III. Research Plan

III.1. Objective and Research Question

The objective of this study is to investigate possible ways to improve the forecasts of the MOHID Tagus Estuary system, contributing to the development of the tool so as to reach an operational system status. In this context, the research is focused in improving the hydrodynamic forecast.

Aiming this objective is considered the following research question:

"Is it possible to improve the existing MOHID Tagus Estuary skill on forecasting hydrodynamic variables through an alternative model configuration or the use of data assimilation techniques?"

The concept of hydrodynamic variables is operationalized by the water level and velocity fields, for the 2D version of the system, and water level and velocity fields and stratification (temperature and salinity) conditions in the 3D version of the system, as these are, according with the theoretical barotropic and baroclinic hydrodynamics, the main indicative properties and also properties forecasted by the MOHID Tagus Estuary numerical modelling tool.

As presented in Chapter II, an Operational Water Quality Forecasting System for the Tagus Estuary and adjacent coast is a very useful tool due to the area sensitivity and anthropogenic pressures which influence it. As seen, the most direct application of the system is the monitorization of the pollution from wastewater discharges and the resulting water quality of the Estoril Coast beaches. It has been referred that this use of the forecasting system alone justifies the need of a relevant forecast for estuarine conditions. However, the operational monitorization of the water quality in the estuary is also a very relevant due to the ecological and economical importance of biological estuarine productivity, which is exemplified by the richness of the Natural Reserve of the Tagus Estuary and the economical impact of the estuarine fisheries, and the extent of industrial and transport activity which potencies the occurrence of hazardous pollution events with large impact. It also has been presented that water quality in the area is predominantly determined by hydrodynamics and hence a reliable hydrodynamic forecast is required.

But a hydrodynamic forecast by itself is also very important in the Tagus Estuary, to support inside estuary navigation, due to the estuarine shallowness and intricate bathymetry, and sediment management and dredging operations planning, due to the intense sediment transport documented in the estuary. Outside the estuary, an operational hydrodynamic forecast system could provide a useful contribute to the minimization of storm surge events effects on the coastline, as these events are reported to have particular incidence the Tagus Estuary mouth.

While contributing to the development of an Operational Water Quality Forecasting System, the proposed research question answers directly to the demand for improving the existing system made by Fernandes (2005), following the approaches modelling and data assimilation proposed by this author and given that the dominant forecasts quality limiting factor identified by the author concerns the hydrodynamics.

In the framework of the knowledge area where it is positioned, Operational Oceanography, this study is also relevant because it contributes to the increase of knowledge of data assimilation techniques behaviour in coastal applications. Adding to this general relevance and besides the large horizontal resolution, the Tagus Estuary Forecasting System domain provides characteristics, namely the large difference of bathymetry throughout the domain, the dominance of tide in dynamics, the inclusion of estuarine, continental shelf and ocean dynamics and the small width of the continental shelf, which are unusual in the coastal data assimilation applications research, more concentrated in the Mediterranean and the North Seas. These conditions are, however, representative of operational forecast systems domain characteristics for the Atlantic Iberian, African and South American coasts. While important in the framework of the international research, this study represents also an innovation because it represents one of the first attempts to apply data assimilation in a Portuguese coast located coastal hydrodynamic model.

III.2. Research Hypotheses

As referred before, the research is oriented in two approaches: improve modelling and use data assimilation techniques.

Concerning modelling, having into account the contrast of the main processes known to influencing estuarine hydrodynamics and advanced modelling operational oceanography practices, on one hand, and the Tagus Estuary and adjacent coast specificities and the MOHID Tagus Estuary deficiencies pointed by Fernandes (2005), on the other hand, it is possible to note that a potential major area for improvement is in the specification of the open boundary conditions.

Concurrently, the hypotheses referring to alternative model configurations are presented in the following.

Hypothesis 1: *Improvement is possible through the use of a larger spatial extent, comprising all the Continental Portuguese and Galician coasts, of the model domain used for tide propagation from the FES 95.2 global tide solution to the local scale domain of the Tagus Estuary and adjacent coast.*

This hypothesis states that is possible to improve the tide description in MOHID Tagus Estuary by including effects of the large scale tidal wave interacting with relatively shallow bathymetry features not included in the present domain. Although the tidal water level analysis made by Fernandes (2005) points that the tidal propagation in MOHID Tagus Estuary is adequate, there is some evidence that tide propagation in this system may eventually be improved.

The first is the reported occurrence of an error of some 0.05m in water level at the estuary mouth (Cascais and Paço de Arcos) in a area where the horizontal grid resolution (approximately 300m) by Fernandes (2005) also supported by the essentially qualitative findings (at Cascais) of Leitão (2003) for a very similar MOHID Water model configuration for the estuary.

This error can be due to insufficient bathymetry resolution but also to lacking continental shelf interaction between the large scale tidal wave and regional bathymetry. However, some extra evidences suggest that the later explanation can be important. On one hand, it is the fact that global tide model solution, such as FES 95.2, are known to be of weak quality at the small depths of the continental shelf: e.g. Sauvaget *et al.* (2000), in their study of the tidal currents in the Portuguese Coast, using the finite-elements model TELEMAC-2D to transform the FES 94 global tide solution to local scale using a domain extending from Sagres to the West Galician coast, point that imposing the global model solution in the shallow continental shelf area can lead to stress in the tide propagation model; the same situation is discussed by Leitão (2003) in the referred work with the Portuguese Coast tidal model. Although the tidal boundary condition choice in both Leitão (2003) and Sauvaget *et al.* (2000) studies seems to solve adequately the noise in the velocity field derived from model domain open boundary intercepting the continental shelf, there still remains to be assessed the effect of the boundary placement in the water level field. Leitão (2003) refers that open boundaries of regional tidal models must preferably be defined in areas where non linear hydrodynamic processes are not important, i.e. in areas of large depth. The Leitão (2003) study finds also that the nesting of a domain with a higher resolution causes no significant change in the water level results at the estuary mouth, which implies that the local bathymetry effect should not be relevant in this area, providing some support that the tide propagation in the regional domain may be a determinant cause of the observed tidal water level error.

An eventual error in tide propagation from the global FES 95.2 solution for the estuary mouth due to a bad model boundary placement should have an important effect in the representation of the tide inside the estuary, as Fortunato *et al.* (1999) and Leitão (2003) find that the amplitudes of the tide components, specially the tide-dominating semi-diurnal components, grow rapidly from oceanic values in the estuary exit channel and also grow, although more steadily, in the upper estuary (Figure II-1) up to Póvoa de Santa Iria. The analysis of the tidal water level error by Fernandes (2005), provide data which can support this fact, as are detected errors in tidal water level of more than 0.1m in the estuary exit channel and of 0.3m in Póvoa de Santa Iria. The test of this hypothesis offers then, indirectly, a strong test for the role played by local bathymetry resolution in deficient tide representation at the Tagus Estuary and adjacent coast.

While contributing to the better understanding of the tidal dynamics in the Tagus Estuary adjacent coast, the consideration of a tidal propagation domain comprising all the Portuguese Coast would cause an increase of the computation cost of the MOHID Tagus Estuary forecast. However, if this cost is not prohibitive to operational use, this model domain could provide a common platform to develop a full 3D domain with baroclinic effects and support the nesting of other operational forecast systems for the Continental Portuguese coastal areas.

