REQUEST FOR A SPECIAL PROJECT 2021–2023

MEMBER STATE: …………Belgium……………………………………………...…

Principal Investigator[1]: ……Luc Vandenbulcke ……………………..……

Affiliation: ……Université de Liège, Belgium………...

Address: ……Bat B5c, Allée du 6 août 19, 4000 Liège, Belgium…………..……

…………………………………………………………………………..……

…………………………………………………………………………..……

Other researchers: … Marilaure Grégoire, Arthur Capet………………………….………

…………………………………………………………………………..……

Project Title: …Black Sea ENSEmble forecasting system………………………
…Acronym : ……BSENSE……………………………

If this is a continuation of an existing project, please state the computer project account assigned previously.	**SP** ____no_____________	
Starting year: (A project can have a duration of up to 3 years, agreed at the beginning of the project.)	2021	
Would you accept support for 1 year only, if necessary?	YES ☐	NO ☒

Computer resources required for 2021-2023: (To make changes to an existing project please submit an amended version of the original form.)		**2021**	**2022**	**2023**
High Performance Computing Facility	(SBU)	650.000	2.400.000	19.200.000
Accumulated data storage (total archive volume)[2]	(GB)	2.000	5.000	46.000

Continue overleaf

[1] The Principal Investigator will act as contact person for this Special Project and, in particular, will be asked to register the project, provide annual progress reports of the project's activities, etc.

[2] These figures refer to data archived in ECFS and MARS. If e.g. you archive x GB in year one and y GB in year two and don't delete anything you need to request x + y GB for the second project year etc.

Principal Investigator: ……Luc Vandenbulcke…………………………………………

Project Title: …Black Sea ENSEmble forecasting system……………………..…

Extended abstract

Scientific context

The MAST lab at the Liege University has a long expertise in the development of ocean numerical models simulating the physics coupled to ecosystem dynamics. In the frame of European (e.g. FP4 INTAS, VENTIL, ARAL-KUM, FP6 Sesame, FP7 PERSEUS, HYPOX, Horizon 2020 ForCoast, JPI Ocean and climate CE2COAST, ESA EO4SIBS project) and national (e.g; BRAIN FaCE-IT, FNRS BENTHOX, Albatros) research projects these models have been applied in different areas mainly the Black Sea, the Southern Bight of the North Sea and the Ligurian Sea. Currently the lab is a production unit of the Copernicus Marine Environment and Monitoring Service (CMEMS) Black Sea monitoring and forecasting center (BS-MFC), providing the 28-year reanalysis and the daily analyse and forecast products. The BS-MFC system underwent various upgrades during the last years, and is now well embedded in the CMEMS operational setup, which includes user feed-back loop and automatic and manual product quality assessment procedures.

In the context of these projects, and with the expected evolution of CMEMS, our aim is to transform the forecasting system by evolving from a deterministic to a probabilistic forecasting sytem using an ensemble configuration. However, before switching the operational system towards an ensemble based production system, different ensemble settings need to be tested, and this will represent a significant computational load that cannot be provided by the Liege University alone.

To that aim the BSENSE project involves the following steps :

(1) **to convert the deterministic NEMO implementation into a stochastic one,** based on the work by Brankart et al (2013, 2015), Bessières et al (2017) and the recent work by Vervatis et al, on the ECMWF computing infrastructure (SCRUM-2 project). Using the stochastic modules and/or different perturbation methods, at least the following will be perturbed: initial conditions, scalar model parameters in the turbulent closure scheme (e.g. the minimum turbulent kinetic energy), the light penetration scheme used in the evolution equation for temperature, the bulk formulae used to derive surface drag and heat fluxes from atmospheric forcings. The surface forcing fields will be obtained from ECMWF-EPS, or the reduced-resolution uncertainty estimates provided with the ERA-5 dataset (depending upon availability to us).

The uncertainty obtained from the ensemble will be quantified, and the ensemble reliability and consistency will be studied (similar to SCRUM-1 and SCRUM-2 projects).

The proposed work will be a significant improvement compared to the Black Sea ensemble model described in Vandenbulcke et al (2015), which remains to our knowledge the only ensemble model of the Black Sea (also including an Ensemble Kalman Filter, EnKF).

