VII. Implementation of First Data Assimilation Configuration for 2D MOHID Tagus Estuary

VII.1. Introduction

To reach the test of Hypothesis 4 the research is conducted by studying possible SFEK and SEEK filters configurations in a twin models experiment framework.

Following the twin experiment description, are presented the setting and results of the required covariance EOF analysis. Next the assimilation trials performed are described and the results presented and discussed, for SFEK and SEEK filters.

The knowledge gained in this twin models experiment is then used for discussing the data assimilation of real tide gauge measurements, establishing the support found for Hypothesis 4 validity.

VII.2. Twin Experiment

VII.2.1. Experiment description

The twin models experiment is designed so that the perturbed model state is unbiased relative to the true model state and that the state error is realistic, to gather information useful for real measurements data assimilation applications in MOHID Tagus Estuary.

Within this framework, the True and Perturbed models are considered to be the following:

- True Model: the best configuration of the MOHID Tagus Estuary 2D nesting 2 domains application (domain 1 and 2) achieved in the research for this thesis which consists in the enlarged domain 1, TVD advection and spatially variable wind; the bathymetry used is the existent system bathymetry for domain 2 (Figure II-3 a)[122];

- Perturbed Model: the True Model configuration with an added Gaussian distributed white noise to the mean sea level at domain 1 boundary tidal points.

The Perturbed Model is derived reckoning that the two most important error sources for water level forecast in MOHID Tagus Estuary evidenced by the research presented in the previous chapters are the bathymetry/horizontal grid resolution and the specification of mean sea level time evolution. Generating a null average error of the first source is not done easily in a twin experiment framework for a SEEK filter, if one wants to follow the traditional assumption that the state variability can be taken as a proxy for the error variability is no longer valid as the same bathymetry is used to generate all historical states and a biased error should certainly be present. This could more naturally be done in an EnKF twin experiment as demonstrated by the work of Auclair *et al.* (2003) and Mourre *et al.* (2006): the ensemble states can be generated to represent a specific error covariance. It is possible to use the same approach to generate the initial covariance structure of the state error for a SEEK filter: a set of model simulations can be made with perturbed bathymetries and the results be used to generate an initial covariance of state error derived from an incorrect bathymetry. This covariance generation scheme is applied by Barth *et al*. (2007) to generate an initial covariance for a state error cause by errors in initial conditions and meteorological forcing. This is very time and computer resource

[122] As referred before this bathymetry is very similar to the improved one (Figure IV-7) outside the estuary while inside the estuary the estuarine channels are better described. In the context of the twin experiment the use of a lower quality bathymetry does not difficult the assessment of support for Hypothesis 4, since both twin models use the same bathymetry. The inverted barometer configuration is not considered at the open boundary of domain 1; this is in order to keep the modelled system not too complex and because the average mean sea level still remains to be tuned.

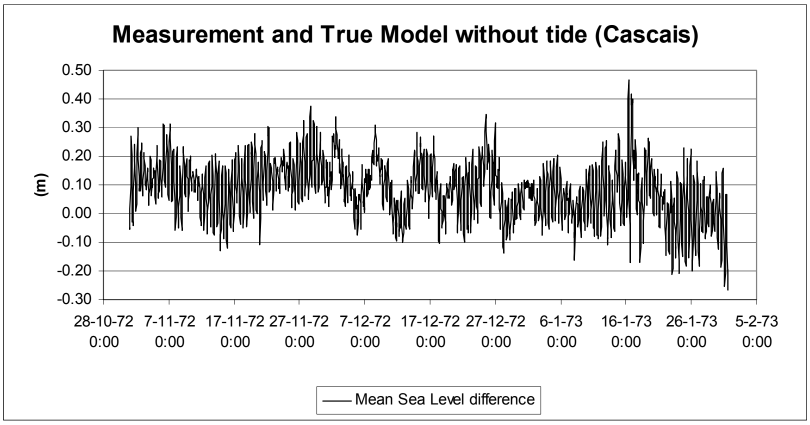
consuming, because as much as 100 or more simulations need to be made, and is because of resources limitation it is not used for this work. In section VII.3 it is discussed how to design such a twin experiment for the SEEK filter in the MOHID Tagus Estuary.

Having this twin models framework, the system state is specified by the water level and the zonal and meridional components of horizontal velocity fields in domain 2 grid computational points, meaning the cell centre for water level and the cell faces for the velocity components; this causes that the dimension of the multivariate state is 47963. For a realistic validation set the state is assessed primarily in the 15 tide gauge locations comprised in the domain 2 spatial extent.

For the characterization of the true state of the system in the twin experiment framework a simulation of the True Model is made for the period from 01/10/1972 to 31/01/1973. The descriptive statistics of the results for state properties at tide gauge locations are presented in Appendix I. This period is chosen because several tide gauges hourly measurements are available and this simulation offers the opportunity for a comparison of the True Model performance relative to the measurements; this comparison is relevant because this model is an improved version of the one used in section IV.2 to test Hypothesis 2, which uses an upwind advection scheme and lacks wind forcing (note that the bathymetry is also very similar). The results of the model-measurements contrast in terms of bias and centered RMSE, presented in Appendix VIII, are qualitatively similar to the ones presented in section IV.2.

The simulation starts from an initial state composed by a 2.08m water level and null velocity and a gradual connection of the imposed tide water level (domain 1) and wind velocity forcing in the first day (both domains). After the first month of simulation, reserved for the system to reach the full dynamics independent of initial state, the True Model detided results for the Cascais tide gauge location, the one more near the domain 2 boundary, are contrasted with the detided tide gauge measurements, to assess mean sea level variability differences. This contrast results in an observed mean sea level error with standard deviation 0.1m (Figure VII-1), which is also the value considered by Cañizares (1999) in a twin experiment designed for the RRSQRT Kalman filter in the North Sea. Consequently, a random normally distributed perturbation with standard deviation 0.1m and null average is added to the hourly mean sea level imposed at boundary of domain 1 of the perturbed, providing a null averaged system error inside the domain.

Figure VII-1: Difference between detided True Model domain 2 results and measurements for Cascais tide gauge.



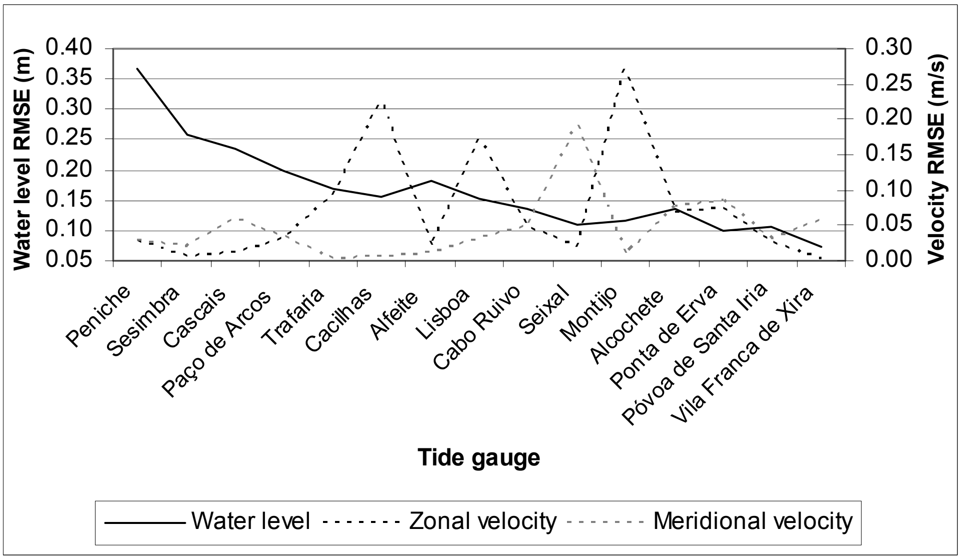
To obtain a reference Perturbed Model state without data assimilation, a simulation of this model is done for the same period and using the same initialization procedure as in True Model simulation. This is called in the context of assimilation trials as the twin experiment Free-run. The descriptive statistics of this simulation results for the tide gauges are presented in Appendix I: the observed variability is in all state properties at tide gauge locations very similar to the one of the true state[123]. An analysis to the Perturbed Model error presented in Appendix IX reveals the Perturbed Model state bias represents less than 10% (typically 5%) of the RMSE (the part of RMSE that the Kalman filter is able to correct) in most locations and state properties, which is concurrent with the hypothesis of a null averaged state error assumed in Kalman filter data assimilation. However, the error probability distribution has important departures from the normal distribution in some tide gauges and state properties, particularly in the velocity field in shallow channel or intertidal locations inside the estuary (Póvoa de Santa Iria, Ponta de Erva, Seixal, Vila Franca de Xira) and in the coastal bays (Cascais, Sesimbra): the SEEK filter is expected to be less performing at correcting the forecast error in these locations.

The RMSE (Figure VII-2) represents on average 0.17m for water level, 0.07m/s for zonal velocity and 0.05m/s for meridional velocity. The spatial distribution of error is different between the state properties: in water level it is generally larger the closer the tide gauge is from the open boundary, a consequence of tidal driven dynamics; in velocity the error distribution seems to be more connected to the specific bathymetric features, being larger in locations with coastline aligned with velocity direction (e.g. estuary exit channel affecting the error in Cacilhas and Lisboa). It is also interesting to note that the magnitude of the velocity error in velocity is directly proportional to the magnitude of the

[123] The variability differences are concentrated in the Estremadura Promontory and Northern continental shelf near boundary suggesting the accumulation of error transported there by tidal wave (which advances from South to North).

velocity variability (see standard deviation in Appendix I), suggesting that the locations with higher variability are places of error concentration; this relationship is not visible in the water level error.