Hypothesis 2: *Improvement is possible through the incorporation of the large scale atmospheric pressure effect on mean sea level in modelling.*

This hypothesis implies that a relevant deficiency of the MOHID Tagus Estuary configuration is the lack of consideration of the mean sea level anomaly originated in the large scale by the effect of atmospheric pressure. In this work mean sea level is considered to be detided water level elevation, i.e. the tidal residuals. It results naturally from the research documenting the importance storm surge phenomena in Portugal and the determinant influence, relative to wind, that the atmospheric pressure has on this phenomena in the Iberian Coast (Fanjul *et al.*, 1998). Also, the contrast between the MOHID Tagus Estuary results and tide gauge measurements made available for the system validation caused the detection of an important mean sea level anomaly.

The previous research on the Tagus Estuary and adjacent coast with MOHID Water (Leitão, 2003; Fernandes, 2005) has been mainly concentrated in the regional and local induced hydrodynamics and validation was performed focusing on the tidal description. Despite this, the work of Leitão (2003) still provides an evidence that makes this large scale hypothesis possible: in the contrast of the tidal model results with ADCP current measurements for a period comprising a month at the Tagus Estuary adjacent coast (near the Guia wastewater outfall location and Cascais), Leitão (2003) detected barotropic current variability in the measurements with period larger than 30 hours which are not of tidal origin and are not reproduced by the tidal model; although the wind effect is not accounted in the model these oscillations could also possibly evidence the influence of large scale barotropic circulation due to atmospheric pressure. Other ADCP measurements for the estuary mouth area analysed by Leitão (2003) and Fernandes (2005) refer to a few days period and due to the short time span do not allow the verification of the existence of similar signatures.

Hypothesis 3: *Improvement is possible through the account of large scale thermal and saline vertical stratification in modelling.*

This hypothesis concerns the 3D hydrodynamics. It gives direct continuity to the work of Leitão (2003) and Fernandes (2005) who find important deficiencies in MOHID Tagus Estuary (or a very similar model in the case of Leitão, 2003) in reproducing the vertical thermal and saline stratification in the border of the continental shelf adjacent to the estuary mouth. The assumption of a constant temperature value for the whole water column in the open boundary condition is specifically referred by Leitão (2003) based on CTD measurements as the reason for a poor performance in temperature representation at the Tagus Estuary mouth. The vertical stratification conditions are, as stated in Chapter I, very important for the adequate representation of upwelling events and water quality in the coastal areas.

Within the approach of data assimilation the following hypothesis is considered in the research.

Hypothesis 4: *Improvement is possible through the use of data assimilation of hydrodynamic measurements with a SFEK/SEEK filter scheme.*

The test of this hypothesis is justified by the operational status aimed for MOHID Tagus Estuary system, as the reduction of the deficiency in representing estuarine hydrodynamics in narrow channel areas is considered to be very costly in terms of computation resources if solely a modelling approach is used. In fact, Fortunato *et al.* (1999) refer that resolution of model horizontal grid in these areas should be in the order of tens of meters in order to adequately solve hydrodynamics, which would involve a potentially tenfold increase of computational time with the finite volumes model MOHID Water[99].

The choice of the data assimilation technique is a relevant one, because as no data assimilation scheme is implemented in MOHID Water code any technique must be programmed from raw. Although the estuarine hydrodynamics are expected to be non linear, namely due to the tide dynamics and the existence of intertidal areas, the SEEK filter is a data assimilation technique extensively used in coastal model applications with capabilities to deal with model non linearities (as EKF derived) and having a lower cost than the non linear models' more performing EnKF, which presents a computational cost ten times larger and hence whose use would be prohibitive in the current MOHID Tagus Estuary setting. The SEIK filter, although reported to provide an improved correction of non linear models relative to SEEK filter at an equal cost, has been less applied and, as shown in Chapter I, presents higher implementation complexity and physical representation problems due to the creation of the interpolated states (Pham, 2001), which could difficult the validation of a first time data assimilation application.

Although the SEEK filter scheme has been predominantly applied in coastal areas where tide is not significant (e.g. Mediterranean Sea), research provides some indirect evidence that it could perform

[99] These authors use for their modelling study a finite element model (ADCIRC) which allows such increase of resolution within acceptable costs.

well with tidal dynamics, as the conceptually similar scheme RRSQRT filter has been extensively used with good results in the North Sea tide dominated coastal areas, provided that intertidal areas are not extensive to a large number of model horizontal grid cells (Verlaan, 1998). Even though the intertidal areas represent a large part of the Tagus Estuary surface, in MOHID Tagus Estuary they concentrate in a domain area where the resolution is smaller, with intertidal areas being described by a relatively small number of cells.

The implementation of a SEEK filter in MOHID Water code does not hamper but rather supports an easy future implementation of EnKF and SEIK filter, when the computational costs become affordable with the ever faster computation technology, due to the common Kalman filter structure. The existence of a validated SEEK filter scheme should also facilitate the new schemes validation work at that stage, as it will be more focused on the Monte Carlo features of both schemes.

Due to its reduced computational costs and reported good performance and usefulness for preparing SEEK filter configurations in some coastal model applications, this hypothesis considers also the SFEK filter.

III.3. Methodology

III.3.1. Research steps

The research is conducted in the following steps:

1. **Development and validation of an alternative 2D MOHID Tagus Estuary model configuration with a larger domain 1 for tide propagation comprising all the Continental Portuguese and Galician coasts:**
    a. Construction of a bathymetry and adequate open boundary conditions for the alternative domain 1;
    b. Making two MOHID Tagus Estuary domain 1 simulations, one with the existent domain and other with the alternative domain, for a fortnight period considering tide as the only forcing effect and contrast the results of each simulation with hydrodynamic measurements distributed throughout the coast; this should **test the support for Hypothesis 1**;
    c. Conducting a one year simulation for the alternative MOHID Tagus Estuary domain 1, with and without the account of tide generation forces in the model domain, and harmonic analyse the water level results and contrast the main tide components, in terms of amplitude and phase, with the ones derived from measurements distributed throughout the coast; this should establish the robustness of the alternative configuration in representing the main tide components.

2. **Construction of an actualized bathymetry for MOHID Tagus Estuary domain 2, considering more detailed and recent coastline and bathymetry information.**

3. **Development and validation of an alternative 2D MOHID Tagus Estuary model configuration accounting for the large scale atmospheric pressure effect on mean sea level:**
    a. Design of a domain 0 configuration comprising all the North Atlantic ocean area and adequate to simulate the large scale atmospheric pressure effect on mean sea level;
    b. Making a MOHID Tagus Estuary domain 1 simulation nested in domain 0 considering only atmospheric pressure and wind forcing and contrast the water level results of both domains with the tidal residuals from measurements; this should test the importance of the effect of meteorological forced large scale and regional circulation in explaining mean sea level variability;

c. Development of an adequate open boundary condition for mean sea level for MOHID Tagus Estuary domain 1 derived from the large scale atmospheric pressure effect;

d. Making two MOHID Tagus Estuary domain 2 (nested in domain 1, using the actualized bathymetry) simulations, with and without the domain 1 mean sea level boundary condition described in the previous point, considering tide as the only forcing and contrasting the two simulations domain 2 water level tidal residuals results with the tidal residuals from measurements; this should **test the support for Hypothesis 2**.

4. **Development and validation of an alternative 3D MOHID Tagus Estuary model configuration accounting for the large scale thermal and saline vertical stratification:**

   a. Design an alternative MOHID Tagus Estuary domain 2 configuration with a thermal and saline vertical stratification open boundary condition, an improved vertical grid resolution, spatially variable meteorological forcing and an improved advection scheme, in order to improve the representation of vertical stratification;

   b. Making two MOHID Tagus Estuary domain 2 simulations, using the existent and the alternative configuration, contrasting the water level and thermal and saline stratification with measurements; this should **test the support for Hypothesis 3**.

