH, LH, LW) for the probability distribution analysis.

In this chapter, we will discuss the probability distribution results based on the PDFs used and their calculated parameter range in terms of statistical inference. We are interested in how these PDF analysis results may be statistically significant in evaluating ocean forecast uncertainty.

1.1 PDF parameters of the atmospheric variables

Through the extended statistical analysis parts, we have captured spatial patterns and distribution ranges for the PDF parameters that help us interpret the probability distribution results. Table 4.1 provides a summary of the PDF parameters of the atmospheric variables with their maximum and minimum ranges.

In this chapter 4, we have introduced an approach to study the computed PDF parameters using a clustering technique which is used in unsupervised machine learning algorithms. The objective is to identify the central clusters and their range along with outliers of the PDF parameter values as a function of the geographic position. The different clusters identify regions of different dynamics which result in different pdf parameters.

To investigate the relationship between shape and scale parameters, a clustering analysis is performed using the K-Mean cluster. The K-Means clustering is performed for the values of the shape and scale parameters calculated for the Weibull PDF over the points of the Mediterranean Sea. Figure 4.1-b shows the scatter plot for the wind amplitude and there are two well identifiable clusters. The range of scale parameters is between 2.5 and 6.5 and represents the strength of the winds (Wang et al., 2018). Cluster-1 falls in the relatively lower wind amplitude range with a shape parameter (κ) in the range of 1.2 to 2, while Cluster-2 peaks at 4.5 to 7 for the scale parameter (λ) and a range of 1.4 to 1.9 for the shape parameter. Therefore, cluster-2 is centred

around higher wind amplitudes supported by lower values of the shape parameter. Before running the clustering analysis, we conducted the silhouette analysis which provides a score indicating the of number of clusters can be used for the clustering of a variable. In those preliminary silhouette scores, we found two clusters are indicative in our clustering analysis.

According to Wang et al. (2018), when the shape parameter is between $2<\kappa<3$, it tends to be a stable wind amplitude, while if $\kappa > 3$, then it becomes a symmetric normal distribution. In a separate clustering of the scale parameter (λ), we find cluster-2 to be larger with stronger picks, indicating a clear Weibull structure (Fig.4.1-a).

We have performed the clustering analysis for the wind components ($\tilde{U}$, $\tilde{V}$) using a similar approach as described for the wind amplitude. It is quite evident that the clusters are now very distinct for both the zonal and meridional wind components (Fig.4.2, Fig.4.3). Here, the central values of the shape parameters for the wind components are almost identical. The shape value for cluster 1 is 1, while for cluster 2 it is -1 (Fig. 4.2-b,). For the scale parameter, the central values for both cluster 1 and cluster 2 are 4 m/s (Fig.4.2-b 4.3-b). Since the scale parameter regulates the dispersion of the PDF, both clusters have the same dispersion. However, the shape parameter controls the direction and magnitude of the skewness, so the two clusters have opposite skewness or opposite anomaly values. The clustering results of only the scale parameter λ have shown a variation of two clusters for wind components ($\tilde{U}$, $\tilde{V}$), where cluster 1 is large for wind $\tilde{U}$, while for wind $\tilde{V}$ cluster 2 is denser and larger than cluster 1 at 4 m/s. for both clusters 1 and 2 (Fig.4.2-a, Fig.4.3-a). Since the scale parameter regulates the spread of the PDF, we understand that both clusters have the same spread. However, the shape parameter controls the direction and the magnitude of the skewness so that the two clusters have opposite skewness or opposite anomaly value tails. The clustering results of only scale parameter λ have shown a variation two clusters for wind components ($\tilde{U}$, $\tilde{V}$), where cluster 1 is large in wind $\tilde{U}$ while for wind $\tilde{V}$ cluster 2 is denser with 4 m/s and larger than cluster 1.

To determine a possible relationship between shape and scale parameters, cluster analysis is also applied to air temperature and dew point temperature. Figure 4.4-b shows two clusters assigned for the scatter of shape and scale parameters of air temperature, one corresponding to the larger cluster (cluster-2) of shape and scale parameter values and the second cluster-1 to a smaller cluster. The cluster-1 is marked with a smaller scale parameter range of 1.5 to 3.5 with a range of 0.5 to 2.5 for the

TABLE 4.1: Range of PDF parameters (shape, location, and scale) for atmospheric variable anomalies over the entire basin and for the period

Atmos. variables (anomalies)	**PDF**	**Location**	**Shape**	**Scale**
Wind $\tilde{U}$	skew-normal	Min -5, Max 7	Min -4.4, Max 2.5	Min 1, Max 9
Wind $\tilde{V}$	skew-normal	Min -4.7,Max 5.4	Min -4.3, Max 2.5	Min 0.8, Max 7.4
Wind amplitude	Weibull		Min 1.2, Max 2	Min 1.5, Max 7
T2M	skew-normal	Min -2.6,Max 5	Min -3.5, Max 2.4	Min 1.4, Max 6.5
D2M	skew-normal	Min 1.3,Max 5.9	Min -6.1, Max -1	Min 2.3, Max 7.6
MSL-P	skew-normal	Min 1.5,Max 6. 2	Min -2, Max 0.7	Min 3.2, Max 9.1

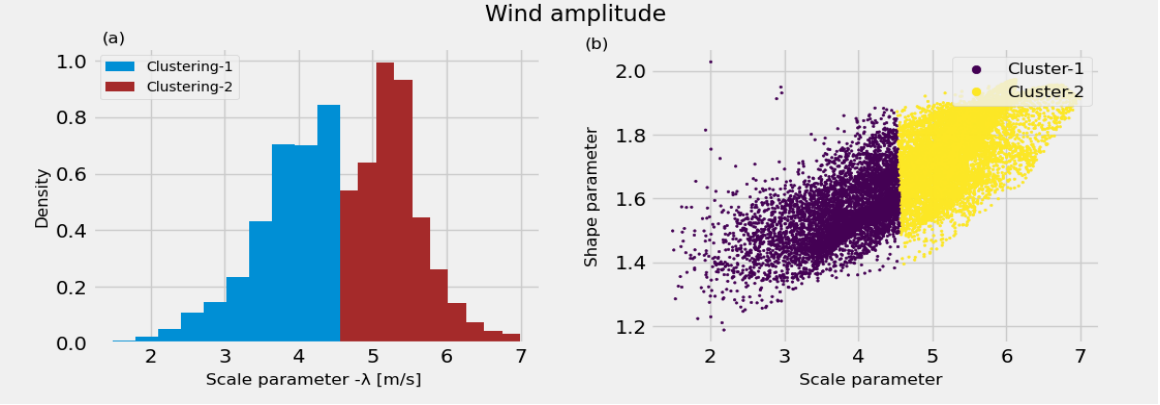

FIGURE 4.1: (a). Clustering of the Weibull PDF scale (λ) parameter (b) Scatter of the plots for Weibull PDF shape (κ) and scale (λ) parameter with assigned clusters