Figure VII-2: Free-run RMSE for each state property. RMSE is calculated considering the hourly results of the True Model and Perturbed Model from 01/11/1972 to 31/01/1973. Tide gauges indicated from open ocean (Peniche) to Tagus River (Vila Franca de Xira).



VII.2.2. EOF analysis

VII.2.2.1. Sampling of states

For the analysis, states from the True Model simulation and the Perturbed Model simulation are sampled every 6 hours starting at 01/11/1972, comprising a set of 245 states distributed in a two months period, following the traditional approach for state covariance calculation. The 6 hours decimation interval is chosen because it corresponds to the Nyquist sampling frequency (Emery and Thomson, 2001) having into account that the dominant tide period in the Tagus Estuary area is 12 hours. The same sampling procedure is done for the Perturbed Model error to assess the true error variability.

To assess the effect of sampling in the EOF analysis results, a higher sampling frequency constituted by sampling every 1 hour is also used for the model results for the first 10 days period starting from 01/11/1972, comprising a total of 217 states. Here the number of sampled states is limited by the available computational memory resources but nevertheless the chosen number seems adequate for the estimation of the covariances in this 2D model, since 200 states have been used for estimating the covariance for SFEK in 3D (potentially with more complex dynamics) models (Barth *et al.*, 2007).

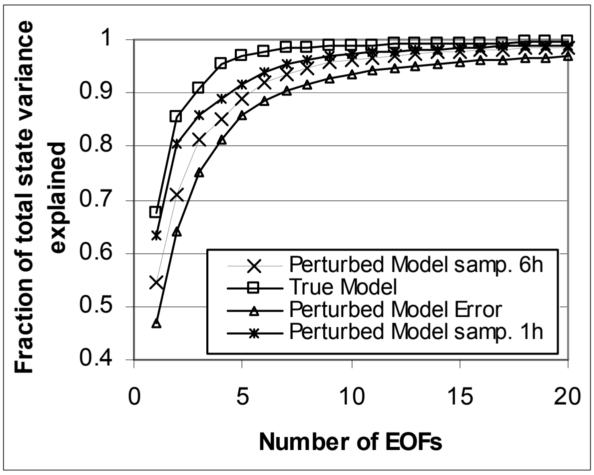
As the state vector is multivariate, the state variables are normalized before the EOF analysis as explained before. The validation of the EOF analysis is made by contrasting the standard deviation in each grid cell obtained from the sampled states and obtained from the states for same time instants reconstructed using the specific set of calculated EOFs, using the pre-processing tool specially developed functionality.

VII.2.2.2. Results

The EOF analysis of the True Model and Perturbed Model simulations domain 2 results sampled every 6 hours reveals that the four EOFs corresponding to the leading eigenvalues represent about 95% and 85% of the total state variability, respectively for the True Model and Perturbed Model (Figure VII-3). Comparatively, the EOF analysis of the Perturbed Model error results leads to a lower explained variance for the same number of leading EOFs, nevertheless still representing more than 80% of total variance. The EOFs arising from the analysis of Perturbed Model states sampled at a higher frequency (1 hour) lead to almost 90% of explanation of total variance in the first 4 EOFs, a 5% increase from the results derived from the states sampled every 6 hours; for a large EOFs number (more than 10) the two sampling frequencies originate almost coincident explained variance results.

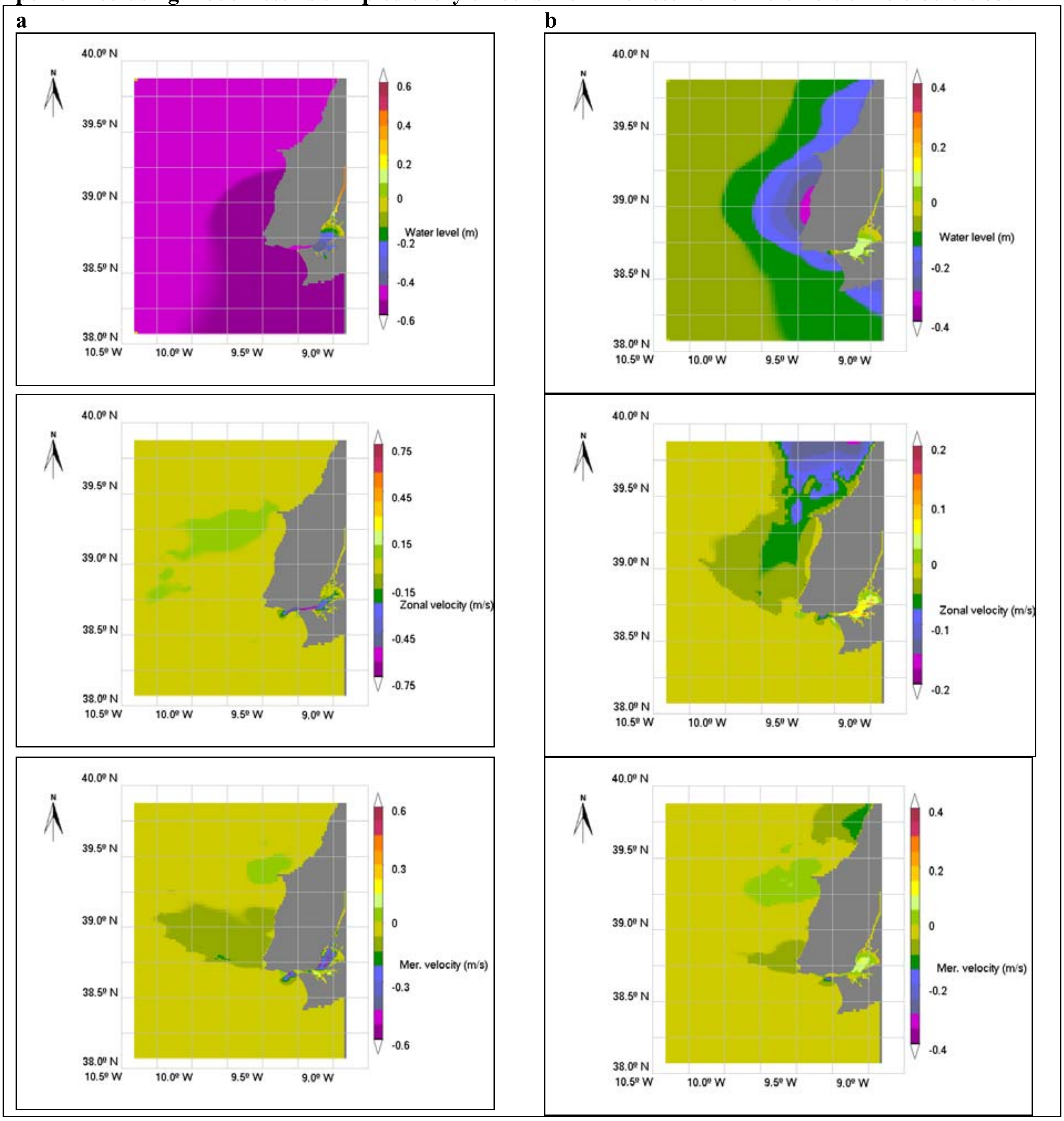
The leading EOF obtained for the Perturbed Model (Figure VII-4 a) presents the signature of tide: an inverse correlation observed in the variability inside and outside the estuary in water level field, derived from the propagation through the domain of the tidal wave, and the visibly high variability in the Tagus Estuary outside channel in the zonal velocity field. The system larger variability in the meridional velocity arises in the estuarine inside channels. It is also interesting to note that while the variability has comparable magnitude, non null, in most of the domain in water level, in velocity (both horizontal components) a large discrepancy exist between the different areas of the domain, with the deepest ocean areas presenting variability approximately null. The True Model leading EOF (Figure VII-5 a) is very similar and the same happens for EOF 2 of both models, providing evidence that is the underlying True Model tidal dynamics that is captured by the first two EOFs of Perturbed Model states. The EOFs of Perturbed Model states sampled every 1 hour show also similar patterns (EOF1 in Figure VII-5 b).

Figure VII-3: Fraction of total state variance explained by EOFs. The EOFs are calculated from True Model states, Perturbed Model states and Perturbed Model state error sampled every 6 hours and for Perturbed Model states sampled every hour. Results from 01/11/1972 to 01/01/1973.



The Perturbed Model error leading EOF (Figure VII-4 b) shows very different patterns of variability, with the most striking difference being the water level field where the variability is now mostly concentrated outside the estuary in the central and north parts of the continental shelf. It is interesting to note also that this EOF 1 of Perturbed Model state error is remarkably similar, although with opposite sign, to EOF 3 of the Perturbed Model state (Figure VII-6 a) and the same happens between EOF 2 of Perturbed Model state error and EOF 4 of Perturbed Model state, which suggests that the first two EOFs of Perturbed Model state are dominated by the representation of state variability, and therefore they are very similar in True Model and Perturbed Model states, and should explain little of the state error variability. This hypothesis is supported by the fact that EOF 3 for True Model (Figure VII-6 b) shows very different spatial patterns from those of the Perturbed Model state EOF 3.

