VI. Implementation of Sequential Data Assimilation in MOHID Water

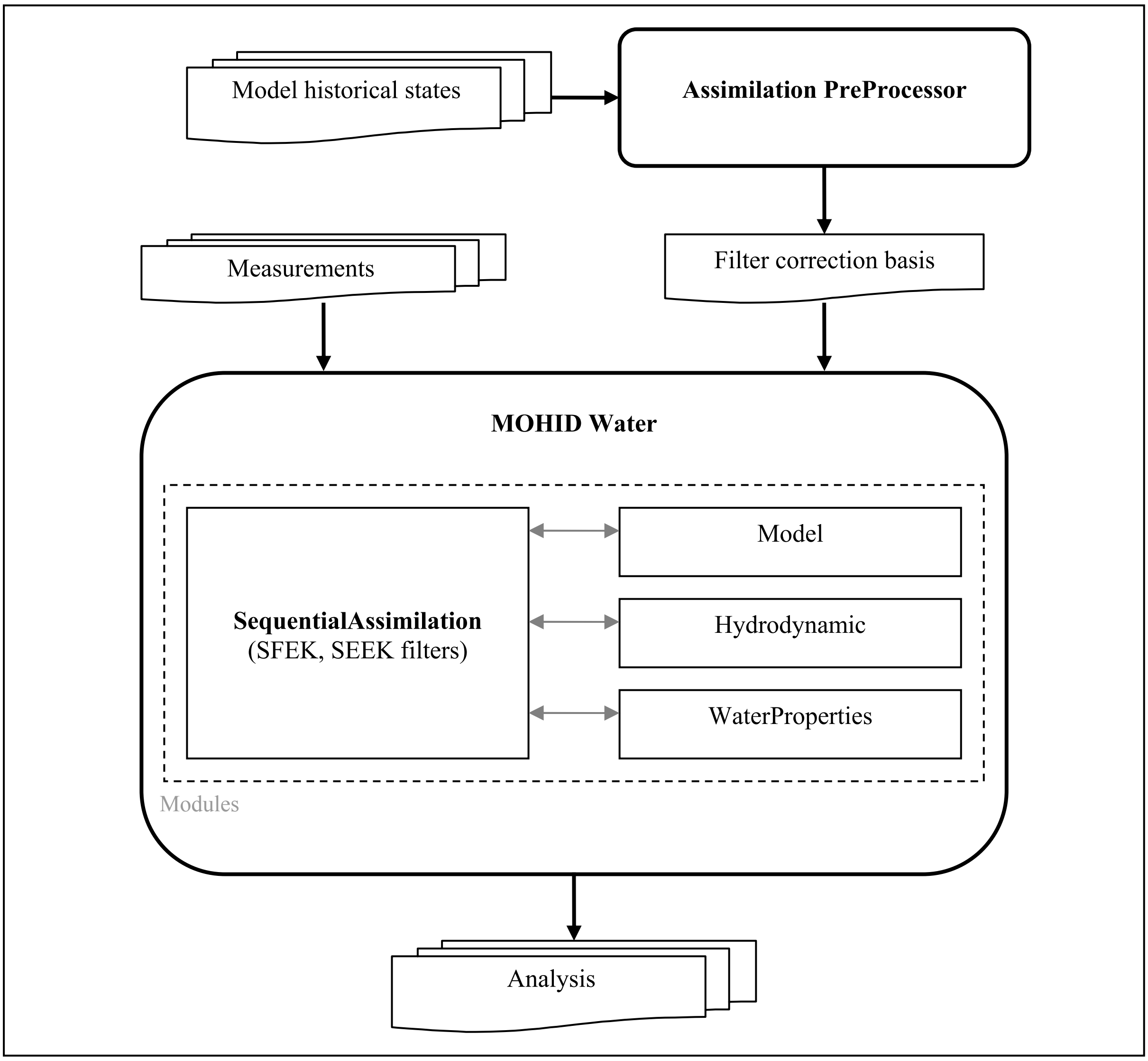
VI.1. Introduction

The implementation of sequential data assimilation in the MOHID Water numerical model is processed by the creation of a pre-processing tool, AssimilationPreProcessor, and of a new module, Module SequentialAssimilation. Both are programmed in Fortran 95 language and in the object oriented programming used in MOHID applications (Leitão, 2003). The interaction between these elements is presented in Figure VI-1.

The AssimilationPreProcessor manages the production of the initial filter correction basis (representing the state error covariance structure) for use in MOHID Water data assimilation applications. Module SequentialAssimilation manages the sequential data assimilation operations in MOHID Water applications. In this development only SFEK and SEEK filters data assimilation is implemented. These are EKF derived simplified Kalman filters in which the correction basis is initially composed by the dominant EOFs of the state estimate error covariance matrix (see Chapter I): in SFEK this correction basis remains fixed in time; in SEEK filter this correction basis evolves in time with model dynamics. As referred in Chapter III, these schemes are chosen for this first data assimilation implementation since they have reported applicability in non linear models, have small computational cost and algorithm complexity and are easily scalable to other more advanced schemes particularly adequate for data assimilation in non linear models.

The functionalities of Module SequentialAssimilation and AssimilationPreProcessor are described in detail in the following sections. A succinct description of code implementation is presented in Appendix X.

Figure VI-1: General scheme of the implementation of sequential data assimilation in MOHID Water framework. In grey arrows are presented the interaction between MOHID Water modules.



VI.2.AssimilationPreProcessor

VI.2.1. Introduction

This pre-processor tool calculates the covariance structure suitable for SFEK/SEEK filters initialization from MOHID Water historical states forecasts in modules Hydrodynamic and WaterProperties results files using the EOF analysis technique, as suggested by Pham *et al.* (1998a).

