Application of data assimilation in portable operational forecasting systems—the DATools assimilation environment

Ghada Y. El Serafy · Herman Gerritsen ·
Stef Hummel · Albrecht H. Weerts · Arthur E. Mynett ·
Masahiro Tanaka

Received: 8 April 2007 / Accepted: 16 August 2007 / Published online: 22 September 2007
© Springer-Verlag 2007

Abstract The first part of the paper describes a portable and flexible data assimilation environment (DATools) for easy application of data assimilation and calibration techniques to models that are used in smaller-scale engineering applications, for example to guide temporary offshore construction works, marine surveying and salvage operations. These applications are characterized by the need for detailed forecasts of often strongly non-linear marine behaviour in shallow-shelf seas under constraints of limited field measurements and absence of large-scale computing facilities. The applications are often prepared, and their results are interpreted by end-user engineers in field offices, not well acquainted with the data assimilation theory. The DATools data assimilation environment has been designed to facilitate such applications. It presently features an ensemble Kalman filter and two particle filters. These can be coupled to any process model in a standardized way through the so-called Published Interface that is used in an increasing number of flood-forecasting applications across Europe. The design of the system and its main modules are discussed, looking at the system from a non-specialist user

perspective and focusing on modularity, transparency, user guidance, intuitive and flexible uncertainty prescription. The second part of the paper describes a typical example application of data assimilation in an engineering environment for which the modelling environment has been developed: the daily forecast of current and salinity profiles to guide construction works in Osaka Bay. An ensemble Kalman filter-based steady-state Kalman filter is developed for assimilation of salinity and horizontal currents into an existing three-dimensional flow model for the highly non-linear stratified shallow bay. Calibration results are summarised for both hindcast mode and forecast mode. After assessment of the results, a second, improved model setup is described, including now assimilation of temperature and water level as well as currents and salinity. The calibration results are shown to improve further. Both applications show that significant forecast improvements can be possible with data assimilation in typical engineering applications, notwithstanding operational constraints such as limited measurement data and restricted computational infrastructure. It is argued that the described portable data assimilation environment is well suited for temporary engineering applications as it provides typical end-user functionalities for data assimilation and will so allow the user to focus on the characteristics of the physical processes and interpretation of the results instead of on details of the mathematical techniques. This should make practical data assimilation accessible for a much wider (end) user community in oceanography and offshore.

Responsible editor: Phil Dyke

G. Y. El Serafy (✉) · H. Gerritsen · S. Hummel · A. H. Weerts ·
A. E. Mynett
WL | Delft Hydraulics,
P.O. Box 177, 2600 MH Delft, The Netherlands
e-mail: ghada.elserafy@wldelft.nl

M. Tanaka
Kajima Technical Research Institute,
Tokyo, Japan

Keywords Data assimilation · Kalman filter ·
Interfacing protocols · Uncertainty prescription ·
Portable modelling systems · Coastal ocean · Stratified flows

1 Introduction

Structured integration of observational data with forecast models through data assimilation techniques is a successful approach to improve operational forecasts of natural phenomena. Well-known examples in the public domain are weather prediction and storm-surge forecasting. Operational forecasting systems for these applications cover a fixed geographical domain and are the long-term responsibility of a permanent (public) organization (commonly a meteorological service). The importance of having good forecasts for the national economy and public life justifies continuing improvement of the operational observation infrastructure and the modelling systems, including their optimization by data assimilation techniques. As a result, most of these systems are long-term dedicated developments for specific processes covering a fixed area of interest. In contrast to this, coastal oceanography projects and offshore activities such as surveying, construction, under water maintenance works and land reclamation are often of a temporary nature. These projects and activities require high-quality forecasts of geophysical processes for economic optimization of operations and/or minimization of environmental impacts. The general characteristics tend to be: temporary projects, globally varying locations, locally often limited public infrastructure, dedicated project measurements (which generally means limited observations), flexibility in techniques, variation in processes and geophysical parameters of interest.

The DATools modelling approach for application of data assimilation and calibration techniques aims to facilitate just these types of application, by providing a portable and flexible data assimilation environment for forecast improvement in time-limited offshore and oceanographic applications. It can be coupled to any process model in a flexible way using the extensible markup language (XML)-based Published Interface (PI), used by many organizations in Europe involved in flood forecasting (Werner et al. 2004; Werner and Heynert 2006). The standardized OpenMI concepts (<http://www.openmi.org>) are presently being considered as a second interface option. The concept has been tested for applications to one-dimensional hydrodynamic models as presented in El Serafy and Mynett (2004) and to a rainfall–runoff model as described in Weerts and El Serafy (2006). The results of these applications showed the feasibility and the easiness of applying this generic assimilation environment in more complex applications.