Figure VII-4: EOF 1 for state properties of Perturbed Model and Perturbed Model error. The fields for water level, zonal velocity and meridional velocity (Mer. velocity) are presented in absolute values, for the Perturbed Model state EOF 1 (a) and the Perturbed Model state error EOF 1 (b). EOF analysis is performed using model results sampled every 6 hours from the results from 01/11/1972 to 01/01/1973.



Contrast between the variability of original states and the variability of reconstructed states using only the leading 4 EOFs for the Perturbed Model state (Figure VII-7 a), confirms the explained variance values provided by the pre-processor (in most of the domain the variability difference is approximately null) but also outlines that the situation is very dependent of location in the domain and also of the state property: while in cells with larger depth the variability is generally well described, the same does not hold in lower depths (in the continental shelf and in the Tagus Estuary), specially for velocity fields where in some areas about 50% of variability is not represented by the first 4 EOFs. A similar situation is obtained for the Perturbed Model state error (Figure VII-7 b), although the continental shelf velocity variability seems to be good represented.

Figure VII-5: EOF 1 for state properties of True Model with 6 hours sampling and Perturbed Model with 1 hour sampling. The fields for water level, zonal velocity and meridional velocity (Mer. velocity) are presented in absolute values, for the True Model state EOF 1 (a) and the Perturbed Model state EOF 1 (b). EOF analysis is performed using model results sampled at 6 and 1 hours, respectively True Model and Perturbed Model, from the results from 01/11/1972 to 01/01/1973.

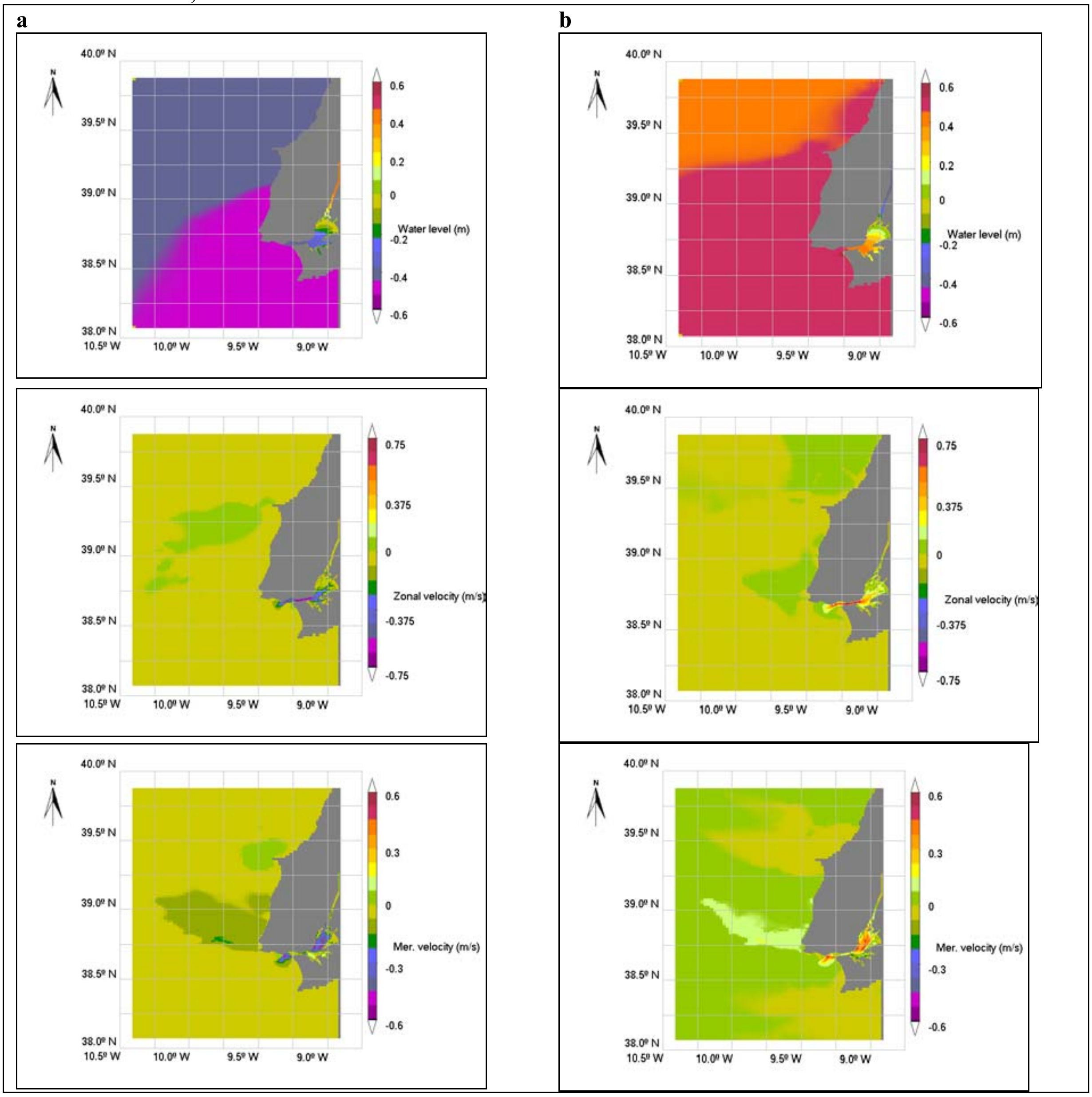


Figure VII-6: EOF 3 for state properties of Perturbed Model and True Model. The fields for water level, zonal velocity and meridional velocity (Mer. velocity), as arising from equation 2 are presented in absolute values, for the Perturbed Model state EOF 3 (a) and the True Model state EOF 3 (b). EOF analysis is performed using model results sampled every 6 hours from the results from 01/11/1972 to 01/01/1973.