Considering the user defined state vector composed of n model variables, the basic equation considered by AssimilationPreProcessor is:

$$P^f\left(t_0\right) = L_0 U_0 L_0^T \qquad \text{(VI.1)}$$

where P^f is the covariance matrix calculated from the historical states forecasts, U_0 is the rxr diagonal matrix containing, in decreasing order, the user defined r highest valued eigenvalues and L_0 is the nxr matrix containing in columns the corresponding eigenvectors or EOFs, which represent the initial filter correction directions basis, and T means the matrix transpose. For simplification of the filter operations in Module SequentialAssimilation, L_0 and U_0 are instead taken

$$L_0 = L_0 \left[U_0\right]^{1/2} \qquad \text{(VI.2)}$$

$$U_0 = I \qquad \text{(VI.3)}$$

so that the new corresponding U_0 is provided by the identity matrix. This last L_0 constitutes the initial correction basis to be used in SFEK and SEEK filters applications. This corresponds to the approach followed by Hoteit (2007a).

The components of the AssimilationPreProcessor operation are presented in Table VI-1. The processing performed by the tool and the interaction with the user, through the input and output steps, are described in the next sections.

Table VI-1: Main components of AssimilationPreProcessor operation. In brackets are presented the format of the input and output files, formats refer to MOHID framework formats (Leitão, 2003).

Input	Processing[a]	Output
User instructions (ASCII format)	1. State statistics calculation	State statistics: average, standard deviation[b] EOFs variance representation Filter correction basis (HDF5 format)
	2. Normalization of states[b]	
Model historical states (HDF5 format)	3. Covariance calculation	EOFs expansion coefficients (Time Series format, one file per EOF)
	4. EOF analysis	Reconstructed states[b] (HDF5 format)
	5. Reconstruction of states[b]	

Notes: [a] cardinal indicates sequential order; [b] optional.

VI.2.2. Interaction with the user

The interaction of the tool with the user is made before processing, through an input file with instructions from the user, and after processing, with the supply of output files.

The instructions in the input file in ASCII format are organized in keywords following the general approach of MOHID Water described in Leitão (2003). These keywords command the operations of data input, state vector definition, processing options and data output as provided in Table VI-2.

This set of keywords allows flexibility in the pre-processor use and in subsequent data assimilation applications in MOHID Water. In the state data input it can be stressed the option to use state decimation avoiding the processing of MOHID Water result files to retain only the desired states and allowing the experimentation with different decimation factors, namely in order to accommodate different availability of computational resources.

As for the state vector, the user can appoint as state variable any Hydrodynamic or WaterProperties Module property (see Leitão, 2003, for a general description of properties) for any model calculus grid point, within a property-specific defined spatial rectangle. Consequently the state vector can consist in different spatial rectangle for each property. The tool assumes that every calculus grid point and boundary grid point within this rectangle which is not land in the land/water map is a state variable. This is in order that intertidal areas in MOHID Water applications can be considered for correction by data assimilation whenever these become covered. Since MOHID Water uses a staggered calculus grid, properties can be defined at grid cells centre, being type Z, or faces, being type U if associated with the zonal velocity grid or type V if associated with the meridional velocity grid. The velocity variables defined in faces between land and water in the land/water map are not considered as state variables as they are assumed null in MOHID Water (no slip condition).

Table VI-2: User instructions supplied in AssimilationPreProcessor input file.

Instruction	Mandatory[a]	Keyword/Block
State data input:		
Name of state containing file (more than one can be supplied)	Yes	<BeginHDF5File> NAME : … <EndHDF5File>
Consider 3D state data	No (2D data)	3D_HDF5
Start and end time comprising the states to be analyzed	Yes	START_TIME : … END_TIME : …
Use state decimation	No (no decimation)	DECIMATION_FACTOR : …
State vector definition:		
Name of data set in state files containing land/water map	Yes	HDF5_MAP_ITEM : …
Definition of state variables per property: - MOHID framework name; - path in HDF5 file; - corners of horizontal grid spatial rectangle; - type of property (Z, U, V); - convert property to faces.	Yes for name and path No (all spatial domain, type Z, not convert to faces)	<beginproperty> NAME : … UNITS : … DIMENSION : … HDF_GROUP : … STATE_WINDOW : … TYPE_ZUV : … CONVERT_TO_FACES : … <endproperty>
Processing options:		
Data assimilation method to supply initial covariance structure	No (SEEK filter)	METHOD : …
Use of state normalization	No (no normalization)	NORMALIZATION : …
Number of EOFs to be estimated	Yes	STATECOV_RANK : …
Reconstruct states with number of EOFs estimated	No (no states reconstruction)	STATE_RECONSTRUCTION : …
Data output:		
Name for output file with correction basis	Yes	OUTPUTFILENAME : …
Name for output file with reconstructed states	No	STATE_OUTPUTFILENAME : …

Notes: [a] in non mandatory instructions are presented in brackets the default options.

The default procedure of MOHID Water modules is to perform the output of every property in cell centre, implying that type U and type V properties values suffer a linear interpolation to achieve cell centre values. However, the state vector should be defined for the model calculus grid, for the simplicity of the data assimilation implementation in MOHID Water code. Having into account this restriction, it is implemented in this work a new option for the output of velocity results in Module Hydrodynamic (keyword OUTPUT_FACES) to allow MOHID Water simulations velocity results to be presented in this module output file also in the respective, zonal or meridional, faces grid (see Appendix X). These results can be directly considered for initial correction basis calculation using AssimilationPreProcessor if the respective type of property (U or V) is provided in the input file

(TYPE_ZUV keyword in Table VI-2). Still to allow the estimation of filter correction basis from results files where the velocity values are referred in the traditional cell centre it is considered in AssimilationPreProcessor the option to convert the cell centre values to faces values (CONVERT_TO_FACES keyword in Table VI-2), which however provides a less representative correction basis than the one obtained from the historical states provided in the respective calculus faces grid.