As an example of the potential of its application in operational marine forecast applications for the optimization of offshore activities, the presently operational forecasting system for offshore activities in the stratified Osaka Bay, Japan, is described. The starting point is a 3D circulation model for the stratified Osaka Bay, based on

Delft3D-FLOW (Tanaka et al. 2006; Lesser et al. 2004). Two different types of data assimilation techniques based on the Kalman filter are used with two different hydrodynamic model setups to improve the daily forecasts of horizontal current and salinity profiles in the bay. The ensemble Kalman filter or EnKF (Evensen 1994) is a generic data assimilation method that is also suited for highly non-linear models. However, for daily operational running of three-dimensional systems, a full EnKF is computationally too demanding for commonly used computer infrastructure in the field office (stand-alone personal computer [PC] or simple PC network), so a simplification was proposed. Instead, the EnKF concept was used to derive a steady-state Kalman filter (SSKF) for this application, assimilating salinity and horizontal currents (El Serafy et al. 2005). The same concept is applied a second time, for a much larger area domain, different measurement period and different type of data, bathymetry adjusted for progress of the works and including assimilation of local water level and temperature in addition to currents and salinity (El Serafy et al. 2006). In the next section, the concept, functionality and modular elements of the DATools data assimilation environment are described. Subsequent sections describe the two oceanographic applications for Osaka Bay. In a final section, conclusions and recommendations are presented on the applications and on the suitability of the described DATools data assimilation environment for these types of application.

2 The DATools data assimilation environment

For structured integration of models and data by data assimilation, several sequential data assimilation techniques have been developed and applied in practice. Most applications have been developed for daily operational forecasting and are integrated in dedicated systems for fixed domains with high-level observation networks (global and regional weather forecasting; storm surge forecasting). The DATools data assimilation environment focuses on flexible and easy-to-realize applications in portable systems for modelling geophysical processes with often only a few, temporary measurement stations and related to specific often temporary (engineering) projects in the coastal zone and rivers. It provides a generic interfacing protocol that allows combination of the implemented data assimilation techniques with, in principle, any time-stepping process model (atmospheric processes, 3D circulation, 2D water level, sea surface temperature, etc.). Therefore, the data assimilation routines and procedures are essentially separated from the process models for which they are used. This avoids entanglement and makes further development and extension of both (process model and toolbox) much clearer

and more efficient. Presently, the modelling environment features two data assimilation techniques, ensemble Kalman filtering as introduced in Evensen (1994) and applied by one of the authors in El Serafy (2003) and residual resampling filtering as introduced in Isard and Blake (1998) and applied by authors in Weerts and El Serafy (2006).

2.1 Role of uncertainties

In the field of surface waters, data assimilation is for example applied to water level and circulation models, models for transport and spreading, prediction of fish stock and algal bloom, models for wave propagation and rainfall–runoff models. By their nature, these models provide schematic representations of the real world. They focus on and represent those phenomena of the real world that are of specific practical interest, characterized by associated temporal and spatial scales of interest. Therefore, these models contain approximations by nature, which are often formulated as ‘errors’ or ‘uncertainties.’ They occur in the model concept as such, in the various model parameters, the driving forces, and in the modelling result. A model uncertainty of general nature is associated with the representativity of model results for observed entities. Equally, field measurements or observations also suffer from errors or uncertainties. The application of structured data assimilation techniques makes essential use of the statistical characteristics of the errors or uncertainties in the model and in the data. By prescribing known (or assumed) uncertainties, their propagation through the model in time is calculated. The better the uncertainty characteristics of the various parameters, data series etc. are known, the more accurate and effective the data assimilation technique can be in estimating the desired result and optimizing the errors and/or uncertainty in that estimate. The need to provide intuitive user guidance to the modeller or forecaster in the prescription of uncertainties, in particular for portable modelling systems that are used by field experts, who are generally less well versed in theoretical issues, is one of the motivations in the design of the DATools modelling environment.

2.2 Structure of the modelling environment

The modelling environment consists of four components or modules (see Fig. 1), each of which has its own function. The sequence manager manages the data flow. The user interacts with the sequence manager through XML input files, and the sequence manager is the engine that launches all other modules. The stochastic model manages the model data flow and introduces uncertainties on the equations of the deterministic model to convert the latter into a so-called stochastic model. The stochastic observer is the module that

Fig. 1 Structure and components

is responsible to collect and identify the observations and their uncertainties (i.e. to communicate with the observation database). Finally, the filter library uses the information to update and estimate the state of the model. The interfacing of the modules has been standardized together with the Delft University data assimilation research group that initiated the COSTA development (Van Velzen 2006). Similarly, the set of functions and tasks that needs to be present within the modules has been defined jointly. This joint definition and rigorous adherence to these standards ensures that new filter developments and modules within COSTA and the various DATools developments can be exchanged in a one-to-one way.

2.3 Interfacing with process models

To apply data assimilation to a time-stepping process model, a dedicated interface needs to be implemented for exchanging the model state (the computed state or state variables plus model forcing and model parameters for that particular model at any required time step) between the model and DATools environment. This exchange protocol can be implemented by directly providing the required functions or by providing XML files that adhere to the flood early warning system (FEWS) PI standard (Werner et al. 2004) or (in the future) the OpenMI standards (<http://www.openmi.org>). Given the definition of the interface for the particular process model, there are no application-dependent features. The interface allows the assimilation or filter algorithm to retrieve (get) the model state for a giving time step and, after adjustment by the filter, to pass it back (put) to the model for continuation of the simulation.

2.4 Sequence manager

The sequence of most of the sequential data assimilation filtering techniques is straightforward. The sequence manager coordinates the assimilation process. According to a set of XML configuration files, the sequence manager starts to operate by initializing the modules involved in the filtering process (i.e. stochastic model, stochastic observer and filter library). It initializes the vectors and matrices according to the problem size at hand. It uses the time information to control the launching of those modules when required and passes or receives the information needed. It is also responsible for the preparation of the final output of the simulation (i.e. user-defined specific output if needed such as expected value, confidence intervals etc.). It is thus considered the main manager and/or controller of the assimilation environment. Much attention has been given to the functional design of the sequence manager and user guidance/assistance, including fall-back options and generation of user warnings and archiving.

