I. Operational Modelling in Estuaries – State of the Art

I.1. Characterization of Estuaries

Estuaries are traditionally defined as semi closed areas where the fresh water coming from rivers and land drainage mixes with salt water (Cameron and Pritchard, 1963).

However, each estuary is a specific case in terms of hydrodynamic processes, a notion which is reflected by the different typologies in which estuaries are classified, linked with the morphology and specific conditions of estuarine areas.

In the next sections are presented the most relevant hydrodynamic and biochemical processes taking place in estuaries and adjacent coastal areas, the most important water quality problems which have been associated with them and the perceived relevance of estuaries.

I.1.1. Hydrodynamics and biogeochemical processes

In ocean faced estuarine areas and adjacent coast the most important hydrodynamic processes comprise the effect of the astronomical tide, density gradients and the effects of wind and atmospheric pressure.

In estuaries with significant astronomical tidal elevation, astronomical tide[1] is usually the dominant driver of currents. Besides the direct effect of the tidal wave, a key factor for the definition of water surface slope (Martins, 1999), tide, affected by bathymetry, density gradients, wind and other atmospheric perturbations, gives rise to a set of non linear mechanisms in coastal areas[2]. These non linear mechanisms cause that the residual circulation in coastal areas affected by tide, in estuaries in particular, cannot be disconnected, in a linear way, from tide itself[3].

[1] Astronomical tide is generated by forces produced by the gravitational attraction of the system composed by the Earth, the Sun and the Moon. Other celestial bodies produce similar forces but with a negligible effect in tide formation (e.g. Parker, 1991). The forcing induced in tide by the different gravitational forces is usually accounted by the decomposition of the total tidal wave in several harmonic components with the frequencies of the generating forces. More than 390 components with frequencies associated to the relative motion of the system composed by the Earth, Sun and the Moon have been identified (Doodson and Warburg, 1961).
[2] These are understood considering the nonlinear terms of the equations of motion (1D): which are the nonlinear continuity term and the inertial and friction terms in the momentum equation (see discussion in Parker, 1991).
[3] As Parker (1991) states, "zero frequency results, i.e., tidally-induced "residual" circulation and changes in mean sea level should be understood to result from non linearly produced asymmetric changes within a tidal cycle, and not be thought of as a generated permanent current or change in mean sea level on which the tidal

The density gradients affect the circulation through the creation of baroclinic forces and the change of turbulent viscosity and diffusion coefficients, due to stratification (Martins, 1999).

The importance of the influence of wind in estuaries is due to these being typically shallow water areas. The wind acts as a driver for water surface shear stress, which can affect the currents in the estuary and the formation of waves. Particularly, the influence of wind direction has been reported by Bertino *et al.* (2002) as being very strong in the hydrodynamics of shallow water areas.

Although the same types of hydrodynamic phenomena are in principle present in estuaries and adjacent coastal areas, the relative importance of each phenomenon is different in each area. Martins (1999) indicates the nature of this difference:

- in estuaries under the influence of tide, this forcing dominates the creation of currents; the relevance of river flow and density gradients is very variable, according to the specific system and the time of the year; the wind effect is comparatively small[4]; bathymetry is very important due to small depths and its effects on tide propagation and shear stress;

- in coastal areas the circulation spatial scale, one to two orders of magnitude larger than in estuarine circulation, determinates the importance of Earth rotation effects and the classification of the horizontal density gradient as the main driver of currents; the wind derived shear stress is also an important driver of surface currents; tide has an associated small effect on average current, which is mostly concentrated in the continental shelf[5] although in certain cases it can generate an important internal tide effect[6].

The combined effect of extreme wind shear stress and reduced atmospheric pressure on the shallow coastal shelf areas gives rise to the storm surge phenomenon (Gill, 1982): this is defined as a meteorological long wave motion, which produces an elevation of the water surface far above the level obtained with astronomical tide.

---

current or tide is super imposed". Martins (1999) also stresses this point by pointing that in an actual tidal wave there is always a residual, although small, which cannot be associated to any of the gravitational tidal forcings composed of frequencies distributed through the frequency spectrum.

[4] This is verified because of a big energetic and time scale difference relative to tide (Martins, 1999).

[5] This small effect is explained by the small magnitude of non linear tidal effects associated with bathymetry, due to depths of coastal areas. Besides this influence on residual current, tide also interferes with the mixing processes (Martins, 1999).

[6] Internal tides are typically generated in presence of stratification when barotropic tides impinge on undersea mountains and ridges, producing energy transfer which is a major sink for the barotropic tidal energy. Submarine canyons have been linked with internal wave energy trapping (Gordon and Marshall, 1976; Hotchkiss and Wunsch, 1982; Hunkins, 1988; Eurostrataform, 2006) and, when narrow, with the amplification of the internal tide (Hotchkiss and Wunsch, 1982; Lafuente *et al.*, 1999). Their influence on energy trapping has been reported as possibly being responsible for an enhancement in turbulence and mixing (Davies and Xing, 2004).

Wind, inertial forces (due to the Earth rotation) may act concurrently with density gradients to cause the occurrence of upwelling events[7]. In these situations a divergence of surface waters in the coast results in the cold water from the ocean depth emerging to the surface (Gill, 1982).

Estuaries are generally the set of high primary biologic production, a factor which is responsible in part for the success of human settlements at their margins. In fact the phytoplankton primary production is the energetic basis of aquatic systems, providing nutrients for higher tropic levels and thus can be understood as the basis of the biological wealth of an aquatic ecosystem.

The high biological productivity of estuaries is connected to its morphological features of having the river discharges and small depths relative to surface width. Together, these features of estuaries are responsible for good light intensity, temperature and nutrient concentration conditions through the water column[8].

When considering the natural control of water quality in estuaries the most important biogeochemical processes to take into account are the turbidity and the nutrient availability, because of their influence on primary biological production. Turbidity is a function of suspended particulate matter and interferes with the light propagation in the water column[9]. The suspended particulate matter concentration is controlled by sediment erosion and deposition processes and production and consumption mechanisms of plankton organisms. Sediments dynamic is driven by currents and by waves[10].

Coastal areas near estuaries are generally classified as ROFI (Regions Of Freshwater Influence)[11] zones, which consist in the main pathway for matter transfer from continents to oceans.

Upwelling events have large relevance in the biologic productivity in coastal areas. In fact, these events and associated filaments provide for the exchange of nutrients from the rich ocean depths to the surface waters, where the phytoplankton has light conditions to thrive its photosynthetic activity based on this nutrient inflow (Gill, 1982)[12]. Evidence has shown that this nutrient source is much larger than the one constituted by anthropogenic sources (Wollast and Chou, 2001).

[7] This is a characteristic phenomenon of the oriental margins of Atlantic and Pacific oceans as stressed by Coelho (2002).
[8] Light intensity depends on turbidity, in which phytoplankton also plays a key role, and nutrient concentration depends on transport processes, natural source and consumption processes and of pollution.
[9] The interference is due to the absorption or dispersion of incident radiation by sediment and plankton, affecting the light extinction.
[10] Currents and waves act by causing bottom shear stress. Waves are important particularly in shallow areas, because they are capable to induce shear stress increments more easily than currents.
[11] Ocean areas influenced by a fresh water discharge.
[12] Upwelling areas are associated with the larger world fishing areas (Gill, 1982).

I.1.2. Water quality problems

Three important water quality problems in shallow high primary production water areas are eutrophication, toxic algae blooms and sediment linked contaminants.

The first two problems are connected with a balance between the photosynthetic and respiratory activity of phytoplankton. In a normal equilibrium situation the oxygen production exceeds its consumption. However, the anthropogenic input of nutrients in the water[13] in some situations[14] cause an exaggerated growth of phytoplankton and algae. This can lead to a situation of oxygen depletion just beneath the surface layer, where the high biological activity is located, which can cause the dead of pelagic (water column) and benthonic (water bottom) organisms: eutrophication. Some of the algae blooms can degrade the water quality also through synthesis of compounds which are toxic to several living organisms from plankton to humans (Sellner *et al.*, 2003); these are called harmful algal blooms.

The problem of the sediment linked contaminants is connected with the fact that it is in the fine particulate matter mass (fraction of suspended sediments) that organic contaminants and heavy metals travel adsorbed. The deposition of the sediments in the water bottom isolates these contaminants from the water column but as the sediments are resuspended, e.g. due to wave activity, contaminants can suffer oxidation and become available in the water column, harming life and water use.

[13] Important sources are discharges of domestic wastewater collecting systems and diffuse discharges from agriculture activity.
[14] This is the case when there are no light limitations.

I.1.3. Relevance of estuaries

The perceived ecological and economical importance of estuaries is expressed by the Portuguese National Water Plan (approved by Decreto Lei nº 112/2002 of April 17):

- are complex and insufficiently known ecosystems which have been perceived as playing a vital role in the global equilibrium of the biosphere;

- are highly productive systems, which in many cases even export that productivity, and supply optimal conditions of reproduction and nursery for approximately 70% of marine fish species stock, contributing positively to replenish coastal areas populations;

- have high biological importance also as places where large populations of birds nidify and inhabit in the winter;

- are important areas where natural water purification processes take place;

- their margins support large human populations and economic activities.

I.2. Operational Modelling in Estuaries and Coastal Areas

I.2.1. Introduction

Operational modeling is included in the knowledge field of operational oceanography. This is defined by Prandle (2000a) as including the acquisition, dissemination and interpretation of measurements of ocean conditions and the performance of real-time modelling, in order to allow forecasts of future conditions.

An operational system for forecast in estuaries and coastal areas is by definition a system management tool. This makes that additionally to a modeling tool, allowing the forecast, and a measurement network, assuring the validation, it is required a data assimilation scheme, to combine model and measurement results[15]. In fact, the literature is generally concurrent in pointing the desirable components of an operational system (e.g. Robinson, 1996; Cañizares *et al.*, 2001):

- numerical model;
- measurement network;
- data assimilation scheme.

Besides these core components, the operational system has also to have efficient validation and data pre and post processing tools, which can include databases for measurements and forecasts storage, information validation mechanisms and tools for supplying the results in a useful format to end-users as stressed by Fernandes (2005).

The whole set of data producing (models, measurement networks), data management (pre and post processing tools) and data presentation (databases, publication devices) components have to be integrated in an efficient way (Fernandes, 2005)[16].

Operational modelling systems, as system management tools, must thus be adapted to the specific needs of the application sites and end-users. These different needs can comprehend different time scales, which are framework in the following classification (Fernandes, 2005):

[15] Nihoul (1984) provides a classification of numerical models according to use which includes the tests-oriented models, process study-oriented models and the system management-oriented models, pointing that in this last case in ocean applications data assimilation is a requirement.
[16] This author provides a simple conceptual diagram documenting the links between each component of an operational system.