In the processing instructions, besides the optional use of normalization, which allows that the differences in units and magnitude of property values in multi properties state vectors to be taken into account in covariance calculation and consequently provide for a better calculation of the correction basis (e.g. Hoteit, 2001), can also be stressed the option to reconstruct the historical states used for covariance calculation with the set of r EOFs calculated. These reconstructed states contain only the variability represented by this subset of EOFs and the contrast of the historical and the reconstructed states can inform the user of the EOFs capacity to represent the variability in different areas of the domain. This information is explored in the Tagus Estuary application in Chapter VII.

The output of the AssimilationPreProcessor comprises information for use in data assimilation MOHID Water applications, but also information describing the historical states and information useful to assess the results of EOF analysis as summarized in Table VI-3.

Table VI-3: Summary of types of information provided in AssimilationPreProcessor output.

Data assimilation support	Historical states description	EOF analysis results assessment
Filter correction basis	State statistics	EOFs variance representation
	Normalization factor[a]	
	EOFs	EOFs expansion coefficients
		Reconstructed historical states[a]
	Eigen values	

Note: [a] optional.

VI.2.3. Processing details

Having described the possibilities of use and the types of results which can be obtained with AssimilationPreProcessor in the previous section, in this section are outlined the most relevant processing details of this tool.

Referring to Table VI-1, the processing made by AssimilationPreProcessor is dominated by two central operations: the covariance matrix calculation and the EOF analysis of this matrix. Optionally the processing includes also the reconstruction of states. These operations are detailed next.

VI.2.3.1. Covariance matrix calculation

To calculate the covariance matrix it is first calculated the average of each state variable. If normalization is chosen the standard deviation of each variable is also calculated and the normalization factor for each property is calculated as the inverse of the average of the standard deviation in all variables of this property. Hoteit (2001) uses as normalization factor the inverse of the square root of average of the variance of the state variables in each property.

The use of the average of the state variables standard deviation instead of the square root of the average of the state variables variance provides little difference in EOF analysis results when the variability is equally distributed within the same property state variables (the Hoteit, 2001, normalization factor in this case provides a slight larger variance explained, which represents bellow 5% in the Tagus Estuary application described in Chapter VII), while in a situation where the variability is unequally distributed the differences in variability are better accounted in the EOF analysis. It is expected that the 2D estuarine models in the Portuguese coast, such as MOHID Tagus Estuary, represent applications where the variability is indeed unevenly distributed through state variables, especially in the velocity fields. Therefore it is chosen in AssimilationPreProcessor to consider this normalization factor.

The differences relative to the average of each state vector are calculated and multiplied by the normalization factor corresponding to each state property, in case normalization is used, or 1, in case normalization is not used.

The covariance matrix (P) is obtained by the multiplication of the matrixes containing the difference relative to the average state vector (SD). This multiplication is usually represented by the following expression:

$$P = \frac{1}{s}(SD)(SD)^T \qquad \text{(VI.4)}$$

This covariance matrix has dimensions nxn. However, usually the number of sampled historical states s is much lower than the number of state variables n. To reduce the computational cost, the EOF analysis is performed with the covariance matrix arranged with dimensions sxs, following the approach of Hoteit (2007a) and using the expression:

$$P^* = \frac{1}{s}(SD)^T (SD) \quad \text{(VI.5)}$$

VI.2.3.2. EOF analysis

EOF analysis is performed by eigenvalues and eigenvectors decomposition of the covariance matrix, using the power method following Hoteit (2007a), obtaining:

$$P^* = L^* L^{*T} \quad \text{(VI.6)}$$

Where L^* is the matrix containing the EOFs in states space multiplied by the square root of the respective eigenvalue. The eigenvalues are the same in states space or in state variables space.

The fraction of state vector variance represented by each EOF is calculated from the respective eigenvalue λ_i and the trace of P^* as:

$$\frac{\lambda_i}{trace\left(P^*\right)} \quad \text{(VI.7)}$$

Each EOF (L^*_i) expansion coefficient (α_i) is calculated by the following expression

$$\alpha_i(t) = \frac{P^* L^*_i}{\lambda_i} \quad \text{(VI.8)}$$

The obtained L^* have the dimension of s instead of n and are after translated from the states space to the state variables space to obtain the EOFs in state variables space, i.e. the filter correction basis L_0, using the following expression:

$$L_{0,i} = \frac{1}{\sqrt{s}}(SD)L^*_i \quad \text{(VI.9)}$$

VI.2.3.3. Reconstruction of historical states

The obtained expansion coefficients (α), the EOFs in state variables space (L_0) and the average state vector ($\overline{x}^f$) are used to reconstruct the original state vector at a specific instant, $x^f(t)$, using the following equation

$$x^f(t) = \left[\alpha_1(t)\ldots\alpha_r(t)\right] L_0^T \sqrt{s} + \overline{x}^f \quad \text{(VI.10)}$$

Where r is the number of EOFs estimated or the rank of the covariance subspace.

VI.3. Module SequentialAssimilation

VI.3.1. Introduction

The SequentialAssimilation Module created in this work is capable of processing SFEK filtering in barotropic and baroclinic MOHID Water applications and SEEK filtering in MOHID Water barotropic applications according with the standard SEEK filter algorithm by Pham *et al.* (1998a) and the SFEK filter algorithm outlined by Brasseur *et al.* (1999). The equations which comprise these algorithms are presented in Table VI-4.

Table VI-4: SFEK and SEEK filter equations for the present application, following Pham *et al.* (1998a) and Brasseur *et al.* (1999). It is followed whenever possible the notation of Ide *et al.* (1997). H and R are constant in time.