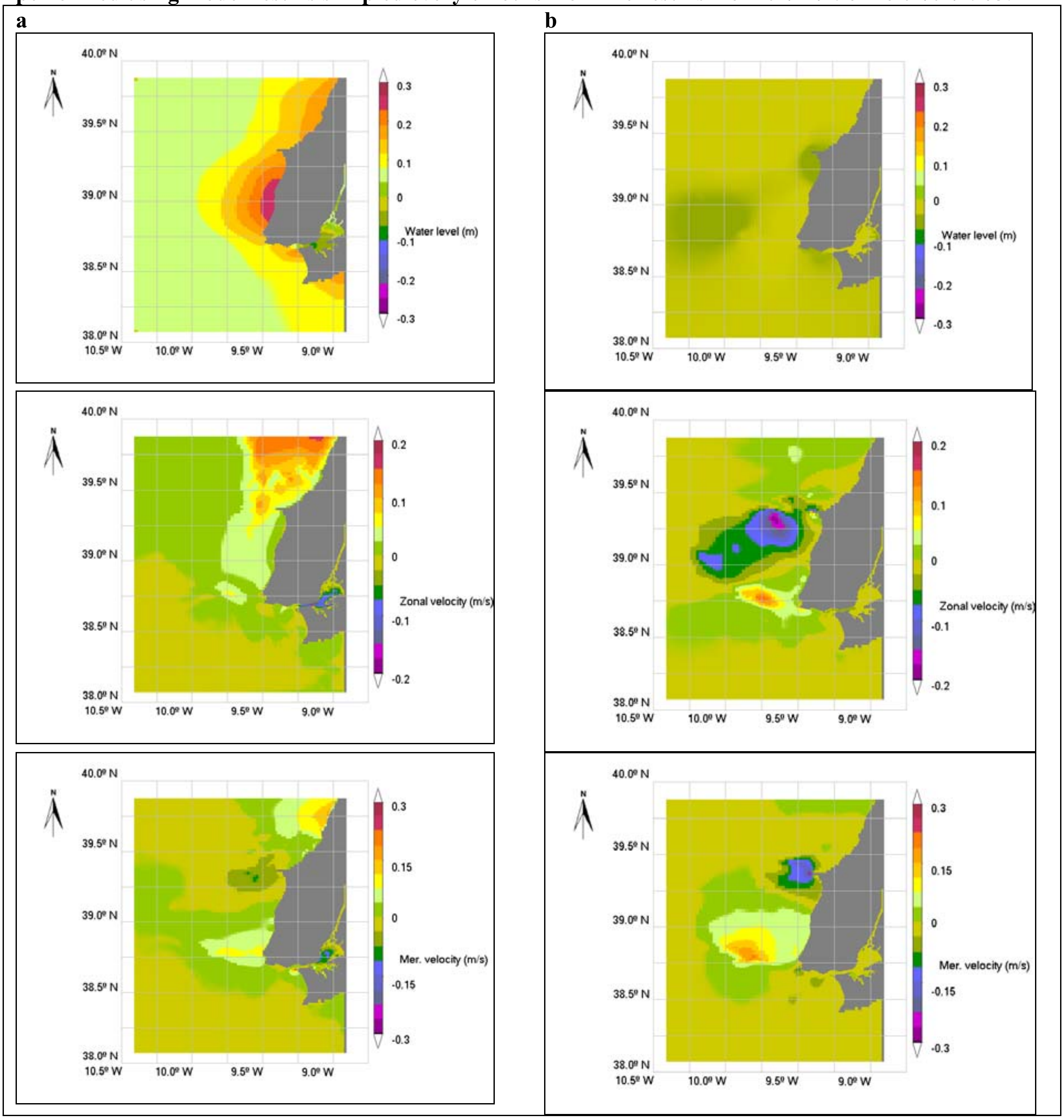
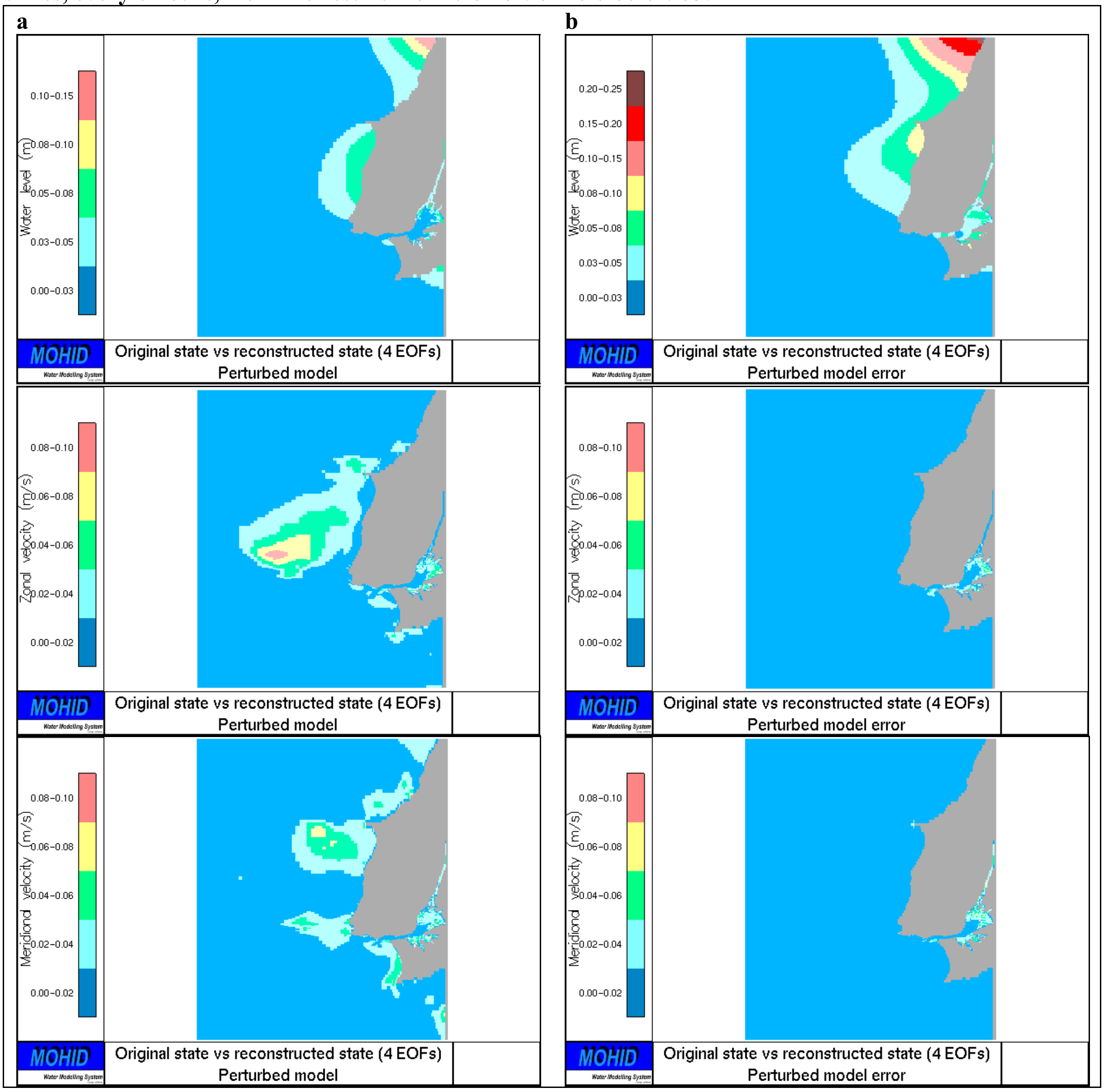


Figure VII-7: Standard deviation difference between originally sampled and reconstructed states using 4 EOFs for Perturbed Model state (a) and Perturbed Model state error (b). The reconstructed fields for water level, zonal velocity and meridional velocity are obtained for the same time instants of the sampled states, every 6 hours, from the results from 01/11/1972 to 01/01/1973.



VII.2.2.3. Discussion

The difference in curves of explained variance function of the number of EOFs can be understood as a difference in the analysed system dynamics complexity: it is noticeable the loss of representativeness of EOFs (more than 10% for the dominant EOF) resulting from a relatively small variance random error introduced at the True Model boundary; the more chaotic Perturbed Model error has an underlying dynamics even less well captured by the leading EOFs. The higher explanatory power of the leading 4 EOFs resulting from the states sampled every hour instead of every 6 hours is probably an effect of the smaller number of states and the shorter time period considered for analysis (only 10 days instead of two months).

The same dynamic complexity argument can be used as a framework to understand why only 4 EOFs represented in this 2D model case the same percentage of state (or state error) variability that are presented for a much larger number of EOFs in published studies of more complex models: e.g. 50 EOFs are required to represent 80% of a 3D baroclinic application for the Ligurian Sea (Mediterranean Sea) in the work of Barth *et al.* (2007).

The simple dynamics of this case cause also that the EOF analysis is able to approximately isolate the state variability and the state error variability in different EOFs. And because of this separation it is possible to see the coarse approximation that is done when considering the state covariance to represent the state error covariance.

The EOFs inform that within each state property the variability is concentrated essentially in the lower depths of the domain: estuary, estuary exit channel and the continental shelf, especially the northern part. They also show that from one state property to another the dynamics are very different. Probably due to this heterogeneous variability the leading 4 EOFs, although explaining more than 4/5 of the total variance of the domain taken as a whole, fail to capture a large part of state variability in some areas of the domain, namely in the interest area of Tagus Estuary. This comprises the velocity variability in the small depth and intertidal parts of the Tagus Estuary and the water level variability in the Tagus and Sorraia rivers discharge areas. The fact that the poor representability is similar in the leading EOFs for Perturbed Model state and for Perturbed Model state error seems to support this hypothesis.

According with the analysis of the probability distribution of the Perturbed Model state error done in Appendix IX it is verified that the areas of EOFs misrepresentation of state variability tend to be the areas where the model is more non linear, which justifies the difficulty in EOF analysis to separating the variability in different orthogonal modes. This variability is therefore local and will tend to be captured only by the EOFs with lower eigenvalue, which calculation is very imprecise.

VII.2.3. Assimilation trials

VII.2.3.1. Introduction

The assimilation trials consist in several Perturbed Model simulations assimilating True Model state derived water level measurements at tide gauge locations with different filter options. The objective is to determine the influence in filter performance of several filtering options, assessing the best filter configuration for improving the Perturbed Model forecast.

All simulations are for the fortnight period from 01/11/1972 0h to 15/11/1972 0h, representing approximately one spring and one neap tide periods, starting from the Perturbed Model state for 01/11/1972 0h. The covariance of the initial state error is the one provided from the EOF analysis of the Perturbed Model state, following the general procedure for SEEK filter initialization (e.g. Hoteit *et al.*, 2005), or of the Perturbed Model state error, according with the specific trial.

The simulations are organized in two broad groups: the ones that use a SFEK filter configuration and the ones that consider a SEEK filter configuration. The base filter configuration and the tested variants in each trials group is presented in Table VII-1 for SFEK filter and Table VII-2 for SEEK filter.

Table VII-1: Filter options for twin experiment SFEK filter trials.

SFEK filter trial	Forgetting factor	Initial covariance		Objective analysis	Number of measurements (water level)	Time interval between corrections
		Origin[a]	Number of EOFs			
Base	0.75	Perturbed Model state	4	No	5	6 hours
Variant 1	0.5	Perturbed Model state	4	No	5	6 hours
Variant 2	1.0	Perturbed Model state	4	No	5	6 hours
Variant 3	0.75	Perturbed Model state error	4	No	5	6 hours
Variant 4	0.5	Perturbed Model state error	4	No	5	6 hours
Variant 5	1	Perturbed Model state error	4	No	5	6 hours
Variant 6	0.75	Perturbed Model state	4	Yes	5	6 hours
Variant 7	0.75	Perturbed Model state	4	No	10	6 hours
Variant 8	0.75	Perturbed Model state	4	No	5	1 hours
Variant 9	0.75	Perturbed Model state	4	No	10	1 hours
Variant 10	0.75	Perturbed Model state	8	No	10	6 hours

Note: [a] States for EOF analysis are sampled at a frequency corresponding to the time interval between corrections.

The results of the simulations are assessed at the end of simulation calculating the simulations RMSE relative to the Perturbed Model without assimilation (Free-run) RMSE.

Table VII-2: Filter options for twin experiment SEEK filter trials.

SEEK filter trial	Forget_ ting factor	Initial covariance		Objective analysis	Number of measure_m ents (water level)	Time interval between correc_ tions	Lineari_ zation factor	Reorthonor_ malization
		Origin	Number of EOFs					
Base	0.75	Perturbed Model state	4	No	5	6 hours	1	Yes
Variant 1	0.75	Perturbed Model state	4	No	5	6 hours	1	No
Variant 2	0.75	Perturbed Model state error	4	No	5	6 hours	1	Yes
Variant 3	0.75	Perturbed Model state	4	No	10	6 hours	1	Yes
Variant 4	0.75	Perturbed Model state	4	No	15	6 hours	1	Yes
Variant 5	0.75	Perturbed Model state	8	No	10	6 hours	1	Yes

VII.2.3.2. Measurements

Measurements are created from the True Model simulation water level results for tide gauge locations by adding a random normally distributed error with null average and $0.0005m^2$ (approximately 0.02m standard deviation), with a consistent specification of the measurement error variance in the filter.