2.5 Stochastic model

The stochastic model consists of a deterministic system model (a model or a set of models) plus an uncertainty

module. The deterministic system model is the computational core of the numerical model that is responsible for propagating the state variable in time. By introducing uncertainties on the deterministic model equations, the model is converted into a so-called stochastic model. The data assimilation procedures use these new stochastic equations to derive the desired optimal result by suitable combination. Within the DATools environment, the uncertainties are defined through a Data Uncertainty Engine or DUE (Brown and Heuvelink 2007).

Through input files, the user defines the model uncertainties on state, parameters and/or driving forces. Accordingly, the uncertainty module calculates the model uncertainty. Figure 2 presents the structure of the stochastic model. The stochastic model passes all this information to the sequence manager. Using XML, the standard PI data exchange mechanism ensures that the data assimilation module can be used straightforwardly in combination with any process/forecast model that also features PI, e.g. any FEWS model (Werner and Heynert 2006).

2.6 Data Uncertainty Engine

The DUE adopted from Brown and Heuvelink (2007) is a flexible user-oriented module that allows the user to define

Fig. 2 Internal structure of the stochastic model

and describe uncertainties in model inputs. Sample data may be used alongside expert judgment to help construct an uncertainty model with DUE. Using DUE, the spatial and temporal patterns of uncertainty (autocorrelation), as well as cross-correlations between related inputs, can be incorporated in an uncertainty model. Such correlations may greatly influence the outcome of any data assimilation technique because models typically respond differently to correlated patterns of uncertainty than to random variation. DUE also supports the quantification of positional uncertainties in geographic objects, represented as raster maps, time series or vector outlines. Most importantly, DUE provides a conceptual framework for structuring an uncertainty analysis, allowing users without direct experience of statistical methods for uncertainty propagation to develop a realistic uncertainty module for their data. This data may be loaded into DUE from file or from a database.

2.7 Stochastic observer

The stochastic observer is the module that provides the sequence manager with the information needed on the observations and their uncertainty. In turn, it has access to two modules, the observer module and the uncertainty module. The observer module's main function is to provide the stochastic observer with the available observations within the assimilation step. The uncertainty module defines the uncertainty attached to the observations returned from the observer. The observation errors or uncertainties are generally a combination of equipment (in)accuracy, instrument drift, equipment fouling or malfunctioning, sampling frequency, data processing and interpretation. In this study, the uncertainties are estimated and prescribed by the user in the sense of their statistical properties.

2.8 Filter library

The filter library receives all information necessary to assimilate the observations and to update the state variable of the computational core. The updated state is returned to the sequence manager, which in turn passes it to the stochastic model to propagate this updated state in time until the next instant when measurements are made available.

Presently, the toolbox features the EnKF and particle filters, while further techniques can easily be added in a flexible way. In an EnKF application, an ensemble of model instantiations is simultaneously propagated step by step like a Monte Carlo simulation, each member differing in stochastically drawn uncertainties that are prescribed for the uncertain states, uncertain model parameters or uncertain model forces.

Figure 3 presents the situation of an ensemble of model instantiations being used.

In the next sections, a case study is described in which a three-dimensional circulation model is used for forecasting current and salinity profiles in the stratified Osaka Bay, Japan. It includes data assimilation for enhancing forecast results. The case study has several characteristics common in small-scale modelling in the marine environment: There is a temporary need for detailed information, the modelling is dedicated to the specific problem at hand, there is often a limited amount of operational measurement of the relevant parameters and there is little or no long-term historic measurement data. Besides, the modelling and forecasting is the responsibility of highly skilled local scientists–engineers, supported by specialists–modelers but with no data assimilation expertise. Time pressure is a common element in commercially motivated activities like these.

Fig. 3 Ensemble Kalman filtering, with an ensemble of model instantiations

The case study is therefore a good example of the type of data assimilation applications that would benefit from using the modelling environment described above. In fact, this particular case served as one of the practical test cases to further define the functionality and to test elements of the first versions of the assimilation environment, jointly with our counterpart Kajima Technical Research Institute.

3 Case study: 3D circulation forecasting Osaka Bay

3.1 User objective and modelling approach

For several years now, construction activities have taken place to create an offshore waste disposal site in Osaka Bay. The engineering contractor has to control the turbidity in the sense that the suspended solids concentration outside the construction area must not exceed the background level by more than 2 mg/l. Silt screens at different depths around the site are applied to minimize the spreading of sediments during the works (Tanaka et al. 2006). For additional support of the activities, a three-dimensional circulation model of the bay was developed to hindcast and forecast the current and salinity profiles at specific locations around the construction site by using the Delft 3D modelling system (Lesser et al. 2004). The forecast current directions and magnitudes are used to estimate the spreading of dredge releases, see Tanaka et al. (2006) for details.

3.2 Osaka Bay stratified basin

Osaka Bay is a shallow bay some 50 km long and some 20 km wide. It is bounded to the North and East by the Japan mainland and to the west by the island Awaji, see Fig. 4.

Osaka Bay is connected to the Sea of Harima and the larger Seto Inland Sea by a wide and deep channel north of Awaji. A constriction with islands in the south connects it to the Kii Channel and the Pacific Ocean. Depths vary from ~70 m in the west to less than 20 m in the east, with an average depth of 28 m.