(2) **To implement the assimilation of satellite SST and in situ temperature and salinity observations, through an EnKF Approach**.. The analysis step of the EnKF will be designed to update a state vector comprising (at least) the model temperature and the perturbed parameters cited above. This EnKF filter is implemented in the Ocean Assimilation Kit (OAK) developed at the Liege University (see the "tools" section).

(3) **To investigate the imprint on the ecological and biogeochemical dynamics of uncertainty of the simulated physical processes (obtained in 2).** The NEMO physical model will be coupled with the BiogeocheMical Model for Hypoxic and Benthic Influenced areas (BAMHBI) that simulates the biogeochemical cycling of carbon, nitrogen, oxygen and phosphorus through the lower trophic levels of the ecosystem up to zooplankton. BAMHBI will be forced by the ensemble-statistics obtained from the ensemble of NEMO models in order to understand and quantify the impact on key aspects of the Black Sea biogeochemical dynamics (e.g. primary production, position of biogeochemical gradients) of uncertainty associated with physical variables. For example, preliminary tests have shown that (as expected) the numerical value of the turbulence module parameters (and hence the vertical diffusion) have a large impact on primary production; as do light penetration and SST estimation. The analysis of physical-biogeochemical coupling and in particular, the assessment of the impact of modifications of the physics on biogeochemical dynamics

are expected to help to better constrain the plausible limits of variability of the physics and associated parameterizations and formulations.

(4) **To identify and alleviate spurious dynamics resulting from state-correction assimilation protocols,** in particular those affecting biogeochemical processes. The impact of (physical) state vector correction by the EnKF, and potential transitory model reactions (spurious vertical velocities is one example to be investigated, but not the only one), on the biogeochemistry, will be analyzed as well. This is a pressing research topic, but in the frame of BSENSE, it will not require supplementary model runs after (3) is completed.

(5) **To assimilate biogeochemical variables in the coupled model**, based on a parameter-estimation approach as presented in Garnier et al. (2016) and Santana-Falcón et al (2020), where parameters may be allowed to vary spatially.

Regarding all topics (1 to 5), the model grid that is currently implemented in the BS-MFC will be used, discarding the resolution increase foreseen for the next versions of BS-MFC products (late 2020 or early 2021), in order to limit the computing resource requirements.

Regarding topics 3 to 5, in this project, the current stable version of BAMHBI will be used without further modifications.

Scientific plan

At the end of the 3-year period, a 25-member ensemble of NEMO-BAMHBI-OAK models of the Black Sea will have run over a 3-year period (e.g. 2019-2021), while assimilating available *in situ* and remote sensed data: temperature and salinity (for the physical model), and potentially nutrients, oxygen and optical reflectances or chlorophyll (for the biogeochemical model). Selected model variables and model parameters will be updated together during the analysis (i.e. state vector correction and parameter estimation).
To reach this goal, a step-by-step plan has been proposed above: first build an ensemble of physical models, then implement an EnKF using this ensemble, then finally study the effect of the ensemble run on a coupled biogeochemical model and eventually also assimilate biogeochemical data (mainly ocean colour provided by satellite, in particular Sentinel).
The obtained ensemble mean is expected to provide a product with a superior quality compared to a single deterministic simulation while the ensemble spread provides an estimation of the uncertainty associated with the mean. The delivery of a sound estimation of the uncertainty associated to the forecast is a real added value particularly relevant for biogeochemical variables suffering from a lack of data. Very often the lack of uncertainty estimates compromise the use of model products in the decision making process. Higher order moments, percentiles, and other statistics can complete the probabilistic picture.

The rich outputs from the ensemble simulation will allow us to address crucial scientific questions, regarding both the methods and the oceanographic processes. Examples of scientific questions are:

- Is the ensemble consistent (especially in the shallow North-Western Shelf area) ? If not, how should the configuration of the stochastic methods be modified ?

- How does the model formulation, and the physical parameters obtained by the EnKF parameter estimation, impact particular processes such as cold water formation or the disappearance of the Black Sea Cold Intermediate Layer in recent years. This may allow to improve the deterministic physical model currently used in BS-MFC.