- nowcasts: results providing the most exact possible information about the current ocean state;

- forecasts: results providing continuous predictions of the future ocean state, for the longest possible time window;

- hindcasts: results providing a description of past ocean state, including time series with trends.

Fernandes (2005) presents possible specific products and services which can be supplied by an operational modeling system according with time scale typology:

- forecast until 10 days: search and rescue, oil pollution evolution, toxic algae blooms, transport and dispersion of particulate matter and water masses (e.g. fish larvae, contaminants, suspended matter, fresh water), coastal climate (e.g. current patterns, turbulence, temperature, salinity, waves, water level, wind);

- month, season or annual hindcasts and due time forecast: primary production, eutrofization (e.g. nutrients and oxygen), nutrient and suspended matter transport, transport and dilution of potentially pollutant discharges, coastal climate;

- several years reconstitution and observation, simulation scenarios: produce and actualized long time series of aquatic systems climate and ecosystems variability, demonstrate the possible state of coastal or regional climate and ecosystem variability facing hypothetical future climate scenarios, quantifying the effect of different coastal environment management scenarios (namely in terms of primary production, eutrofization and nutrient transport).

Two important types of operational systems in estuaries and coastal areas answer to the needs of storm surge forecasting and water quality forecast.

Referring to the end users specific needs, Castellucci (2002) details the importance of operational oceanography[17]:

[17] Within the same general framework, Fernandes (2005) provides more cases where operational oceanography can give an important contribute.

- for large companies having impact in aquatic resources (due to their location in space or the adopted technology) it can provide a certification on the impact forecast at short and long terms, useful to answer to ecologist groups or concurrent companies demands;
- for government administration organisms it can provide useful information to reduce the dependence of their economical and environmental interventions on lobby pressure;
- in specific cases, such as:
    - assessing the risk of particular decisions (e.g. in safety studies);
    - technology adoption disputes;
    - regulation validation;
    - decision support in policy implementation.

To tackle efficiently the demands, in terms of scales, processes and forcing, posed on operational systems have been pointed some desirable features of the tools to be included in operational systems:

- use of a modular architecture, encapsulating different tasks in modules, and the standardization of input data formats (e.g. bathymetry and tide boundary conditions), so as to make the more flexible and easily scalable (Prandle, 2000b);
- flexibility in operating in diverse computational machines and operating systems, namely using parallel computation such as the Message Passing Interface (MPI) protocol, a technique extremely useful when working with high computational demand systems such as the coupling of a complex biologic model with a 3D hydrodynamic model (James, 2002).

I.2.2. Representative cases

As seen previously, estuarine and coastal system conditions are dependent of large scale conditions. For that reason, to understand estuarine and coastal operational systems, typically of regional or local scale, it is important to known existing global/large scale operational systems.

Global operational systems include the case of HYCOM and MERCATOR[18]:

[18] Other large scale systems are the FOAM and TOPAZ, operated respectively in the United Kingdom and in Norway.

- HYCOM[19]: was founded by the National Ocean Partnership Program as part of the United States Global Data Assimilation Experiment (GODAE) and comprises a model with data assimilation expressed in hybrid vertical coordinates (isopicnic in the stratified deep ocean, sigma in coastal shallow areas and pressure type in mixing layer and non stratified areas) for the North-Atlantic; grid resolution is 1/12º geographic coordinates (Bleck, 2002);

- MERCATOR: founded by six French organizations (Cnes, CNRS/Insu, Ifremer, IRD, Météo France, Shom), is maintained by the work of CERFACS and CLS and occasionally by external computer technology and engineering companies; operational since early 2002, provides real time predictions of North-Atlantic and Mediterranean conditions with 1/15º resolution in the coordinate z system with rigid lid (Zodiatis *et al.*, 2003).

Operational or pre-operational systems developed for coastal areas, with regional or local scale, include:

- GoMMOOS (Gulf of Maine Observing System): one of the most advanced operational models[20], is managed by the University of Maine, the Woods Hole Oceanographic Institution and the Bigelow Laboratory for Ocean Sciences; provides surface[21] and depth[22] real-time, remote sensed[23] and radar[24] measurements, through the internet and telephone, and model forecasts of waves (using several models) and oceanic circulation (using Princeton Oceanographic Model - POM); this information is mainly oriented to support fishing activities but it aims in the future to support other end-users, such as scientific researchers, educators, search and rescue teams and public health authorities and to integrate the national system called Integrated Sustainable Ocean Observation System[25] (Hankin *et al.*, 2002);

- PORTS (Physical Oceanographic Real-Time System): composed of several operational systems for coastal areas of the U.S.A., such as the Cheasepeake Bay and the New York/New Jersey Port systems, directed to support local navigation but also occasionally oil pollution incidents and water quality and ecosystem monitorization (NOAA, 2002); these

[19] An abbreviator for Hybrid Coordinate Ocean Model.
[20] Accessible in the Internet through the site www.gomoos.org.
[21] Wind, waves, air temperature and fog.
[22] Currents, temperature, salinity, color, turbidity, dissolved oxygen and other parameters.
[23] SST, color, and surface wind.
[24] Currents, through the use of CODAR system.
[25] This national system will integrate the Global Ocean Observing System.

systems include in some cases the acquisition and supply of water level measurements and modeling work using the POM, real-time measurements and meteorological models;

- Juan de Fuca Straight Operational Model: defined for the area southwest of British Columbia in Canada and used in the Juan de Fuca Straight[26], forecasts surface currents with POM model for two different resolution domains (Tinis and Thomson, non dated);

- MERMAID – OPMOD system: operating in the North Sea, integrated in the German Network of Marine Environment Monitorization (MARNET), it is composed of the measurement and monitorization system MERMAID and the operational modelling system OPMOD, measures water temperature, conductivity, turbidity, pH, dissolved oxygen and other chemical and biochemical parameters, and forecasts water level, currents, salinity, temperature and density (Pfeiffer *et al.*, non dated);

- Farvands system: developed by DHI – Water & Environment, in Danmark, integrates the Danish Program of Acquatic Monitorization and Assessment, uses the numerical model MIKE 3 for forecasting currents, water level, salinity and temperature for the Danish inner waters, Baltic Sea and great part of the North Sea; includes a data assimilation module;

- Mediterranean Forecasting System: developed based on the Mediterranean Forecasting System (MFSPP) pilot project, forecasts in near real-time the ecosystem variability in Mediterranean coastal zones, with a resolution of 12.5 Km; it is composed by a local measurements acquisition service, a meteorological data center and a modeling and data assimilation center; measurements, both remote sensed and in-situ, are used for model initialization and modeling using data assimilation techniques (Zavatarelli and Pinardi, 2003); it provides boundary conditions, through the use of nested modeling, to high resolution models located at the Adriatic Sea (ADRICOSM), Cyprus coast, Aegean Sea, Gulf of Lyon and Malta coast;

- Norwegian Meteorological Institute system: based on the 3D ECOM (Estuarine, Coastal and Ocean Model, a version from POM), forecasts currents and storm surges for an horizontal 20 Km resolution grid comprising the Norway and the Barents seas (Flather, 2000);

- Danish Meteorological Institute (DMI) system: composed of a 2D model with three nested domains (18, 6 and 2 Km resolutions), forecasts water levels for flooding alerts in the Danish coast comprising North Sea, Baltic Sea, Skagerrak and Danish Belts (Vested *et al.*, 1995);

[26] Located in the boundary with the U.S.A. state of Washington.

- Dutch Continental Shelf Model (DCSM): a 2D model operated by KNMI, forecasting tides and storm surges for the Northwest European continental shelf, with a horizontal grid of ¼º resolution in longitude and a 1/6º resolution in latitude (about 16 Km) (Gerritsen *et al.*, 1995);

- Management Unit of the North Sea Mathematical Model (MUMM): a 2D model, with domain covering all the Northwest European continental shelf, with a resolution 5' in longitude and 2.5' latitude (Flather, 2000);

- Meteorological Office Model: operated in United Kingdom this system forecasts tide and storm surges in the Northwest European continental shelf using CS3 model with 1/6º resolution in longitude and 1/9º resolution in latitude (about 12 Km) (Flather, 2000);

- Nivmar: a 2D tide water level forecasting model, developed and operated by the Clima Marítimo/Puertos del Estado, supplying forecasts in all Spanish coast with ¼º longitude and 1/6º latitude resolutions (Flather, 2000).

Apart from the Mediterranean Forecasting System, oriented to biogeochemical and water quality conditions forecast, most of the regional or local systems concentrate on hydrodynamics forecast, specially water level and storm surge[27] oriented. This fact reflects that it is on navigation activities and floods and storm surges emergency management that a real-time forecast is most valuable. Nevertheless, new needs are arising and coastal operational oceanography is currently concurrently expanding its main field of application for the ecosystems description and water quality (Prandle, 2000a).

I.2.3. Modelling tools

Regardless the final end-use for the operational system results, the modelling tool component of an operational system should be always based in system description in physical terms, since the physical system, expressed in system hydrodynamics, is the basis of the other systems, such as the chemical or biological system as stressed by James (2002). In the next sections are describe the main issues to be taken into account in hydrodynamic and in biogeochemical/water quality modelling.

[27] A comprehensive review of storm surge models used for the North Sea and Baltic Sea is presented in Cañizares *et al.* (2001).

I.2.3.1. Hydrodynamic modelling

This modeling tool has to be chosen or developed taking into account the main modeling issues relevant for the hydrodynamic modeling of coastal, shallow, waters (James, 2002), which reflect the main aspects of coastal hydrodynamics referred before:

- turbulence parameterization: as the flow is usually turbulent and the understanding of turbulence, and of the way to parameterize the diffusion processes, is regarded as a major conditioning of physical models;
- waves: provide for a better description of the bottom shear stress, making use of a wave model, or using an simplified approximation of wave activity;
- accounting of baroclinic instability: which constitutes an important mechanism for the transport of river matter to the sea, through coastal circulation;
- large scale circulation effect: namely the tide, wind and density gradients contributions.

The quality of the hydrodynamic models results is perceived as being limited essentially by boundary conditions: results are acceptable for water level, currents, temperature and salinity for a horizontal grid resolution of 10 Km, since the boundary conditions for this scale are of good quality (James, 2002). Boundary and initial conditions can actually be a constraint of the spatial and time resolutions of modelling forecasts (Prandle, 2000a). The issue of boundary conditions dependence advocates the use of good numerical schemes for the boundary conditions parameterization in the model but also the use of data from other models for boundary, especially the open boundary, conditions (James, 2002).