5. **Study of the base Kalman filter algorithm in a linear model schematic case:**

   a. Development of a linear model in MATLAB programming environment;

   b. Development of a Kalman filter in MATLAB programming environment;

   c. Design of a data assimilation configuration for the model;

   d. Conducting several model simulations with data assimilation for assessing the filter performance under several filter operating conditions.

6. **Implementation of the SEEK and SFEK filters original algorithms, respectively from Pham *et al.* (1998a) and Brasseur *et al.* (1999), in the MOHID Water Fortran 95 code framework:**

   a. Development of a new MOHID Water module for performing sequential data assimilation, incorporating the SFEK and SEEK filter schemes;

   b. Development of a data assimilation pre-processor to calculate the covariance derived from MOHID Water results files (HDF5 format), perform EOF analysis and produce the initial correction directions set in a form usable for data assimilation applications of SEEK and SFEK filters with MOHID Water.

7. **Development and validation of the first 2D MOHID Tagus Estuary configuration using data assimilation in domain 2:**
    a. Design of a twin experiment for MOHID Tagus Estuary;
    b. Performing EOF analysis for twin models results without data assimilation;
    c. Making several MOHID Tagus Estuary simulations with data assimilation using SFEK and SEEK filters for assessing the filters performance under several operating filter conditions; this should **test the support for Hypothesis 4**.

The model domains and the modelling conditions in terms of vertical resolution and forcing used for the research are summarised in Table III-1. Some details of the methodology are discussed in the next paragraphs.

Table III-1: Model domains configurations used in simulations for research hypotheses testing.

<table>
<tr><th rowspan="2">Hypothesis</th><th colspan="5">Model domains vertical resolution and forcing</th></tr>
<tr><th>Domain 0</th><th colspan="2">Domain 1</th><th colspan="2">Domain 2</th></tr>
<tr><td>1</td><td>-</td><td colspan="2">2D
tide</td><td colspan="2">-</td></tr>
<tr><td>2</td><td>2D
atmospheric pressure
spatially variable
wind</td><td>2D
atmospheric
pressure
spatially
variable wind</td><td>2D
tide
open boundary
condition from
large scale
atmospheric
pressure effect</td><td>-</td><td>2D
river flow</td></tr>
<tr><td>3</td><td>-</td><td colspan="2">2D
tide
wind (spatially variable in
alternative configuration)</td><td colspan="2">3D
river flow
wind and radiation (spatially
variable in alternative
configuration)
open boundary condition with
thermal/saline stratification</td></tr>
<tr><td>4</td><td>-</td><td colspan="2">2D
tide
spatially variable wind</td><td colspan="2">2D
river flow
spatially variable wind</td></tr>
</table>

As seen, the test of the improved modelling hypotheses is based on running model simulations of alternative configurations of MOHID Tagus Estuary and of the existent MOHID Tagus Estuary configuration. The need to run simulations of the existent MOHID Tagus Estuary configuration is not trivial and is justified by the fact that the MOHID Water numerical model code suffers continuous actualizations in time and so the same up to date version of the code must be used in the alternative and the existent MOHID Tagus Estuary configuration to efficiently test the hypotheses.

Although MOHID Tagus Estuary comprises several model domains, as described in Chapter II, not all are considered for the test of the research hypotheses. For economy, only the domains directly related to the hypotheses are considered, but is assumed that if improvement is observed in the assessed domains the subsequent nested domains model solution should also improve.

In the case of the research concerning Hypothesis 1 (step 1 in this work Methodology), the choice of a configuration with only the domain 1 is trivial, since the specific effect which is being tested only refers to this domain; also only tide is used as forcing because it is the tide propagation that is to be studied. The open boundary of the alternative domain 1 is established so as to avoid interception of the continental shelf in the direction of the tidal wave propagation from South to North and to extend in the West direction to take into the domain the important bathymetric feature constituted by the Gorringe seamount[100]. In the one year simulation experiment the effect of tide generating forces inside the domain is assessed in order to extend the knowledge of regional tidal dynamics and contrast the findings of Sauvaget *et al.* (2000) in this respect.

For testing Hypothesis 2 (step 3 in this work Methodology) an adequate representation of the Tagus Estuary and adjacent coast bathymetry is required. Hence and as new bathymetric and coastline data is available, an actualization of the domain 2 bathymetry (step 2 in this work Methodology) is undertaken.

For the research on this second hypothesis, both atmospheric pressure and wind forcing must be taken into account in domain 0 because previous research suggests that non linear effects in circulation can result from the interaction of these two forcing effects, for consistency the same forcing is considered in the nested domain 1; however tide is not considered, as the tide boundary condition from the FES 95.2 global solution at the south of the domain in a previous test did not provided acceptable tide results at a low spatial resolution. Concurrently, domain 2 is not simulated because the tide interaction with bathymetry is expected to generate important non linear effects in the Tagus Estuary and the interpretation of mean sea level contrasts between model results and measurements should be therefore more complex in the absence of tide. The MOHID Tagus Estuary configuration using a domain 0 is not a realistic alternative configuration for an operational MOHID Tagus Estuary. A large scale solution of mean sea level to be used in that operational framework should instead be provided by an operational system with forecasts for the North Atlantic, such as HYCOM or MERCATOR systems; however at the time this work is developed no solutions are available for the test of the effect of such open boundary condition. Hence, the domain 0 is used in an indicative test.

[100] Located at -11.5ºE and 36.5ºN.

For developing an adequate open boundary condition for mean sea level is followed the approach of the inverted barometer formulation. In this formulation the mean sea level time evolution is determined from the atmospheric pressure time evolution at the domain open boundary, as in Cañizares (1999). Since this formulation provides for a direct connection between the atmospheric pressure and the water level acting in the large scale, it is particularly adequate for the test of Hypothesis 2. It is also a very common approach to provide adequate large scale open boundary conditions to coastal storm surge models as the work of Vries *et al.* (1995), Cañizares (1999), Choi *et al.* (2003) and Zhang *et al.* (2007) shows. This approach involves the production of a software application for producing the mean sea level time series. For the validation of the mean sea level forecasts within Tagus Estuary and immediate adjacent coast the tide forcing must be taken into account due to the referred non linear effects resulting from the shallow bathymetry. No local atmospheric pressure and wind forcing is considered in domains 1 and 2 for this estuary validation in order to determine with more precision the influence of large scale atmospheric pressure in mean sea level. It is however acknowledged the potential importance of the non linear interaction of tide and local meteorological forcing in the shallow areas (Fanjul *et al.*, 1998).

For Hypothesis 3 test (step 4 in this work Methodology) and since it refers to stratification, it is important to have a detailed vertical resolution and also to consider solar radiation, air temperature, heat fluxes, relative humidity (which affect water temperature) and wind spatially variable forcing (wind effect is important since a 3D configuration is used for modelling). Additionally, it is used an improved advection scheme – TVD Superbee – in order to more adequately represent advection in momentum and temperature and saline conservation equations: both Fortunato *et al.* (1999) and Leitão (2003) find that advection is very important at the mouth of the estuary for the dispersion of fresh water plume. The TVD Superbee scheme is chosen following the findings of Riflet (2005) in a fresh water cylinder case that this scheme is, from the ones implemented in the MOHID Water, more adequate to assess the capabilities of a hydrodynamic model to show realistic meandering of river plume fronts, coastal currents and baroclinic instabilities in oceanic frontal systems, which are modelling issues relevant in MOHID Tagus Estuary.