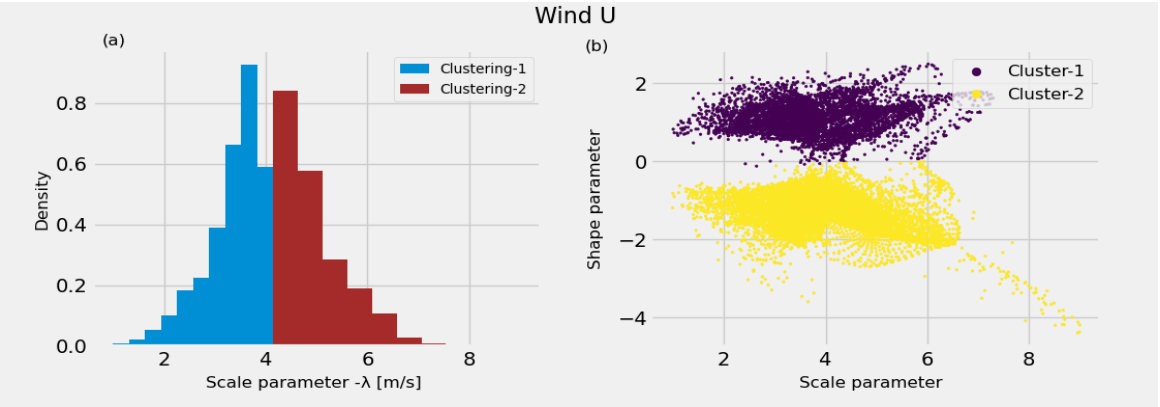

FIGURE 4.2: (a) Clustering of the scale parameter (λ) of the skew-normal PDF for the zonal component of the wind anomaly ($\tilde{U}$) (b) Scatter plot for the shape (α) and scale parameter (λ) with assigned clusters

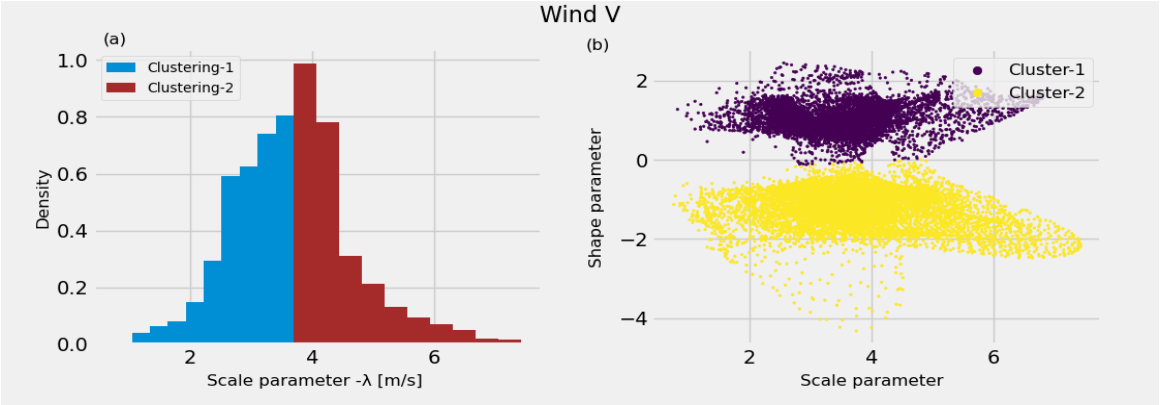

FIGURE 4.3: (a). Clustering of the skew-normal PDF scale parameter (λ) for the meridional component of the wind anomaly ($\tilde{V}$) (b) Scatter plot for shape (α) and scale parameter (λ) with assigned clusters

shape parameter. This time the temperature anomalies do not have the same scale central value for clusters 1 and 2. This means that the spread is different at different locations in the domain. The shape is also different meaning that the PDF has also different slopes in the domain in the positive and negative skewness directions. Cluster 2 is denser with 3 (Kelvin) and has more spread in the clustering result of scale parameter only (Fig. 4.4-a).

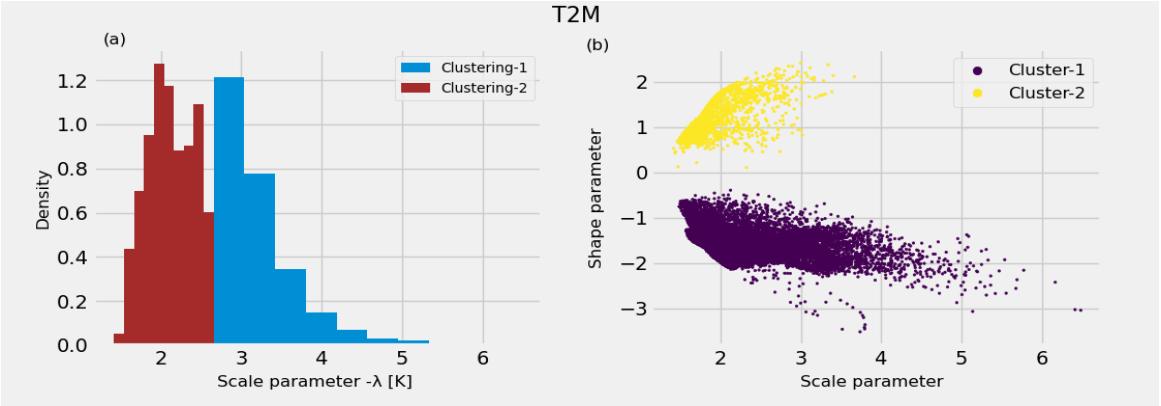

FIGURE 4.4: (a). Clustering of the skew-normal PDF scale parameter for air temperature (b) Scatter of the plots for shape α and scale (λ) parameters with assigned clusters

Under the cluster analysis of the dewpoint temperature PDF parameters (4.5-b), we find that cluster-1 is smaller than cluster 2 and cluster-2 values in the scale parameter range $\lambda > 3.2$. So, we expect distributions with larger spread and larger negative skewness since the shape parameter central value is -3. Thus cluster-2 of scale parameter (λ) dominates the dew-point distribution with long negative tails.

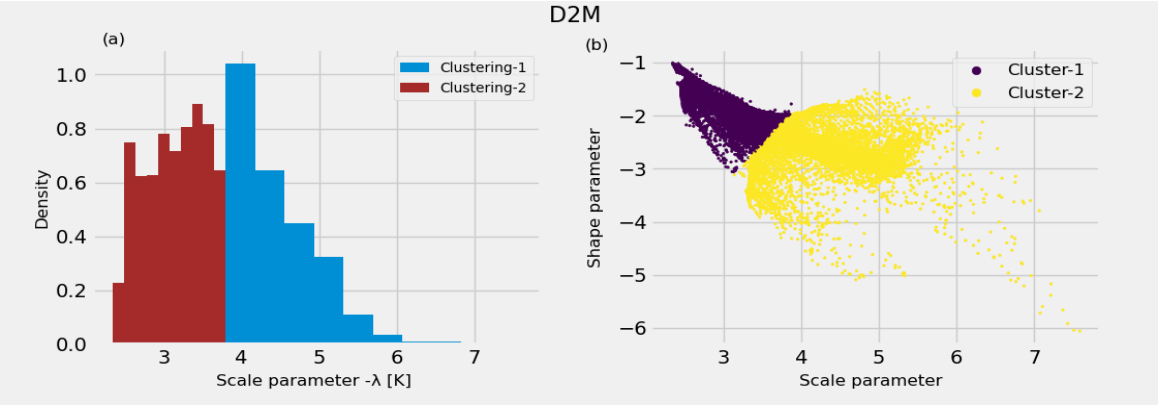

FIGURE 4.5: (a). Clustering of the scale parameter of the skew-normal PDF for the dew point temperature (b) Scatter plots for shape (α) and scale parameters (λ) with assigned clusters

Fig. 4.6-b shows the results of clustering for the PDF parameters of mean sea level air pressure. Cluster-2 is larger than cluster-1, and they are centered around different values of the scale and shape parameters, while cluster-1 is centered around smaller values of the scale and shape parameters. When in the clustering of the scale parameter λ, two clusters are almost equally distributed (Fig. 4.6-a).