Predictor equations	Corrector equations
Forecast of state vector: $\mathrm{x}^{f}\left(t_i\right)=M_{i-1}\left[\mathrm{x}^{a}\left(t_{i-1}\right)\right]$ (VI.11)	*Actualization of U:* $\mathrm{U}_{i}^{-1}=\rho\mathrm{U}_{i-1}^{-1}+\mathrm{L}_{i}^{T}\mathrm{H}^{T}\mathrm{R}^{-1}\mathrm{HL}_{i}$ (VI.12)
Forecast of correction basis (only in SEEK filter): $\mathrm{L}_{i}=\mathrm{M}_{i-1}\mathrm{L}_{i-1}$ (VI.13)	*Calculation of analysis:* $\mathrm{x}^{a}\left(t_i\right)=\mathrm{x}^{f}\left(t_i\right)+\mathrm{L}_{i}\mathrm{U}_{i}\mathrm{L}_{i}^{T}\mathrm{H}^{T}\mathrm{R}^{-1}\left[\mathrm{y}_{i}^{o}-\mathrm{H}\left[\mathrm{x}^{f}\left(t_i\right)\right]\right]$ (VI.14)

Notes: x^{f} is state vector forecast; x^{a} is state vector analysis; y_{i}^{o} is the measurements vector; H is the observation operator; R is the measurements covariance matrix; M_{i-1} is non linear model operator performing the evolution from time i-1 to time i and M_{i-1} is the model linearized operator.

This new module supports the inclusion in the state vector definition of any properties from MOHID Water modules Hydrodynamic and WaterProperties (see Leitão (2003) for description of available properties). The measurements can comprise any of those properties as long as the measurement locations represent state variables. This version requires that the measurements are supplied in MOHID TimeSeries format and the measurement number and the measurement locations are fixed in time.

In case a nesting modelling MOHID Water application is considered for providing the forecast for data assimilation, it is allowed that independent data assimilation is performed in each of the model domains. This means that for instance in the case of MOHID Tagus Estuary three data assimilation applications can operate simultaneously, one for each of the system three domains. This approach is chosen because it is consistent with several previous data assimilation studies (e.g. Bell *et al.*, 2000; Martin *et al.*, 2002), although it is not necessarily preferable to the use of only a state vector for all model domains as pointed by Barth *et al.* (2007).

The present version of SequentialAssimilation module does not allow that data assimilation specific processing variables are stored from one MOHID Water simulation to the next.

Following the filter equations in Table VI-4, the components of the Module SequentialAssimilation operation are presented in Table VI-5. The processing performed by the tool and the interaction with the user, through the input and output steps, are described in the next sections.

Table VI-5: Main components of Module SequentialAssimilation operation. In brackets are presented the format of the input and output files; formats refer to MOHID framework formats (Leitão, 2003).

Input	Processing[a]	Output
User instructions (ASCII format)	***Initialization step:***	Analysis fields and correction directions[c] (HDF5 format)
	1. Objective analysis[c]	
	Forecast step:	
Initial filter correction basis (HDF5 format)	2. Set initial state for model forecasting[b] (real and virtual simulations)	Analysis time series[c] (TimeSeries format)
	Correction step:	
	3. Actualization of correction basis[b]	
Measurements (TimeSeries format)	4. Actualization of U	Forecast and analysis error time series for measurement state variables[c] (TimeSeries format)
	5. Calculation of analysis	
	6. Reorthonormalization[b,c]	

Notes: [a] cardinal indicate sequential order; [b] only in SEEK filtering; [c] optional.

VI.3.2. Interaction with the user

The interaction between the Module and the user is processed by the reception of the user instructions from an input file and the supply, after processing, of results files.

Similar to other MOHID Water modules the instructions are organized in the input file in ASCII format through a set of keywords (which are in some cases organized in blocks), following the general approach described in Leitão (2003). The keywords specify operations of data input, state vector definition, processing options and data output as described in Table VI-6.

Table VI-6: User instructions supplied in module SequentialAssimilation input file.

Instruction	Mandatory[a]	Keyword/Block
Data input:		
Name of initial correction basis containing file (produced by AssimilationPreProcessor)	Yes	INITIAL_STATECOV_FILE : …
Definition of measurement information per measurement location: - type of file; - file name; - file data column (TimeSeries files) with measurement values; - time interval between measurements in file; - variance of measurement error (constant in time); - time tolerance to consider measurement for correction steps; - measurement location in model grid indexes.	Yes for file name, data column, time interval between measurements and measurement location No (TimeSeries type, 0 seconds time tolerance)	<<begin_measure>>[b] FILE_IN_TIME : … FILENAME : … DATA_COLUMN : … DT : … VARIANCE : … TIME_TOLERANCE : … LOCALIZATION_I : … LOCALIZATION_J : … LOCALIZATION_K : … <<end_measure>>
State vector definition:		
Definition of state variables per property: - MOHID framework name; - corners of horizontal grid spatial rectangle; - type of property (Z, U, V); - property with measurements.	Yes for name No (all spatial domain, type Z, no measurements)	<beginproperty> NAME : … UNITS : … DIMENSION : … STATE_WINDOW : … TYPE_ZUV : … MEASURE : … <endproperty>
Processing options:		
Data assimilation method	No (SEEK filter)	METHOD : …
Time interval between correction steps (starting at initial time)	Yes	DT : …
Rank of correction basis	Yes	STATECOV_RANK : …
Forgetting factor value	No (1.0)	FORGETTING_FACTOR : …
Use objective analysis	No (no Objective Analysis)	OBJECTIVE_ANALYSIS : …
Perform evolution of correction basis (SEEK filter)[c]	No (no evolution)	STATECOV_EVOLUTION : …
Linearization factor value[c]	No (0.5)	LINEARIZATION_FACTOR : …
Perform reorthonormalization of correction basis[c]	No (no reorthonormalization)	REORTHONORMALIZATION : …
Minimum time interval between reorthonormalization steps[c]	Yes	REORTHONORMALIZATION_DT : …
Data output:		
Indication of time instants for analysis fields output	No (no output)	OUTPUT_TIME : …
Make analysis time series output	No (no output)	TIME_SERIE : …
Name of file with locations for analysis time series output	Yes	TIME_SERIE_LOCATION : …
Make forecast and analysis error output for state variables with measurements	No (no output)	ERROR_TIME_SERIE : …
Indication of time instants for correction directions output	No (no output)	EOF_OUTPUT_TIME : …