- How do the stochastic physics and assimilation of physical variables impact the biogeochemical variables ?
  It is well known that the instantaneous modification of the density field (by assimilation of temperature and salinity) can lead to spurious horizontal currents, and vertical velocities, that can affect the nutrient distribution in the water column. The problem is more severe in the case that sea surface elevation is assimilated. Furthermore, the stochastic equation of state has also been shown to lead to larger vertical currents, at least in areas with large horizontal gradients of temperature and salinity (Brankart, 2013). In the Black Sea, this would be the margins of the Rim Current and of the semi-permanent eddies, the shelf break, the area around the Bosphorus mouth and the river plumes. These enhanced vertical currents are probably realistic, on the contrary of some spurious currents generated by the assimilation of physical variables; yet the effect on biogeochemical tracers needs to be analyzed carefully.

- Is the coupled model improved, if critical parameters (e.g. optical parameters governing the light penetration scheme) are allowed to have spatial and temporal variations?

In the case of assimilating and correcting biogeochemical parameters (step 5), what are the seasonal and spatial patterns of retrieved parameters? Do they correspond to well known dynamics such as seasonality or a North-Western shelf vs. open sea spatial pattern? The identification of such patterns would constitute important information for improving parameterization of the biogeochemical model, even for deterministic runs.

- Does the stochastic context allow a better understanding of physical and biogeochemical processes not directly constrained by the assimilation process such as a better estimation of the impact of river plume physics on pH and on deoxygenation in the North Western Shelf; impact of cross-shelf exchanges on the oxycline shoaling in the deep basin ?

- When a higher-resolution nominal product will be used in the CMEMS BS-MFC, how will it be used together with the lower-resolution ensemble? Using a lower-resolution ensemble (compared to a higher-resolution deterministic forecast) is a common practical way to deal with limited computing resources (both in atmospheric and oceanographic models); it is expected that the results obtained during BSENSE will still be relevant even after the foreseen upgrades.

Technical characteristics

NEMO (Nucleus for a European Model of the Ocean, see Madec et al, 2017) is a widely-used model for research, operational oceanography, seasonal forecasts and climate studies. In the proposed project, NEMO v3.6 will be used. The passive tracer (“my_trc”) module of the OPA ocean model will be used to drive the BAMHBI pelagic and benthic variables. Thus, no coupler is used but the biogeochemical model is run directly inside NEMO. The 28 pelagic tracers are transported by OPA, whereas the 6 benthic tracers interact only with the pelagic tracers, but are not transported. Apart from OPA, no other components of NEMO are used. XIOS-2 is used to manage disk outputs.
NEMO and XIOS are both parallelized using MPI. In the proposed set-up, the Multiple Process Multiple Data (MPMD) paradigm is used, NEMO using 178 processes and the XIOS server using 14. The NEMO computational domain is split into 264 square sub-domains, out of which 86 are discarded because they contain only land.
The Ensemble capabilities of NEMO (Bessières et al, 2017) have been tested already with the NEMO v3.6 Black Sea configuration, as well as the stochastic module (Brankart et al, 2015).

The **OAK** (**Ocean Assimilation Kit**, see Barth et al 2007, 2008, and Vandenbulcke et al 2006, 2010, 2015, 2017, 2018) is an implementation of sequential data assimilation filters (SEEK, EnKF, EnKT, Equal Weight Particle Filter...) in Fortran-90, parallelized with MPI. OAK was first developed in the early 2000’s, then modernized during the SANGOMA FP7 project (2011-2015). It was further improved afterwards, a.o. adding the capability of using non-diagonal observation error covariance matrices.
OAK is currently used in stand-alone mode (also called offline mode), requiring to sequentially run NEMO and OAK, and leading to a large disk i/o overhead as each component reads and writes all the restart files. Current development (to be completed by the end of 2020) aims at including OAK “online” into NEMO v3.6, to reduce this overhead.. However, the online assimilation can only take place if all ensemble members are run together, i.e. if it is possible to submit a job using $N*m$ cores, where N is the ensemble size (30) and m the domain parallelization of NEMO (e.g. 192). Alternatively, the assimilation could still be done offline.

The Biogeochemical Model for Hypoxic and Benthic Influenced areas (**BAMHBI**) is a model of the pelagic and benthic foodwebs, from bacteria to gelatinous carnivores through 34 state variables (both pelagic and benthic). The model simulates oxygen, nitrogen, silicate, phosphate and carbon cycling. In addition, an innovation of this model is that it explicitly represents processes in the anoxic layer. The model also comprises a carbonate module as in Soetaert et al (2007). When coupled to NEMO, all facilities of NEMO (such as parallelization) are automatically available to the tracer model.