In order to achieve the first application of data assimilation in MOHID Tagus Estuary and due to the complexity of MOHID Water code and the SFEK/SEEK filter schemes algorithms, it is considered the Kalman filter application to a schematic linear model case (step 5 in this work Methodology). The advantages of working first with a simpler model include increasing of knowledge on the practical implementation of the Kalman filter schemes algorithms, namely in the operationalization of the main concepts involved in data assimilation, and also gaining a sensibility of the operation options which mostly influence the filter performance. The use of MATLAB for programming environment is a result of the bold functionalities there existent to deal with matrix operations, which as seen in Chapter

I compose a major part of the Kalman filter and derived schemes algorithms. Having into account the dominant tide dynamics in MOHID Tagus Estuary, the linear model considered for this schematic test is constituted by the one dimensional linear advection equation for the water level evolution: following the study of Evensen (1992) one aims to use the Kalman filter to make the model state approach the analytic solution of the advection equation, taken as measurement source. The results of this schematic data assimilation application are used to assist the MOHID Tagus Estuary application design.

For the implementation of the SFEK and SEEK filter schemes in the MOHID Water code framework (step 6 in this work Methodology) it is chosen to create a new MOHID Water module for the sequential data assimilation. This option is preferred to include the new functionalities in the existing module which processes simple data assimilation operations based on newtonian relaxation as described by Leitão (2003) - Module Assimilation - because of two main reasons:

- Module Assimilation responds hierarchically to MOHID Water modelling processes modules, such as Module Hydrodynamic or Module WaterProperties (see Chapter II), which is an relationship not adequate for advanced data assimilation schemes: they should occur in the numerical model architecture at a higher level than modelling processes modules because by definition these schemes command changes in the memory state of the modelling processes modules (through the correction of model solution produced by several modules);
- the implementation of the advanced data assimilation schemes in Module Assimilation while maintaining the option of the simultaneous use of newtonian relaxation and advanced data assimilation techniques in MOHID Water applications would involve a lack of code clarity and potentially increasing the difficulty in the creation of input instructions by the user.

The choice of developing a pre-processor tool for covariance creation operations is a natural result of the analysis of MOHID Water simulations result files, drawn from the fact that the most common procedure for this generation to be done is by calculating the covariance based on model historical states variability (e.g. Hoteit *et al.*, 2005). Since this is usually a very time and computer resources consuming operation it is more practical to perform it in preprocessing phase: the same produced covariance structure can then be used for several MOHID Water simulations. Additionally, it is known from previous research presented in Chapter I that the production of covariance structure adequate for the success of a SFEK/SEEK filter data assimilation application is not a trivial process, involving a considerable amount of choices for the user to make and frequently a trial and error process. Therefore, its implementation in a preprocessing tool allows the user to try different options in the covariance calculation process before, based in the covariance results, choosing the define covariance structure for using in the interest data assimilation application. Additionally, because having EOF analysis functionalities the pre-processing tool can also be useful outside the scope of data assimilation applications in studies of variability of MOHID Water results.

The configuration of MOHID Tagus Estuary chosen for the test of Hypothesis 4 (step 7 in this work Methodology) is straightforward since the validation of the system by Fernandes (2005) indicates that the most important forecast problem that could not be solved operationally by modelling are the water level results inside the Tagus Estuary, which are modelled in domain 2. This fact indicates that the most relevant application of data assimilation MOHID Tagus Estuary would be in this domain. It is chosen to process the first data assimilation application to MOHID Tagus Estuary in a twin experiment framework, the common practice in data assimilation research (see Chapter I), in order to validate the data assimilation code implemented in MOHID Water in a case where the true state is completely known and the error is more consistent with the state estimate error assumptions of the Kalman filter schemes, so that the filters have good conditions to operate.

III.3.2. Creation of bathymetries

All new bathymetries are created from MOHID XYZ format bathymetric data and MOHID XY (polygon) format coastlines using the DigitalTerrainCreator MOHID support tool. Raw bathymetric data are converted into XYZ MOHID format using the ConvertToXYZ MOHID support tool. Priority is given to more detailed and more recent data.

III.3.3. Validation of model results

Harmonic analysis and the calculation of tidal residuals from tide gauge measurements or model water level results are processed using the MATLAB environment T_Tide V1.1 tool presented in the Pawlowicz *et al.* (2002) work and available in the Internet[101]. The tidal residuals are obtained by performing harmonic analysis, reconstructing the tidal water level and subtracting the tidal water level from the original water level.

The reconstitution of water level time series from tidal components available previously to this work (not generated by T_Tide harmonic analysis) is processed using the TidePrev MOHID support tool.

[101] http://www2.ocgy.ubc.ca/~rich/#T_Tide.

The contrast between model results and measurements is based on the calculation of the Bias, i.e. the average error (the difference between the model and the measurement), and the centered Root Mean Square Error (RMSE), considering the following expressions:

$$\text{RMSE}=\sqrt{\frac{\sum(\text{error})^2}{n}}$$

$$\text{Centered RMSE}=\sqrt{\text{RMSE}^2\text{-Bias}^2}$$

Where n is the number of error values.

The Bias expresses systematic departures from the model results relative to measurements and the centered RMSE represents non systematic variability differences. Since the measurements usually represent a point in space while the model results are referred to a model grid cell with a specific resolution, the correlation coefficient statistic measure (e.g. Emery and Thomson, 2001) is not generally used for this purpose.

III.4. Data

III.4.1. Summary of requirements and choices

The data used to test the research hypotheses refer to two types:

1. Data used to support modelling using alternative model and data assimilation configurations.
2. Measurements of the system state used for model or data assimilation results validation.

The data used in both types for the research of each hypothesis is presented in Table III-2.

Table III-2: Data requirements and choices for research hypotheses testing.

Hypothesis	Data	
	Modelling (type 1)	**Validation (type 2)**
1	Bathymetry (domain 1): ETOPO 2'[a], METEOGALICIA[b], existent domain 1 bathymetry	Tide gauge harmonic components for 13 locations
2	Bathymetry (domain 0): METEOGALICIA Bathymetry (domain 2): bathymetric surveys in Tagus Estuary and adjacent coast, ETOPO 2'[a], existent domain 1 bathymetry Sea level atmospheric pressure and wind: ERA40 reanalysis[c] River flow: daily averages of Ómnias hydrometric station[d] for Tagus, long term average for Sorraia and Trancão[e]	Tide gauge measurements (1972-1973) for 11 locations
3	Thermal/saline open boundary stratification: Levitus climatology Meteorological forcing (wind, solar radiation, temperature, relative humidity, heat fluxes): MM5 River flow: hourly values of Ómnias and Coruche hydrometric stations[d] for Tagus and Sorraia, daily average for Trancão	CTD campaigns MODIS (level 2) SST images[f]
4	Wind: ERA40 reanalysis[c] River flow: daily averages of Ómnias hydrometric station[d] for Tagus, long term average for Sorraia and Trancão	-

Notes: [a] bathymetry raw data with 2' resolution available from the Internet[102]; [b] bathymetry raw data with variable resolution for the areas in front of Vigo (Cies Miño) and eastward of Coruña (San Ciprian Ortegal and Estaca Cabo Peñas) produced by Instituto Hidrográfico de Cádiz; [c] provided by ECMWF in the Internet[103]; [d] provided by INAG in the Internet[104]; [e] 39.5m^3/s for Sorraia and 6m^3/s for Trancão; [f] provided by NASA in the Internet[105].