For the trials considering 5 and 10 measurements the locations are chosen from the available tide gauge network considering the following criteria with decreasing importance:

1. Located in Tagus Estuary or immediate adjacent coast;
2. Large Free-run RMSE considering all state variables: these are the locations where the correction of state is most required;
3. Large variability of Free-Run state and in True Model state, as observed in the leading EOFs and Perturbed Model state standard deviation: in order for the filter to be sensible to measurement dynamics;

The verification of the criteria for each tide gauge location is presented in Table VII-3. Following this classification, giving more weight to water level RMSE and considering the relative magnitude of the RMSE between tide gauge locations, the chosen set of measurements locations is:

- 5 measurements trials: Cascais, Trafaria, Lisboa, Cacilhas and Alcochete;
- 10 measurements trials: Cascais, Trafaria, Lisboa, Cacilhas, Alcochete, Paço de Arcos, Cabo Ruivo, Seixal, Montijo and Ponta de Erva.

For the trial with 15 measurements all tide gauge locations are considered.

Table VII-3: Measurement choice criteria verification for each tide gauge location. An x indicates that the location belongs to the 10 locations where each criterion is most verified.

Tide gauge location	**Criteria verification per state property**						
	1	2			3		
		Water level	Zonal velocity	Meridional velocity	Water level	Zonal velocity	Meridional velocity
Peniche		x	x	x	x	x	x
Sesimbra		x					
Cascais	x	X		x			x
Paço de Arcos	x	X	x			x	x
Trafaria	x	X	x		x	x	
Cacilhas	x	X	x		x	x	
Alfeite	x	X	x		x		
Lisboa	x	X	x	x	x	x	x
Cabo Ruivo	x	X		x	x	x	x
Seixal	x		x	x	x		x
Montijo	x		x		x	x	
Alcochete	x	X	x	x	x	x	x
Ponta de Erva	x		x	x	x	x	x
Póvoa de Santa Iria	x			x		x	x
Vila Franca de Xira	x			x			x

VII.2.3.3. Results

VII.2.3.3.1. SFEK filter

VII.2.3.3.1.1. Base trial

The relative RMSE for the trials with variable forgetting factor value are presented in Figure VII-8. For the average forgetting factor value considered, 0.75, the filter is able to correct the water level in most tide gauges with 5%-25% error reductions, more expressive at measurement or nearby locations. The tide gauges located at the Tagus and Sorraia rivers mouths (Vila Franca de Xira, Póva de Santa Iria, Ponta de Erva), which have the smaller Free-run error and are located at areas of small system variability in the correction basis, show, however, a large (near 50%) error increase.

While the water level analysis results are encouraging for the usefulness of data assimilation, the velocity results are disappointing: most of the locations show no improvement or even an increase of the error, especially in locations outside the estuary or near the river mouths. As in water level, these tend to be the locations having smaller error in the Free-run and where the variability is less represented in the EOFs basis. The determinant factor seems to be the magnitude of variability as present in the correction basis, which is exemplified by the good filter performance in Lisboa, in an area with large variability in the velocity correction basis, showing 25% error reductions in both components of horizontal velocity, and a poor filter performance in the meridional velocity at Seixal, where the dominant EOF does not show much variability although it is the location with the largest Free-run error.

VII.2.3.3.1.2. Forgetting factor effect

A reduction of one third in the forgetting factor value (i.e. considering a forgetting factor valued 0.5), translated in a perceived larger state error in the SFEK algorithm, leads to more expressive improvements in the water level error in most locations, now reaching 45% error reduction in some measurement locations (Figure VII-8). However in places whose water level variability is not well represented in the correction basis, as Póvoa de Santa Iria, Ponta de Erva, Vila Franca de Xira, the analysis shows a large error increase (above 50%) relative to the Free-run. A poor performance is also observed for velocity with most locations presenting a large (25% or higher) increase in the error relative to the Free-run.

The simulation with a forgetting factor equal to 1 presents a general negligible influence of the filter in the state error for all properties, an expected result since the underlying assumption of a perfect model is obviously wrong in this twin experiment framework. An exception is the water level error degradation at the locations near the river mouths (Póvoa de Santa Iria, Vila Franca de Xira, Ponta de Erva), where the correction basis misrepresentation of variability is high.

VII.2.3.3.1.3. Initial covariance from state error

The simulations considering as correction basis the results of the EOF analysis of Perturbed Model state error variability, instead of the Perturbed Model state variability, and a forgetting factor valued bellow 1 lead to generally large degradations (above 100%) of the state property fields, especially for velocity, and are not shown here. A striking exception is the location of Lisboa, where error relative to the Free-run reductions amount to 25%, 20% and 10% respectively for water level, zonal velocity and meridional velocity. The results obtained for forgetting factor valued 1 are presented in Figure VII 8 and are very similar to the ones obtained with the correction basis derived from the perturbed model state variability. It is not observed, however, the degradation of the near river locations Ponta de Erva, Póvoa de Santa Iria and Vila Franca de Xira in water level error.

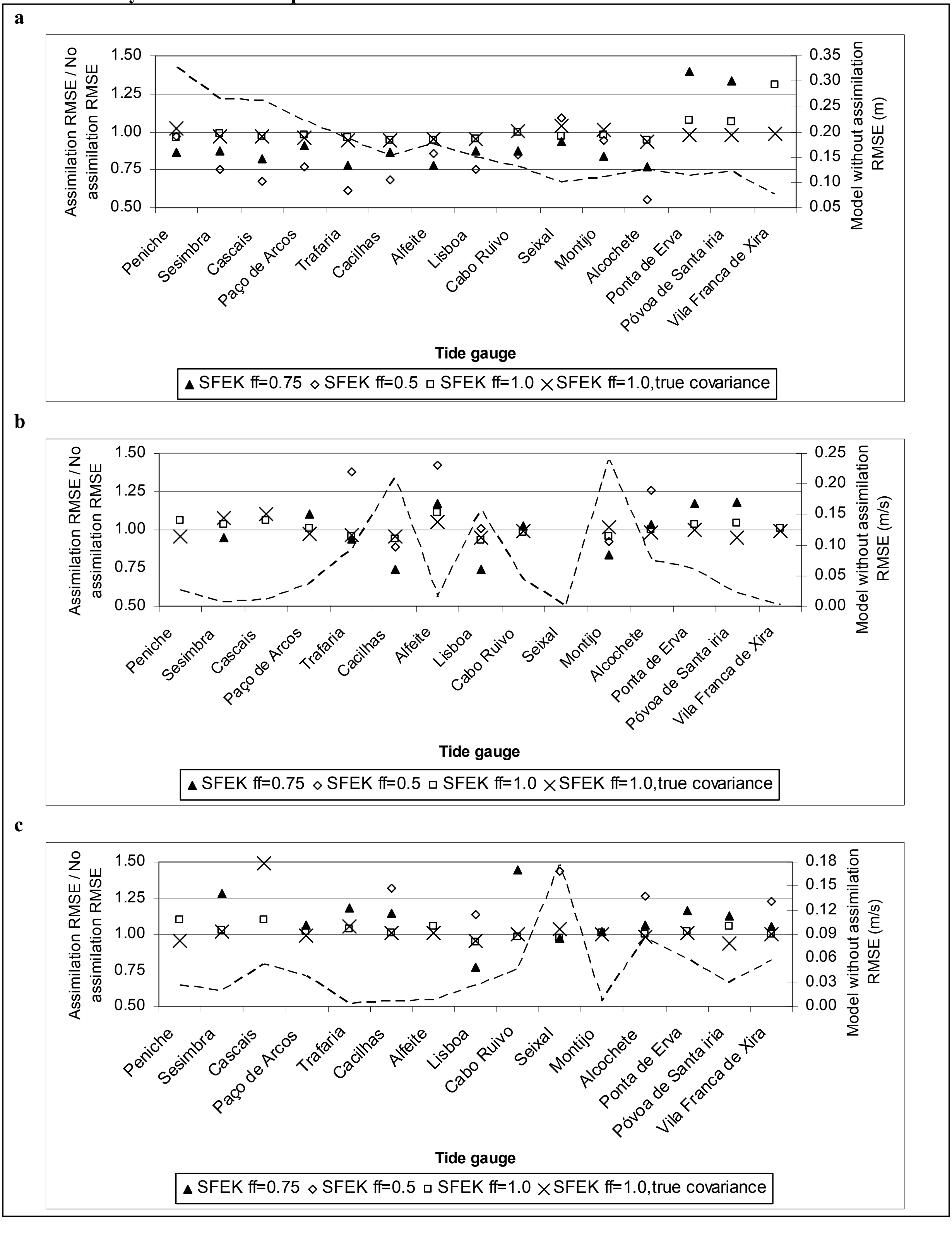
VII.2.3.3.1.4. Objective analysis

The analysis error results using objective analysis to correct the initial state are presented in Figure VII-9. The usage of objective analysis does not affect the overall filter performance in general; this behaviour is expected as the error is introduced by the model and the correction basis is fixed. The only relevant exception is the case of meridional velocity at Cabo Ruivo, whose error is reduced with the use of initial state correction.

