SPECIAL PROJECT FINAL REPORT

Project Title:	PSAS Data Assimilation for the Adriatic Sea using Regional Ocean Modelling System (ROMS)
Computer Project Account:	spcrjane
Start Year - End Year :	2018- 2020
Principal Investigator(s)	Ivica Janekovic
Affiliation/Address:	Rudjer Boskovic Institute, ZIMO, Bijenicka c 54, 1000 Zagreb, Croatia. Current affiliation: The University of Western Australia, Ocean Graduate School, Cnr Fairway and Service Road 4, M470, Crawley, WA 6009, Australia.
Other Researchers (Name/Affiliation):	hr1e - Clea Lumina Denamiel, ZIMO, Rudjer Boskovic Institute, Bijenicka c 54, 1000, Zagreb, Croatia.

Summary of project objectives

Objective of this project was to explore method, usefulness and consequences of optimal 4D-Var ocean state estimation in the Adriatic Sea. Within the project we setup year-long experiment and used all available ocean observations from many different observation platforms (HF radars, Acoustic Doppler Current Profiler - ADCP data, Temperature and Salinity profiles, Surface temperature of the sea based on Satellite measurements, ARGO buoys, Sea level anomaly from Altimeter Satellites) together, and at the same time, with full nonlinear 3D ocean dynamics from solving model equations. Combining both information (observations and model) in a dynamically consistent way was made possible within 4D-Var data assimilation technique available inside the ROMS model.

Summary of problems encountered

This was research project with many unknowns and had steep learning curve. At the beginning of the project I had authentication problem using token, resolved by mid of February (delay in the project). After that I started to setup the modelling system and use ECMWF HPC facility and discovered significant bug https://www.myroms.org/forum/viewtopic.php?f=19&t=4864 within ROMS ocean model. Formulation of the river inflows inside the version of the model was wrong, and as I have 48 river inputs in the Adriatic Sea, my first results were terribly wrong. Later, I managed to fix the problem by changing the source code of the ROMS model and was able to get statistical representation for model error background covariance matrix. Within ROMS model current PSAS 4D-Var architecture is memory allocation scheme inefficient - every MPI tile is allocating memory for all (!) observations used during the assimilation whether they belong to that specific tile or not. That memory allocation is posing burden when having large amount of observation data and fitting it in the memory of each individual node. To that end I had to use more nodes as memory was bottleneck. Idea to use ALADIN atmosphere model within 4D-Var was not possible because of license restriction – i.e. not public domain model, so I had to implement WRF-ARW atmospheric model instead and use it as the forcing component for ROMS model.

Experience with the Special Project framework

As mentioned before, at the beginning I had problem with auth and token which was resolved. Setup of WRF-ARW model at the ECMWF HPC was resolved with great help from the ECMWF staff (Bojan Klasnic). Additional storage was requested and quickly approved to store WRF model results. During the project new promising direction emerged – using 4D-Var in coupled ocean-atmosphere constellation and was pursued further during the last year. As I had to use my quota for simulations of atmosphere model (WRF) along with ROMS ocean 4D-Var model my CPU quota was depleted quickly, and I was out of the SBU during the last year of project. Overall, I was really happy and impressed with the ECMWF computing environment and support.

Summary of results

Within this project we performed a comprehensive study to (a) quantify possibility and usefulness of advanced 4D-Var data assimilation (DA) in a such dynamic environment as found in the Adriatic Sea, (b) limitations due to linear constraints inherent inside Tangent Linear and Adjoint models, (c) quantify how long the improvements achieved by DA corrections persist in the model, and (d) quantify the improvements in reproducing the ocean state, i.e. to reach the best ocean state description in relation to different observing platforms, all used inside the Adriatic Sea. Later in the project we start to explore new directions emerged from the previous results, i.e. 4D-Var DA of the ROMS ocean model coupled with the atmospheric model (WRF-ARW), trying to propagate needed corrections during model integration in time and that way correct lower boundary conditions in the atmosphere model providing feedback to the full WRF dynamics. The last research direction will be continued and pursued in the future.