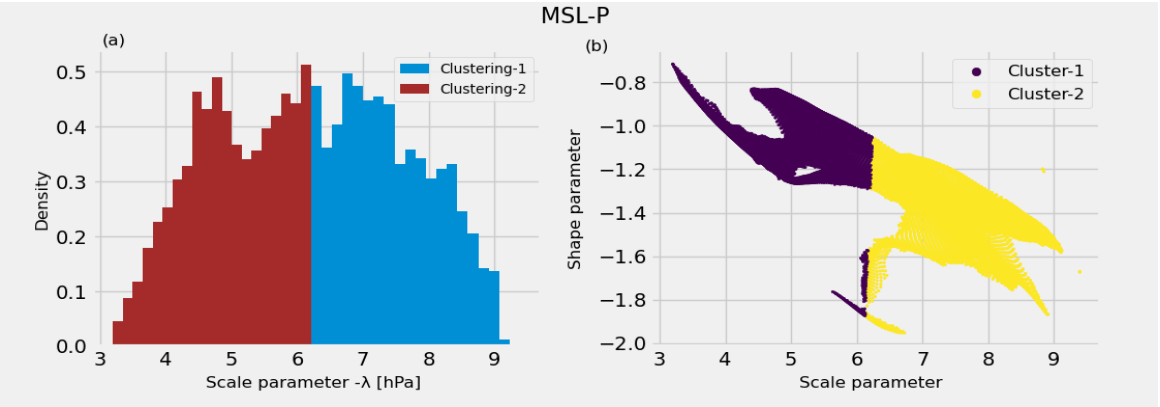

FIGURE 4.6: (a) Clustering of the skew-normal PDF scale parameter for mean sea-level pressure (b) Scatter of the plots for shape (κ) and scale (λ) parameter with assigned clusters

1.2 Significance of the PDF parameter values

In our investigation, we have added the Wind components (U, V), which are not a common approach within the related atmospheric wind field study, rather wind speed is the most studied one in published literature. Moreover, most of the studies have dealt with winds over land, but in the real-world surface ocean winds are stronger

than land because of less surface friction (Wallace and Hobbs, 2006). In the published studies, it's proven that the Weibull PDF with 2 parameters is a powerful and convenient PDF to measure the density of surface wind (Pérez et al., 2007; Wais, 2017). But for wind components, it's normally assumed to follow a Gaussian distribution, which is not true always and the Gaussian curves some time fail to shape the tails in the distribution in a histogram (Drobinski et al., 2015). We have observed similar non-Gaussianity in our histograms for wind components (Chapter 2, section3.2). Therefore, it's an important part of this study and a newer approach for the Mediterranean Sea; and we have applied a three-parameter skew-normal PDF to cover the non-Gaussian features along with retaining the Gaussianity.

From the climatological analysis for February and August (Chapter 2, section 2.2), we have observed the difference in the distribution pattern of both wind components ($\tilde{U}$, $\tilde{V}$). This pattern is divisible between the western and eastern Mediterranean Seas. Particularly, we have identified two local wind regimes in the Alboran Sea and the Aegean Sea which are known as Mistral and Etesian winds (Brossier et al., 2011; Drobinski et al., 2005; Guéenard et al., 2006). We find a difference in scale parameter values between wind $\tilde{U}$ and wind-$\tilde{V}$ which is noticeable (Table 4.1). As the scale parameter indicates the speed of the components, $\tilde{U}$ or zonal component speed is distributed in a larger range than the meridional component ($\tilde{V}$). On the contrary, $\tilde{V}$ is in a more stable range with a maximum value of 7.43 (m/s). So far to our best knowledge, we have So far to our best knowledge, we have not found any study that is done particularly for the Mediterranean Sea on the PDF analysis of wind components. By referring the clustering analysis, we can approximate an average distribution speed range of the wind components.

Under table 4.2, we present a summary result to infer the line between an average distribution value range and a probable uncertainty range that could make variables unstable in their distributions. In both cases, for ($\tilde{U}$, $\tilde{V}$) cluster 2 is estimated with a higher value range with an indication of spread in the distribution than the average one. Through these clustering results, we can approximately assume the clusters -1 is mainly composed of the normal distribution range.

The PDF distribution results for wind amplitude have been verified well through the statistical moment comparison and can be referred to in the literature also. In

TABLE 4.2: PDF parameters value range (shape and scale) for the two clusters found for each atmospheric variable analyzed

Atmospheric variables (anomalies)	Cluster -1 range		Cluster range -2	
	Shape	Scale	Shape	Scale
Wind $\tilde{U}$	0 to 2.5	1.7 to 7	0 to -4.4	1 to 9
Wind $\tilde{V}$	0 to -2.5	0 to 6.5	0 to -4.3	0 to 7.4
Wind amplitude	1.5 to -4.5	1.2 to 2	1.4 to 1.9	4.5 to 7
T2M	0.5 to 2.2	1.4 to 3.5	0 to -3.5	1.4 to 6.5
D2M	-1 to 3	2.4 to 3.9	-1.5 to 6.1	3.2 to 7.9
MSL-P	-0.7 to -1.9	3.2 to 6.1	-1.1 to 1.9	6.1 to 9.1

the Mediterranean Sea basin, the average wind amplitude is 6 m/s (Pettenuzzo et al., 2010), while our Weibull PDF parameter analysis provides a range of 1.49 to 6.99 m/s in the scale parameter range. If we segregate this range into clusters, again cluster-2 comes with a higher range of 4.5 to 6.99 for a possible uncertain range. Monahan (2006) studied sea-surface winds on a global scale and confirmed the previous results that surface wind amplitude fits well with 2 parameter Weibull PDF. If we just refer to a Weibull PDF scale parameter globally, Monahan (2006) presented a result of 1-6 (m/s) for surface ocean wind.

Likewise, for wind components, for our study area, no particular study is reported on the probability distributions of air temperature, dew point temperatures, and mean sea-level pressure. In numerical ocean forecasting, air temperature and dew-point temperature are important forcing variables and are used as inputs in computing air-sea fluxes internally. Several studies have dealt with surface air temperature but mostly focused on the extremes or tail of the temperature PDF. Approximately, temperatures are normally distributed, but they may differ at the near-surface distribution with a heavy tail, and it can be skewed positively or negatively (Von Storch and Zwiers, 2002). The non-Gaussian behavior of temperatures with heavy tails can be linked to Generalized Extreme Value (GEV) distribution (Kyselỳ, 2002; Yiou et al., 2008). In our study, we are not particularly interested in the extremes but rather more focused on the density of the distributions.

We have explored that a two-parameter PDF is might not a good fit for the air temperature and dewpoint temperature but fits well with a three parameters skew-normal PDF. The scale parameter (λ) is identical to the observed temperatures. The air temperature and dew-point anomaly ranges are detected for a value of 1.42 to

TABLE 4.3: Range of the PDF parameters (shape, location, and scale) for the atmospheric heat fluxes over the Mediterranean Sea basin (1991-2020)

Heat fluxes	**PDF**	**Location**	**Shape**	**Scale**
SH	ALD	Min -4.9,Max 19.7	Min 0.7, Max 1.4	Min 6.2, Max 33.8
LH	skew-normal	Min 20.3, Max 97.5	Min -10.1, Max -1.3	Min 28.0, Max 128
LW	skew-normal	Min -3.5 , Max 0.29	Min 6, Max 22.9	Min 18.1, Max 32.4

6.48 K and 2.34 to 7.59 respectively (Table 4.1). In terms of an average and probable extreme, the clustering analysis provides an overview of the temperature distribution threshold. For instance, cluster 1 estimates the average range of anomalies of air temperature and dewpoint temperature with 1.42 to 3.5 and 2.39 to 3.9 respectively. Cluster-2 has larger anomaly values, and we have observed the asymmetry in the distribution range. We assume that cluster 2 is composed of probable extremes that can be tested in future experiments of ocean forecasting. Similarly, cluster 2 tends to be an unstable range in the mean sea-level pressure anomaly, whereas we can consider cluster 1 for an average range in the Mediterranean. This average range of 3.2 to 6.1 (hPa) mainly in the eastern part divides the distribution pattern from the western Mediterranean Sea.