As described in the previous section, in the estuarine hydrodynamics the adjacent coast hydrodynamics plays generally an important role, due to tide but also due to the existence of important ROFI zones. The forecasting systems for estuaries frequently should thus include also the adjacent coast, but be able to deal with the specificities of relevant hydrodynamic factors of each area (estuary/coast). The account of these specificities includes the integration of different spatial and time scales while limiting the computational and time costs to a level acceptable to operational use, as complicated bathymetries require the use of high resolution grids which increases the computational time.

In terms of spatial scales the two most important issues are grid horizontal resolution and grid geometry (vertical resolution). Grid horizontal resolution is frequently associated with an issue of bathymetry representation: high resolution in small depth areas where bathymetry has an important effect on circulation, comparatively smaller resolution in high depth areas where it is not so important. In 3D models the grid geometry choice is very important (Martins, 1999):

- in estuaries the importance of bathymetry and the dominance of tide relative to wind recommends the use of a sigma coordinate geometry[28];

- in coastal areas the importance of density gradients, acting on the horizontal plane, objects to the use of a sigma coordinate geometry and recommends the use of cartesian geometries[29]; except in seasonal situations when the water column is well mixed.

One of the most common and efficient approaches to deal with the different spatial and time scales requirements in different areas of the modelling forecast domain and to adequately provide open boundary conditions is the nesting of domains[30], by which more detailed resolution can be used only where needed (James, 2002). Several of the forecast systems referred before use this technique: the ADRICOSM system (Zavatarelli and Pinardi, 2003) and the DMI system are two examples (Flather, 2000). The relevance of nesting techniques is expressed in the development of projects such as GODAE, which aims at integrating several systems in larger ones (see discussion in James, 2002).

Two main types of nesting can be used: one-way, when the information travels from the coarser domain to the finer one, and two-way, the most advanced form, when the finer domain results also affect the coarser domain results (e.g. James, 2002). However, the use of nesting still provides some constraints to the system; some of the most important, so as to ensure model stability and smooth the transition between areas, are summarized by Cañizares (1999):

- the open boundaries can only be defined in the coarsest grid;

- spatial resolution from one level to another is reduced by a fixed factor equal to 3;

[28] This geometry provides a smooth representation of bathymetry allowing an accurate treatment of the vertical momentum transfer (Bode and Hardy, 1997).

[29] The drawbacks, in the form of numerical calculus errors, of the use of a sigma coordinate geometry in these cases are linked with the baroclinic forcing, associated with the density gradients, being oblique to the grid in areas with large bathymetry transitions. The cartesian coordinates approach avoids these drawbacks and due to its simplicity is generally superior to other coordinate types in terms of computational use (Martins, 1999).

[30] The nomenclature of nesting is controversial. Cañizares (1999) states that nesting is different from the boundary transfer from a coarse model to a finer one; the latter is also called one-way nesting. In the nesting definition presented by this author the information between grids travels in both directions.

- water depths in common grid points along the internal boundaries should be equal in both the coarse and the fine grid[31];

- water depths in coarse grid has to be equal in three points orthogonal to the internal boundary[32].

In cases where the constraints of nesting are unacceptable a possibility is to use a spatial step variable grid and only one domain (James, 2002).

Another important issue for hydrodynamic modeling in estuarine and coastal areas operational systems is the meteorological forcing. Besides the wind forcing, which, as indicated previously, constitutes usually an important forcing of coastal and estuarine hydrodynamics, another important meteorological forcing in storm surge models is sea level atmospheric pressure, which can be accounted in hydrodynamics namely through the use of the inverted barometer formulation for defining boundary conditions, the most usual procedure as in Cañizares (1999)[33]. The inclusion of good surface (meteorological derived) boundary conditions is recognized as critical in hydrodynamic operational modeling (James, 2002).

For that reason and the common forecast aim of operational systems it is common that this meteorological forcing is provided by a meteorological model, being one of the most important developments in improvement meteorological forcing the development of coupled ocean-atmosphere models (Prandle, 2000b). This coupling is perceived as very relevant especially for operational storm surge forecasting systems, where it constitutes the main core of the systems (Bode and Hardy, 1997). Usually the coupling is performed in one-way mode, with the atmosphere influencing the ocean and not the opposite, but the use of two-way coupling is reported as being very advantageous in operational systems (Prandle, 2000b; James, 2002).

From the main systems presented before the examples of the use of meteorological forcing from models are diverse (Flather, 2000): the Nivmar system uses wind and sea level atmospheric pressure forcing from HIRLAM model operated by the *Instituto Nacional de Meteorologia*; the DCSM system uses wind and sea level atmospheric pressure results from the KNMI HIRLAM; the DMI nested model system uses the LAM (Limited Area Model, operated by the British Meteorological Office)

[31] This is accomplished by interpolating water depths from the coarse grid to the fine grid: "between the common points along the internal boundary, the water depths in the fine grid are linearly interpolated using the values at the common points" (Cañizares, 1999).
[32] The points are at the border and one point at each side.
[33] In the model used by Cañizares (1999) the open boundaries are defined using tidal data and a correction term associated with changes in the atmospheric pressure relative to a reference value.

forcing; the Norwegian DNMI system operates with operational wind and sea level atmospheric pressure forecasts from HIRLAM; the United Kingdom tide and storm surge forecast system uses also the wind and sea level atmospheric pressure British Meteorological Office LAM results.

Another very relevant boundary condition, which must be adequately defined in hydrodynamic models of coastal and estuarine operational systems, is constituted by the fresh water, namely river, discharges (James, 2002).

Due to the often strong hydrodynamics of coastal and estuarine areas, in the context of operational systems adequate attention must be given also to model bathymetry representability. This issue goes beyond the aspect of horizontal grid resolution to contemplate, in some cases, the frequent bathymetric data acquisition to account the fast bathymetry changes in the model hydrodynamics (James, 2002).

In the last years several important developments have been achieved in ocean and coastal hydrodynamic modelling which can benefit operational systems results. Some examples, still not accounted in several operational systems, are presented by Bleck (2002):

- improvement and diversification of open boundary conditions;
- vertical coordinate generalization;
- higher order schemes for the advection term discretization;
- baroclinic pressure discretization;
- two-way nested modelling;
- non-hydrostatic pressure;
- new parameterizations for vertical turbulence;
- vertical mix induced by waves.

I.2.3.2. Biogeochemical/water quality modelling

As referred before, the use of tools for biogeochemical and water quality modeling in operational systems is still an expanding field. Two specific needs of this type of modelling are relevant in explaining the constraints imposed by this type of modeling in operational systems (James, 2002):

- adequate data for model test and validation[34];
- higher computational power requirements than hydrodynamic modelling, due to the consideration of a larger number of variables and processes, namely through the use of ecosystems complex models such as the European Regional Seas Ecosystem (ERSEM)[35].

In systems intended to forecast water quality in domains where important point pollutant sources are present it is adequate to use lagrangean transport models (Neves and Martins, 1996). This happens because these allow to known the patterns of transport of water masses associated with not only time scales larger that the tidal cycle, which is allowed by the eulerian transport models, but also of time scales in the tidal cycle, both time scales important for pollutant dispersion for point sources.

It is also relevant in biological and water quality modeling to adequately take in consideration the ecosystem diversity, where phytoplankton and zooplankton are important components. This diversity is usually taken in consideration in modeling through the representation of several biologic groups, each with specific variables and physiologic processes, and the description of population dynamics in terms of carbon and nutrient fluxes between groups and between the groups and the environment. This approach provides good modeling results for lower trophic levels but should be substituted by other approaches, such as population and individuals models, for higher trophic levels (James, 2002).

[34] According to Prandle (2000b), one of the main challenges in this issue consists in associate data from difference sources to minimize the individual deficiencies.
[35] According to James (2002), only recently the required computational power and the parallelization systems were made available at an acceptable cost.

I.2.4. Measurement networks

The measurement networks in the context of operational systems can support two types of use (Allen *et al.*, 2002)[36]:

- model results validation, offering an opportunity to assess, validate and refine models (Carmillet *et al.*, 2001);
- establishing of initial conditions and regular correction of model state variables, through data assimilation.

If for data assimilation real-time data are desirable, for model validation it is important to have an historical measurements data base. Therefore, operational systems should provide for an efficient organization and storage of measurements data.

Essentially two types of measurements can constitute a measurement network (e.g. Cañizares *et al.*, 2001):

- continuous measurements, supplied essentially by moored/in-situ instruments (ADCP, XBT, tide gauges): these have as advantage the providing of continuous data for diverse parameters (currents, dissolved oxygen, temperature, salinity, chlorophyll, nutrients, water level) but as disadvantage the covering of only one single point in horizontal;
- remote sensing satellite data: have the advantage of covering a large area with a large spatial resolution but the disadvantage of presenting data which is very sparse in time and covering a thin band near surface.

The choice in using or not each type of data is dependent on the time and space requirements of the specific application of the operational system. Thus, as pointed by Prandle (2000b), coastal systems should rely essentially on *in-situ* measurements, while remote sensed data should compose the

[36] The choice of the type of data use can depend on the specific parameters aimed in the measurement use process. Philippart and Gebraad (2002) refer the case of the water level forecasts and point the possible off-line use in model calibration, and on-line use in data assimilation. Calibration is particularly relevant in model parameters that are known with only limited accuracy. Philippart and Gebraad (2002) address in the context of water level and storm forecast the case of depth (bathymetry), bottom friction and boundary conditions such water levels and geometry. It is important to keep in mind the limitation imposed to the maximum number of parameters that can be calibrated due to the equilibrium between the number of measured data and the number of degrees of freedom.

measurement network of large scale systems. However, these typologies should be considered indicative and not mutually exclusive.

In fact, as stressed by Fernandes (2005), the use of remote sensed data in coastal areas in an expanding field, which is fuelled by the significant advances in data representability achieved with new sensors such as MODIS-Aqua (NASA) and MERIS (ESA) and new data post processing algorithms applied to old sensors such as Seawifs (NASA).

A remote sensed measurement system with apparent high potential for coastal areas operational systems is the use of radar devices to measure wind, waves, surface currents, bathymetry and oil discharges. The use of this type of device is expected, in particular, to contribute significantly to a better understanding of wave propagation processes in shallow water (Prandle, 2000b).

Also, in large scale 3D models it would be advantageous to use vertical profiles of in-situ measurements (Cañizares *et al.*, 2001).