Notes: [a] in the case of non mandatory instructions in brackets are presented the default options; [b] measurement block included in the respective property block; [c] instruction only for SEEK filter.

In what refers to data input, it should be noted that the information concerning measurements has been organized so that several types of measurement files can be supported in the future (as referred before the only type currently supported is TimeSeries). Also, there is some flexibility in the time distribution of measurements: they can be provided in files with different time intervals between the measurements and the user can establish the acceptable tolerance time interval for the measurements to be considered in a specific filter correction step. This flexibility is considered in the implementation driven by the aim to assist operational forecast systems needs, where measurement devices can temporarily become unavailable or measurement times are not equal to the desired correction step times.

The state vector definition is in its essential equal to the one made in the AssimilationPreProcessor input instructions (see Table VI-2): the specification details are discussed in section VI.2.2.. To the instructions provided there it is only required to add the indication of property with measurements if any are available. The choice of the same set of keywords for the two tools is, besides the obvious simplicity gain, driven by two concerns: reduce the errors derived with incorrect typing of a large number of keywords in file; avoid the errors derived from the initial correction basis referring to a state vector of different dimension than the state vector specified in the user instructions. It should be noted however that here no conversion to faces operation occurs because the state variables are assumed to be in the grid as the model calculus grid; therefore the type of property (keyword TYPE_ZUV in Table VI-6) should be specified according with the model calculus grid.

The version of Module SequentialAssimilation developed in this work has capabilities to consider as state variables for correction the properties of modules Hydrodynamic and WaterProperties, which are the relevant properties MOHID Water simulations involving the description of barotropic and baroclinic hydrodynamic and water properties advection processes. These properties are provided in Table VI-7.

Table VI-7: Hydrodynamic and WaterProperties modules properties that can be considered in the state vector defined in Module SequentialAssimilation.

Module Hydrodynamic	Module WaterProperties
Water level Zonal velocity Meridional velocity Vertical velocity Zonal water flux Meridional water flux Vertical water flux	Density Generic water property (temperature, salinity, phytoplankton, zooplankton, nitrate, fecal coliforms, etc.)

Depending on the processes being modelled measurements of one state property can have or not a visible effect on the model evolution of another state property: e.g. the assimilation of phytoplankton measurements should not cause a significant change in the hydrodynamics, as hydrodynamic processes are not determined by phytoplankton state and the correlations between these properties detected in historical states should not be significant or only coincidental. It is up to the user, through a thoughtful specification of the state vector at the instruction input files of Module SequentialAssimilation and of AssimilationPreProcessor, to ensure that the reasonable properties are considered as state variables.

As for the processing options, these are mainly organized in instructions of operation of the filter regardless the variant, SFEK or SEEK, and instructions concerning only the SEEK filter operation. The choice between the SFEK and SEEK filter variants is signalled by the user through the absence or presence of the STATECOV_EVOLUTION keyword, respectively. In both variants the method specified in the METHOD keyword is SEEK, since the SFEK filter is just a variant of SEEK filter. The default value of forgetting factor (1.0) is the assumption of a perfect model, since this is a value which minimizes the potential data assimilation shock caused by inconsistencies between analysis and model dynamics. It should be stressed, as referred in Chapter I, that the forgetting factor is in the SEEK (or SFEK) algorithm in representation of not only model errors but also the filter deficiencies. As for reorthonormalization it is introduced a keyword specifying the minimum time interval between reorthonormalization steps (keyword REORTHONORMALIZATION_DT in Table VI-6) because this is an operation which involves extensive use of computer resources and may cause a considerable increase of MOHID Water simulation time if performed too often. It is for this reason also that the reorthonormalization is not performed always in SEEK filter applications but only by user request.

In order for the user instructions in the input file to be considered by Module SequentialAssimilation in a particular MOHID Water simulation the user must signal the use of sequential data assimilation in that simulation in the Module Model input file and in the input files of the modules (Hydrodynamic and/or WaterProperties) to which refer the state properties. This signalling is made by the inclusion in these files of the keyword SEQUENTIAL_ASSIMILATION.

The output of data performed by Module SequentialAssimilation contemplates the traditional options for output of MOHID Water modules, for HDF5 format and for TimeSeries format, which are applied to the analysis results of state variables, and also output options specific for this module. These specific options comprise, as seen in Table VI-6, the output of time series of forecast and analysis error at state variables with measurements, which is a valuable resource for a quick assessment of a data assimilation application performance, and the output of the filter correction basis, which allows the assessment of the time evolution of the correction basis. These two types of output were used extensively in the early phases of this Module development to detect programming errors.