1.3 PDF parameters of heat fluxes

In addition to addressing the net heat budget closure problem, in chapter 3 we have analysed the probability distribution of heat flux components – sensible heat (SH) flux, latent heat (LH) flux, and longwave radiation. We present a summary of the heat fluxes PDF parameter results in table 4.3

To our best knowledge, for the Mediterranean Sea, no dedicated study is reported on the probability distribution of heat fluxes. Therefore, this study can be considered as an initial approach to investigate the probability distributions of heat fluxes. From our PDFs, we can signify the scale parameter as identical to the structure of flux components. Likewise, for the atmospheric variables, an additional shape parameter is necessary to fit on the flux distribution because we consider the anomaly of the flux distribution distributed with positive and negative values. We assume it will be

further work to verify this PDF results with longer time series using higher resolution atmospheric input data. However, clustering analysis provides an overview of the average heat fluxes and probable extremes in the distribution.

We have applied the K-mean clustering analysis on heat fluxes – SH flux, LH flux, and longwave. Fig. 4.7-b shows the two clusters in between a scatter plot of shape (α) and scale parameter (λ) for SH flux fitted with ALD PDF.

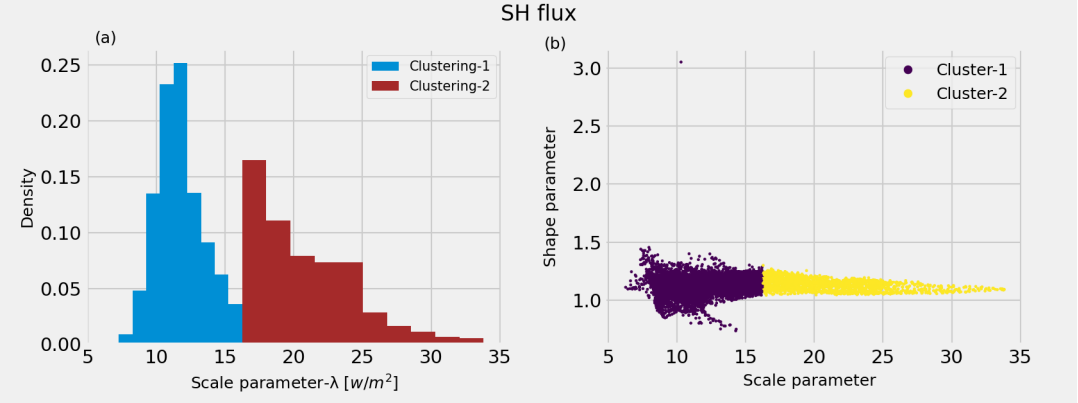

FIGURE 4.7: (a). Clustering of ALD PDF scale parameter for SH flux (b). Scatter of the plots for shape (α) and scale (κ) parameter with assigned clusters

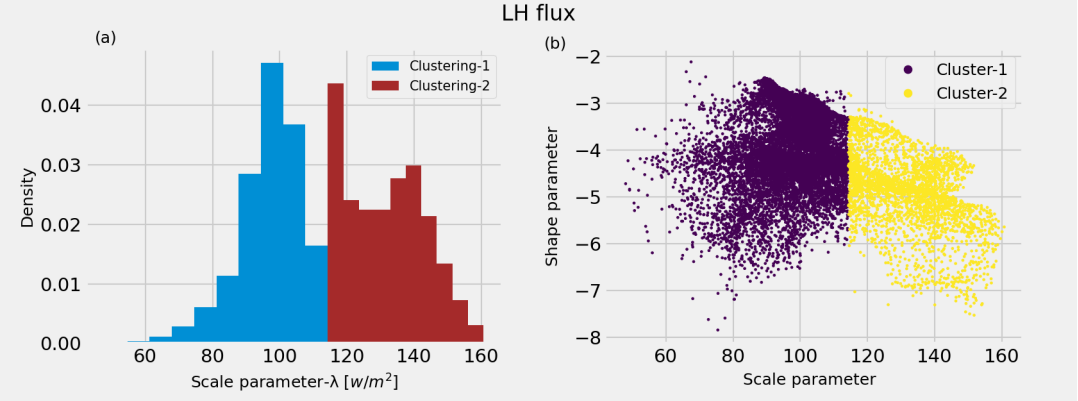

FIGURE 4.8: (a). Clustering of the skew-normal PDF scale parameter for LH flux (b). Scatter of the plots for shape (α) and scale (λ) parameter with assigned clusters

Clearly, the cluster 2 shows a large spread in the distribution with λ values > 16 with lower range of α values 1.0-1.25. We can consider the cluster 1 as our average SH flux. Clearly, cluster 2 shows a large spread in the distribution with λ values > 16 with a lower range of α values 1.0-1.25. We can consider cluster 1 as our average SH flux PDF distribution with a λ value range 6-16 w/m^2. This feature is also verified in the clustering of scale parameter only (Fig. 4.7-a), where cluster 1 clearly

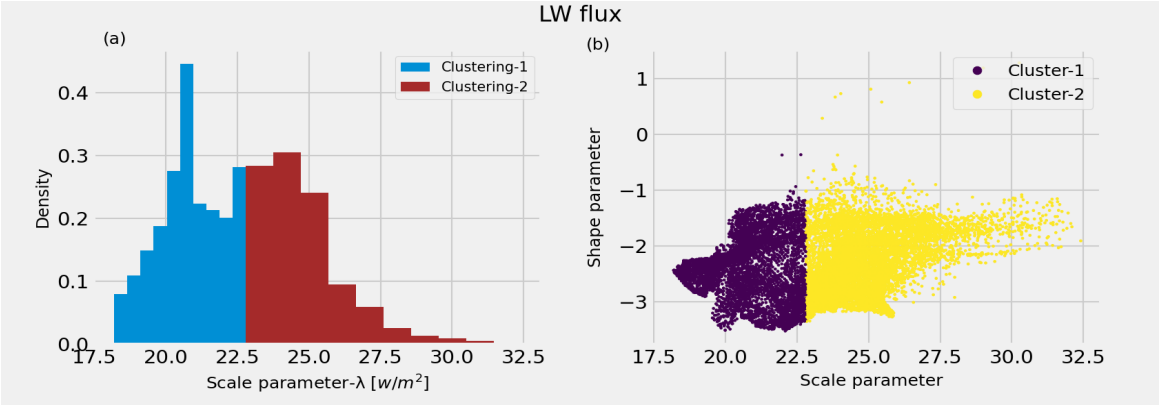

FIGURE 4.9: (a). Clustering of the skew-normal PDF scale parameter for longwave (b). Scatter of the plots for shape (α) and scale (λ) parameter with assigned clusters

shows a structure of double pick that fitted with ALD PDF. In this relation, we can hypothesize that cluster 2 is featured with extreme SH flux values (Fig. 4.7-a).

For the LH flux PDF parameter, cluster 1 is large but has spread in both ranges of scale and shape parameter values (Fig. 4.8-b). On the contrary, cluster 2 is ranged with higher LH flux values ($\lambda > 110$), while -5 is central for both clusters. In the clustering of scale (λ) parameter (Fig. 4.8-a), we find cluster 1 as denser in the center with 100 w/m^2 and 120 w/m^2 visible with a long pick for cluster 2.