Results from the project showed direct benefits of using 4D-Var DA in estimating optimal ocean state. For a year-long experiment we assimilated more than 20 million, high-resolution observations, most of which were sea surface temperatures (SST) measured by satellite radiometers (L2P granule swath SST data from MODIS-AQUA platform at the native resolution of 1 km and available from NASA Jet Propulsion Laboratory together with AVHRR L3C mono sensor gridded product only for night scenes available from E.U. Copernicus Marine Service Information, http://marine.copernicus.eu/). In addition we assimilated surface ocean currents measured with high-frequency (WERA HF) radars, *insitu* observations for vertical temperature and salinity profiles coming from various ocean platforms (Argo profiling floats, shipborne towed profiler, autonomous seaglider, Conductivity-Temperature-Depth probes and multiple moorings with vertical current profiles from moored Acoustic Doppler Current Profilers).

The ROMS ocean model numerical grid was set to horizontal resolution of 2 km, with 20 vertically distributed “s” layers as compromise to CPU demand and still realistically represent the Adriatic Sea basin. In order to avoid numerical instabilities, the raw bathymetry was iteratively smoothed using linear programming technique. A nonlinear no-assimilation was integrated for 8 years (2008-2015) and results were used to compute a statistical representation of monthly model background error covariance for initial, boundary and forcing fields (following Weaver and Courtier, 2001). Isotropic decorrelation error covariance scales for initial and boundary fields, used during the normalization computations, were set to 30 km for horizontal and 20 m in vertical direction, while atmospheric forcing used 100 km in horizontal. All scales were based on the precomputed statistical values from the long 8-year run.

Atmospheric forcing at the ocean surface has been taken from the hydrostatic version of the numerical weather prediction model ALADIN/HR (Aire Limitée Adaptation dynamique Développement InterNational), operationally run by the Croatian Meteorological and Hydrological Service. The model resolution was 8 km in horizontal, with 37 vertical sigma layers in vertical. The winds were further dynamically downscaled to horizontal resolution of 2 km. All meteorological variables were provided with a 3 h time interval. The meteorological model was integrated 4 times a day, initial conditions were computed using meso-scale data assimilation (3D-Var for upper-air fields and optimal interpolation for screen level parameters). Lateral boundary conditions for ALADIN model were obtained from operational global forecast runs of the IFS (Integrated Forecast System) model in the ECMWF (European Centre for Medium-Range Forecasts) where global analysis was done using 4D-Var assimilation technique. Bulk parameterization has been used for transferring surface variables into the ocean ROMS model forcing.

During year-long experiment we used 4D-Var DA PSAS approach, splitting the one-year simulation into 91 four-day assimilation cycles, each starting from the previous cycle using saved initial conditions. This window cycling was needed to insure validity of Tangent Linear model assumption within the 4-day window. The window length of 4-days was determined using multiple Tangent

Linear model simulations propagating perturbations forward in time (the same as in the DA) and computing appropriate spatial correlations (in time) with fully nonlinear model solutions propagating the exact same perturbations. When those correlation values begin to decay, linear assumption is no longer valid, and the timing was used in setting the assimilation window length. Moreover, high nonlinearity usually occurring inside vertical turbulent scheme was not linearized but instead saved during the first nonlinear background model integration (as spatial and temporal variable vertical viscosity and diffusivity coefficients) and then later used inside Tangent Linear and Adjoint models within cycle.

To compare the performance of the DA analysis to that of the "free" nonlinear non-assimilative model, three experiments were undertaken: (1) free non-assimilative nonlinear simulation – from here on referred to as baseline simulation; (2) free non-assimilative nonlinear simulation initiated by the previous 4D-Var data assimilation 4-day cycle – from here on referred to as background simulation, and (3) fully assimilative nonlinear simulation that used all available observations during the 4-days assimilation cycle – from here on referred to as analysis simulation.