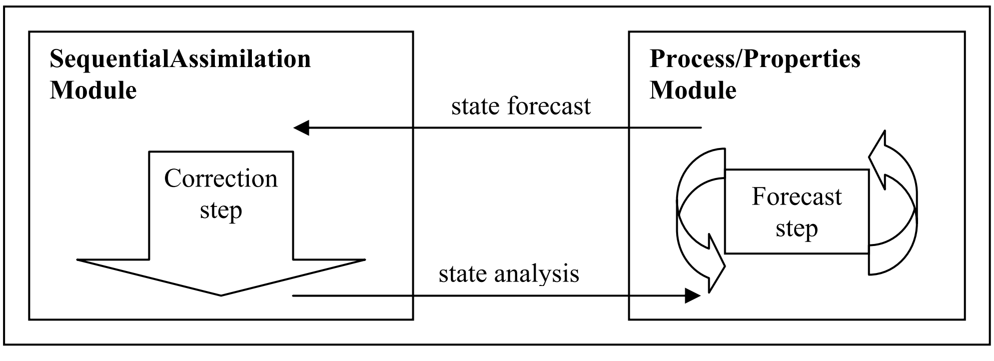
VI.3.3. Interaction with the other MOHID Water modules

The Module SequentialAssimilation occupies a special place in MOHID Water code architecture. As described by Leitão (2003), most of the MOHID Water modules control operations referring to realistic processes or properties (physical/chemical/biological): due to the nature of they operation these modules are located in the MOHID Water architecture at a comparable level. While this is still a reasonable assumption for the placement of the relatively simple newtonian relaxation operations controlled by Module Assimilation referred in section VI.1., this cannot be considered for Module SequentialAssimilation due to the invasive action, through the state correction step, of sequential data assimilation in processes and properties modelling.

Module SequentialAssimilation is therefore placed at a superior level relative to processes and properties modelling module, responding hierarchically only to Module Model, which controls the making of model simulations.

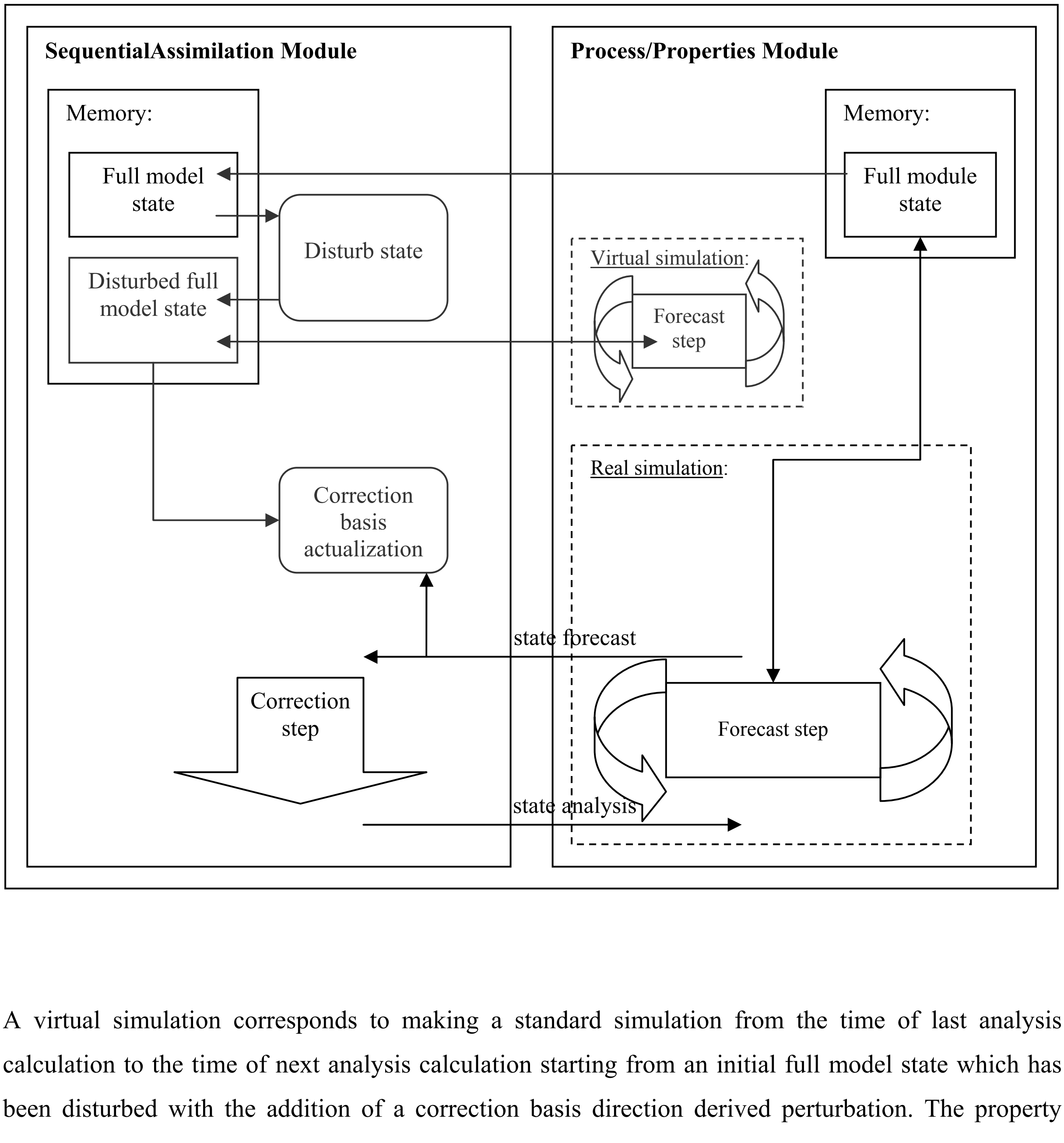
As referred in section VI.3.2., the current version of Module SequentialAssimilation allows the consideration for state vector definition of properties from modules Hydrodynamic and WaterProperties. The Module SequentialAssimilation interacts, therefore, more closely with these processes/properties modules. The interaction is always commanded by Module SequentialAssimilation and is presented in Figure VI-2 for SFEK filtering and Figure VI-3 for SEEK filtering.