Figure Fig. 4.9-b shows the two clusters for longwave PDF parameters fitted with skew-normal PDF. Cluster 1 is compact and dense in the density distribution while cluster 2 has a large spread and α values are almost equally distributed in both clusters with some spread for cluster 2. In the clustering of scale parameter (λ), we observe the high density for 20 w/m^2 in cluster 1 considering the average range, but probable extremes present in cluster 2 with a long tail (Fig. 4.9-a)

Since we have converted the observed heat fluxes into a daily mean anomaly time series, it cannot alone provide spatial patterns for extreme values. Regarding the extreme values distribution in the whole basin, we compute the 90th (q_{90}) and 10th (q_{10}) percentile of daily heat flux anomaly (Fig. 4.10). For the turbulent heat fluxes, q_{10} indicates the extremes of large negative (upward) fluxes, and q_{90} indicates the extremes of large positive (downward) fluxes. In this connection, we have pointed notably several key hot spots of intense air–sea interaction, which considerably regulates the heat and water budgets. From Figure. 4.10, we can mark the north-western

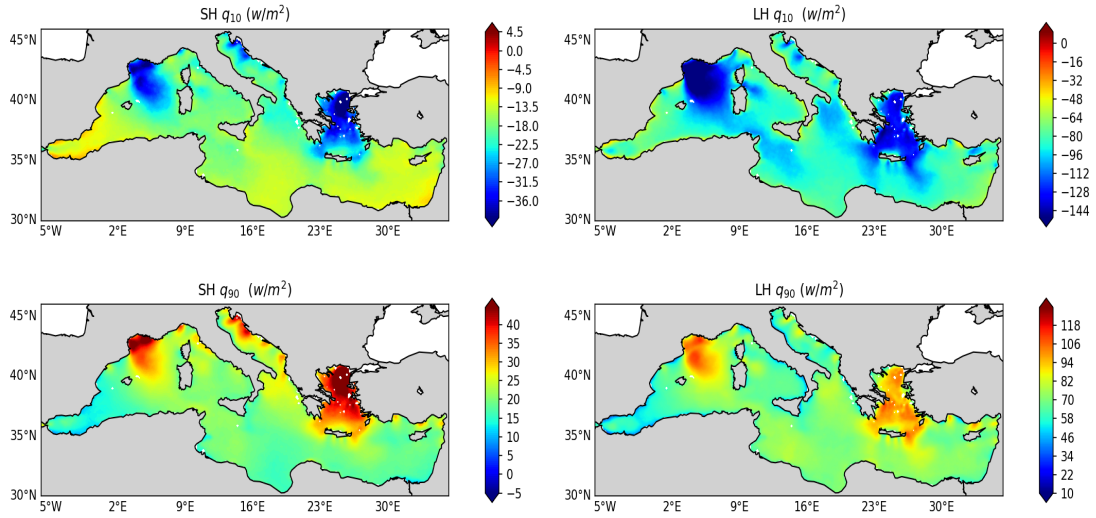

FIGURE 4.10: Distribution maps of 90th and 10th percentile values for turbulent heat fluxes in the Mediterranean Sea

Mediterranean, the Adriatic Sea, and the Aegean Sea which illustrates the regions for large-scale variability in net heat flux (Brossier et al., 2011).

1.4 Application of PDF results and future works

The probability analysis of the atmospheric variables provides a scope to utilise the result in the atmospheric forcing part for ocean forecasting. We have referred to some details of the background of this study in the background chapter (Chapter 1.3). Here we present an overview of how we can apply the PDF results in the experiments of uncertainty analysis.

In our next approach, we aim to analysis the state of possible uncertainty that arises in a short-term ocean forecasting system. This investigation is based on the Mediterranean Ocean Forecasting system which is called MFS-EAS1 short. The MFS system is an operational one using core ocean model NEMO and the atmospheric is used from the ECWMF atmospheric analysis fields. Using the atmospheric surface fields-surface wind, air temperature, dew-point temperature, mean sea-level pressure, and cloud coverage that combines with SST input the EAS1 system provides momentum and heat fluxes to the ocean model. The EAS1 system provides forecast outputs of ocean temperature, salinity, ocean currents, and sea level anomaly (SLA).

We have obtained the probability distribution results that have set the next step of future investigation. The climatological PDF study is the baseline for applying the atmospheric forcing in an ocean forecast. In short, the pdf at each grid point will be sampled to find a realistic perturbation for the specific instantaneous atmospheric forcing variable forcing the ocean predictive model. The resulting perturbed heat fluxes will be compared with the estimated heat flux pdf to ascertain their realistic values with respect to the heat flux pdf. We plan to build our atmospheric forcing input covering the range of statistically assessed PDF results to force the initial conditions in perturbation. In a forecast, variation in different atmospheric variable distributions individually can be a potential source of uncertainty. For instance, we have observed comparatively more uncertain areas for temperatures (air and dew point) in our study area. Therefore, initial experiments by regulating atmospheric variables will be a key approach to mark the major source of hampering more reliable forecasts. The other major approach, thermocline intensified random perturbation (TIRP), developed by Pinardi et al. (2008) for initial condition perturbation using temperature and salinity fields will be also combined in the experiments.

Chapter 5

Summary and outlook

This thesis has extensively studied for the first time the statistical distributions of atmospheric surface variables and heat fluxes for the Mediterranean Sea. Climate is often understood as the characteristics of the Earth system observed year after year: the average state, the range of variability, and the frequency of occurrence of extreme events. This thesis starts these climate investigations using a PDF approach that allows us to understand the first few moments of the surface atmospheric variables and the heat fluxes that drive ocean circulation. The PDFs are applied to long time series of 20-30 years of reanalyses and analyses that form the data set for climate studies. The PDF will be also used in the future to generate perturbations for ensemble predictions.

From the extended statistical analysis of atmospheric variables, we have discovered spatial patterns of the PDF parameters for the different variables. After removing the large seasonal cycle in all the surface atmospheric variables our analysis has concentrated on the anomalies. In our study, we have explored that a two-parameter PDF is might not a good fit for the air temperature and dewpoint temperature but fits well with a three parameters skew-normal PDF.

The wind components, dew point temperature and mean sea level pressure are seldomly studied while the wind amplitude and air temperature are more commonly analysed in the literature. Moreover, most of the studies have dealt with the surface wind in terms of the evaluation of wind power for offshore wind farms. In the literature, it is reported that the Weibull PDF with 2 parameters is a powerful and

convenient PDF to represent wind amplitude. For wind components, it is normally assumed a Gaussian PDF, which we show is insufficient because data are asymmetric and have fat tails (Drobinski et al., 2015). Therefore, we have applied a three-parameter skew-normal PDF to cover the non-Gaussian structure of the distribution of the wind components as well as the other atmospheric surface fields.

From the analysis of the wind component PDF parameters values in the Mediterranean Sea, we found a major difference between the Western and Eastern Mediterranean Seas. The range of the scale parameter (related to the PDF spread) for wind components is large and bigger for the zonal component. Using cluster analysis, we can find two main clusters different for the shape, i.e. the skewness of the PDF.

The mean value of 6 m/s was reported in the literature (Pettenuzzo et al., 2010) while our Weibull PDF scale parameter shows a range of 1.5 to 7 m/s (Table 4.1). Monahan (2006) studied sea-surface winds on a global scale and confirmed that surface wind amplitude fits well with 2 parameters Weibull PDF.