The assimilation significantly improved the modelling system performance, especially in SST, with time average RMSE equalling 0.9, 0.7 and 0.5°C and BIAS equalling 0.4, 0.15, 0.01°C for the baseline, background and analysis simulations, respectively. These reductions were mostly achieved during the wintertime outbreaks of a cold and dry bora wind, caused by an underestimation of heat and momentum fluxes by the ALADIN atmospheric model.
For the assimilated vertical temperature and salinity measurements, which were mostly carried out in the shallow northern Adriatic, salinity RMSE has been reduced by 50%. The RMSE for the baseline, background and analysis simulations was 0.75, 0.81 and 0.39, respectively. However, there was no reduction in temperature RMSE (1.2°C, 1.5°C, 1.4°C for the baseline, background and analysis simulations, respectively). This was presumably a consequence of the domain-wide SST data that dominated in the cost function minimization, whilst assimilated vertical profile measurements were exclusively made in the northern shallow Adriatic (less than 50 m depth). Yet, validation of simulations on the independent vertical temperature measurements from 85 stations in the deeper regions of the eastern Adriatic (observation depths up to 270 m) results in a reduction of the BIAS for temperature that is almost 3 times lower than in the baseline simulation, while the respective RMSE has been reduced by about 20%. However, in the case of salinity, analysis results didn't show any improvement over the baseline solution.
During the minimization of the cost function we found that the values vary with different assimilation cycles, falling in range between 30% and 70%. SST and HF radar surface currents apparently contributed the most to the cost function, both in the background and analysis simulations. However, when scaling their contributions to the cost function taking into account the total number of used observations, it appears that vertical temperature and salinity observations for example (when present in specific assimilation cycle) contributed between 37% and 68% of the cost function, while SST contributed only about 17%. This finding may allow (a) future reductions in the use of vast numbers of SST observations as their contribution to the innovation vector is relatively modest, and (b) point towards the most important observation platform.
Mechanism to get ocean model closer to the observations was predominantly due to better estimates of surface atmosphere-ocean boundary conditions, where the largest correction was found for the wind stress. It showed as the most effective mechanism for minimizing the difference between the model and observed HF radar surface circulation. During the winter/spring season, corrections were driven by the bora cold wind events. The bora-driven wind stress is responsible for strong downwind ocean currents (up to 1 m/s in the coastal Senj Jet), precisely where the 4D-Var DA wind stress corrections are important. The consequences were far going, the Senj Jet, together with the Trieste Jet, drives the bipole cyclone-anticyclone circulation in the whole northern Adriatic, delimiting the area where, for example, dense water is formed during wintertime. Thus, the introduction of a proper wind stress component is crucial for a more accurate reproduction of the northern Adriatic dynamics. These corrections reflect a handicap of the bora dynamics captured in the ALADIN/HR model, which, using a horizontal resolution of 8 km for heat fluxes and 2 km for downscaled winds, cannot resolve bora-

driven jet changes, vortices and sub-mesoscale features that are known to appear during the severe bora wind episodes.
The finding during the last year steered the project in the direction of propagating those ocean-atmosphere boundary corrections back into the atmospheric model in fully coupled way to further improve full atmosphere dynamics, again used as ocean forcing. Our hope was that propagated corrections would converge - minimize during coupled simulations while getting more balanced ocean-atmosphere communication and at the same time closer to the ocean observations driving 4D-Var.

List of publications/reports from the project with complete references

Janeković, I., Mihanović, H., Vilibić, I., Grčić, B., Ivatek-Šahdan, S., Tudor, M., & Djakovac, T. (2020). Using multi-platform 4D-Var data assimilation to improve modelling of Adriatic Sea dynamics. Ocean Modelling, 146, [101538]. https://doi.org/10.1016/j.ocemod.2019.101538.

In review at Ocean Modelling: Janekovic, I., Rayson, M.D., Jones, N.L., Watson, P., Pattiaratch C. 2021. 4D-Var data assimilation using satellite sea surface temperature to improve the tidally-driven interior ocean dynamics in the Indo-Australian Basin.

Future plans

In the future I plan to continue to use and further develop 4D-Var DA ocean systems with focus on improving their usage on dedicated HPC architecture. Results from this project already generated huge amount of data that will need time to analyse and would result in further publications later in time.
Personally, I permanently moved (from Croatia) to Australia and continuing to use similar HPC Cray system at ECMWF here at the Pawsey supercomputer centre (Perth, Australia). Results along the lines of the ECMWF research project will focus on better optimization of 4D-Var algorithms on the HPC environment (memory allocation, parallel writing etc). Unfortunately, as far as I know, Australia is not ECMWF research partner, and I can not apply for similar research project in the future at the ECMWF centre.