A mainly diurnal tide from the Pacific Ocean is the primary driving force for the flow, with a range of the order of 2 m. Seasonal northeast–southwest monsoon-induced tilting of the ocean basin leads to seasonal variation of the mean sea level in Osaka Bay, which are effectively represented in the annual and semi-annual tidal variations. Further, five rivers discharge into Osaka Bay in the area of interest. The total river discharge varies in time from less than 50 to more than 500 m³/s, leading to a salinity-stratified three-dimensional circulation system. In calm-wind situations, vertical salinity levels of Osaka Bay may vary from a minimum of ~25 psu at the surface to a maximum of ~32.5 psu near the bed (with more or less

Fig. 4 Osaka Bay and its location relative to the Pacific Ocean and the Seto Inland Sea

fully mixed conditions at the transition to the Sea of Harima). Also present but less important is the seasonal temperature stratification. Varying wind and river discharges are the main drivers of the local salinity and temperature stratification, while storms can lead to full vertical mixing of the water column.

3.3 The deterministic circulation model

A coarse grid 3D limited area flow model was designed with tidal open boundaries beyond the channels and outside Osaka Bay (Tanaka et al. 2006). The model grid, channels, modelled islands, locations of open boundaries and rivers are shown in Fig. 5. The grid size is chosen to be 1,000 m.

In this overall model, a detailed model for the most northeastern area is nested. For local geometry and bathymetry representation, a curvilinear orthogonal planar grid has been designed with a 45° clockwise rotated grid orientation (149×104 grid points and with ten equidistant vertical so-called sigma layers). Grid sizes vary from ~750 m along the nesting interface to less than ~100 m near the construction site. Water level results from the overall model are prescribed along the southwest boundary, while along the northwest boundary, the current profile is prescribed, to constrain both water level and currents in the nested model. Daily river discharges are prescribed in m³/s and derived from local water level gauge recordings. Further, a spatially uniform wind forcing is prescribed in the form of hourly samples. Finally, the model was initialized using an assumed simple stratification state and

Fig. 5 Model setup for Osaka Bay application of 2002. Grid of the overall and nested flow model for Osaka Bay

was run for 2 months to create a dynamic salinity equilibrium. The model calculates the water level, the salinity and the horizontal velocity components in north and east directions in all points of the 3D model grid.

3.4 Considerations on data assimilation for Osaka Bay

Initially, operational forecasts were made based on the nested three-dimensional model results, using forecasts of wind and river discharges. The predictions of the current and salinity profiles of the nested model were used as forecasts. For further improvement of these predictions, application of a Kalman filter in combination with the regional flow model was proposed. The Kalman filter gives an optimal solution of the state estimate with least variances (i.e. high accuracy). It was developed for linear problems (see Kalman 1960; Kalman and Bucy 1961), but for non-linear models, the sequential extended Kalman filter (EKF) can be used, which is fully described in Jazwinsky (1970). However, the implementation of the linearization-based

EKF is essentially a model-specific implementation, integrated with the process model. The generic EnKF on the other hand can be easily implemented for use with complex highly non-linear models (Evensen 1997). The EnKF introduced by Evensen (1994) is a sequential data assimilation method where the error statistics are predicted using Monte Carlo or ensemble integration. It does not require adjustment of the model implementation as such. One of the disadvantages of the EnKF is the computational effort. Assuming a time invariant system, the SSKF described by Morf et al. (1974) can be used to reduce the computational effort. The steady state assumes a steady-state covariance matrix.

For the Osaka Bay case, with general access to regular computer facilities (state-of-the-art powerful PCs), the SSKF approach was considered a good and practically feasible approach. The SSKF is based on the correlation scales of the covariance matrix calculated by the EnKF. Based on the model sensitivity assessment, horizontal correlation scales for water level and salinity were found to be around 800–900 m. Similarly, the correlation scales

Fig. 6 Salinity measurements, those modelled by the model without assimilation and the hindcast salinity including data assimilation at station 4 at the vertical depth of 3 m from the surface

Fig. 7 Measured north velocity that modelled by the model without assimilation and the hindcast including data assimilation at station 4 at the vertical depth of 3 m from the surface

for currents were around 500–600 m. The uncertainty induced on all variables was assumed uncorrelated in the vertical. To avoid numerical artefacts, the correlations in the modelled river reaches and between islands were locally reduced to zero by applying suitable tapering. The ensemble simulations and the subsequent processing and calibration of the SSKF took place in a research environment, with the tailored SSKF result transferred for a first operational use in the field (El Serafy et al. 2005). After evaluation of the preliminary results and associated recommendations for improved modelling and assimilation, the same technique was used on a new model setup of Osaka Bay, including now the assimilation of temperature and water level (El Serafy et al. 2006). The results of both applications are presented in the next two sections.