Wind components are clustered around positive and negative values of the shape parameter indicating opposite skewness values for the two main parameter clusters (The air temperature and dew-point anomaly clusters are different in terms of scale and shape indicating very different PDF distributions in some regions of the Mediterranean Sea, notably the south-eastern Mediterranean.

After the analysis of atmospheric variables, we then carried out the first extensive analysis of the heat budget of the Mediterranean Sea. We have found that neglecting the first 10 years of the time series, affected by low resolution in the original data set, the neat heat budget of the basin for the case of ECMWF analyses data is -3 W/m2. This result confirms that the net heat balance calculated from the ECMWF data can satisfy the closure problem of the Mediterranean Sea heat budget: since the basin gains heat at Gibraltar, heat has to be lost at the surface if a steady state has to be achieved on the long term. Our analysis shows for the first time the skewness and standard deviations of the seasonal anomaly flux components as used to force an ocean model simulation. Furthermore, it shows that the longwave upward flux can be larger than previously found providing a solution for the heat budget closure problem. The comparison with heat fluxes computed with lower resolution reanalysis fields, ERA5, cannot reproduce the correct negative heat flux thus pointing out that at least the resolution of 0.125 degrees is required.

Through this extended statistical analysis of atmospheric variables, we have obtained the PDF that will be used in future investigations of the Mediterranean Sea Ocean forecast uncertainties. In the future step, we plant to build atmospheric forcing perturbations for an ensemble forecasting system for the Mediterranean Sea with the PDF extracted from the long time series for atmospheric surface variables. Pinardi et al. (2011) already showed the importance of perturbing the winds to obtain a realistic forecast spread and the assessment of forecast uncertainty. With our analysis, we will be able to perturb each atmospheric surface variable entering heat and momentum fluxes thus sampling more effectively the uncertainty due to atmospheric forcing.

This methodology has shown a significant result in terms of statistical inference on the probability distributions of atmospheric variables. Similar way, we can analyse the ocean variables retrieved from the ocean model or if we have a long enough time series of in-situ observations in the ocean.

Bibliography

Ali, S., Lee, S.-M., & Jang, C.-M. (2018). "Statistical analysis of wind characteristics using weibull and rayleigh distributions in deokjeok-do island–incheon, south korea". *Renewable Energy*, *123*, 652–663.

Amador, J. A., Alfaro, E. J., Lizano, O. G., & Magaña, V. O. (2006). "Atmospheric forcing of the eastern tropical pacific: A review". *Progress in Oceanography*, *69*(2-4), 101–142.

Azzalini, A. (1985). "A class of distributions which includes the normal ones". *Scandinavian journal of statistics*, 171–178.

Beena, B. S., & Von Storch, J.-S. (2009). "Effects of fluctuating daily surface fluxes on the time-mean oceanic circulation". *Climate dynamics*, *33*, 1–18.

Bernie, D., Guilyardi, E., Madec, G., Slingo, J., & Woolnough, S. (2007). "Impact of resolving the diurnal cycle in an ocean–atmosphere gcm. part 1: A diurnally forced ogcm". *Climate Dynamics*, *29*, 575–590.

Béthoux, J.-P., Gentili, B., & Tailliez, D. (1998). "Warming and freshwater budget change in the mediterranean since the 1940s, their possible relation to the greenhouse effect". *Geophysical Research Letters*, *25*(7), 1023–1026.

Bignami, F., Marullo, S., Santoleri, R., & Schiano, M. (1995). "Longwave radiation budget in the mediterranean sea". *Journal of Geophysical Research: Oceans*, *100*(C2), 2501–2514.

Brossier, C. L., Béranger, K., Deltel, C., & Drobinski, P. (2011). "The mediterranean response to different space–time resolution atmospheric forcings using perpetual mode sensitivity simulations". *Ocean Modelling*, *36*(1-2), 1–25.

Castellari, S., Pinardi, N., & Leaman, K. (1998). "A model study of air–sea interactions in the mediterranean sea". *Journal of Marine Systems*, *18*(1-3), 89–114.

Chu, P. C. (2008). "Weibull distribution for the global surface current speeds obtained from satellite altimetry". *IGARSS 2008-2008 IEEE International Geoscience and Remote Sensing Symposium*, *3*, III–59.

Chu, P. C. (2009). "Statistical characteristics of the global surface current speeds obtained from satellite altimetry and scatterometer data". *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *2*(1), 27–32.

Cronin, M. F., Gentemann, C. L., Edson, J., Ueki, I., Bourassa, M., Brown, S., Clayson, C. A., Fairall, C. W., Farrar, J. T., Gille, S. T., et al. (2019). "Air-sea fluxes with a focus on heat and momentum". *Frontiers in Marine Science*, *6*, 430.

De Dominicis, M., Pinardi, N., Zodiatis, G., & Lardner, R. (2013). "Medslik-ii, a lagrangian marine surface oil spill model for short-term forecasting–part 1: Theory". *Geoscientific Model Development*, *6*(6), 1851–1869.

Dey, D. (2010). *Estimation of the parameters of skew normal distribution by approximating the ratio of the normal density and distribution functions*. University of California, Riverside.

Drobinski, P., Bastin, S., Guénard, V., Caccia, J.-L., Dabas, A., Delville, P., Protat, A., Reitebuch, O., & Werner, C. (2005). "Summer mistral at the exit of the rhône valley". *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, *131*(605), 353–375.

Drobinski, P., Coulais, C., & Jourdier, B. (2015). "Surface wind-speed statistics modelling: Alternatives to the weibull distribution and performance evaluation". *Boundary-Layer Meteorology*, *157*, 97–123.

Elderton, W. P. (1902). "Tables for testing the goodness of fit of theory to observation." *Biometrika*, *1*(2), 155–163.

Fairall, C. W., Bradley, E. F., Hare, J., Grachev, A. A., & Edson, J. B. (2003). "Bulk parameterization of air–sea fluxes: Updates and verification for the coare algorithm". *Journal of climate*, *16*(4), 571–591.

Flecher, C., Naveau, P., Allard, D., & Brisson, N. (2010). "A stochastic daily weather generator for skewed data". *Water Resources Research*, *46*(7).

Gao, F., & Han, L. (2012). "Implementing the nelder-mead simplex algorithm with adaptive parameters". *Computational Optimization and Applications*, *51*(1), 259–277.

Gille, S. T. (2005). "Statistical characterization of zonal and meridional ocean wind stress". *Journal of Atmospheric and Oceanic Technology*, *22*(9), 1353–1372.

Guéenard, V., Drobinski, P., Caccia, J.-L., Tedeschi, G., & Currier, P. (2006). "Dynamics of the map iop 15 severe mistral event: Observations and high-resolution numerical simulations". *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, *132*(616), 757–777.

Gulev, S. K., & Belyaev, K. (2012). "Probability distribution characteristics for surface air–sea turbulent heat fluxes over the global ocean". *Journal of Climate*, *25*(1), 184–206.

Gumbel, E. J. (1958). "Statistical theory of floods and droughts". *Journal of the Institution of Water Enigneers and Scientists*, (12), 157–184.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al. (2020). "The era5 global reanalysis". *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049.

Jayaraman, A., Venkat Ratnam, M., Patra, A., Narayana Rao, T., Sridharan, S., Rajeevan, M., Gadhavi, H., Kesarkar, A., Srinivasulu, P., & Raghunath, K. (2010). "Study of atmospheric forcing and responses (safar) campaign: Overview". In *Annales geophysicae* (pp. 89–101).

Kang, D., Ko, K., & Huh, J. (2018). "Comparative study of different methods for estimating weibull parameters: A case study on jeju island, south korea". *Energies*, *11*(2), 356.