VII.2.3.3.1.5. Larger measurement number

The assimilation of 10 measurements conducts generally to a better performance of the filter relative to the 5 measurements case (Figure VII-9): while improving further, in most cases, the state in locations already well corrected by the filter it also avoids or controls the error increase in locations with worse results than the Free-run using just 5 measurments. The velocity state for most locations present now a lower error than the free run with reductions in the range 5%-45%, particularly for zonal velocity, and the bad performance of the filter in correcting water level at rivers nearby locations (Póvoa de Santa Iria, Ponta de Erva, Vila Franca de Xira) has been controlled (Ponta de Erva presents an error reduced by nearly 20%).

Figure VII-8: SFEK filter analysis RMSE for different forgetting factor values and different specification for initial state error covariance. RMSE results for water level (a), zonal velocity (b) and meridional velocity (c), obtained with forgetting factor (ff) values of 0.75, 0.5 and 1 and the state error covariance (true covariance) as the initial covariance, are presented as a fraction of the Free-run RMSE (values not shown present larger RMSE). In dashed line is presented the RMSE of the Free-run. RMSE is calculated with the analysis values for the period from 01/11/1972 to 15/11/1972.



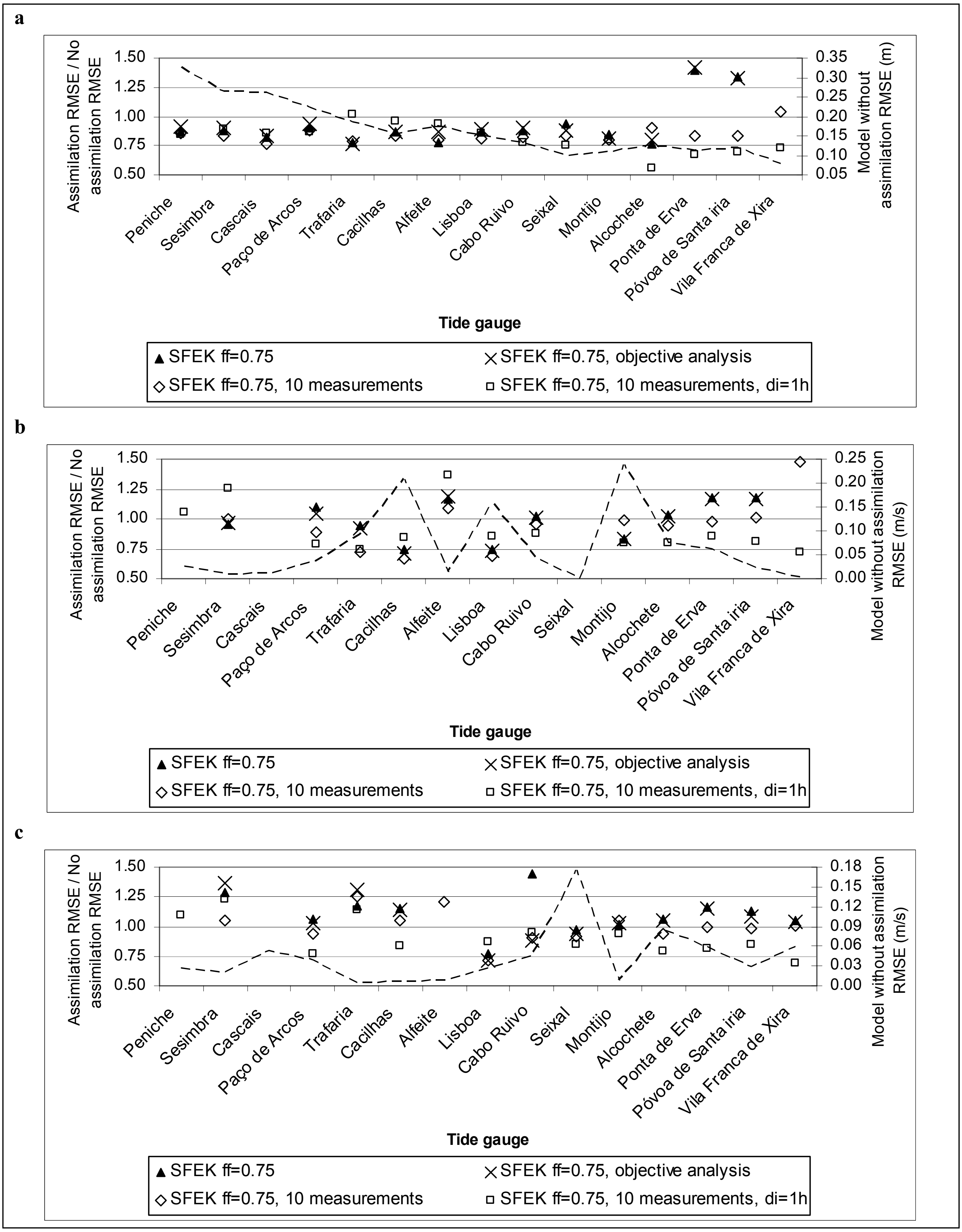
VII.2.3.3.1.6. Assimilation every hour

The results for the simulation making the correction with 10 measurements every hour (Figure VII-9) show a better filter performance essentially inside the estuary, particularly in Alcochete, Ponta de Erva and Vila Franca de Xira, but a smaller error reduction, relative to the Free-run, in the estuary exit channel locations. The simulation correcting the forecast state every hour with just 5 measurements results in a large increase (100%) of the perturbed model error in all state variables, and is not shown here.

VII.2.3.3.1.7. Number of correction directions

The trial using 8 EOFs in the correction basis resulted in the model collapse after 1 day of simulation, suggesting that the use of a large number of EOFs, which provide little additional state variance explanation as seen before, can introduce spurious errors in analysis.

Figure VII-9: SFEK filter analysis RMSE using objective analysis, assimilating 10 measurements and assimilating every hour. RMSE results for water level (a), zonal velocity (b) and meridional velocity (c) are presented as a fraction of the Free-run RMSE (values not shown present larger RMSE). In simulation assimilating every hour (di=1h) the initial covariance is derived from states sampled every hour; in the other simulations are considered the covariance from states sampled every 6 hours. In dashed line is presented the RMSE of the Free-run. RMSE is calculated with the analysis values for the period from 01/11/1972 to 15/11/1972.



VII.2.3.3.2. SEEK filter

VII.2.3.3.2.1. Base trial

The base SEEK filter trial RMSE results are present in Figure VII-10. It is observed that Free-run RMSE is generally not reduced by the data assimilation, or, in the case of velocity outside the Tagus Estuary, it is further increased, especially in the velocity fields.

While the bad filter performance on the velocity field outside the estuary is also observed in the SFEK filter results, the SEEK results are more homogeneous inside the estuary, denouncing a better control of RMSE degradations resulting from a bad specification of the initial correction directions, affecting in SFEK the velocity at Cabo Ruivo and the velocity and water level in the near river/estuarine channel locations of Ponta de Erva and Vila Franca de Xira. However, the good RMSE reductions observed in Cacilhas and Lisboa in SFEK filter are no longer present. The exception is the water level in Cascais, which RMSE shows improvements with the SEEK filter.

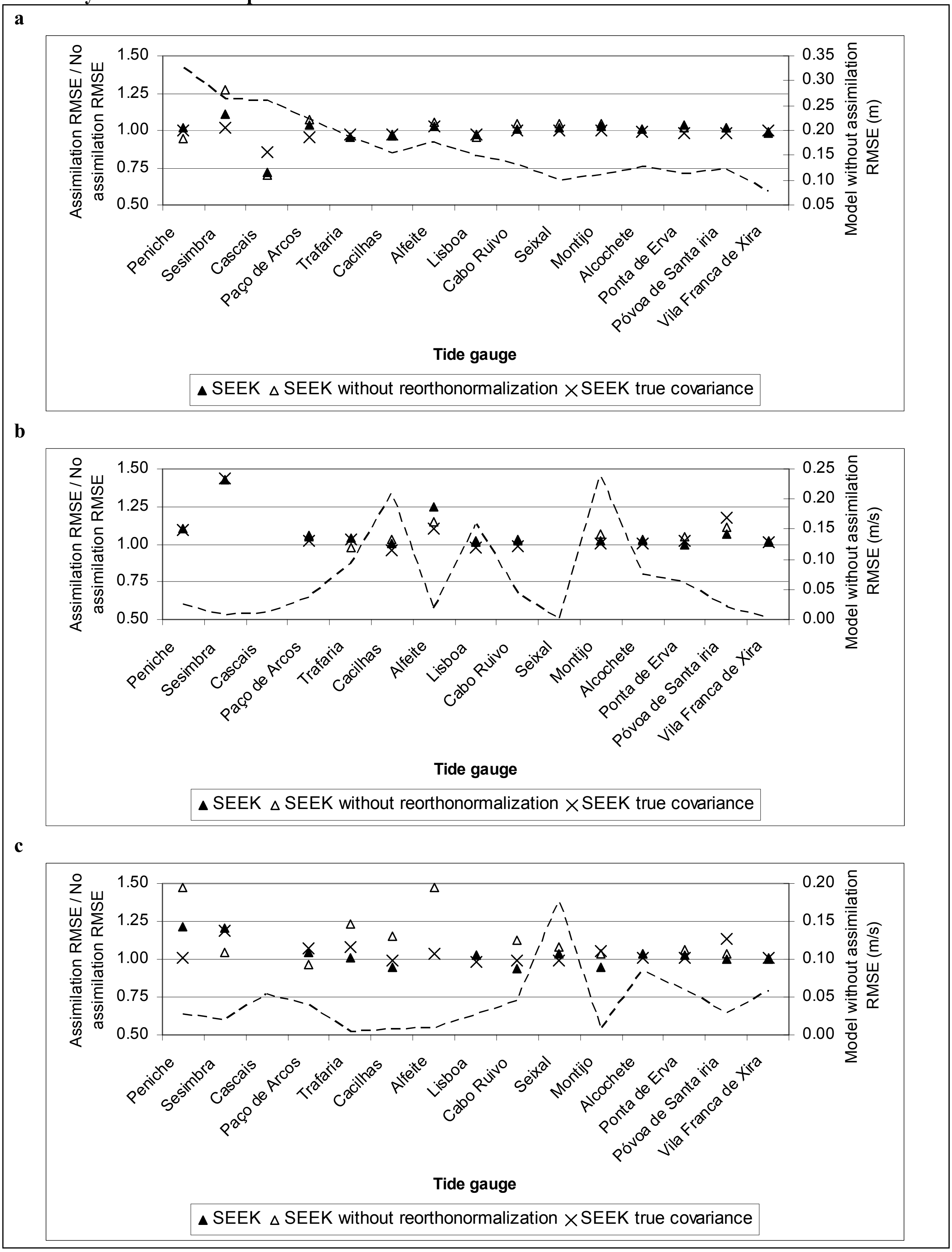
VII.2.3.3.2.2. Reorthonormalization effect