4 Data assimilation for Osaka Bay 2002 (OBFS 2002)

4.1 Calibration of the Kalman filter (hindcasting)

In line with the focus of the practical application objective, the SSKF filter was designed for the nested model. For setting up and calibration of the filter, a comprehensive set of field data (Osaka Bay Regional Offshore Environmental Improvement Center, 2002) for the period 13th–27th February, 2002 (15 days=360 h) was used. In stations 1, 2, 3, 4 and 5 shown in Fig. 5, measurements of salinity and velocity components at four different vertically fixed levels (1, 3 and 6 m below surface water and 1 m from the bed level) were available. Because the data at the stations 1, 2 and 5 were not measured during the operational phase, only data at stations 3 and 4

Table 1 Percentage model errors and percentage reduction in the model error because of application of SSKF for stations 3 and 4 at all four vertical layers over 15 days=360 h

	1 m		3 m		6 m		1 m	
	Station 3	Station 4	Station 3	Station 4	Station 3	Station 4	Station 3	Station 4
Percent error in salinity								
Model without assimilation	4.41	3.51	2.01	1.44	1.29	0.63	0.52	0.68
SSKF-Hindcast	1.08	0.97	1.42	0.49	0.90	0.25	0.09	0.10
% reduction in model error	75.5	72.4	29.4	66.0	30.2	60.3	82.7	85.3
Percent error in north velocity component (for velocities >5 cm/s)								
Model without assimilation	97.4	59.5	80.5	88.8	75.2	101.6	49.7	84.1
SSKF Hindcast	34.8	44.9	50.8	23.3	21.4	28.7	17.3	17.1
% reduction in model error	73.3	60	22.2	70.3	33.9	73.8	66.6	61.0
Percent error in east velocity component (for velocities >5 cm/s)								
Model without assimilation	122.7	125.8	66.8	61.0	56.5	58.4	43.5	34.4
SSKF Hindcast	34.8	44.9	50.8	23.3	21.4	28.7	17.3	17.1
% reduction in model error	71.6	64.3	24.0	61.8	62.1	50.9	60.2	50.3

were used in the calibration of the SSKF and the subsequent operational forecasting. The model time step was 1 min, with the measurement interval being 60 min. The measurements were assimilated in a hindcast run. The results of the SSKF-hindcast, the measurements used and the model run with no assimilation are shown in Figs. 6 and 7 for the salinity and the North velocity, respectively, for station 4 at the vertical depth of 3 m from the surface. The East velocity showed a behaviour that was similar to that of the North velocity. From the figures, it is clear that the hindcast including assimilation of recent measurement data follows the measurements better than the model without assimilation. The results at other layers show a similar improved behaviour. At the surface, i.e. at the -1 m level, results are even significantly better than those shown here (El Serafy et al. 2005). For the deterministic model, we introduce the error measure $\varepsilon_{\text{model}}$ as the difference between the model and the measurements normalized to the measurements. For measured velocities less than 5 cm/s, the normalized error was disregarded from the calculations. Similarly, for the hindcast with assimilation, the percentage error, $\varepsilon_{\text{hindcast}}$, is calculated. The percentage error reduction because of application of the SSKF is then defined as $(\varepsilon_{\text{model}} - \varepsilon_{\text{hindcast}}) / \varepsilon_{\text{model}} \%$. The percentage errors and the resulting error reductions because of data assimilation are shown in Table 1, for the two measured stations (i.e. stations 3 and 4) and all four measured layers. The overall reduction in the error varies between 29.4 and 85.3% for the salinity, 22.2 and 73.8% for the north velocity component and 24.0 and 71.6% for the east velocity component.

4.2 The forecasting capability of the model

The potential SSKF improvement in forecast mode is analysed by simulating a possible (semi-)operational application in Osaka Bay. This prediction capability of the

SSKF was compared to the results of the model without assimilation. For the test, it is assumed that the measurement data are available until 19 February 0:00 hours and that a 24-h forecast is required till 20 February. The SSKF was provided with the set of measurements every hour from the beginning of the simulation time (13 February, 2002) until 19 February 0:00 hours. From that time, the state is then predicted in forecast mode till the 20 February 0:00. Because the model is driven by diurnal tide, the main improvement of the forecasted state is expected within the first 12 h and its effect rapidly dying out with time. However, because the stratification in the salinity has a longer timescale, the improvement in the forecast state can last longer. The three model simulation results and the measurements are shown in Figs. 8 and 9 for salinity and the North velocity component, at station 3, at the vertical depth of 6 m below the surface and for the period (16–20 February, 2002). From the figures, it is seen that forecast improvements are realized within the first 16 h forecast for the salinity and almost 20 h for the north velocity. For the east velocity (not shown here), this improvement is only within the first 6 h. The SSKF improvement in the velocities is reduced by the tidal boundary condition of the nested model. This is strongest for the east velocity component during this period. Similar results hold for the other levels and for the results at station 4. For all three parameters, the results of the SSKF forecast after 24 h is again better than the deterministic model result. It is observed that the improvement in the forecast of the salinity within the 24-h forecast is stronger than that of the velocity components because of the effect that short-term tidal variations are damped in their effect on salinity.

Due to operational constraints, an operational quasi-forecast cycle equivalent to a 6 hour hindcast cycle was proposed (Tanaka et al. 2005). They showed an improve-

Fig. 8 Salinity measurements, those predicted by the model without assimilation, the full period SSKF-hindcast and the 24-h SSKF-forecast salinity at station 3 at the vertical depth of 6 m from the surface

Fig. 9 Measured north velocity component, those predicted by the model without assimilation, the full-period SSKF-hindcast and the 24 hour SSKF-forecast at station 3 at the vertical depth of 6 m from the surface

ment in the forecast during those 6 hours better than the deterministic model with an overall reduction in error that varies between 6.6 and 50.3% for the salinity, and 8.8 and 18.2% for the resultant velocity.