[102] http://dess.ucar.edu/datasets/ds759.3/.
[103] http://data.ecmwf.int/data/d/era40_daily/.
[104] http://snirh.inag.pt.
[105] http://oceancolor.gsfc.nasa.gov.

In the identification of data, of both types, used on research, are investigated data sources which could be used by an operational MOHID Tagus Estuary. This was not possible in all cases. The details are discussed in the following. One of the main difficulties encountered in this work is the measurement availability for validation, since as stated previously the MOHID Tagus Estuary measurement network is very sparse in hydrodynamic measurements. Relevant measurements are only available for short time periods. Since this is the most relevant data type it is discussed first.

III.4.2. Validation data

The most consistent data set for validation of hydrodynamic conditions in the estuary and adjacent coast is constituted by tide gauge measurements time series collected mostly for a year in 1972-1973 in the Tagus Estuary and at the estuary mouth (Table III-3). The location of these gauges are presented is presented in Figure III-1. This is used as validation data for the test of Hypothesis 2, which required accurate measurements of mean sea level. In order to facilitate the use of this data, it is converted from the original format to MOHID framework TimeSeries format using a MOHID support tool specially developed for the purpose in this work called ConvertToTimeSerie (see Appendix X).

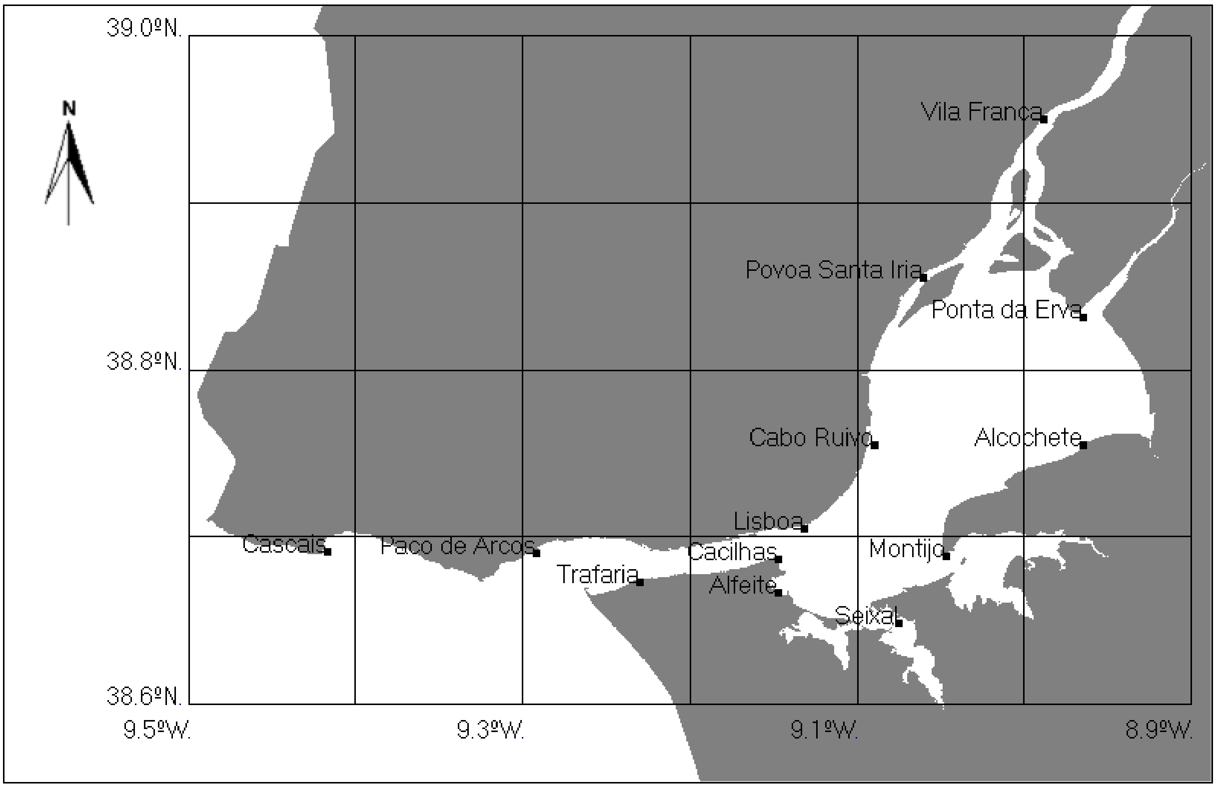
Table III-3: Tide gauge measurements available to validate the 2D configuration of MOHID Tagus Estuary for Hypothesis 2 test. Location and time span are shown. Tide gauges are presented from ocean to Tagus River.

Tide gauge	Geographical coordinates		Period of data
	Latitude (ºN)	Longitude (ºE)	
Cascais	38.6917	-9.4167	25/04/1960 – 30/06/1991
Paço de Arcos	38.6900	-9.2917	26/01/1972 - 25/01/1973
Trafaria	38.6733	-9.2300	01/01/1972 - 31/12/1972
Cacilhas	38.6867	-9.1467	01/01/1972 - 31/12/1972
Alfeite	38.6667	-9.0267	09/04/1981 – 14/05/1981
Lisboa	38.7050	-9.1317	01/01/1972 - 31/12/1972
Cabo Ruivo	38.7550	-9.0150	01/01/1972 – 31/12/1972
Seixal	38.6483	-9.0750	01/01/1972 - 31/12/1972
Montijo	38.6883	-9.0467	01/01/1972 - 31/12/1972
Alcochete	38.7550	-8.9650	05/01/1972 – 31/12/1972
Ponta da Erva	38.8317	-8.9650	01/01/1972 - 31/12/1972
Póvoa de Santa Iria	38.8550	-9.0600	08/04/1972 – 16/10/1972
Vila Franca de Xira	38.9500	-8.9883	01/01/1972 - 31/12/1972

Note: data provided from Instituto Hidrográfico Português.

However, because these measurements concentrate in the inner estuary and mouth they cannot be used for the validation of tide for Hypothesis 1 test. For this case are collected a set of tidal components files for several tide gauge locations in the Portuguese and Spanish coasts (Table III-4 and Figure III-2). As the test of this hypothesis is conducted with only tide forcing (see section III.3.) and the dominant tidal components in terms of amplitude are present in each location set, no relevant drawback is associated to not using tide gauge measurements for validation. Nevertheless, the number of components for each tide gauge is variable[106] and this must be taken into consideration when interpreting the validation results.

Figure III-1: Location in the Tagus Estuary and at the estuary mouth of tide gauges available to validate the 2D configuration of MOHID Tagus Estuary for Hypothesis 2 test.



In the case of the research on Hypothesis 3 the most important validation data should describe the real stratification conditions to be reproduced by MOHID Tagus Estuary. This imply that measurements should include temperature and salinity profiles. The measurement network of MOHID Tagus Estuary includes, as presented in Chapter II, field multiparametric sensor campaigns providing CTD measurements of temperature and salinity at several depths for the Guia outfall location and adjacent areas (integrated in the Monitorization Program for the Estoril Coast). As these are much localized measurements a useful complement of the spatial distribution of temperature can be found in remote sensed satellite SST images of the all Tagus Estuary adjacent coast. Having these issues into account, it is considered the data acquired in February 2005, a month where the wind is predominantly from the

[106] The number of available harmonic components is linked with the length of the time period considered for the harmonic analysis used to generate the components.