One of the most important in-situ measurement resources for operational modeling in coastal area is tide gauge data. In fact, tide gauges measurements are included in the measurement network of several of the operational systems presented before, such as the DMI, DCSM and the Meteorological Office systems (Flather, 2000). According to Mourre *et al.* (2006), these data provide a unique monitoring of sea-level variability along the coasts of the world ocean and the sea-level “gathers the signature of most of the processes at work at the difference temporal and spatial scales”. This resource is used primarily for model validation purposes and its use for data assimilation is still not widespread due to complexity of coastal hydrodynamics. In the case of the large scale MERCATOR model Mourre *et al.* (2006) refer that tide gauge data are not assimilated partly because “data assimilation at the coast still requires further research before it can be used operationally”, since “the shape and evolution of model error statistics required for data assimilation experiments have been shown to be more complicated in coastal areas over the open ocean”. The same situation prevents the use of satellite altimetry data near the coast in the Atlantic and Mediterranean basin models.

If the objective of a measurement network is to provide data for data assimilation as well as for model validation it should be considered when selecting such a network that measurements should have a long time series so as to allow a proper computation of statistics (Mourre *et al.*, 2006).

Another important component of the measurement network of an operational system for coastal areas and specially estuaries is constituted by river flow measurements. The usual practice, according with the review of operational systems performed by James (2002), is to consider average flows, assessed based on diverse time periods, according with the specific case. This is the case of the ADRICOMS system, incorporated in the Mediterranean Forecasting System, which uses a monthly climatology of river flow (Zavatarelli and Pinardi, 2003). However, as noted by Fernandes (2005) this practice could inadequate when the river flow is characterized by sudden changes.

I.2.5. Data assimilation tools

The major driver for the use of data assimilation tools in operational systems is the understanding that the modeling tools used to provide forecasts in these systems are affected by diverse possible error sources, such as the ones provided in Table I-1.

Table I-1: Possible error sources in an ocean model (Carmillet *et al.*, 2001).

- Misspecification of initial conditions
- Biologic processes parameterization
- Specification of external forcing and associated constraints
- Spatial discretization of the numerical scheme used to solve evolution equations
- Coupling mechanisms between physical and biological processes, which are in some cases solved with different time steps
- Approximations considered in model equations

Data assimilation provides a way to minimize the problems associated with deficiencies in real systems representation by modelling results in operational systems. In fact, data assimilation consists in a tool for optimal combination of real world information from measurements and dynamical processes information from models (e.g. Brusdal *et al.*, 2003).

In particular, two major modeling problems where data assimilation has been pointed as being a good solution for improving forecasts in coastal areas are the modelling with poor horizontal grid resolution (Moure *et al.*, 2006), where tide gauge measurements can be very useful, and storm surge forecasting (Cañizares, 1999; Cañizares *et al.*, 2001).

Despite its recognized usefulness in coastal areas forecasting, data assimilation, while generally used in large scale ocean operational systems, is still not very disseminated in operational systems for coastal areas. Several reasons have been pointed for this situation: the relative lack of suitable measurements in coastal areas comparatively to ocean areas (James, 2002), high computational power demands, doubtful acceptability of simplifications of data assimilation methods, difficulty in connecting measurement variables to model variables and in the definition of measurement and model errors (Minster, 1998). It should be noted that the availability of measurements in coastal areas is a very relevant issue as the space and time variability of coastal areas conditions is much larger than in ocean conditions (James, 2002).

Following this succinct introduction to data assimilation tools relevance in operational systems and taking into account the importance of this field of knowledge and technology for this thesis, a more detailed review of data assimilation methods is made in a subsequent section of this chapter.

I.2.6. Validation techniques

Regarding validation techniques in operational modeling one has to distinguish two sets: validation for modeling and validation specific for data assimilation applications.

In terms of validation for modeling, one of the most useful and used techniques is the **harmonic analysis** of water level results, both provided by a measurement device (e.g. tidal gauge) or a model. A usual part of the water level data analysis is the separation of tidal from non-tidal components of the data signal. The tidal signal can then be discarded or its characteristics described in some fashion useful for further analysis.

One of the most relevant techniques for model validation making use of maps of spatially and time variable data is the **Empirical Orthogonal Function (EOF) analysis** (Emery and Thomson, 2001). With this technique are identified modes of spatial variability, their time evolution and the relative importance of each pattern. The modes produced by this analysis are orthogonal in space and time and the first mode is the mode that maximizes the explained variance over the total dataset. These modes are not physically linked and derive from statistical analysis. The results of EOF analysis can be used for variability comparison between model fields and spatially variable measurements and for reconstruction of model or measurements fields with only some of the modes of variability, usually to remove variability due to noise. The representation quality of EOF analysis results can be assessed facing observations.

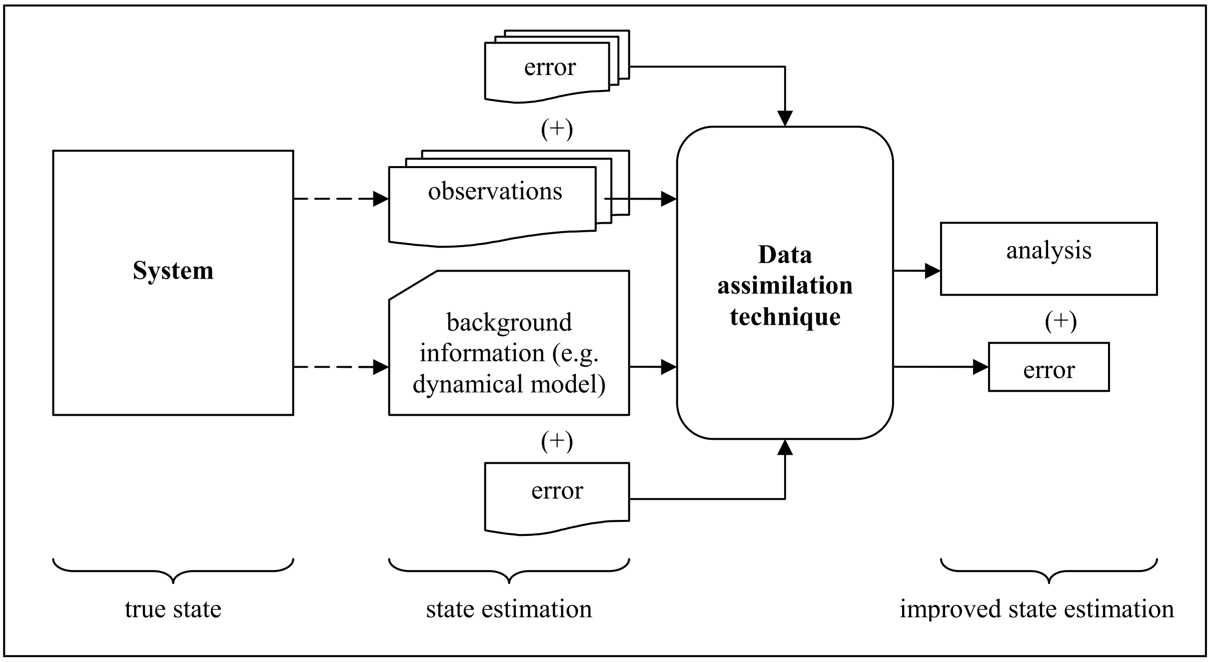
Validation of data assimilation methods is generally made using a **twin model experiment** framework. This is an experiment where one model provides measurements to be assimilated in another model, derived from the first with the introduction of a perturbation. This is a validation which must be made off the operational setting. Examples of twin experiments can be found e.g. in Cañizares (1999). The use of the twin experiment allows the test of the data assimilation method in a situation of perfectly known error statistics.

I.3. Data Assimilation Techniques

I.3.1. Main concepts

The main elements of the data assimilation problem are presented in Figure I-1. In the presence of a system whose true state should be estimated, such as the hydrodynamic conditions in an estuary, data assimilation techniques allow the combination of the incomplete state estimations provided by **observations** of the system, e.g. tide gauge measurements, and **background information**, such as an hydrodynamic model solution in order to create an improved state estimation, which is called the **analysis**.

Figure I-1: Main elements of the data assimilation problem.



In hindcasting and nowcasting the intervention of the data assimilation technique is required if the system state is undetermined by observations: if the system state is overdetermined by observations then the analysis can be produced simply through interpolation. Usually the system state is underdetermined by observations, with data being sparse and only indirectly related to the true state variables. In this case is necessary to rely on the background information, which consists in an *a priori* estimate of the system state. In forecasting data assimilation can be useful even in a case of the system state being overdetermined by observations, since it can improve the initial state to perform the forecast using a model.

The processing of data assimilation is based on the formulation of the system evolution in time according with the background information and the formulation of the observations evolution in time. The system is represented in these formulations in the form of a **state vector**, which is the least collection of variables considered to represent the state of the system in a specific time in order to forecast its future state (Kalman, 1960). This state vector is defined according with the degree of approximation it is intended between the state estimate and the true state: the elements of this vector relate to the true system state through the choice of a discretization.

Based only on the state estimates provided by the background information and the observations the true system state estimation problem is ill-conditioned and without solution. Only if **random errors** are assumed in both the system time evolution and the observations time evolution formulations it can be found an optimal combination the different state estimates, through the consideration of the problem as an optimization problem (Bertino *et al.*, 2002).

The analysis represents the optimal estimation of true state but has also an associated error due to the error in the state estimates of observations and background and also errors associated with the data assimilation technique. These sources of errors are summarized in Table I-2.

Table I-2: Error sources in the three elements used for calculation of analysis in the data assimilation problem (Cañizares, 1999).

Background information	Observations	Data assimilation technique
Dynamic model (e.g. Table I-1)	Instrumental noise	Approximations
State vector discretization	Interpretation of observed variables as state sector variables	Simplifications

The relationship between the observations and the state vector is established through the **observation operator** (H) which is a function from state vector space to the observations space. This operator can change with time if the observing network characteristics (locations, types of measurements) are not constant in time.

The discrepancies between observations and the background derived state vector are gathered in the **innovation vector** at the observation points.

The solution of the optimization problem can be achieved following one of two equivalent approaches (e.g. Malanotte-Rizzoli and Tziperman, 1996; Robert *et al.*, 2006): the statistical estimation theory, with the use of **sequential methods**, the optimal control theory, through the use of **variational methods**. These approaches are detailed in the next sections.

I.3.2. Sequential assimilation

I.3.2.1. Introduction

In sequential methods the correction of the system state estimated by the model with the information contained in the observations is performed at specific time instants (in which the observations are available), according with the corresponding accuracy of the model estimate and the observations estimate (Carmillet *et al.*, 2001). A major difference relative to variational methods of data assimilation is that observations only have influence on the estimated state at the observation time or future times and not at past times (Cañizares, 1999).

Sequential methods may be classified as **simple** and **advanced** (Bertino *et al.*, 2002) based on the way the model results correction is evaluated. Examples of each category of methods are presented in Table I-3.