4.3 Evaluation of the OBFS 2002 application

The full hindcast results show that even in a shallow stratified bay with strongly non-linear dynamics, the SSKF-based data assimilation can already significantly improve the model results. Table 1 shows that the improvement at intermediate levels in the water column tends to be smaller than at the surface or near the sea bed. This may be due to the strong effect on the currents of the prediction of the salinity interface depth. The results also show the potential forecast improvement because of assimilation compared to the deterministic simulation, with the update effect disappearing in time. However, the application of the SSKF

improved the operational forecasts less than expected based on theory. The results in forecast mode were shown for a single day only, although, while the results obviously will differ from day to day. Therefore, statistical results based on many forecast periods will allow a reliable assessment of the forecast improvement. In a further analysis, it was concluded that the domain on which the SSKF filter was active needed to be enlarged to the full Osaka Bay (even though the practical interest was very local). The longer propagation time of deterministic forcing from the boundary will reduce its effects in the region of interest and increase the duration of the assimilation improvement. Similarly, the deterministic model results with the new single model domain should be critically validated to ensure proper propagation characteristics of the uncertainties through the model. A new application would also give the opportunity to reconsider the specifications of the measurement data for filter calibration and operational

Fig. 10 Horizontal curvilinear grid of the Delft3D-FLOW model for Osaka Bay of 2004

Fig. 11 Temperature measurements against that modelled without assimilation, the hindcast with assimilation and the model results without assimilation at station 2 at the mid vertical level (22 September–19 November 2004)

forecast use. Based on this assessment, the OBFS 2004 application was designed to verify the hypothesis that the redesign of the model and the steady-state covariance matrix avoids the earlier drawbacks and can further improve the forecast capability of the model.

5 Revised model and assimilation (OBFS 2004)

5.1 Model configuration and oceanographic forcing

The second model and data assimilation configuration differs from the first in the following ways: (1) a single, curvilinear model grid for the whole Osaka Bay, (2) adjustment of the local model bathymetry to 2004 data to

account for the progress in the construction activities, (3) rigorous re-calibration of the model for tide, salinity and temperature, (4) measurement data for a different season (period 22 September–30 November 2004) in two measurement stations labelled stations 2 and 3 either side of the construction site (these are stations 3 and 4 of the OBFS 2002 application, respectively) and (5) assimilation of local temperature and water level in addition to salinity and horizontal currents. The new curvilinear orthogonal planar grid overall 3D model has 211×117 grid points with ten equidistant vertical sigma layers. This model size is barely acceptable for operational use at the engineering office. The new model was first calibrated on tidal stations in the area, while monthly mean salinity and temperature profiles based on data from the Japan Oceanography Data Center

Fig. 12 Salinity measurements against that modelled without assimilation, the hindcast with assimilation and the model results without assimilation at station 2 at the mid-vertical level (19 October–10 November 2004)

Fig. 13 North velocity measurements against that predicted by the model without assimilation, the hindcast with assimilation and the model results without assimilation at station 2 at the mid-vertical level (19 October–10 November 2004)

were prescribed along the open boundary (www.jodc.jp). Figure 10 presents the new model grid.

Along the ocean boundary, the model tide (composed from the 12 constituents S_a , S_{sa} , M_m , M_f , Q_1 , O_1 , K_1 , P_1 , N_2 , M_2 , S_2 and K_2) and monthly mean profiles of salinity and temperature are prescribed. In addition to that, the model is forced by time varying daily river discharges and uniform hourly wind. A sensitivity analysis to variations in forcing was performed, with evaluation of tidal representation in the frequency domain and against time series of available currents, salinity and temperature in stations 3 and 4. The overall tidal representation improved by 34% (El Serafy and Gerritsen 2006).

5.2 Calibration results of the new SSKF

At stations 2 and 3 shown in Fig. 10, new measurements are available, now for an autumn period, with different temperature characteristics than the 2002 measurement period. The data consists of Acoustic Doppler Current Profilers current data with vertical resolution 50 cm and hourly water level (at stations 2, 70 days, and 3, 47 overlapping days) plus for the same period's salinity and temperature at positions 1, 3 and 5 m below the surface and 1 m above the sea bed (at station 2 only). Again, the EnKF approach was applied to derive a SSKF, similar to the approach outlined in the previous section. From preliminary

Fig. 14 East velocity measurements against that predicted by the model without assimilation, the hindcast with assimilation and the model results without assimilation at station 2 at the mid-vertical level (19 October–10 November 2004)

Table 2 Reduction of error because of application of SSKF of the new model OBFS 2004 compared to that of OBFS 2002

Variables	OBFS 2004 (%)	OBFS 2002 (%)
water depth	10.5–40.2	–
Salinity	30.4–89.6	29.4–85.3
Temperature	78.2–93.4	–
North velocity	31.3–82.6	22.2–73.8
East velocity	29.0–81.2	24.0–71.6

simulations, horizontal correlation scales for water level, temperature and salinity were found to be around 5.7 km. Similarly, the correlation scales for currents were around 4.7 km. The uncertainty induced on all variables was again assumed uncorrelated in the vertical. To avoid numerical artefacts in the modelled river reaches and between islands, locally, the covariances were set to zero by applying a tapering from the open bay. The results of the full-period SSKF-hindcast, the measurements used and the model run without assimilation, are shown in Figs. 11, 12, 13 and 14 for temperature, salinity, the north velocity and the east velocity components, respectively, at the mid-vertical level at station 2.

Because of the high variability of the salinity and the velocity components, only a period of 21 days is shown in the figures.