Northeast direction, increasing the influence of the Tagus Estuary exit fresh water plume in the stratification conditions at the estuary mouth: the data comprises CTD measurements taken at 01/02/2005 at 11 locations (Table III-5) at the border of the continental shelf and deep ocean and several SST very complete images for the period 11/02/2005-25/02/2005 from the MODIS (level 2) sensor (Table III-6).

Table III-4: Tide gauge tidal components used to validate domain 1 MOHID Tagus Estuary results for Hypothesis 1 test. Location and number of harmonic components are shown. Tide gauges are presented following the coastline from South to North.

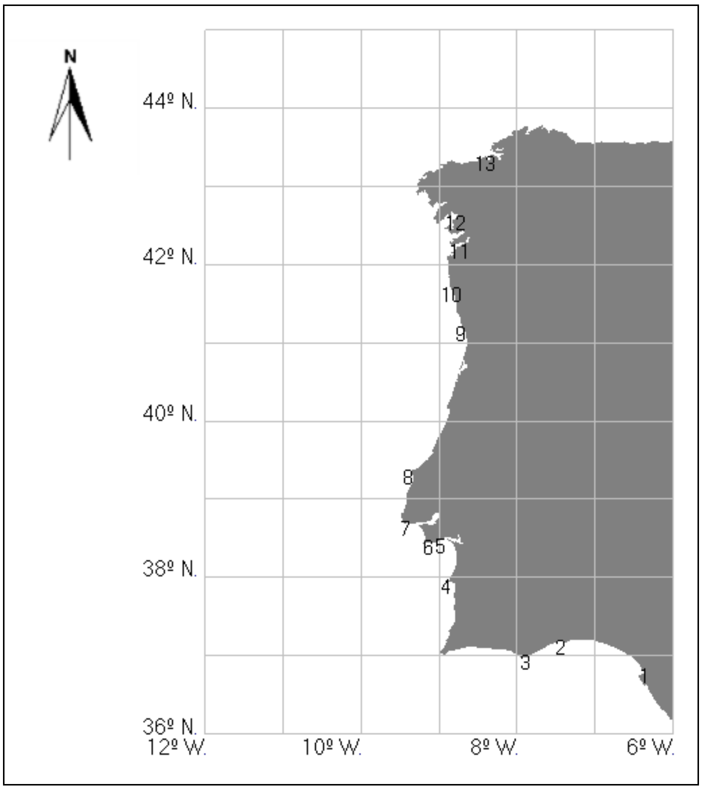
Tide gauge	Geographical coordinates		Number of harmonic components
	Latitude (ºN)	Longitude (ºE)	
Bonanza[a]	36.80	-6.34	37
Guadiana[b]	37.17	-7.42	17
Barra de Faro[b]	37.10	-7.60	22
Sines[c]	37.95	-8.89	30
Sado Baliza[b]	38.46	-8.96	22
Sesimbra[b]	38.44	-9.11	24
Cascais[b]	38.69	-9.42	59
Peniche[b]	39.35	-9.37	9
Leixões[b]	41.19	-8.70	35
Viana do Castelo[b]	41.68	-8.83	9
Vigo[a]	42.24	-8.73	31
Villagarcia[a]	42.60	-8.77	34
Corunha[a]	43.36	-8.39	24

Notes: [a] from Puertos del Estado – Ministérico de Fomento – available in the Internet[107]; [b] from MARETEC, based on historical tide gauge measurements by Instituto Hidrográfico Português; [c] obtained from harmonic analysis of Sines tide gauge measurements[108].

In order to assist the validation of simulation results it is developed in this work a software application for extracting data in MOHID TimeSeries format from MOHID HDF5 format files, which is called HDF5Exporter (see Appendix X).

[107] http://www.puertos.es.
[108] Measurements have hourly frequency and comprehend the periods from 07/03/1997 0h to 20/05/1998 23h and from 27/05/1998 to 11/03/2001 23h. Harmonic analysis is performed with annual time series resulting in a set of 30 statistically significant harmonic components.

Figure III-2: Location in the Western Iberia Peninsula coast of tide gauges used to validate domain 1 MOHID Tagus Estuary results for Hypothesis 1 test.



1 – Bonanza
2 – Guadiana
3 – Barra de Faro
4 – Sines
5 – Sado Baliza
6 – Sesimbra
7 – Cascais
8 – Peniche
9 – Leixões
10 – Viana do Castelo
11 – Vigo
12 – Villagarcia
13 – Corunha

Table III-5: Location and time of CTD measurements used to validate the 2D configuration of MOHID Tagus Estuary for Hypothesis 3 test. All measurements are taken for 01/02/2005.

CTD point	Geographical coordinates		Time 01/02/2005 (hour:minutes)
	Latitude (ºN)	Longitude (ºE)	
G01	38.66958	-9.45705	15:14
G05	38.67767	-9.42845	15:36
G06	38.68022	-9.4623	14:48
G08	38.67465	-9.46617	14:30
G10	38.6685	-9.47235	13:55
G12	38.6831	-9.48913	13:20
G14	38.68017	-9.47517	13:33
G15	38.69745	-9.50187	13:05
G16	38.69027	-9.51023	12:49
G17	38.67417	-9.47155	14:17
GMET	38.66723	-9.57032	12:17

Figure III-3: Location in the Tagus Estuary mouth of CTD measurements used to validate the 2D configuration of MOHID Tagus Estuary for Hypothesis 3 test. Guia outfall is located at point G08.