The results of the trial without the use of reorthonormalization after each forecast step (Figure VII-10) show that the RMSE is not affected in most of the locations, predominantly inside the Tagus Estuary. The effect is only noticeable near the estuary exit channel (Trafaria, Cacilhas), where the reorthonormalization is able to control the RMSE increase in meridional velocity, observed in the base SFEK and SEEK trials, and outside the Tagus Estuary, in Peniche and Sesimbra, where the meridional velocity field and the water level field, respectively is less degraded by the use of reorthonormalization.

VII.2.3.3.2.3. Initial covariance from state error

The use of the Perturbed Model state error EOFs to specify the initial correction directions of the filter is able to control and reduce the RMSE degradations relative to the Free-run in the velocity field at the borders of the estuary outside channel (Alfeite, Trafaria) and in the northern continental shelf (Peniche), as shown in Figure VII-10. This is a very different effect than the one observed with the SFEK filter, where the state error covariance provided increased RMSE in most of the locations. It is also interesting to observe that contrarily to these locations the water level field in Cascais verifies a smaller RMSE improvement with the state error covariance representing the filter correction directions at initial time.

Figure VII-10: SEEK filter analysis RMSE using or not reorthonormalization and different specification for initial state error covariance. RMSE results for water level (a), zonal velocity (b) and meridional velocity (c), obtained with or without reorthonormalization and the state error covariance (true covariance) as the initial covariance, are presented as a fraction of the Free-run RMSE (values not shown present larger RMSE). In dashed line is presented the RMSE of the Free-run. RMSE is calculated with the analysis values for the period from 01/11/1972 to 15/11/1972.



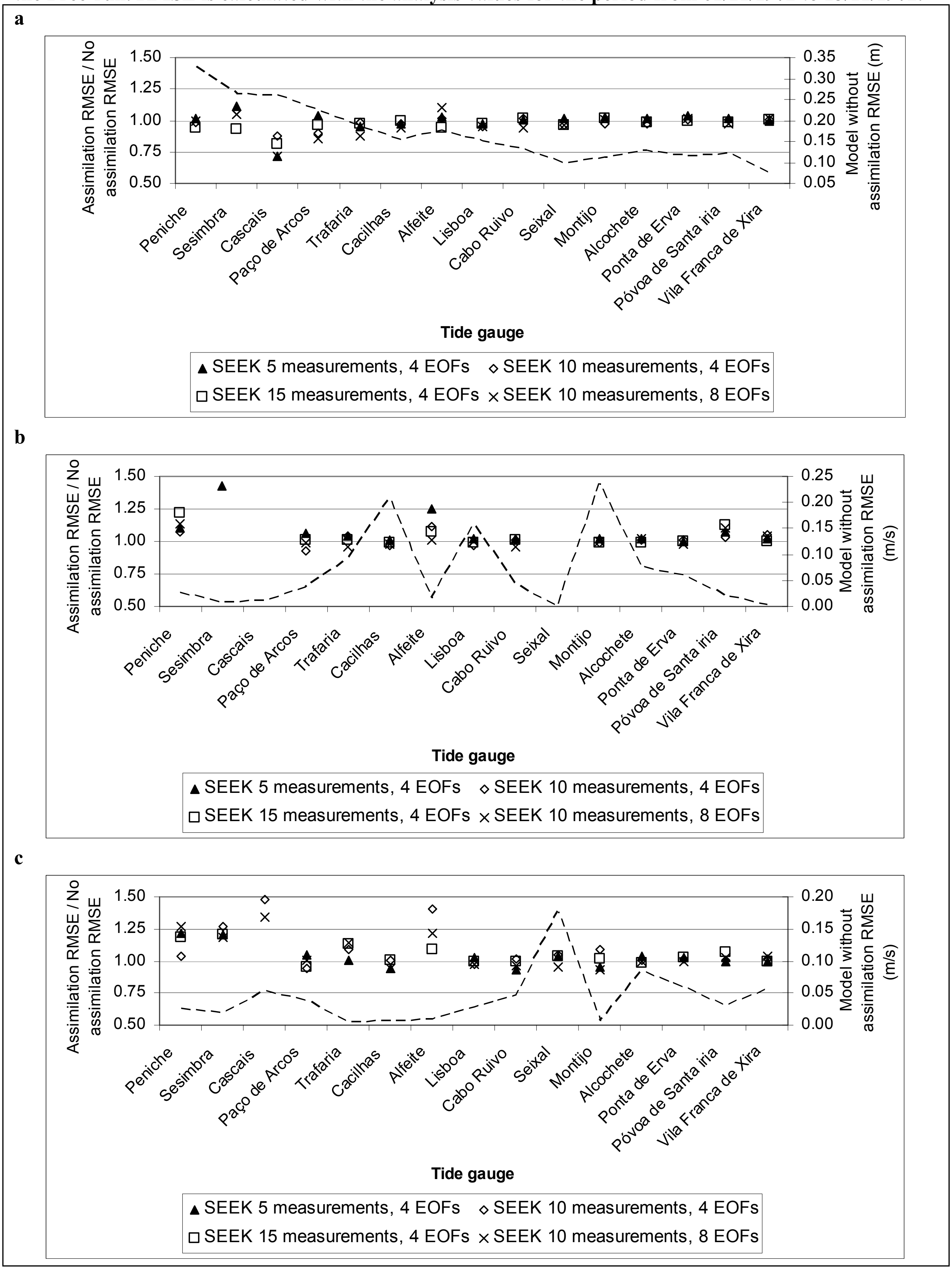
VII.2.3.3.2.4. Measurements number

The use of an increased measurements number affects the state variable field only punctually (Figure VII-11): mild improvements of RMSE (in the range of 10%) can be achieved in the water level outside the estuary (Peniche, Sesimbra) and near the estuary exit channel (Alfeite) through the use of measurements in these locations; velocity RMSE degradations observed sensibly in the same areas can also be more controlled. It is also observed that the providence of extra measurements in other locations can degrade the filter analysis in some locations, which is a consequence of the smaller weight given to each measurement in the correction with increasing measurement number: this is the case of Cascais and Paço de Arcos. It should also be noted that inside the Tagus Estuary in the near river/estuarine channel locations, where the SFEK filter performance increased with the use of local measurements, the SEEK filter is not sensible to those measurements.

VII.2.3.3.2.5. Number of correction directions

The use of a correction basis with the double of directions provides for 10% water level RMSE reductions relative to the Free-run in the estuary exit channel (Figure VII-11), being the tested configuration option to which the filter is most sensitive to in this area, and in meridional velocity in Seixal. The extra correction directions are also capable to control significantly the RMSE degradations in punctual cases such as the meridional velocity field in Cascais and Alfeite. This is a significant different response from the similar SFEK filter trial. However, as in SFEK, it is visible that the extra correction directions introduce errors in some state variable fields, namely in the zonal and meridional velocity in Peniche and in the water level in Alfeite.

Figure VII-11: SEEK filter analysis RMSE for different measurements number and number of correction directions. RMSE results for water level (a), zonal velocity (b) and meridional velocity (c), obtained with assimilating 5, 10 and 15 measurements and 4 and 8 correction directions, are presented as a fraction of the Free-run RMSE (values not shown present larger RMSE). In dashed line is presented the RMSE of the Free-run. RMSE is calculated with the analysis values for the period from 01/11/1972 to 15/11/1972.



VII.2.3.3.3. Discussion

The results obtained in this application show that the implemented SFEK and SEEK filters respond to filter parameter changes according with the theory and the known state error statistics but that, nevertheless, have difficulty to correct the state error. The main reason for this weak performance seems to be the weak representability in the correction basis of the system variability.