Kara, A. B., Rochford, P. A., & Hurlburt, H. E. (2000). "Efficient and accurate bulk parameterizations of air–sea fluxes for use in general circulation models". *Journal of Atmospheric and Oceanic Technology*, *17*(10), 1421–1438.

Korolev, V., Gorshenin, A., Gulev, S., & Belyaev, K. (2015). "Statistical modeling of air-sea turbulent heat fluxes by finite mixtures of gaussian distributions". *Information Technologies and Mathematical Modelling-Queueing Theory and Applications: 14th International Scientific Conference, ITMM 2015, named after AF Terpugov, Anzhero-Sudzhensk, Russia, November 18-22, 2015, Proceedings 14*, 152–162.

Kuhlbrodt, T., & Monahan, A. H. (2003). "Stochastic stability of open-ocean deep convection". *Journal of physical oceanography*, *33*(12), 2764–2780.

Kyselỳ, J. (2002). "Probability estimates of extreme temperature events: Stochastic modelling approach vs. extreme value distributions". *Studia Geophysica et Geodaetica*, *46*, 93–112.

Large, W., & Yeager, S. (2009). "The global climatology of an interannually varying air–sea flux data set". *Climate dynamics*, *33*, 341–364.

Liguori, G., Di Lorenzo, E., & Cabos, W. (2017). "A multi-model ensemble view of winter heat flux dynamics and the dipole mode in the mediterranean sea". *Climate dynamics*, *48*, 1089–1108.

Lima, L., Pezzi, L., Penny, S., & Tanajura, C. (2019). "An investigation of ocean model uncertainties through ensemble forecast experiments in the southwest atlantic ocean". *Journal of Geophysical Research: Oceans*, *124*(1), 432–452.

Lorenz, E. N. (1963). "Deterministic nonperiodic flow". *Journal of atmospheric sciences*, *20*(2), 130–141.

Marchenko, Y. V., & Genton, M. G. (2010). "A suite of commands for fitting the skew-normal and skew-t models". *The Stata Journal*, *10*(4), 507–539.

Marullo, S., Pitarch, J., Bellacicco, M., Sarra, A. G. d., Meloni, D., Monteleone, F., Sferlazzo, D., Artale, V., & Santoleri, R. (2021). "Air–sea interaction in the central mediterranean sea: Assessment of reanalysis and satellite observations". *Remote Sensing*, *13*(11), 2188.

Milliff, R. F., Bonazzi, A., Wikle, C. K., Pinardi, N., & Berliner, L. M. (2011). "Ocean ensemble forecasting. part i: Ensemble mediterranean winds from a bayesian hierarchical model". *Quarterly Journal of the Royal Meteorological Society*, *137*(657), 858–878.

Moccia, B., Mineo, C., Ridolfi, E., Russo, F., & Napolitano, F. (2021). "Probability distributions of daily rainfall extremes in lazio and sicily, italy, and design rainfall inferences". *Journal of Hydrology: Regional Studies*, *33*, 100771.

Mohammadi, K., Shamshirband, S., Motamedi, S., Petković, D., Hashim, R., & Gocic, M. (2015). "Extreme learning machine based prediction of daily dew point temperature". *Computers and Electronics in Agriculture*, *117*, 214–225.

Monahan, A. H., He, Y., McFarlane, N., & Dai, A. (2011). "The probability distribution of land surface wind speeds". *Journal of climate*, *24*(15), 3892–3909.

Monahan, A. H. (2006). "The probability distribution of sea surface wind speeds. part i: Theory and seawinds observations". *Journal of climate*, *19*(4), 497–520.

Morgan, E. C., Lackner, M., Vogel, R. M., & Baise, L. G. (2011). "Probability distributions for offshore wind speeds". *Energy Conversion and Management*, *52*(1), 15–26.

Mudholkar, G. S., & Srivastava, D. K. (1993). "Exponentiated weibull family for analyzing bathtub failure-rate data". *IEEE transactions on reliability*, *42*(2), 299–302.

Myhre, G., Shindell, D., Breon, F., Collins, W., Fuglestvedt, J., Huang, J., Koch, D., Lamarque, J., Lee, D., Mendoza, B., et al. (2013). "Anthropogenic and natural climate forcing". *Climate Change*.

Organization, W. M. (2017). Wmo guidelines on the calculation of climate normals.

Pérez, I. A., Sanchez, M. L., & Garcia, M. A. (2007). "Weibull wind speed distribution: Numerical considerations and use with sodar data". *Journal of Geophysical Research: Atmospheres*, *112*(D20).

Perron, M., & Sura, P. (2013). "Climatology of non-gaussian atmospheric statistics". *Journal of Climate*, *26*(3), 1063–1083.

Pettenuzzo, D., Large, W., & Pinardi, N. (2010). "On the corrections of era-40 surface flux products consistent with the mediterranean heat and water budgets and the connection between basin surface total heat flux and nao". *Journal of Geophysical Research: Oceans*, *115*(C6).

Pezzulli, S., Stephenson, D., & Hannachi, A. (2005). "The variability of seasonality". *Journal of Climate*, *18*(1), 71–88.

Pinardi, N., Bonazzi, A., Dobricic, S., Milliff, R. F., Wikle, C. K., & Berliner, L. M. (2011). "Ocean ensemble forecasting. part ii: Mediterranean forecast system response". *Quarterly Journal of the Royal Meteorological Society*, *137*(657), 879–893.

Pinardi, N., Bonazzi, A., Scoccimarro, E., Dobricic, S., Navarra, A., Ghiselli, A., & Veronesi, P. (2008). "Very large ensemble ocean forecasting experiment using the grid computing infrastructure". *Bulletin of the American Meteorological Society*, *89*(6), 799–804.

Pinardi, N., Cavaleri, L., Coppini, G., De Mey, P., Fratianni, C., Huthnance, J., Lermusiaux, P. F., Navarra, A., Preller, R., & Tibaldi, S. (2017). "From weather to ocean predictions: An historical viewpoint". *Journal of Marine Research*, *75*(3), 103–159.

Pinardi, N., Cessi, P., Borile, F., & Wolfe, C. L. (2019). "The mediterranean sea overturning circulation". *Journal of Physical Oceanography*, *49*(7), 1699–1721.

Pinardi, N., Estournel, C., Cessi, P., Escudier, R., & Lyubartsev, V. (2023). "Dense and deep water formation processes and mediterranean overturning circulation". In *Oceanography of the mediterranean sea* (pp. 209–261). Elsevier.

Pishgar-Komleh, S., Keyhani, A., & Sefeedpari, P. (2015). "Wind speed and power density analysis based on weibull and rayleigh distributions (a case study: Firouzkooh county of iran)". *Renewable and sustainable energy reviews*, *42*, 313–322.

Rabier, F., Järvinen, H., Klinker, E., Mahfouf, J.-F., & Simmons, A. (2000). "The ecmwf operational implementation of four-dimensional variational assimilation. i: Experimental results with simplified physics". *Quarterly Journal of the Royal Meteorological Society*, *126*(564), 1143–1170.

Rosati, A., & Miyakoda, K. (1988). "A general circulation model for upper ocean simulation". *Journal of Physical Oceanography*, *18*(11), 1601–1626.

Sanchez-Gomez, E., Somot, S., Josey, S., Dubois, C., Elguindi, N., & Déqué, M. (2011). "Evaluation of mediterranean sea water and heat budgets simulated by an ensemble of high resolution regional climate models". *Climate dynamics*, *37*, 2067–2086.