The figures show that the full period hindcast including assimilation of recent measurement data gives again a far better representation of the measurements than the model without assimilation. The results at other layers show the same behaviour. Note that no salinity and temperature data are available at station 3.

The measures introduced in paragraph 4.1 above are again used to quantify the improvement because of the new model setup in combination with the steady state filter.

Fig. 15 Measured north current against that modelled without assimilation, the full period hindcast with assimilation and the 18-h forecast at station 2 at a mid-vertical level (30–31 October)

Table 2 presents the percentage improvement $(\epsilon_{\text{model}} - \epsilon_{\text{hindcast}})/\epsilon_{\text{model}}$.

It is concluded from the results of the hindcasting that the new model setup, its rigorous calibration and the more comprehensive assimilation has further improved the enhancement of the results by applying SSKF. The SSKF also improves the total water depth at the measurement stations (i.e. the water level), which was considered already well represented.

5.3 Improvement in forecasting mode

Similar to situations in operational systems, it is assumed that the measurement data are available for the hindcast part of the model simulation (here: until 31 October 06:00), after which a 18-h forecast is required (here: 31 October 6:00 until 31 October 24:00). During the hindcast part, the SSKF was provided with the set of measurements every hour from the beginning of the simulation time until the beginning of the forecast period (31 October 06:00). The state is then predicted in a forecast mode until the 31 October 24:00. The main improvement of the forecast state is expected within the first 12 h because of the tidal timescale of variation, with its effect rapidly disappearing with time. However, because the stratification in the salinity and the temperature extends over a longer timescale, the improvement in the forecast state should last longer. The SSKF-forecast, the model without assimilation, the SSKF-hindcast and the measurements are shown in Figs. 15 and 16 for the north velocity and the east velocity components, respectively, at the mid-vertical level for the period (30 October 06:00–31 October 24:00), both for station 2.

Forecast improvements are realized for the first 18 h for both the north and east velocity components. In this particular case, notably, the east velocity component

Fig. 16 Measured east current against that modelled without assimilation, the full-period hindcast with assimilation and the 18-h forecast at station 2 at a mid-vertical level (30–31 October)

strongly benefits from the SSKF application. However, as stated before, these improvement results are valid for this specific case of 18-h forecast and cannot be given a generic statistical meaning. The results for another 18-h period may be different. Statistical analysis of the forecasting improvement statistics to the results obtained by applying a moving time window over the full period of the measurement period unfortunately was beyond the scope of the present study.

6 Summary and conclusions

6.1 On the data assimilation case study

In Section 4, a SSKF was applied as a data assimilation technique to analyse the improvement of operational forecasts of velocity and salinity profiles in the stratified and strongly non-linear Osaka Bay. Both a full hindcast situation (the filter calibration stage) and a forecast case were simulated. The structure of the uncertainties in the model and the typical correlation scales were assumed based on those calculated by the EnKF. A critical re-design of the initially available 3D circulation model and better measurement specifications were realized based on the evaluation of the earlier SSKF application for the same practical application (see Section 5). Besides, the model now also assimilates water levels and temperature, in the expectation that this can further enhance the improvement of forecast results. The results show a further improvement in the velocity components in the full period SSKF hindcast, used for calibration of the SSKF filter. Equally, the benefit of SSKF application during the first 12–18 h forecast is illustrated, even for this highly dynamic and non-linear stratified flow field. The particular case shows that

improvement in the forecast of the velocity components can be up to 18 h and more but can also be much less.

In a more general sense, the experience of the first application in this case study in Section 4 has direct practical relevance as it once again stresses that data assimilation in operational forecasts such as SSKF can only be optimal if the design of the deterministic flow model takes into account the characteristic demands of the data assimilation technique, such as a sufficiently large domain, good (error) propagation properties in the model and well-designed types and locations of the measurement data used in correlation scales calibration and in the operational forecasting. The operational skill of the data assimilation technique is directly linked to the skill or quality of the deterministic model with which it is applied. This will be even stronger in highly dynamic non-linear environments. These are the areas where improved forecasts are often needed for project-based oceanographic and offshore activities.

6.2 On the DATools data assimilation environment

In Section 2, the stand-alone DATools modelling environment of data assimilation and calibration techniques was described (we remark that the calibration aspects will be treated in a separate paper). It has been developed to provide a portable and flexible data assimilation environment for improvement of forecasts in smaller scale, often temporary geophysical applications such as flows and transports for construction and salvage in the marine environment or, for that matter, in river and estuarine systems. In the view of the authors, there are two main benefits of providing the data assimilation functionality in a separate modelling environment such as DATools with its standard model interfacing and data exchange protocols. On

the one hand, (developments in) the assimilation routines are kept separated from (those in) the process models. On the other hand, the same model environment is easily applicable to other process models, without the need for software adjustment apart from a one-time linking using the standard interface protocol. In addition to the above, the provision of modules for flexible and easy user prescription of a range of uncertainties on parameters and model input (Stochastic modeller; DUE) and the modules for post-processing of results into optimal results plus statistics and confidence information will likely make application of data assimilation more accessible to non-specialists in the field.

The above features can be of practical use to any one applying data assimilation techniques to real life problems. In the view of the authors, they are particularly relevant for application of data assimilation in engineering type applications. In this study, the parameters of interest (processes) and the model area often vary from case study to case study, operational data is often limited, there is much time pressure for results and computing capacity is less than comprehensive. Commonly, it will be scientists–engineers who do not have a background in data assimilation who will be applying the techniques to enhance their model forecasts. User-oriented functionalities such as implemented in the DATools environment and described in this paper are expected to prove their added value especially for this category of users and for those types of applications.