Table I-3: Classification of sequential methods (Malanotte-Rizzoli and Tziperman, 1996; Cañizares, 1999; Bertino *et al.*, 2002).

Simple	Advanced
Direct insertion	Kalman filter
Blending	Optimal interpolation
Nudging/newtonian relaxation	

Simple methods evaluate the correction based on empirical criteria. These have the advantage of being easy to implement but they do not provide any information about the uncertainty of the estimate resulting from data assimilation (Malanotte-Rizzoli and Tziperman, 1996). Direct insertion consists in the substitution of the value of variables in the model state estimate by observations value. Blending is the substitution of model variables value with a value obtained by weighting the values of both the model state estimate and observations (application is presented by Matsumoto *et al.*, 2000). In newtonian relaxation the dynamic model relaxes towards the observations through the addition to the state evolution equation of a term proportional to the difference between observed and model calculated variables (Cañizares, 1999).

The direct insertion method and the Kalman filter represent the two extremes of updating schemes (Evensen, 1992). Each has specific disadvantages: while by the first method spikes and discontinuities appear in the model state estimate, on the Kalman filter the error covariance matrices of both the state estimate and the observations must be known. Simple methods, such as the newtonian relaxation, have been referred as having limited usefulness for complex models of regional and local scale and scattered measurement networks (Brasseur *et al.*, 1999). Due to their importance, Kalman filter method and its most important derived methods are presented in the following sections.

Sequential methods, specially those derived from Kalman filter, are perceived as especially more suited for operational forecast system applications than variational methods due to their flexibility, as presented in Table I-4.

Table I-4: Advantages of sequential methods relative to variational methods in operational ocean models data assimilation applications (Hoteit, 2001).

- Support different degrees of degradation, allowing the adaptation of the data assimilation scheme to specific availability of resources (e.g. computational, human)
- Allow a modular implementation of the numerical code, with the model and data assimilation codes being independent

I.3.2.2. Kalman filter

The Kalman filter is developed by Kalman (1960) for application in the field of control theory and in managing industrial production systems. It is later, in the end of 1980 decade, introduced in oceanography applications for data assimilation purposes.

Its algorithm assumes the following system of equations, which formulates the time evolution of the system state according with the background information (equation (I.1)) and the time evolution of observations (equation (I.2)).

$$\mathrm{x}^{\mathrm{t}}(t_i) = M\left[\mathrm{x}^{\mathrm{t}}(t_{i-1})\right] + \nu(t_{i-1}) \qquad \text{(I.1)}$$

$$y_i^o = H\left[\mathrm{x}^{\mathrm{t}}(t_i)\right] + \omega_i \qquad \text{(I.2)}$$

Where $\mathrm{x}^{\mathrm{t}}(t_i)$ is the true state vector of the system at time t_i, M is a model describing system dynamics, ν is the error of the state estimation by the model, y_i^o is the observation vector, H is the observation operator and ω is the error of representation of the true state vector in the observations space.

The solving of this system of equations is based on the least squares estimator (Emery and Thomson, 2001). The Kalman filter algorithm, organized in one forecast step and one correction step (Cañizares, 1999), then results in the set of equations provided in Table I-5.

Table I-5: Kalman filter algorithm equations (e.g. Hoteit, 2001). Equations are presented under the notation of data assimilation established by Ide *et al.* (1997):

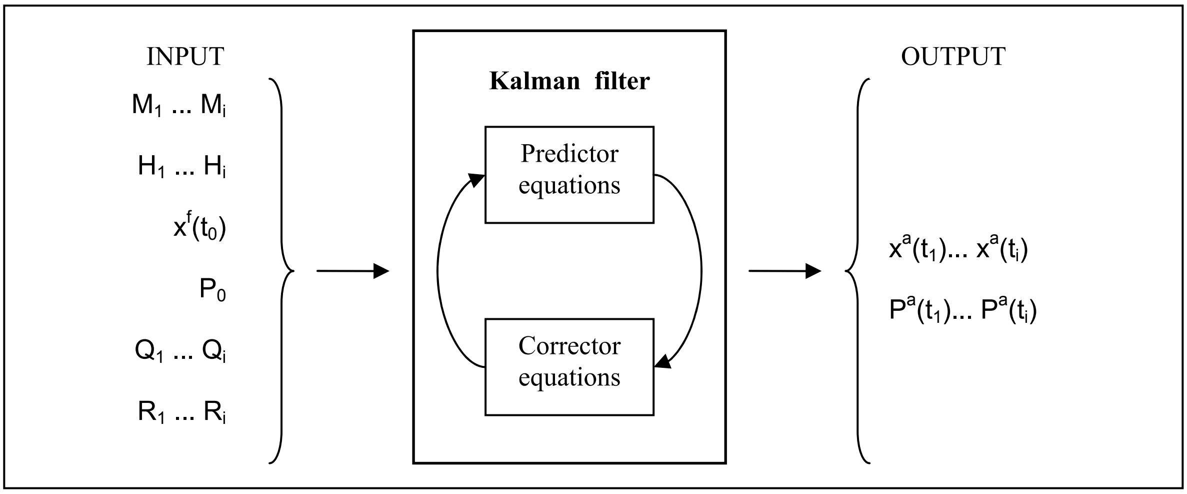
Forecast step	Correction step
$x^f(t_i) = M_{i-1}\left[x^a(t_{i-1})\right]$ (I.3)	$K_i = P^f(t_i)H_i^T\left(H_i P^f(t_i)H^T_i + R_i\right)^{-1}$ (I.5)
$P^f(t_i) = M_{i-1}P^a(t_{i-1})M^T_{i-1} + Q(t_{i-1})$ (I.4)	$x^a(t_i) = x^f(t_i) + K_i\left[y_i^o - H_i\left[x^f(t_i)\right]\right]$ (I.6)
	$P^a(t_i) = (I - K_i H_i)P^f(t_i)$ (I.7)

Notes: P is the covariance matrix of error of state estimate by (I.3), Q is the covariance matrix of ν, K is the Kalman gain, R is the covariance matrix of ω, x^f is forecast state vector, x^a is analysis state vector.

Relatively to the ordinary least squares estimator (Emery and Thomson, 2001) the Kalman filter is distinguished by the formulation of the Kalman gain K. This matrix describes the weights given to the innovation vector in the calculation of the analysis: each column of K contains the influence of each observation in all variables of the state vector; each line of K contains the influence of all observations in each variable of the state vector. If p observations are available and the state vector has dimension n then K is a matrix with dimensions nxp.

The order of these two steps in algorithm is not relevant as long as the correction for a given time instant is performed based on a state estimate at that instant provided by a previous forecast step or from a user defined first guess. The process diagram of the Kalman filter with necessary inputs and provided outputs is presented in Figure I-2.

Figure I-2: Process diagram of the Kalman filter.



The algorithm provides the unique optimal state vector estimation under the validity of a set of underlying assumptions, which are summarized in Table I-6.

Table I-6: Assumptions of the Kalman filter algorithm (Kalman, 1960).

Errors ν and ω	*M*	*H*	Covariance matrixes P, R and Q
Random Null averaged Gaussian distributed Independent from each other Independent within them Known covariance	Linear	Linear	Positive definite

It should be noted the importance of the assumptions of Gaussian distribution and null average of errors and the linearity of the model: these assumptions ensure that, in one hand, the error processes ν and ω are completely described by their covariance matrixes (a Gaussian distribution is completely described by the average, the first moment, and the variance, the second moment), in another hand, the use of a linear model ensures that the Gaussian distribution is maintained in the state estimation error and so it is sufficient to represent this error through matrix P.

In real applications the assumptions of the Kalman filter are normally violated by non linear model dynamics and/or observation operator, unknown or partly known initial conditions for state vector and P, unknown or partly known noise statistics assigned by Q and R (Cañizares, 1999). The Gaussian distribution of the errors may also not be true: in this respect Brown and Hwang (1997) state that if it is not known *a priori* if this is true then it can always be devised a corresponding Gaussian process with the same covariance to which the Kalman filter method can be applied and conclusions can be drawn; however the extension of these conclusions to the original process are always to be regarded as risky.

As is seen in Table I-5, M, H, Q and R can change with each time step. This capacity to accommodate variable parameters is one of the advantages of the Kalman filter method according to Brown and Hwang (1997).

A weak version of the Kalman filter is the Optimal Interpolation, where the state forecast error covariance remains fixed in time. One example of the application of this methodology for coastal areas can be found in Besiktepe *et al.* (2003).

I.3.2.3. Filters for non linear system dynamics

As seen in section I.3.2.2. the Kalman filter method is not applicable for providing the state estimate if *H* and/or *M* are not linear operators. Essentially two approaches been developed to apply the Kalman filter methodology to non linear operators. These are the Extended Kalman Filter (EKF), established by Jazwinsky (1970) still in the field of control theory, and the Ensemble Kalman Filter (EnKF), developed by Evensen (1994) in the field of oceanographic applications. These filters no longer provide the optimal state estimate. Their characteristics in terms of forecast and correction steps are provided in Table I-7.

The **EKF** is based on the linearization of model and the observation operator. As visible in Table I-7 its algorithm is very similar to the one of Kalman filter, only replacing in the equations for the forecast of P and the analysis the full non linear model and observation operators by linear equivalents. Linearization is accomplished by a Taylor series expansion of the model/observation operator, around the current estimate of the state vector (Cañizares, 1999).

In the standard EKF algorithm the terms of the Taylor series expansion representing moments of order higher than first are neglected in linearization even if it is recognized that higher moments may be critically important for strongly non linear systems (Evensen, 1992; Cañizares, 1999; Bertino *et al.*, 2002). The use of higher order approximations is rarely done in practice (second-order and forth-order filters are referred by Verlaan, 1998) due to the complexity of the equations and the computational cost associated with the storage and evolution of higher order statistics (Verlaan, 1998, Bertino *et al.*, 2002). The neglecting of high order statistical moments by EKF causes that matrix P is not correctly estimated and errors originated propagate in time in P and the analysis (affected by the correction step), eventually causing divergence in strongly non linear system. This behaviour is found in practical applications by Evensen (1992) and in several studies reviewed by Triantafyllou *et al.* (2005).

The stability of the filter assumes an important role in the absence of production of the optimal state estimation because it ensures that the state estimate error although not minimum remains bounded: the root of the EKF instabilities is found in the extreme dependence of the stability of the filter on the values of K, M and H (and not so much on the values of Q and R), as found in the stability analysis of Pham *et al.* (1998a) which implies that the linearization of the *M* and *H* operators is potentially a sensible issue. Evensen (1992) presents two approaches to handle instability in EKF: use only the linear part of the dynamic (model) matrix, which may work well if the state estimate is close to the true state, and increase the value in Q to account for some of the error growth which is being eliminated; control the instability with measurements as the instability will increase for long time integrations of P, thus if enough measurements are available errors will not be allowed to increase too much.