Figure VI-2: Interaction between Module SequentialAssimilation and MOHID Water process/properties modules in SFEK filtering. The process/properties modules can be Module Hydrodynamic and/or Module WaterProperties in the current version.



It can be observed that this interaction consists basically in a two-way exchange of state vector information between the Module SequentialAssimilation and the processes/properties modules, for the realization of the analysis, to which is added the making of virtual simulations and memory displacement for the evolution of the correction basis in SEEK filtering.

Figure VI-3: Interaction between Module SequentialAssimilation and MOHID Water process/properties modules in SEEK filtering. The process/properties modules can be Module Hydrodynamic and/or Module WaterProperties in the current version. The set of virtual or real simulation status is performed by Module Model. In grey are operations referring to the evolution of the filter correction basis. A virtual simulation is made for each correction basis direction.



A virtual simulation corresponds to making a standard simulation from the time of last analysis calculation to the time of next analysis calculation starting from an initial full model state which has been disturbed with the addition of a correction basis direction derived perturbation. The property values output in process/properties modules is not made during virtual simulations.

A full model state comprises all the model variables whose value from the last time step is required to make a new time step of the model. The present version of Module SequentialAssimilation can only consider a full model state suitable for barotropic applications. This full model state is constituted by the properties of several MOHID Water modules which are summarized in Table VI-8. These properties are always considered, regardless if they are present in the state vector considered for data assimilation operations, since they are required to ensure that the model is starting from the same initial condition in virtual (before the operation of state disturbance presented in Figure VI-3) and real simulations. The restriction of applicability of the SEEK filter code implemented in this work to barotropic MOHID Water simulation is due to the fact that the hydrodynamic properties included in Table VI-8 only suffice to represent the initial full state in barotropic simulations.

Table VI-8: Properties comprising full model state formulation in Module SequentialAssimilation used to make virtual simulations.

Module Hydrodynamic	Module WaterProperties	Module Geometry
Water level	Density	Cells vertical dimensions (centre and faces)
Zonal velocity[a]	Relative density	Cells faces areas
Meridional velocity[a]	Water properties[c]	Cells centre depth
Vertical velocity		Cell volumes[a]
Zonal water flux		
Meridional water flux		
Vertical water flux		
Zonal water flow from father domain[b]		
Meridional water flow from father domain[b]		
Bottom shear stress coefficient		

Notes: [a] resulting from the two last time steps; [b] only on nested modelling; [c] all water properties calculated in Module WaterProperties are considered for full model state.

In virtual simulations the model uses and alters the full model state values stored in the Module SequentialAssimilation specific memory, while in the real simulation the model uses and alters the memory of the process/properties modules. This memory exchange is possible due to the use of programming with pointers and is an approach motivated by the concern that the memory of the process/properties modules remains isolated from the virtual simulation processing. However, this approach represents a multiplication of memory space which can be considerable if the model calculus points are extensive. This causes that SEEK filtering in MOHID Water is more computationally expensive than SFEK filtering not only in terms of time, for the making of the additional one virtual simulation per correction direction considered in the correction basis, but also in terms of memory consumption. The managing of this memory exchange is processed in the point 2 of Module SequentialAssimilation presented in Table VI-5.

VI.3.4. Processing details

The processing in Module SequentialAssimilation is based on the Predictor and Corrector equations presented in Table VI-4 and on the points provided in Table VI-5. However, due to its importance some of the processing operations of this module deserve an extra insight: objective analysis, actualization of correction basis, the calculation of analysis and reorthonormalization. These are described in the next sections.

VI.3.4.1. Objective analysis

The formulation for performing this operation is the one used by Hoteit *et al.* (2004), which comprises the two equations

$$\mathrm{U}_0 = \left[\mathrm{L}_0{}^{\mathrm{T}}\mathrm{H}^{\mathrm{T}}\mathrm{R}^{-1}\mathrm{H}\mathrm{L}_0\right]^{-1} \quad \text{(VI.15)}$$

$$\mathrm{x}^{\mathrm{a}}\left(t_0\right) = \mathrm{x}^{\mathrm{f}}\left(t_0\right) + \mathrm{L}_0\mathrm{U}_0\mathrm{L}_0{}^{\mathrm{T}}\mathrm{H}^{\mathrm{T}}\mathrm{R}^{-1}\left[\mathrm{y}_0^{\mathrm{o}} - \mathrm{H}\left[\mathrm{x}^{\mathrm{f}}\left(t_0\right)\right]\right] \quad \text{(VI.16)}$$

It can be observed that the second equation is the same as the analysis calculation equation presented in Table VI-4. This is the usual approach for performing Objective Analysis in sequential data assimilation studies.

VI.3.4.2. Actualization of correction basis

The actualization of the correction basis is specified in the formulation of the Predictor equations of Pham *et al.* (1998a) as requiring the use of a model linearized operator. However, as referred in Chapter I, this accomplishment is difficult in complex models such is the case of MOHID Water. Hence, the first order finite difference approximation of this operator is used instead, following Hoteit (2001) and Nerger *et al.* (2005):

$$\mathrm{L}_{\mathrm{i}} = \frac{1}{\varepsilon}\left\{M_{i-1}\left[\mathrm{x}^{\mathrm{a}}\left(t_{i-1}\right) + \varepsilon\mathrm{L}_{\mathrm{i-1}}\right] - M_{i-1}\left[\mathrm{x}^{\mathrm{a}}\left(t_{i-1}\right)\right]\right\} \quad \text{(VI.17)}$$

The parameter ε is the linearization factor (see Table VI-6).