Computer resources requested

Using 192 cores, out of which 178 for the model and 14 for the XIOS-2 input-output manager, the NEMO forecasting system of the BS-MFC at 3 km horizontal resolution simulates 1 year in slightly under 10 hours on the Zenobe Tier-1 calculator at Cenaero (Belgium) (https://tier1.cenaero.be/), which possesses similar processors as the ECMWF cluster. The coupled NEMO–BAMHBI system (at the same resolution) takes about 48 hours computing time.

The NEMO BS-MFC config scales very well up to 192 cores (and probably more). Using more cores actually decreases the SBU consumption as the finer sub-domains allow to eliminate more land-only sub-domains. However, the parallelization can be easily changed to match the ECMWF cluster possibilities.

Computing units for 1 year simulation of physics	P * (10*3600) * 192 ≃ 30.000 SBU
Computing units for 1 year coupled simulation (physics+bio) :	P * (48*3600) * 192 ≃ 149.000 SBU

The following working plan is then proposed

Year 1 : 500.000 SBU

4 runs of 1-month NEMO ensemble run with 40 members	400.000 SBU
4 runs of 1-month coupled ensemble run with 5 members	250.000 SBU

Year 2 : 1.500.000 SBU

1 run of 1-year NEMO ensemble run with 40 members, and D.A.	1.200.000 SBU
2 runs of 1-month coupled ensemble run with 40 members, and D.A.	1.200.000 SBU

Year 3 : 12.000.000 SBU

2 runs of 1-month coupled ensemble run with 40 members, and D.A.	1.200.000 SBU
1 run of 3-year coupled ensemble run with 40 members, and D.A.	18.000.000 SBU

Concerning the first year, we chose to perform 4 runs (2 in winter and 2 in summer) using 40 members. The 5 members of the coupled run, would use as physical forcings the results from the physical ensemble central member, and given percentiles (e.g. +/- 1 standard deviation and +/- 2 standard deviations).

This work plan explains why we would not accept support for only 1-year: the resources requested during the 1st year are much smaller than the other 2 years; and each year functions as a preliminary step for the next year.

1 year of daily outputs of the NEMO config takes about 60 GB disk space. The coupled NEMO-BAMHBI outputs consume 370 GB / year. Therefore, outputs generated during the 1st year would consume 1800 GB, during the second year 3640 GB, and during the third year 45.800 GB.
On top of these amounts come the model and its datafiles, but some parts of the results can be deleted after the first and second years, so that in the end, we request 2.0 TB, 5 TB and 46 TB respectively for the 3 successive years.