Even if instabilities can be controlled, EKF still provides an important drawback for practical use in oceanographic applications: the manipulation of the P matrix with dimensions nxn is impossible in most practical cases, due to the great dimension of the ocean states being considered (e.g. Triantafyllou *et al.*, 2005).

Table I-7: EKF and EnKF algorithms equations (Jazwinsky, 1970; Evensen, 2003).

Scheme	Forecast step	Correction step
EKF	$x^f(t_i) = M_{i-1}\left[x^a(t_{i-1})\right]$ (I.8) $P^f(t_i) = M_{i-1}P^a(t_{i-1})M^T_{i-1} + Q(t_{i-1})$ (I.9)	$K_i = P^f(t_i)H_i^T\left(H_iP^f(t_i)H^T_i + R_i\right)^{-1}$ (I.10) $x^a(t_i) = x^f(t_i) + K_i\left[y_i^o - H_i\left[x^f(t_i)\right]\right]$ (I.11) $P^a(t_i) = (I - K_iH_i)P^f(t_i)$ (I.12)
EnKF	$x^f{}_j(t_i) = M_{i-1}\left[x^a{}_j(t_{i-1})\right] + \eta_j(t_{i-1})$, j=1…N (I.13) $\overline{x}^f(t_i) = \frac{1}{N}\sum_{j=1}^{N} x_i^f(t_i)$ (I.14) $P^f(t_i) = S^f(t_i)\left(S^f(t_i)\right)^T$ (I.15)[a] $s_i^f(t_i) = \frac{1}{\sqrt{N-1}}\left(x_i^f(t_i) - \overline{x}^f(t_i)\right)$ (I.16)[a]	$K_i = P^f(t_i)H_i^T\left(H_iP^f(t_i)H^T_i + R_i\right)^{-1}$ (I.17) $x_j^a(t_i) = x_j^f(t_i) + K_i\left[y^o_{i\,j} - H_i\left[x_j^f(t_i)\right]\right]$ (I.18) $\overline{x}^a(t_i) = \frac{1}{N}\sum_{j=1}^{N} x_i^a(t_i)$ (I.19) $P^a(t_i) = S^a(t_i)\left(S^a(t_i)\right)^T$ (I.20)[a] $s_i^a(t_i) = \frac{1}{\sqrt{N-1}}\left(x_i^a(t_i) - \overline{x}^a(t_i)\right)$ (I.21)[a]

Notes: [a] detailed formulation of Cañizares (1999); M is the gradient of *M* assessed at $x^f(t_{i-1})$ and H is the gradient of *H* assessed at $x^f(t_{i-1})$; N is the number of ensemble members.

The **EnKF** scheme is developed to avoid the linearization error in the forecasting of P. This has a more vast applicability, assuming the Kolmogorov equation for the time evolution of Gaussian errors statistics (Evensen, 2003), and the EKF can be derived as a special case of it (Bertino *et al.*, 2002). It is based on the use of an ensemble prediction (or Monte Carlo) method for the evolution of the error statistics: the state estimate error is represented by an ensemble of N individual possible states which evolve with the full non linear dynamics of the model. Table I-7 evidences that the forecast and correction steps act on each state individually: this causes that no higher order truncation occurs in the forecast of the P implicit in this state ensemble, accounting for non linear dynamics in the evolution of covariance (Evensen, 2003). However, as a Kalman filter method the linearity is still present in the analysis equation (I.18) which still removes the optimality of the state estimation as highlighted by Bertino *et al.* (2002) and Natvik and Evensen (2003).

A very important parameter influencing the performance of this method is the size of the ensemble N: forecast and correction steps are repeated as many times as there are members in the ensemble. A large ensemble number, usually from 100 to 200, is needed to compensate for information redundancy (Bertino *et al.*, 2002), causing EnKF applications to have a high computational cost, representing more than 100 times the computational cost of making the model simulation alone.

Still, comparatively with EKF, the EnKF represents a drastic reduction of the computational cost and memory storage requirements and also allows the model errors to be directly related to elements in the numerical model that are expected to be poorly described. Cañizares (1999) evidences that in traditional Kalman filter if these elements are not present in the state vector the filter performance is not the most correct.

Also the published research suggests that the EKF is best suited for linear or weakly non linear models and EnKF for strong non linear models (Evensen, 1992; Cañizares, 1999; Verlaan and Heemink, 2001) although it is not clear when can be assumed a weakly or a strong non linear dynamics.

I.3.2.4. Simplified EKF schemes

These schemes are based on EKF algorithm and explore the time evolution of the error statistics in a way more efficient in the use of computational resources. An important branch of these schemes is provided by a set of algorithms in which the state estimate error covariance matrix P is approximated by a lower rank matrix obtained usually from eigenvalue and eigenvector decomposition using EOF analysis, which introduces the suboptimality of theses schemes relative to EKF. The most widely used members of this branch are the Reduced Rank Square Root Kalman Filter (RRSQRT), established by Verlaan and Heemink (1995), the Singular Evolutive Extended Kalman Filter (SEEK), developed by Pham *et al.* (1998a), and the Singular Evolutive Interpolated Kalman Filter (SEIK), developed by Pham *et al.* (1998b). The algorithms of these schemes are presented in Table I-8.

In **RRSQRT** the initialization of the filter is made considering a lower rank representation of P_0, represented by the square root matrix S_0, obtained from the r leading eigenvalues and eigenvectors. The algorithm then progresses in forecast and correction steps with this square root matrix. The use of a square root matrix is explained by Cañizares (1999): while the use of a lower rank covariance matrix can originate that the required property of positive-semidefinitness assumed in Kalman filter general algorithm is not preserved, causing instabilities in the filter due to inadequate calculation of P, with a square root matrix it is ensured that the matrix is positive-semidefinite always because no negative eigenvalues can appear (except when the calculation of the square root matrix is performed with inadequate precision, as stated by Verlaan, 1998). The convenient rank of the reduced P has to be determined from experiments (Verlaan, 1998)[37]. The yield in computational cost of this scheme relative to EKF is not as good as could be thought considering only the reduced rank because the eigendecomposition of a large covariance matrix, performed in the forecast step presented in Table I-8 is still a very costfull operation. Some important studies with this method can be found in the works of Verlaan (1998) and Verlaan and Heemink (2001).

Table I-8: RRSQRT, SEEK and SEIK algorithms equations (Bertino *et al.*, 2002; Cañizares, 1999; Pham *et al.*, 1998a; Pham *et al.*, 1998b; Triantafyllou *et al.*, 2003).

Scheme	Initialization	Forecast step	Correction step
RRSQRT	$P^a(t_0)=L_0U_0L_0^T$ [a] (I.22) $S_0=L_0U_0^{1/2}$ (I.23)	$x^f(t_i)=M_{i-1}\left[x^a(t_{i-1})\right]$ (I.24) $S^{*f}_i=\left[M_{i-1}S^a_{i-1}\middle\vert Q_{i-1}^{1/2}\right]$ (I.25)[b] $S^{*fT}_iS^{*f}_i=L_iU_iL_i$ (I.26)[c] $S^f_i=S^{*f}_iL_i$ (I.27)[d]	$A_i=S_i^{fT}H_i^T$ (I.28) $B_i=\dfrac{1}{A_i^TA_i+R_i}$ (I.29) $K_i=S_i^fA_iB_i$ (I.30) $x^a(t_i)=x^f(t_i)+K_i\left[y_i^o-H_i\left[x^f(t_i)\right]\right]$ (I.31) $S_i^a=S_i^f-K_iA_i^T\left(1+\sqrt{B_iR_i}\right)^{-1}$ (I.32)
SEEK	$P^a(t_0)=L_0U_0L_0^T$ [a] (I.33)	$x^f(t_i)=M_{i-1}\left[x^a(t_{i-1})\right]$ (I.34) $L_i=M_{i-1}L_{i-1}$ (I.35)	$U_i^{-1}=\rho U_{i-1}^{-1}+L_i^TH_i^TR_i^{-1}H_iL_i$ (I.36) $K_i=L_iU_iL_i^TH_i^TR_i^{-1}$ (I.37) $x^a(t_i)=x^f(t_i)+K_i\left[y_i^o-H_i\left[x^f(t_i)\right]\right]$ (I.38)
SEIK	$P^a(t_0)=L_0U_0L_0^T$ [a] (I.39)	$x^a(t_{i-1})=\overline{x}^a(t_{i-1})+\sqrt{r+1}L_{i-1}\left(\Omega_{i-1}C_{i-1}^{-1}\right)$ (I.40) $x^f_j(t_i)=M_{i-1}\left[x^a_j(t_{i-1})\right]$, j=1…r+1 (I.41) $\overline{x}^f(t_i)=\dfrac{1}{r+1}\sum_{j=1}^{r+1}x_i^f(t_i)$ (I.42)	$L_i=\left[x_1^f(t_{i-1})\ldots x_{r+1}^f(t_{i-1})\right]T$ (I.43) $U_i^{-1}=\rho(r+1)T^TT+(HL)_i^TR_i^{-1}(HL)_i$ (I.44) $(HL)_i=\left[H_i\left[x_1^f(t_{i-1})\right]\ldots H_i\left[x_{r+1}^f(t_{i-1})\right]\right]T$ (I.45) $K_i=L_iU_i(HL)_i^TR_i^{-1}$ (I.46) $\overline{x}^a(t_i)=\overline{x}^f(t_i)+K_i\left[y_i^o-H_i\left[\overline{x}^f(t_i)\right]\right]$ (I.47)

Notes: [a] U_0 is a diagonal matrix containing the r larger eigenvalues with decreasing order, L_0 contains the corresponding r eigenvectors; [b] the square brackets indicate that S^{*f}_i is a matrix composed of two blocks; [c] eigenvalue and eigenvector decomposition; [d] calculation of the r columns approximation of S^{*f}_i ; Ω_{i-1} is any (r+1) x r matrix with orthonormal columns and zero column sums drawn randomly; C_{i-1} is the Cholesky decomposition of U_{i-1}; T is a (r+1)r matrix with zero column sums (Hoteit *et al.*, 2002).

[37] As stressed by Bertino *et al.* (2002), the reduction of the rank of the state estimate covariance matrix can be validated by comparing the estimated standard deviation of the state estimate error, provided by the data assimilation method, with the true one verified at measurement sites: if the chosen rank is not adequate then the standard deviation would be underestimated by the data assimilation method.