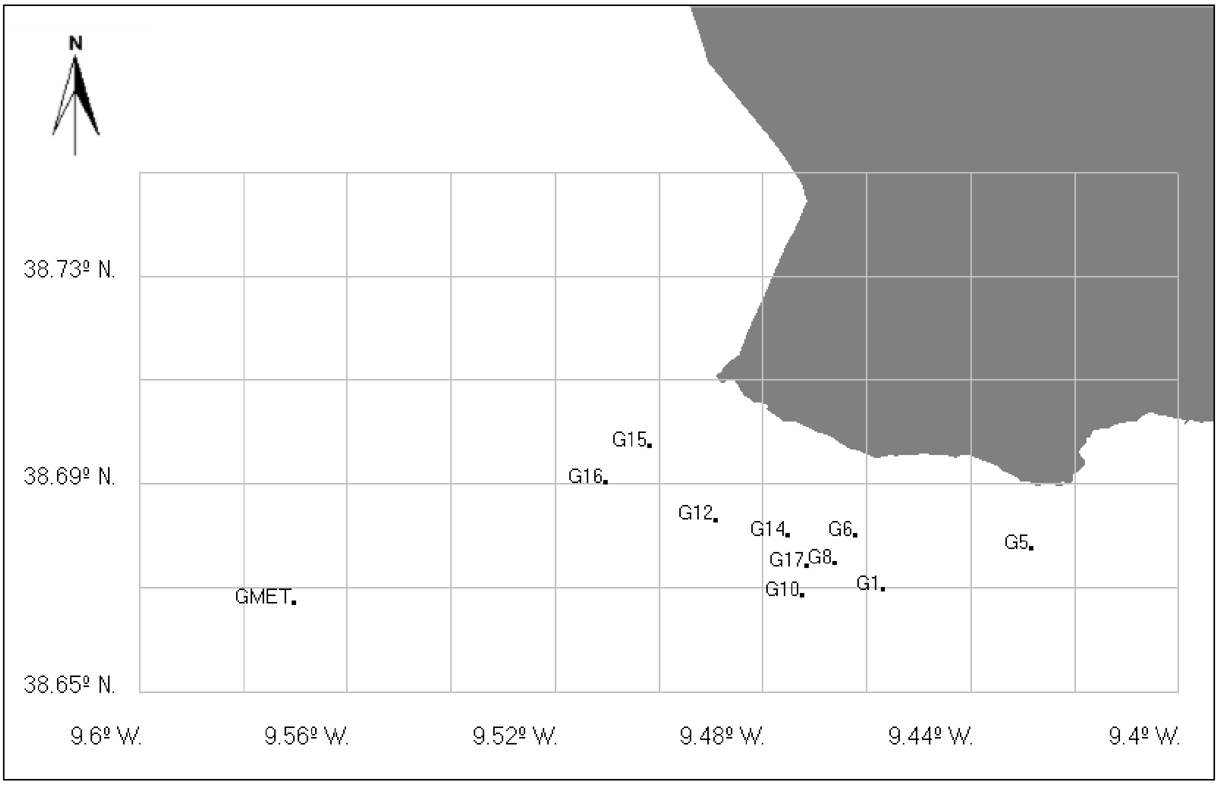


Table III-6: Date and time of MODIS (level 2) SST images used to validate the 2D configuration of MOHID Tagus Estuary for Hypothesis 3 test.

Date	Time (hour:minutes)
11/02/2005	14:15
14/02/2005	13:07
15/02/2005	13:50
16/02/2005	12:55
17/02/2005	13:38
18/02/2005	14:21
19/02/2005	13:25
25/02/2005	14:27

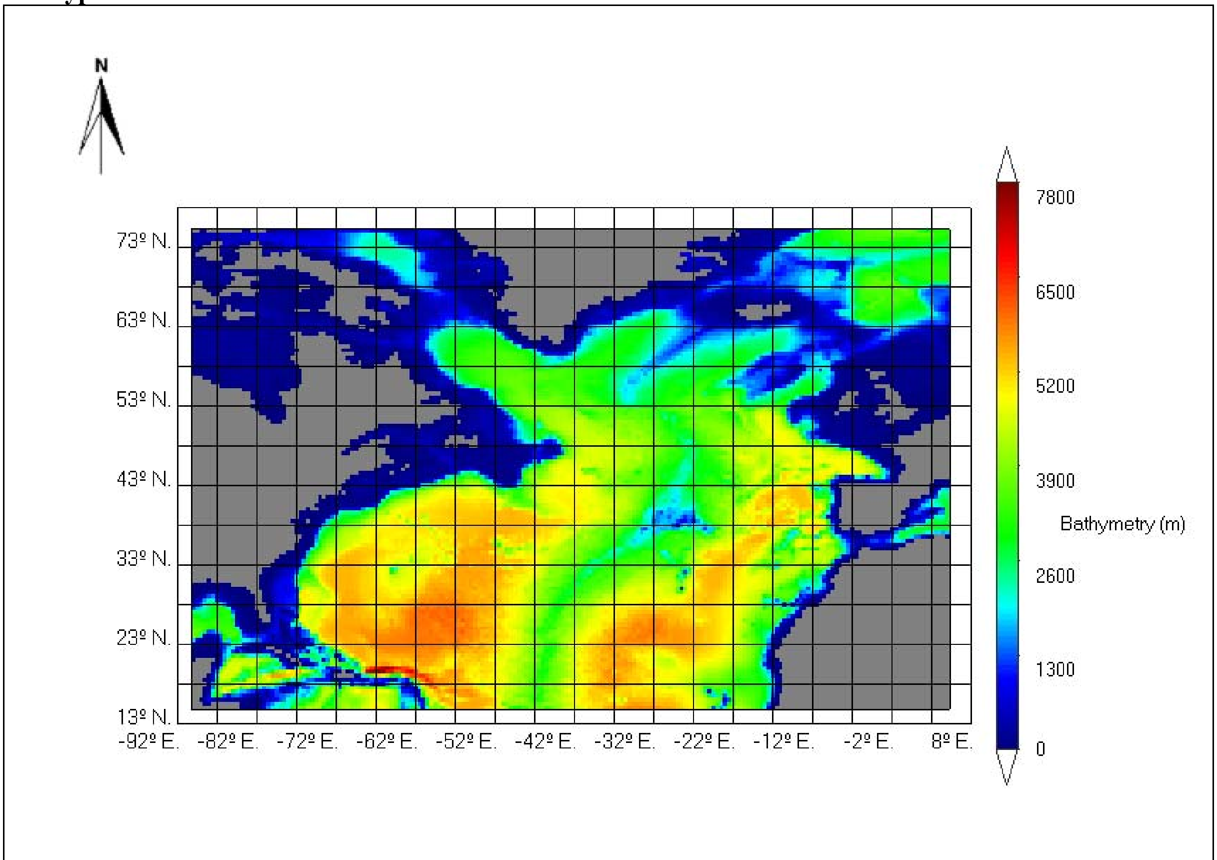
III.4.3. Modelling support data

In what refers to data required for modelling purposes, it has to be distinguished the situation of the time fixed data, which comprises the bathymetry information, from that of the time variable data, corresponding to the meteorological forcing, stratification and river flows.

As for bathymetry, the required information for the creation of domain 0 is chosen, for economy, to be fulfilled with the North Sea bathymetry created by METEOGALICIA and used currently for the wave forecast system operated at MARETEC[109] (Figure III-4). The domain 0 characteristics determined by this bathymetry are provided in Table III-7.

For the creation of the alternative domain 1 bathymetry the ETOPO 2' data is considered whenever high resolution data (from the existing domain 1 bathymetry or the METEOGALICIA data) is not available. Initially higher resolution GEBCO 1' data[110] is considered for this purpose but this data source is abandoned after some wrong bathymetry information was found in several areas of the domain. For the actualization of domain 2 bathymetry in areas where the bathymetric surveys are not available is used the existent domain 1 bathymetry for ocean areas and the ETOPO 2' data for coastal areas due to the unrealistic high depths at coastal areas present in the domain 1 bathymetry; the existent domain 2 bathymetry is not used to fill the gaps in bathymetric raw data because the coastline has been detected to have a small displacement from the actualized coastline considered for bathymetry construction.

Figure III-4: North Atlantic Ocean bathymetry considered for domain 0 of MOHID Tagus Estuary used for Hypothesis 2 test.



[109] System description and forecasts are available at the Internet in site http://www.maretec.mohid.com/ww3/index.htm.
[110] http://www.ngdc.noaa.gov/mgg/gebco/gebco.html.

Table III-7: Bathymetry determinated characteristics for domain 0 of MOHID Tagus Estuary used for Hypothesis 2 test.

Characteristics	Value
Spatial span	92.25ºW – 7.25ºE 12.75ºN – 72.75ºN
Grid resolution	0.5ºx0.5º
Number of cells	191x121

The meteorological forcing data is determined by the time period of the validation data. For the test of Hypothesis 2 it is required sea level atmospheric pressure and wind for 1971-1972. This time requirement does not allow the use of the MM5 model results, since these results were only available since the early 2000's. Instead historical reanalysis are used. Two sources of data in the Internet are envisaged: NCEP reanalysis and ECMWF ERA40 reanalysis. The spatial and time resolution of both data is similar but previous use in MARETEC group supported the use of ERA40 data, since it is considered more reliable facing measurements. In the scientific research this seems also to be the most used data source (e.g. Hoteit *et al.*, 2001). ERA40 data has smaller time and spatial resolution than MM5 results, with 2.5ºx2.5º and 6 hours spatial and time resolutions. For the test of Hypothesis 4 is used the same forcing because, although the test is conducted in a twin experiment framework, the test conclusions are intended to support the future data assimilation of the tide gauge measurements which are available only currently for the 1971-1972 period. The data is available from the producer in NetCDF format while it must be in HDF5 format to be usable in MOHID Water framework. It is developed in this work a special module of the MOHID support tool ConvertoToHDF5 to convert it from the specific NetCDF format to HDF5 format, which is called Module ERA40Format (see Appendix X). For the test of Hypothesis 3, due to the use of recent measurements for validation, the MM5 results are used for forcing.