This calculation is performed immediately previous to the correction step, what means that the disturbed states, expressed initially by $\mathrm{x}^{\mathrm{a}}\left(t_{i-1}\right) + \varepsilon\mathrm{L}_{\mathrm{i-1}}$, evolve as an ensemble of states in the forecast step.

VI.3.4.3. Calculation of analysis

This operation involves considerable computational cost due to the required matrix U in equation (VI.14). While this matrix is taken to be unity in the initialization of the filter, following equation (VI.2), and its inversion for equation (VI.12) is simple, following the first application of the same equation (VI.12) (or the use of Objective Analysis at the initial time) matrix U ceases to be diagonal and the inversion of its inverse becomes a complex process.

In Module SequentialAssimilation, as in Hoteit (2007b), this inversion is avoided through the use of Cholesky decomposition to solve the linear system of equations:

$$U_i^{-1}X = Y \qquad \text{(VI.18)}$$

Where $Y = L_i^T H^T R^{-1}\left[y_i^o - H\left[x^f(t_i)\right]\right]$.

The intertidal areas, which are part of the state vector as referred in section VI.2.2, are accounted in the correction step by performing the correction of the forecast state only on variables in cells covered with water at that time. This involves the use of a covered state operator, Λ_i, which has value 1 in state variables covered and 0 in state variables not covered. Equation (VI.14) then becomes:

$$x^a(t_i) = x^f(t_i) + \Lambda_i L_i U_i L_i^T H^T R^{-1}\left[y_i^o - H\left[x^f(t_i)\right]\right] \qquad \text{(VI.19)}$$

In this way the intertidal areas are forecasted by the model dynamics and not altered by the data assimilation procedure, in order to avoid the specification of inadequate values for state variables there located (e.g. large water levels or non null velocities).

VI.3.4.4. Reorthonormalization

Because the MOHID Water is a non linear model it is possible that the reorthogonal characteristic of the correction basis directions may not be conserved in the actualization procedure described in the previous section, at the same time that their value becomes large. To avoid these cases, Nerger *et al.* (2005) recommend the use of a reorthonormalization step, just after the correction step. In Module SequentialAssimilation is implemented the reorthonormalization step of these authors, which is constituted by the equations presented in Table VI-9.

Table VI-9: Reorthonormalization equations in Module SequentialAssimilation, following Nerger *et al.* (2005). Cardinal indicates sequential order.

1. Cholesky factorization of U_i	2. Calculus of covariance matrix	3. EOF analysis in r space	4. Translation of EOFs to n space
$U_i = C_i C_i^T$ (VI.20)	$B_i = C_i^T L_i^T L_i C_i$ (VI.21)	$B_i = V_i X_i V_i^T$ (VI.22)	$\hat{L}_i = L_i C_i V_i$ (VI.23) $\hat{U}_i = I$ (VI.24)

Notes: V_i is matrix with eigenvectors (EOFs) on its columns; X_i is diagonal matrix with the V_i corresponding eigenvalues; I is the identity matrix; r is the number of correction basis directions; n is the number of state variables.

In equations (VI.23) and (VI.24) of Table VI-9 the algorithm of Nerger *et al.* (2005) is altered to arrive to a definition of $\hat{L}_i$ and $\hat{U}_i$ which is consistent with equation (VI.2) and (VI.3). This altering does not affect the reorthonormalization procedure in the same way that equations (VI.2) and (VI.3) do not alter the filter performance relative to the use of the matrixes resulting from (VI.1).

While the performance of EOF analysis is a computational resources intensive step, the use of a covariance matrix specified in r space makes that it is not so resource consuming operation as the one performed in AssimilationPreProcessor.

VI.4. Conclusive Remarks

The AssimilationPreProcessor application developed in this work provides for the easy calculation of correction direction basis to be used in SFEK or SEEK filtering in MOHID Water applications from historical results files while providing the user with information describing these historical states and information useful to assist the decision making involved in the initial correction basis specification. The application code allows the future expansion of its processing to contemplate the production of initial state covariance structures for other sequential data assimilation schemes which rely on the analysis of historical model results.

Further, due to the supply of state descriptive information, from general statistics to EOF analysis results, and the use of generic MOHID formats this application can be used outside the scope of data assimilation studies or projects. Any data not produced by MOHID Water applications, e.g. produced by other models, are supported for analysis by AssimilationPreProcessor as long as the HDF5 format is MOHID format and the properties names are consistent with the names registered in MOHID Water Module GlobalData. A powerful tool for conversion from several data formats to MOHID HDF5 format exists, ConvertToHDF5, which can be used or adapted to attain specific format conversion.

As for Module SequentialAssimilation, it is equipped to perform adequate SFEK and SEEK filtering in high resolution coastal forecast MOHID Water applications, namely through the consideration of intertidal areas in state vector and in correction step and the possibility to use reorthonormalization only at specific time intervals and controlling the computational and time costs. These functionalities are explored in the MOHID Tagus Estuary application of Chapter VII. Another important feature of this module is the possibility to use it to improve the initial conditions of MOHID Water simulations through the use of Objective Analysis technique, even outside the framework of data assimilation applications. This is not explored in this work but it should suffice to establish a very large correction step time interval for Module SequentialAssimilation operation to be resumed to the Objective Analysis step.

Like AssimilationPreProcessor, the module is programmed in a way so that it can be extended with simplicity in the future to accommodate other Kalman filter based methods, which present, as seen in Chapter I, algorithms with considerable similarities with SEEK filter algorithm. Also, the module can be updated to support other formats for the input of measurements and a number of measurements changing with time: this is not attempted in this work solely due to time economy reasons. It should also be possible to extend the module code easily for the assimilation of measurements of properties not considered state variables, through allowing the reception of detailed user instructions for the construction of the observation operator.