The **SEEK** scheme algorithm is based on making the error correction only along directions or modes of error propagation parallel to a linear subspace of dimension r, smaller than state vector space dimension. These directions, composing the correction basis L, are established between the ones for which the state estimate error is mostly amplified by the model. The consideration of a reduced space for error evolution has justification because the natural systems dynamics is generally dominated by a relatively small set of attractors, which cause that error in a model system will increase more in some directions of the state space than in others (see explanation in Brasseur and Verron, 2006).

Since a low rank P, usually obtained through eigenvectors and eigenvalues decomposition, is used in SEEK this method is similar to the RRSQRT. The main difference between these two methods is the way the working space is reduced (Pham *et al.*, 1998a): while RRSQRT requires, with a non linear system, the decomposition of the error space in the main modes before every filter calculation, in SEEK no decomposition is usually necessary (besides possibly the one used to generate the initial P) since a low rank P is always produced by the algorithm and the method is specially intended to use within a non linear time variant system, or at least with a non linearity and time evolution not too much big, as Pham *et al.* (1998a) stress. The forecast of the correction basis L is achieved by ensemble forecast, i.e. one column of L at a time, similarly to the EnKF scheme.

A formulation considering explicitly the Q covariance matrix of model error is provided by Pham *et al.* (1998a). In this case, if Q is not projected into the space of the filter correction modes then SEEK algorithm would also need the decomposition in the dominant modes after each correction step (Hoteit, 2001), as in RRSQRT. Due to the complexity of considering Q, the model error is generally accounted in SEEK applications (as in Table I-8) through the use of the amplification factor ρ in equation (I.36), called the forgetting factor. This factor is valued between 0 and 1: if ρ is more close to 0 than to 1 analysis state depends more on the observations and less on the model forecast. This implies a weighting scheme where older states have generally smaller weight[38] and, with no extra cost, provides a two-fold advantage (Hoteit *et al.*, 2002): it increases the filter capability to follow changes in the system while attenuates the undesirable propagation in time of errors due to filter deficiencies (Triantafyllou et al., 2005), thus enhancing the stability of the filter. Hoteit (2001) advocates the use of this factor to deal with the practical difficulties of using Kalman filter methods, namely badly known model errors and non linearity of models.

As pointed by Hoteit *et al.* (2004), the computational cost of SEEK is drastically smaller than in EKF due to the fact that the forecast and correction of the estimate error covariance matrix are not required by the algorithm. The most expensive calculation in the SEEK scheme is the time integration of the

[38] If $\rho=1$ then all data have the same weight, if $\rho<1$ then recent data is exponentially more weighted than old data.

different error modes. This causes the SEEK to, while implying a significant cost reduction relative to the EKF, still be costly for real operational assimilation (Hoteit *et al.*, 2001).

In practice the SEEK algorithm, while reducing the computational cost, may also improve the EKF stability (Pham *et al.*, 1998a). Hoteit (2001) refers that if non linearity is sufficiently weak one can hope that the evolving correction basis L still captures most of the errors not adequately attenuated by model dynamics. As Hoteit *et al.* (2001) point, a built-in weakness of SEEK is that the correction basis may not contain enough directions to capture the range variability of the model dynamics. This problem is not adequately solved by the increase of the number of correction base modes because of the EOF analysis considered initially to generate these modes: while costly in computational terms this increase can only marginally increase the representativeness of the basis, as the last EOFs usually cannot be accurately evaluated, due to statistical fluctuations and the historical simulation used for EOF analysis being not as long as needed. An EOF analysis based on a model simulation affected by errors would consequently be affected by errors too and it is though that this error influences mostly the last EOFs, since they are the most numerically unstable. For an extensive review of applications and developments of SEEK scheme see Brasseur and Verron (2006).

The **SEIK** scheme is a SEEK variant in which linear interpolation is used instead of linearization in model and observation operators for the propagation of the state estimate error covariance matrix (Pham *et al.*, 1998b). As stressed by Pham (2001) this method can also be regarded as a variant of EnKF using a minimum number of ensemble members. The r+1 interpolating state vectors, which should be of the lowest possible number, are usually created by second order exact sampling, as presented in equation (I.40) (Hoteit *et al.*, 2002). As in RRSQRT and SEEK the rank r is obtaining by performing trials. The advantages of this scheme (Pham *et al.*, 1998b; Pham, 2001) are the avoidance of the *M* and *H* gradients calculation (through equations (I.43) and (I.45) in Table I-8), which can be very complex numerically, and through the use of interpolation seems to be taken more into account the model non linearity: it is verified that when large deviations from linear dynamics are present the interpolation provides a lower error than linearization. As for SEEK, it is possible to use a formulation of the equation of calculation of U (equation (I.44) in Table I-8) with the explicit consideration of Q (Pham *et al.*, 1998b).

Pham (2001) outlines two important problems with SEIK algorithm. The first is the fact that an ensemble member or the analysis state can result physically unrealizable, which while common to all Kalman filter methods, is more severe in SEIK due to the random nature of the resampling or creation of interpolating states. A solution is to replace the anomalous state with a nearby realizable state. The second problem arises when considering model error and projecting it on the correction basis space. The issue is that while it is admissible that the model error lies in a reduced space it is not obvious that this space should be the state estimate error space.

All these three simplified schemes should be affected by underestimation of the state estimate error, as described by Pham (2001), which decreases the filter performance. This underestimation happens because the state estimate error covariance matrix reflects only, through the way it is computed, the error from model and observations. But in reality other errors are present in all these three schemes: the linearization error in RRSQRT and SEEK, the interpolation errors in SEIK. This author refers also that problem should also affect EnKF, due to the sampling fluctuation in the sampling of the ensemble members. When the model and observation errors are small the neglected errors are significant and so the data assimilation scheme can severely underestimate the actual error of the state estimate. This can dealt with by considering a time variable R, adapted during filter operation to reflect the actual state estimate errors (Pham, 2001; Hoteit *et al.*, 2002; Triantafyllou *et al.*, 2005). Also, the RRSQRT and SEEK, which perform linearization using a first order approximation, are referred to be especially prone to instabilities (Bertino *et al.*, 2002).

Since the creation of SEEK and SEIK filters research has originated several variants mainly to reduce the costs, which are called degraded forms, and to deal with specific flaws of these methods. The most important of these variants are succinctly presented in Table I-9.

From these variants it should be referred the importance of SFEK. The approach of this variant can be justified by a slowly evolving ocean state, in which case the model operator is almost equal to unity (Hoteit *et al.*, 2002). While considered to be a weak version of SEEK (Verron *et al.*, 1999) is widely recognized in research as a good way to assess the usefulness of different correction direction basis and to tune parameters in SEEK and SEIK and their degraded forms due to the low computational costs it involves (e.g. Hoteit *et al.*, 2001, Hoteit *et al.*, 2002, Triantafyllou *et al.*, 2005), and it can even provide good results if the correction basis represents sufficiently well the model variability (Hoteit *et al.*, 2002).

The use of SFEK implies an important specificity, which does not seem to apply in the SEEK or SEIK schemes: Triantafyllou *et al.* (2005) stress that it has been found in literature that "increasing the number of modes in the error covariance matrix approximation can sometimes deteriorate the performance of the filter". This is not observed in SEEK and SEIK although as stated before a stagnation of the filter performance is to be expected at some point. The reason for this behaviour is the transient characteristic of the optimality of a correction basis: as the authors explain the fixed correction directions are only optimal in a time average sense and over the period for which they have been computed; thus the lower eigenvalue correction directions, which generally represent fine-scale information, can be therefore irrelevant during the assimilation period.

Table I-9: Variants of SEEK and SEIK schemes.

Scheme	Description	Advantage
SFEK Singular Fixed Extended Kalman Filter (Brasseur *et al.*, 1999)	Time independent state estimate error subspace (L).	Small cost
Apprentice filter (Brasseur *et al.*, 1999)	Progressively enriches the correction basis with information from analysis residuals (the contrast between observations and analysis), relying in the principle of adaptivity.	Corrects of an inadequate specification of the initial correction basis
SEPLEK Semi-Evolutive Partially Local Extended Kalman Filter (Hoteit *et al.*, 2001)	Use a spatially specific correction basis, which remain fixed, and a global correction basis which evolves with model dynamics.	Use more correction basis directions at specific areas of state vector with smaller cost
SSEEK Singular Semi-Evolutive Extended Kalman Filter (Hoteit *et al.*, 2000) SSEIK Singular Semi-Evolutive Interpolated Kalman Filter (Hoteit *et al.*, 2002)	Only some directions of correction basis, the ones which contribute the most to error representation, evolve with model dynamics.	Small cost
SIEIK Singular Intermitent Evolutive Interpolated Kalman filter (Hoteit *et al.*, 2002)	Draw interpolating states only in unstable periods of model simulations, otherwise perform correction and forecast steps as SFEK.	Small cost

I.3.2.5. Typology of application of schemes

Following the characteristics of the most important sequential schemes, presented in the previous two sections, and the scheme comparison studies, which are presented in published research, it is possible to summarize in Table I-10 the field of applicability of each scheme.

The most representative applications of each scheme in coastal areas, which are present in Table I-10, indicate that the important research on Kalman filter derived schemes is made essentially in European coasts. In the East North Atlantic coastal areas the dominant schemes are the EnKF and the RRSQRT, although in this later case no intensive application is apparent in the recent years (since 2002). These schemes are mostly applied for improving storm surge forecasts (Verlaan, 1998; Cañizares, 1999; Cañizares *et al.*, 2001; Mourre *et al.*, 2006). The applications of SEEK and SEIK schemes are mainly confined to the Mediterranean Sea coastal areas and are mainly used for improving biogeochemical forecasts (Triantafyllou *et al.*, 2003; Hoteit *et al.*, 2004; Triantafyllou *et al.*, 2005; Raick *et al.*, 2007). It should be noted that, with the exception of SFEK, the variants of SEEK and SEIK do not appear to be applied in realistic coastal modelling.

Table I-10: Typology of application of sequential schemes derived from the Kalman filter method.