The MM5 model results used for Hypothesis 3 test require the use of the several MM5 domains (1, 2 and 3) results due to the MOHID Tagus Estuary domains extent and the required use of spatially variable data, as the higher resolution MM5 domain data is not available everywhere. The MM5 model higher resolution domain 3 is used as forcing source where available and smaller resolution domain 2 or domain 1 results are used where the domain 3 data lacks. To make this kind of hierarchical interpolation possible was developed in this work a new interpolation module, called PatchHDF5Files, included in the software application ConvertToHDF5 (see Appendix X). Forcing is obtained from interpolation of MM5 best results of simulations starting at 0h, as the four times a day simulations are present (see Chapter II). The first 6h data from each MM5 results file is discarded as this time interval corresponds to the spin up of this meteorological domain. To support this extraction of data in HDF5 format from HDF5 files it is developed in this work a software application called HDF5Extractor (see Appendix X).

As for the stratification conditions to be used as reference for the open boundary condition for Hypothesis 3, the preferred option would be to use a solution of a large scale model, such as MERCATOR or HYCOM. However, as referred in section III.3., this is not possible option at the time of this work. As such, it is chosen to use climatology stratification. The best climatology for this purpose is the monthly climatology of Levitus (1982). This climatology grid resolution is 0.25ºx0.25º. Since it is coarse relative to the domain 2 resolution, the same vertical profile of temperature and salinity is used for the entire domain boundary. This profile is obtained from Levitus climatology in a location near domain 2 boundary where the data depths are representative of the depths existent in the model bathymetry, so that the profile data has the highest resolution available. The Levitus data is converted to HDF5 format using the MOHID support tool ConvertToHDF5 through Module LevitusFormat.

In what refers to the last kind of modelling support data required in this work, river flows, it is observed a great data scarcity for the period of 1971-1972 when the Hypothesis 2 validation data are available. From a succinct analysis of the water discharges in Tagus Estuary, presented in Appendix II, it is established that a good representation of the Tagus River flow in this period is required. As such for Hypothesis 2 and Hypothesis 4 model simulations are used the daily average flow for Tagus River at Ómnias hydrometric station, available from 01/10/1972 onwards and the default long term average flows for Sorraia and Trancão rivers. For the model simulations conducted for Hypothesis 3 research, hourly flow values exist for the Tagus River from the Ómnias hydrometric station, which is used in the existent MOHID Tagus Estuary configuration, and Sorraia from the Coruche hydrometric station, which is an innovation relative to the existent MOHID Tagus Estuary configuration. Hence, these data are used together with the Trancão River daily averages, already used in existent MOHID Tagus Estuary configuration. These approximations lead to the fact that only the Tagus River flow is realistic in the Hypothesis 2 and Hypothesis 4 model simulations and the Tagus and Sorraia rivers flow in Hypothesis 3 model simulations. Due to the fact that the wastewater discharges inside Tagus Estuary is not representative relative to river flows (Appendix II), these discharges are not considered in the model simulations.

III.4.4. Discussion of research limitations imposed by data

The availability of data, especially the one required for model results validation, implies that some limitations are imposed to this work research. The most visible of those limitations are the time periods for simulation. In Table III-8 are presented the time periods for simulation used for the research conducted in the several hypotheses. In the case of hypotheses 2 and 3 the simulation time period is a direct consequence of data availability: for hypotheses 2 and 3 it is chosen due to validation data and river flow data availability; in Hypothesis 3 test is mainly limited by the availability of

MODIS (level 2) SST images. As presented previously, for Hypothesis 4 test is also chosen a period based on validation data availability, although in this case the linkage between the data and the simulation is indirect. Also, the period chosen for the robustness of the alternative domain 1 configuration is due to the availability of tide gauge measurements at Sines.

Table III-8: Time periods of model simulations used in research on hypotheses.

Hypothesis	Model simulations time period	
1	Test of hypothesis: 10/11/2004 0h → 30/11/2004 0h	Test of robustness of alternative domain 1 configuration: 01/03/1999 0h → 15/03/2000 0h
2	Test of meteorological forced large scale circulation effect on mean sea level: 15/01/1972 0h → 07/10/1972 0h	Test of hypothesis: 01/10/1972 0h → 31/01/1973 0h
3	15/01/2005 0h → 01/03/2005 0h	
4	01/10/1972 0h → 31/01/1973 0h	

It is possible to observe that these data limitations cause the test of the hypotheses to be performed always in the autumn/winter season. In the case of Hypothesis 1 this does not cause any limitation of the test conclusions because the tide propagation which is studied is only affected by the astronomical forces and bathymetry, which are independent of the year season. The non linear interactions between tide and meteorological forcing, which could be dependent of the year season, are not addressed in this study.

For Hypothesis 2 the use of winter conditions in the hypothesis test is not a limitation but instead an advantage: the most comprehensive study of storm surge phenomena in the Portuguese Coast of Gama *et al.* (1994) through the analysis of 2 years of tide gauge data reveals that meteorological storms tend to concentrate in the autumn/winter season. Therefore an autumn/winter period should cause that the mean sea level anomalies to be more easily analysed.

As for Hypothesis 3, the use of winter conditions seems *a priori* to imply a weaker test than one conducted in summer season conditions, as suggested by the review of circulation conditions at Tagus Estuary mouth presented in Chapter 1 and by the study of Leitão (2003) of the circulation in this area. In fact, although the salinity gradients are expected to be larger in the winter due to the increased Tagus River flow due to precipitation, the general water column stratification at the large depths outside the continental shelf is expected to be more important in the summer due to solar heating of sea surface. This causes that stratification is mostly conditioned by the salinity of the river in the winter, while in the summer the large scale depth thermal and saline stratification is more important, bring into play the occurrence of upwelling events, with the characteristic summer winds from Northwest which should be adequately represented by the model. However, the chosen simulation time period is characterized by predominance of winds from the Northeast direction which can propitiate the occurrence of mild upwelling events. The results of test of Hypothesis 3 should be therefore considered valid only for winter conditions.

For Hypothesis 4, the use of winter conditions, which should affect the wind forcing conditions, favours the increased turbulence due to the winter characteristic strong winds which should potentially increase the non linearity of circulation. Also the large winter flows from Tagus should make the advection at the estuary mouth more important. This should make the model representation of the actual system dynamics a more difficult task, showing more clearly the adequateness of data assimilation schemes.

Another important limitation imposed by the validation data is the fact that validation concentrates in the water level representation, for Hypotheses 1 and Hypothesis 2, and thermal and saline stratification representation, for Hypothesis 3. Although the water level is considered an indicator of barotropic hydrodynamics an important property to characterize both barotropic and baroclinic hydrodynamics is velocity. It is only possible to validate the velocity in Hypothesis 4 because it is used a twin experiment. This causes that the conclusions of the test of hypotheses 1 and 2 are only valid for the water level. Also, the conclusions of Hypothesis 3 test are only valid for the sea surface temperature outside the estuary and the thermal and saline stratification in the Tagus Estuary mouth and immediate adjacent continental shelf.