The main evidence that the correction basis is not capturing efficiently the dominant modes of the error dynamics is the fact that the SEEK filter has a generally poorer performance than the SFEK filter in reducing the Free-run error, which is not concurrent with the general findings of published studies using these schemes. In fact, this behaviour can be understood by considering that if the correction directions at the initial time do not represent the state error variability in a large part of the interest area, namely inside the Tagus Estuary, as is evidenced by the results of reconstruction of states using the dominant EOFs, the variability introduced by these modes contain will tend to be dissipated in the forecast step. This causes that the analysis shows little improvement relative to the Free-run error, because the correction directions should decrease in magnitude. This helps to explain in part why adding extra measurements in the near river/estuarine channel locations (Ponta de Erva, Póvoa de Santa Iria and Vila Franca de Xira) does not provide any appreciable change in the filter performance or contributes to the error reduction: this is an evidence that the correction directions show very little state error variability in these areas. This is only a partial explanation because it is known from the Free-run error analysis that error at these locations tends to be not normally distributed and so a linear filter should have more difficulty to correct it.

In the domain areas where error degradations are achieved by using SEEK filter data assimilation this it is possibly the effect of the introduction of noise in the system by inadequate corrections at the first correction steps, where the variability of the correction directions still has a significant magnitude. There is some evidence supporting this hypothesis: the model dynamics attenuates the bad early corrections inside the estuary through larger kinetic energy diffusion due to the shallow depths shear stress, while the error introduced by the correction steps is allowed to increase in the larger, outside the estuary, depth areas. The fact that the use of the state error estimated EOFs (instead of the state estimated EOFs) as initial correction directions and that the lack reorthonormalization (which should help the correction direction inadequate patterns to be destroyed by the forecast step) provides some relief to the RMSE degradations at some of these larger depth locations (Sesimbra, Peniche) also supports this hypothesis. In the case of the coastal bays of Cascais and Sesimbra, where the variability seems to be relatively well captured by the leading EOFs, it is expected that the linearization of the model used in the forecast of the correction basis should degrade the correction basis influencing some of the degradation observed.

While the SFEK filter correction basis should also not represent adequately the dominant error dynamics, the filter perceives a larger error of the forecasted state, conserving also correlations between the state variables, which causes that more weight is given to innovation information in the correction of state. The disastrous effect of using a small forgetting factor value supports this hypothesis, revealing that the EOFs do not represent well the state, particularly the velocity, error dynamics.

Both SFEK and SEEK filters results confirm the conclusion from EOF analysis results that the use of initial correction directions coming from the state error EOF analysis instead of those coming from the state analysis does not provides a solution for the Tagus Estuary area even if it was a valid option in a real data assimilation application (which is not). The same is valid for the option to increase of the correction directions number, particularly dangerous in the SFEK filter context, because of the small precision of the calculation of the smaller eigenvalue EOFs as referred in the published studies of SEEK filter applications.

The results indicate that for this twin experiment it is preferable to use a SFEK filter with a large number of measurements to compensate for the inadequate fixed correction directions of the filter In fact, the limiting factor appears to be the ratio between the number of correction directions and the number of measurements, which should be adequately small so as to provide the filter with more degrees of freedom for the statistical interpolation. They also show the potential of large measurements sources, such as those provided by remote sensed altimetry, for correcting state using SFEK filters (e.g. as shown by the studies of Brasseur *et al.*, 1999 and Barth *et al.*, 2007) which use however remains difficult in estuarine areas. The importance of the availability of measurements stressed in the present study is concurrent with the findings of the work of Mourre *et al.* (2006). These authors find that even with the more advanced EnKF scheme, which captures more adequately the system error dynamics, in a 2D system more homogeneous in bathymetry (all domain is comprised in the North Sea continental shelf) the assimilation of only 4 tide gauge measurements provides little correction to the state but that great improvement in state correction is possible by the assimilation of additional tide gauge measurements.

These results also indicate that the response of the SFEK filter to data assimilation configuration choices cannot always be considered as a worse quality predictor of a SEEK filter response. From the published studies this seems very much to be advocated for the choice of the number of correction directions and the value of the forgetting factor to use, but the present study shows that in systems with scarce measurements and very distinct dynamics in the model domain and state properties the response of SFEK and SEEK filters to extra measurements can be different and the SFEK filter can arise as a better choice for assimilation. The reason why this is not evidenced in previously published studies using both SFEK and SEEK filters (e.g. Brasseur *et al.*, 1999) seems to be the use of a large number of measurements in assimilation, mostly derived from remote sensing, in more dynamically homogeneous systems, in terms of bathymetry and of vertical discretization.

Finally, this twin experiment results also provides evidence that an increased grid resolution accentuates the SEEK filter deficiencies in correcting state forecasted by non linear models. In this sense the present work, while exploring essentially the relevance of vertical grid resolution (from 4km in the ocean to 1m or less in the Estuary), provides an interesting complement of the study of Hoteit *et al.* (2005) with small (1km) and high (500m) horizontal grid resolution non linear models for the Mediterranean Sea, more focused on the horizontal grid resolution effect, which shows that a SEEK filter is inadequate for the high resolution, due to its drawbacks in the evolution of the state error covariance.

VII.3. Hypothesis 4 Discussion

For MOHID Tagus Estuary this first data assimilation application results suggest that if the available tide gauge network is the only measurement source for operational use the tested SFEK and SEEK filter configurations are not expected to provide support for Hypothesis 4, since already poor results are observed in the twin experiment and inside estuary model non linearities are expected to be more expressive in the state error dynamics due to the determinant influence of bathymetry and horizontal grid resolution in the observed error between True Model and the real tide gauge measurements.

However, the twin experiment results suggest that the use of data assimilation could be successful in correcting the model forecast if any of two approaches is followed: either by revision of MOHID Tagus Estuary through the development a nested model domain only for the Tagus Estuary modelling, increasing the homogeneity of the domain dynamics variability, in order to facilitate the representativeness of the dominant EOFs to be used for correction basis of a SFEK/SEEK filter, or the improvement of the available data assimilation module of MOHID Water to allow the correction basis evolution to be better performed.

Following the second approach, two issues must be dealt with: model non linearities, which hamper a proper evolution of the correction basis of the filter, and initial correction basis poor representability of the state error covariance.

The first issue is important because model non linearities seem to degrade the SEEK filter performance in some specific locations (Cascais, Sesimbra) where the initial correction basis is expected to be a relatively good representation of the state error. The published research, particularly the work of Hoteit *et al.* (2005), evidences that a SEIK filter would provide a more adequate correction directions forecast (and eventually also a better representability of the initial state error covariance) in this case.

The second issue indicates that, if the initial state error covariance is to be estimated from the historical state variability (the traditional approach) a mechanism must be present in the data assimilation scheme for the improving of the representability of the correction directions, in order for adequate quality of the analysis. An adequate mechanism is provided by the apprentice filter introduced by Brasseur *et al.* (1999) which was applied with success, in a pre-operational setting, in a North Atlantic model by Brusdal *et al.* (2003). Both the SFEK and SEEK versions of the filter could be applied and tested. In a situation of relatively unconstrained time and computational resources it also seems worthwhile to improve the estimate of the initial state error covariance of the SFEK/SEEK filter through an ensemble method such as used by Barth *et al.* (2007).

In any of the approaches, it is recommended the use of error covariance splitting in coloured and white noise fractions such as used by Cañizares (1999) as the state error is not expected to be null averaged, as indicated by the descriptive statistics of the error between the MOHID Tagus Estuary and the tide gauge measurements. Consequently the model bias should be studied and a model of the coloured noise fraction must be developed.

VII.4. Conclusive Remarks

The implementation, in a twin experiment framework, of a first data assimilation configuration for MOHID Tagus Estuary 2D version assimilating tide gauge measurements from the estuary and coastal tide gauge network with SFEK and SEEK filters configurations provides evidence that the traditional filter algorithms proposed by Pham *et al.* (1998) for SEEK filter and Brasseur *et al.* (1999) for the SFEK filter cannot be viewed as a solution for reducing the model water level forecast error inside the Tagus Estuary.

While the SEEK filters are demonstrated to be ineffective in correcting the state, the SFEK filters provide only slight improvements (mainly 10% RMSE reductions relative to the Free-run) in water level but at the expense of velocity field error increasing in some areas of the domain. However, through the analysis conducted of the state and state error variability and the test of several configurations for filters it is possible to indicate, with a considerable amount of support, the main causes of this poor performance of these state-of-the-art Kalman filters for operational coastal forecast systems in MOHID Tagus Estuary. These are the poor estimate of the initial state error covariance and the extreme non linear dynamics of error at some areas of the domain, and also for the SFEK filters the small number of measurements available for assimilation. All these causes result from the heterogeneous model domain 2, in terms of dynamical effects and of vertical resolution, considered in MOHID Tagus Estuary.

Despite the poor performance of the tested SFEK and SEEK filter configurations, considering the twin experiment results and the simplified EKF alternative formulations described in published literature it is considered to be evidence that an improved forecast of MOHID Tagus Estuary through the data assimilation of tide gauge measurements is possible if assimilation is made only on a nested domain for the Tagus Estuary or if the assimilation module of MOHID Water is further improved to contemplate an apprentice SFEK/SEEK filter or a SEIK filter.

This study is found to provide very useful information both on the MOHID Tagus Estuary forecast state error dynamics and on SFEK and SEEK filters performance in correcting the state provided by 2D high resolution hydrodynamic tidal coastal/estuarine models in the presence of large changes in bathymetry in the domain and scarce measurement network. It's findings are shown to complement the available published studies on SFEK and SEEK filters schemes in these cases and are envisaged as particular relevant for modelling applications for areas with small width continental shelf, as the western Iberia Peninsula and African and South American mid-latitude Atlantic coasts.