Sardeshmukh, P. D., Compo, G. P., & Penland, C. (2015). "Need for caution in interpreting extreme weather statistics". *Journal of Climate*, *28*(23), 9166–9187.

Sardeshmukh, P. D., & Penland, C. (2015). "Understanding the distinctively skewed and heavy tailed character of atmospheric and oceanic probability distributions". *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *25*(3), 036410.

Small, R. J., Bryan, F. O., Bishop, S. P., & Tomas, R. A. (2019). "Air–sea turbulent heat fluxes in climate models and observational analyses: What drives their variability?" *Journal of Climate*, *32*(8), 2397–2421.

Song, X., & Yu, L. (2017). "Air-sea heat flux climatologies in the m editerranean s ea: Surface energy balance and its consistency with ocean heat storage". *Journal of Geophysical Research: Oceans*, *122*(5), 4068–4087.

Tian, F., von Storch, J.-S., & Hertwig, E. (2017). "Air–sea fluxes in a climate model using hourly coupling between the atmospheric and the oceanic components". *Climate Dynamics*, *48*, 2819–2836.

Toth, Z., & Kalnay, E. (1997). "Ensemble forecasting at ncep and the breeding method". *Monthly Weather Review*, *125*(12), 3297–3319.

Tsimplis, M. N., Zervakis, V., Josey, S. A., Peneva, E. L., Struglia, M. V., Stanev, E. V., Theocharis, A., Lionello, P., Malanotte-Rizzoli, P., Artale, V., et al. (2006). "Changes in the oceanography of the mediterranean sea and their link to climate variability". In *Developments in earth and environmental sciences* (pp. 227–282). Elsevier.

Tye, M. R., Stephenson, D. B., Holland, G. J., & Katz, R. W. (2014). "A weibull approach for improving climate model projections of tropical cyclone wind-speed distributions". *Journal of Climate*, *27*(16), 6119–6133.

Von Storch, H., & Zwiers, F. W. (2002). *Statistical analysis in climate research*. Cambridge university press.

Wais, P. (2017). "A review of weibull functions in wind sector". *Renewable and Sustainable Energy Reviews*, *70*, 1099–1107.

Walin, G. (1982). "On the relation between sea-surface heat flow and thermal circulation in the ocean". *Tellus*, *34*(2), 187–195.

Wallace, J. M., & Hobbs, P. V. (2006). *Atmospheric science: An introductory survey* (Vol. 92). Elsevier.

Wang, S., Zhang, Y., Waring, M., & Lo, L. J. (2018). "Statistical analysis of wind data using weibull distribution for natural ventilation estimation". *Science and Technology for the Built Environment*, *24*(9), 922–932.

Weijer, W. (2005). "High-frequency wind forcing of a channel model of the acc: Normal mode excitation". *Ocean Modelling*, *9*(1), 31–50.

Yiou, P., Goubanova, K., Li, Z., & Nogaj, M. (2008). "Weather regime dependence of extreme value statistics for summer temperature and precipitation". *Nonlinear Processes in Geophysics*, *15*(3), 365–378.

Yu, K., & Zhang, J. (2005). "A three-parameter asymmetric laplace distribution and its extension". *Communications in Statistics—Theory and Methods*, *34*(9-10), 1867–1879.

Appendix

Nelder-Med function

To estimate the PDF parameters, we apply the maximum likelihood estimation (MLE) method. Let's assume the random variable values of $X = (x_1, x_2, x_3, \ldots, x_n)$ are independent samples from a skew-normal (α, μ, λ) , where the parameters α, μ, λ are real numbers and then the likelihood function is expressed by –

$$L(x, \alpha, \mu, \lambda) = \frac{2^n}{\lambda^2} \Pi_{i=1}^{n} \, \phi(\frac{x-\mu}{\lambda}) \, \Phi(\frac{x-\mu}{\lambda}) \tag{5.1}$$

Where the MLE method estimate for the parameters which use maximize equation for a random variable X and can be expressed with θ

$$\theta = argmax \;\; f(X, \theta) = argmin \, L(\theta|) \tag{5.2}$$

As we have detected an issue in the initialization values for the MLE function referred from chapter 2.3, particularly with the skew or shape parameter, where the MLE method is not itself enough to estimate the skew parameter correctly. This complexity arises in using of MLE method due to the presence of asymmetry in our dataset where the normality assumption is not valid. To solve this problem, for max $f(x)$ or min $f(x)$, in this study we use an optimization function because the MLE function can't be maximized or minimized without an optimization algorithm. Here we apply the algorithm which is called the Nelder-Mead or simplex algorithm and a widely used search method to solve an optimization problem.

In simple text, the Nelder-Mead function generates a sequence of simplices to approximate an optimal point by iteration and at each iteration, the vertices the vertices $(x_{j=1}^{n=1})$ of the simplex are ordered according to the objective function values (Gao and Han, 2012) and a simplex with vertices $(x_1, x_2, ..x_(n+1))$ i is expressed by

$$f(x_1) \leq f(x_2) \leq ..... \leq f(x_{(n+1)}) \tag{5.3}$$

Where we can refer x_1 is the best vertex and $x_(n+1)$ is the worst vertex

Transformation of the SH flux distribution

Here, we show how the SH flux distribution is transformed within the computation using atmospheric variables input. To compute the SH flux, we use the following formula for the Mediterranean Sea.

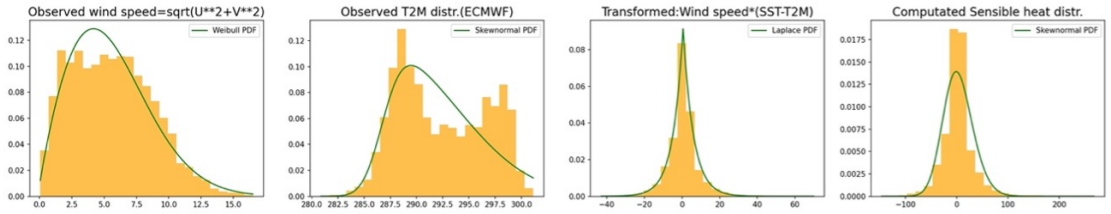

FIGURE 5.1: Histograms presenting a single grid location distributions of observed wind speed and air temperature, and their transformation using a part of the SH flux formula(third histogram). The fourth one is the resulting SH flux distribution using the formula

SH flux requires atmospheric variables as inputs along with fixed parameters. In chapter 2, we observed and analyzed the wind amplitude with Weibull PDF, while the air temperature is fitted with 3 parameters skew-normal PDF. In our investigation, computed SH flux resulted in a different shape than a Gaussian one. In the histogram panel (Fig. 1), from left to right, we notice the difference in their distribution wind speed (m/s) and surface air temperature (kelvin). shape visually, and their transformed distribution is presented in the third histogram. The third histogram show explicitly the multiplication result of wind speed and the difference of SST and air temperature required for SH flux computation. The 4th histogram shows an SH flux computation distribution using the formula where skew-normal PDF cannot cover the picks completely. While the transformed distribution in the third histogram is fitted well with the Asymmetric Laplace Distribution (ALD). The ALD PDF fit shapes the picks well. Therefore, the transformed SH flux distribution is not fitted with skew-normal, but we have applied the ALD PDF for the probability analysis.

Acknowledgment

First of all, I am thankful to Professor Nadia Pinardi for her continuous monitoring, guidance, and advice.

I would like to express my sincere thanks to Dr. Francesco Trotta, Dr. Giovanni Liguori, and Dr. Silvia Bianconcini for their technical supports, for sharing valuable scientific knowledge, and for guiding in solving critical issues.

I wish to extend my thanks to the people, colleagues, and researchers I met around during my Ph.D journey.