Scheme	System dynamics		Operational setting (computational cost)[b]	Representative coastal applications
	Nearly linear or weakly non linear	Highly non linear		
EnKF	x	x (Verlaan, 1998; Madsen and Cañizares, 1999)	x/- (N~100 model simulations) (Brusdal *et al.*, 2003)	2D North and Baltic seas (Cañizares *et al.*, 2001) 2D North Sea shelf (Mourre *et al.*, 2006) 1D Ría de Vigo (Torres *et al.*, 2006)
RRSQRT	x	- (small intertidal areas can be supported, Verlaan, 1998)	x/- (r+1~10 model simulations + eigendecomposition cost)	2D North Sea (Verlaan, 1998) 2D Danish coast (Cañizares., 1999) 3D Baltic Sea coastal lagoon (Bertino *et al.*, 2002)
SEEK	x	-	x (r+1~10 model simulations) (Brusdal *et al.*, 2003)	3D Pagasitikos Gulf (Triantafyllou *et al.*, 2005 [a]; Hoteit *et al.*, 2005) 1D Ligurian Sea (Raick *et al.*, 2007) 3D Ligurian Sea (Barth *et al.*, 2007)[c]
SEIK	x	x (Nerger *et al.*, 2005)	x (r+1~10 model simulations)	3D Cretan Sea (Triantafyllou *et al.*, 2003) 1D Cretan Sea (Hoteit *et al.*, 2004) 3D Pagasitikos Gulf (Triantafyllou *et al.*, 2005; Hoteit *et al.*, 2005)

Notes: [a] application of SEEK and SFEK versions; [b] a detailed cost analysis of the several steps of EnKF, SEEK and SEIK algorithms is presented by Nerger *et al.* (2005); [c] SEEK is applied under the form of SFEK.

I.3.2.6. Implementation issues

The implementation of the Kalman filter schemes presented in the previous sections involves the consideration of several practical issues. From these perhaps the most relevant for coastal applications concern the account of non null average state error, the estimation of the initial state error covariance matrix, the creation of the initial state estimate for model simulations through Objective Analysis technique and the linearization of the model. These are discussed in the following paragraphs.

As seen in section I.3.2.2. one of the basic assumptions for the applicability of Kalman filter methods is that the errors involved in the estimation are null averaged, so that the state estimate is unbiased relative to the true state. However, in realistic applications this may prove not to be true and the state estimate bias can be important. The approach used in this case it to introduce the non null averaged part of the error in the state vector and the consideration of a time evolution dynamics for this error part (Brown and Hwang, 1997). This approach is used by Cañizares (1999) in data assimilation in coastal models.

The estimation of the initial covariance matrix for the state estimate is potentially a sensible issue. The standard Kalman filter method is robust to a bad specification of P due to control theory of linear systems. In the presence of a non linear system this may not happen or the convergence to the adequate P may be too slow to be of practical interest: this may occur in SEEK applications as discussed by Pham *et al.* (1998a) and Hoteit *et al.* (2002). The traditional way to obtain the initial P for SEEK and SEIK applications is to calculate the covariance from historical dynamic model states (e.g. Hoteit, 2001). The state vectors are usually sampled using decimation (Pham *et al.*, 1998a): only one state in a set of successive states is considered, to reduce calculation costs as successive states are quite similar and to reduce the effect of autocorrelation degradation of covariance estimation. Further, in the subsequent process of EOF analysis if state variables are not all of the same nature, i.e. typically have different magnitudes or units, it is useful to prior the analysis normalize each variable value using a convenient factor (e.g. Hoteit *et al.*, 2004). The rank r of the reduced covariance space is estimated by experience and considering the fraction of total variance represented by the sum of leading EOF eigenvalues (Pham *et al.*, 1998a; Triantafyllou *et al.*, 2005).

Brasseur *et al.* (1999) stress the consequences of the initial estimation of P in the SEEK scheme: if the error statistics specified by L_0 are consistent with actual model-observations discrepancies then assimilation is successful, with the RMSE being strongly reduced in the first assimilations; if error statistics are not consistent with the actual discrepancy between first guess and observations then assimilation fails even when the EOF is updated accordingly to model dynamics. The important characteristic of the filter in this issue is that it will discard the innovation components which are orthogonal to the error subspace constituted by the correction directions basis.

In some applications it may be useful to perform a correction step at the initial time of the model simulation. This can be made with the Objective Analysis technique (e.g. Hoteit, 2001; Hoteit *et al.*, 2004): the state estimate is corrected with the observations at the initial time using the standard equation for the calculation of analysis (I.6) using $U_0 = \left(L_0{}^T H_0^T R_0^{-1} H_0 L_0\right)^{-1}$.

The linearization of the model and observation operators, which is theoretically done by calculating the Jacobian matrix is potentially complex. To avoid this, a finite difference approximation of the tangent linear operator is generally used (Verlaan and Heemink, 1997; Verlaan, 1998; Cañizares, 1999; Hoteit, 2001; Bertino *et al.*, 2002) where the columns of S, in RRSQRT scheme, or L, in SEEK scheme, being interpreted as a state vector perturbation. The results of the evolution by the model of the state vector and of the perturbed state vector are then considered in a first-order finite difference considering a linearization parameter ε. This should be an infinitesimal real parameter. However, a small value of this parameter is reported as the cause of divergent error growth in RRSQRT by Bertino *et al.* (2002). This operation involves making extra simulations of the model every time step, justifying the computational cost of SEEK provided in Table I-10. This cost explains the appeal of the SFEK scheme, as demonstrated in the works Brasseur *et al.* (1999), Hoteit *et al.* (2002) and Hoteit *et al.* (2004).

I.3.3. Variational assimilation

Variational methods consider the observations acquired during a long time interval, adjusting a set of control variables, representing undetermined model parameters, through the whole period (Carmillet *et al.*, 2001; Hoteit, 2008). In this way, estimation in all times is influenced by observations in all times (Cañizares, 1999), an important difference relative to sequential methods. The processing of these methods is performed through the minimization of a cost function considering all available state estimates, such as observations and a numerical model, with the model equations representing usually a strong constraint, i.e., these equations are imposed to be completely satisfied in the optimization process (Devenon *et al.*, 2001; Zhang *et al.*, 2002; Hoteit, 2008).

One of the methods in this category most used in coastal forecast systems is the adjoint method (Devenon *et al.*, 2001; Zhang *et al.*, 2002). In oceanography, it has been successfully used to assimilate satellite observations of sea-surface heights as referred by Bertino *et al.* (2002). Recent applications of this method in coastal systems include the estimation of open boundary conditions and the turbulent viscosity coefficients in a 3D model by Devenon *et al.* (2001) and the assimilation of tide residuals from tide gauge measurements for the estimation of wind shear stress coefficients in a 2D version of POM model for the North American East coast through the work of Zhang *et al.* (2002). The main components of the general adjoint method as described by these authors are summarized in Table I-11.

Table I-11: Main components of an adjoint method data assimilation according with Zhang *et al.* (2002).

Forward model	Observations	Control variables	Cost function	Adjoint model
Numerical model	Chosen according with physical connection with the control variables	Set of numerical model parameters, initial conditions, open boundary conditions, bathymetry	Formulation of differences between numerical model results and observations, to be minimized	Derived from the forward model and its discretization for calculating the gradient of the cost function with respect to control variables

The forward model is used to provide the initial conditions for the adjoint equations, which represent an inverse model propagating the state back in time while assimilating the observations. At the end of this inverse simulation is calculated the gradient of the cost function with respect to the control variables. The general steps of the adjoint method algorithm are summarized in Table I-12.

Table I-12: General steps of adjoint method algorithm according with Zhang *et al.* (2002).

1. Make the simulation of the forward model for the time period comprising the data assimilation window with the initial values of control variables.
2. Calculate the model results – observations differences and the cost function.
3. Make the simulation of the adjoint model backwards in time forced by the model observations data differences to calculate the adjoint variables (replicas of forward model variables) and calculate the gradient of the cost function with respect to control variables.
4. Employ an algorithm to calculate the optimal control variables estimates.
5. Verify if the convergence criterion is satisfied: if yes the iteration is stopped; if not steps 1 to 5 are repeated with the new control variables estimates.
6. Make the simulation of the forward model with the optimal control variable estimates.

The study of Zhang *et al.* (2002) can be taken as an example of the calculations and assumptions considered in the application of this method. In this case the control variables are the wind shear stress coefficients specified for several areas of the model domain. This specification of the control variables, together with the assimilation of measurements tide residuals, implies that all the errors relative to observations are treated as being due to an incorrect specification of wind shear stress coefficients. This is not exactly true in every areas of the domain, where other error sources such as the representation of large scale features as the Gulf Stream are considered to be important. Following the specification of the control variables and in order to simplify the application computational cost (due to operational constraints) and the derivation of the adjoint equations the numerical model is linearized. The estimation of the control variables involves, in the reported work, the making of the cycle composed of steps 1 to 5 presented in Table I-12 several times and the use of penalty terms in the cost function so that estimated control variables values remain within realistic values.

The model linearization required in adjoint method is reported to limit the time interval in which the data assimilation is performed in strongly non linear models, which restricts the number of observations taken into account (Hoteit, 2008). A recent study by Hoteit (2008) avoids the adjoint method, and its deficiency in consideration of non linear dynamics, in a variational data assimilation application through the use of a Monte Carlo method (Simulated Annealing) optimization in a reduced space representation of the control variables space, at a comparable computational cost.

The work of Zhang *et al.* (2002) exemplifies reported important drawbacks of variational methods which hamper their common use ocean forecast systems, as presented in Table I-13.

Table I-13: Drawbacks of variational methods in ocean forecast system applications (Cañizares, 1999, Carmillet *et al.*, 2001).

▪ Difficult to apply in on-line forecasting procedures
▪ Complex software development, mostly connected with adjoint methods
▪ High computational cost (more than ten times the amount of calculus assigned for numerical model integration)

However, variational methods have also important advantages relative to sequential methods which make them particularly interesting in non operational data assimilation applications for the ocean, which are summarized in Table I-14.

Table I-14: Advantages of variational methods relative to sequential methods in ocean models data assimilation applications.

▪ Better adaptation to disperse measurements
▪ Model adjusted trajectory not verifying abrupt changes
▪ No requirement for the calculation of the state estimate error covariance matrix (Malanotte-Rizzoli and Tziperman, 1996)

I.4. Knowledge Deficiencies

The research conducted in this chapter reveals that operational forecast systems constitute an important resource for the adequate management of the ecological, economical and social relevant estuarine areas. Currently, several advanced modelling and data assimilation techniques are available for use in forecast improvement in operational systems.

However, some knowledge deficiencies are present in the field of hydrodynamics forecast in coastal areas. While modelling techniques are well established, both in large/meso scale and in local scale, and its application with full yield in coastal operational systems is only a matter of time, the data assimilation techniques application to coastal areas constitute still a recent and experimental field, with in a big gap between the large number of operational applications in large/meso scale and the few in local scale. Also, the coastal applications have been concentrated in mostly in the Mediterranean Sea and in the North and Baltic Sea, causing scarce data assimilation applications in the North Atlantic small width east continental shelf, where tide and bathymetric changes pose an interesting challenge for data assimilation techniques.