Acknowledgements The authors wish to thank Martin Verlaan and Nils van Velzen for the in-depth discussions on the architectural concepts. They thank Juzer Dhondia for his contribution in designing and programming and Hanneke van der Klis for aspects of user requirements. The authors thank Dr. Henk van den Boogard for his useful input during the research and in the preparation of this manuscript. Kajima Technical Research Institute, Japan, is acknowledged for permission to use the materials pertaining to the case study in Osaka Bay. Finally, the authors thank the reviewers for their constructive and critical comments, which have much improved the logic and readability of the paper.

References

- Brown JD, Heuvelink GBM (2007) The Data Uncertainty Engine (DUE): a software tool for assessing and simulating uncertain environmental variables. *Computers & Geosciences* 33(2):172–190
- El Serafy GY (2003) Comparison of EKF and EnKF in SOBEK River. Case study Maxau–IJssel. Delft Hydraulics Research Report X0281.30, pp 56
- El Serafy GY, Gerritsen H (2006) Redesign and calibration of a 3D flow model of Osaka Bay in preparation of data assimilation application. Delft Hydraulics Research Report X0333, pp 28
- El Serafy GY, Mynett AE (2004) Comparison of EKF and EnKF in SOBEK RIVER: case study Maxau–IJssel. In: Liong SY, Phoon KK, Babovic V (eds) *Proceedings of the 6th International Conference on Hydroinformatics*. World Scientific, Singapore, pp 1205–1212, ISBN 981-238-787-0
- El Serafy GY, Gerritsen H, Mynett AE, Tanaka M (2005) Improvement of stratified flow forecasts in the Osaka Bay using the steady state Kalman filter. In: Jun B-H, Lee S-I, Seo IW, Choi GW (eds) *Proceedings of the 31st IAHR Congress*, Seoul, pp 795–804
- El Serafy GY, Gerritsen H, Mynett AE, Tanaka M (2006) The use of Kalman filters in improving stratified flow forecasts for engineering purposes. In: Gourbesville P, Cunge JA, Guinot V, Liong S-Y (eds) *Proceedings of the 7th International Conference on Hydroinformatics*, 4. Research Publishing, Chennai, India, pp 2415–2422, ISBN 81-903170-2-4
- Evensen G (1994) Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J Geophys Res* 97(17):905–924
- Evensen G (1997) Advanced data assimilation for strongly non-linear dynamics. *Mon Weather Rev* 125(6):1342–1345
- Isard M, Blake A (1998) CONDENSATION—conditional density propagation for visual tracking. *Int J Comput Vis* 29(1):5–28
- Jazwinsky AH (1970) *Stochastic processes and filtering theory*. Academic, New York
- Kalman RE (1960) A new approach to linear filtering and prediction problems. *J Basic Eng* 82:35–45
- Kalman RE, Bucy RS (1961) New results in linear filtering and prediction theory. *Basic Engineering* 83D:95–108
- Lesser GR, Roelvink JA, Van Kester JATM, Stelling GS (2004) Development and validation of a three-dimensional morphological model. *Coast Eng* 51:883–915
- Morf M, Sidhu GS, Kailath T (1974) Some new algorithms for recursive estimation in constant Linear Discrete-time Systems. *IEEE Trans Automatic Control* AC 19(4):315–323
- Osaka Bay Regional Offshore Environmental Improvement Center (2002) Report of preliminary environmental impact assessment on construction of “New Island” (in Japanese), No. 3
- Tanaka M, El Serafy G, Gerritsen H, Adachi T (2005) Development of a new real time flow forecast model in estuary based on Ensemble Kalman Filter (in Japanese), *Annual Journal of Coastal Engineering*, JSCE, Vol.52, pp 321–325
- Tanaka M, Inagaki S, Adachi T (2006) A current and turbidity forecasting system for the environmental control of construction activities in a stratified sea. In: Gourbesville P, Cunge JA, Guinot V, Liong S-Y (eds) *Proceedings of the 7th International Conference on Hydroinformatics*, 3. Research Publishing, Chennai, India, pp 1577–1584, ISBN 81-903170-2-4
- Van Velzen N (2006) COSTA, A problem solving environment for data assimilation. In: *Proc Computational Methods in Water Resources*, 18–22 June 2006, Copenhagen, Denmark
- Weerts AH, El Serafy GY (2006) Particle Filtering and Ensemble Kalman Filtering for runoff nowcasting using conceptual rainfall runoff models. *Water Resources Research* 42(9):W09403 DOI 10.1029/2005WR004093
- Werner M, Heynert K (2006) Open model integration—a review of practical examples in operational flood forecasting. In: Gourbesville P, Cunge JA, Guinot V, Liong S-Y (eds) *Proceedings of the 7th International Conference on Hydroinformatics*, 1. Research Publishing, Chennai, India, pp 155–162, ISBN 81-903170-2-4
- Werner MGF, Van Dijk M, Schellekens J (2004) DELFT-FEWS: an open shell flood forecasting system. In: Liong SY, Phoon KK, Babovic V (eds) *Proceedings of the 6th International Conference on Hydroinformatics*. World Scientific, Singapore, pp 1205–1212, ISBN 981-238-787-0