Bibliography

A. Barth, A. Alvera-Azcárate, J.-M. Beckers, M. Rixen, L. Vandenbulcke, 2007: Multigrid state vector for data assimilation in a two-way nested model of the Ligurian Sea, Journal of Marine Systems, Volume 65, 1–4, Pages 41-59
A. Barth, A. Alvera-Azcárate, and R. H. Weisberg, 2008: Assimilation of high-frequency radar currents in a nested model of the West Florida Shelf. Journal of Geophysical Research, 113:C08033
Bessières, L., Lerous S., Brankart JM, Molines JM, Moine MP, Bouttier PA, Penduff T., Terray L., Barnier B., Sérazin G., 2017: Development of a probabilistic ocean modelling system based on NEMO 3.5: application at eddying resolution, Geosci. Model Dev., 10, 1091-1106
Brankart, JM., Candille, G., Garnier, F., Calone, C., Melet, A., Bouttier, PA., Brasseur, P, Verron, J., 2015: A generic approach to explicit simulation of uncertainty in the NEMO ocean model, Geosci. Model Dev., 8, 1285-1297
Brankart, 2013
Capet, A., Stanev, E. V., Beckers, J.-M., Murray, J. W., and Grégoire, M., 2016: Decline of the Black Sea oxygen inventory, Biogeosciences, 13, 1287-1297, https://doi.org/10.5194/bg-13-1287-2016
Capet A. Meysman F., Akoumianaki I., Soetaert K., Grégoire M, 2016.: Integrating sediment biogeochemistry into 3D oceanic models: A study of benthic-pelagic coupling in the Black Sea, Ocean Modelling
Capet, Arthur, Evan Mason, Ananda Pascual, and Marilaure Grégoire. 2017. Contribution of Mesoscale Eddies to Black Sea Ventilation. Vol. 19. EGU General Assembly Conference Abstracts. adsabs.harvard.edu
Garnier, F., Brankart, J. M., Brasseur, P., and Cosme, E.: Stochastic parameterizations of biogeochemical uncertainties in a 1/4◦ NEMO/PISCES model for probabilistic comparisons with ocean color data, Journal of Marine Systems, 155, 59–72, 555 https://doi.org/10.1016/j.jmarsys.2015.10.012, 2016.
Grégoire, M., & Soetaert, K. 2011. Carbon, nitrogen, oxygen and sulfide budgets in the Black Sea: A biogeochemical model of the whole water column coupling the oxic and anoxic parts. Ecological Modelling, 15.
Grégoire, M., C. Raick, and K. Soetaert. 2008. "Numerical Modelling of the Central Black Sea Ecosystem Functioning during the Eutrophication Phase." Progress in Oceanography 76 (3): 286–333
F. Lenartz, C. Raick, K. Soetaert, M. Grégoire, 2007: Application of an Ensemble Kalman filter to a 1-D coupled hydrodynamic-ecosystem model of the Ligurian Sea, Journal of Marine Systems, Vol 68, 3–4, pages 327-348
Madec G., Bourdallé-Badie R., Bouttier PA., Bricaud C., Bruciaferri D., Calvert D., Chanut J., Clementi E., Coward A., Delrosso D., Ethé C., Flavoni S., Graham T., Harle J., Iovino D., Lea D., Lévy C., Lovato T., Martin N., Masson S., Mocavero S., Paul J., Rousset C., Storkey D., Storto A., Vancoppenolle M. 2017. NEMO ocean engine (Version v3.6). Notes Du Pôle De Modélisation De L'institut Pierre-simon Laplace (IPSL). Zenodo.
Raick,C., A. Alvera-Azcarate, A. Barth, J.M. Brankart, K. Soetaert, M. Grégoire, 2007: Application of a SEEK filter to a 1D biogeochemical model of the Ligurian Sea: Twin experiments and real in-situ data assimilation, Journal of Marine Systems, Vol 65, 1–4, pages 561-583,
Santana-Falcón, Y., Brasseur, P., Brankart, J. M., and Garnier, F.: Assimilation of chlorophyll data into a stochastic ensemble simulation for the North Atlantic ocean, Ocean Sci. Discuss., https://doi.org/10.5194/os-2020-6, in review, 2020.
Soetaert, K., Middelburg, J., Herman P., Buis K., 2000: On the coupling of benthic and pelagic biogeochemical models, Earth-Science Reviews, vol 51, 1–4, pages 173-201,
Soetaert K., Hofmann A., Middelburg J., Meysman F., Greenwood J., 2007: Reprint of "The effect of biogeochemical processes on pH", Marine Chemistry, Volume 106, Issues 1–2, pages 380-401,
Vandenbulcke L., A. Barth, M. Rixen, A. Alvera-Azcarate, Z. Ben Bouallegue, and J.M. Beckers. Study of the combined effects of data assimilation and grid nesting in ocean models - application to the Gulf of Lions. Ocean Science, Volume 2, 213-222, 2006.
Vandenbulcke L., A. Capet, M. Grégoire, J.-M. Beckers and S. Besiktepe An operational implementation of the GHER model for the Black Sea, with SST and CTD data assimilation. Journal of Operational Oceanography, Volume 3-2, 2010, 47-54
Vandenbulcke L., A. Barth A stochastic operational forecasting system of the Black Sea: technique and validation. Ocean Modelling, Volume 93, 22-39, 2015
Vandenbulcke L., A. Barth, J.-M. Beckers. Correction of inertial oscillations by assimilation of HF radar data in a model of the Ligurian Sea. Ocean Dynamics, Volume 67, 117-135, 2017.
Vandenbulcke L. and A. Barth, Upscaling of regional models into basin-wide models, Ocean Science, special issue CMEMS, 2018