https://doi.org/10.5194/egusphere-2026-3182
Preprint. Discussion started: 3 September 2026

A novel Gauss-Hermite High-Order Sampling Hybrid ensemble filter for computationally efficient data assimilation in geosciences - Part 2: OGSTM-BFM-GHOSH

Simone Spada[1], Anna Teruzzi[1], Stefano Maset[2], Stefano Salon[1], Cosimo Solidoro[1], and Gianpiero Cossarini[1]

[1]National Institute of Oceanography and Applied Geophysics - OGS, 34010 Trieste, Italy
[2]University of Trieste, 34127 Trieste, Italy

Correspondence: Simone Spada (sspada@ogs.it)

Abstract. Data assimilation (DA) methodologies play a crucial role in integrating observational data with numerical models to provide accurate state estimates of geophysical systems. A novel ensemble algorithm, the Gauss-Hermite High-Order Sampling Hybrid (GHOSH) filter, aims to improve DA performances by featuring an approximation order higher than in other ensemble DA filters like SEIK (Singular Evolutive Interpolated Kalman filter) or ETKF (Ensemble Transform Kalman Filter). Building upon the successful application in idealized twin experiments (Part 1), we propose a parallel implementation of the GHOSH filter which has been coupled with a realistic geophysical application: the assimilation of surface satellite chlorophyll in a physical-biogeochemical model of the Mediterranean Sea (OGSTM-BFM).

This application was selected due to its complexity and potential relevance to operational settings, such as those of the Copernicus Marine Service, where the uncertainty of prediction can also be crucial. By assimilating satellite chlorophyll data, GHOSH aims to improve the accuracy of the state estimation, not only for the assimilated variable but also for other non-assimilated variables of interest (nutrients), thereby enhancing our understanding of marine ecosystem dynamics.

The simulation results are validated using both semi-independent (satellite chlorophyll) and independent (nutrient concentrations from an in-situ climatology) observations. Results show that: i) the GHOSH implementation in a realistic three-dimensional application is just as computationally feasible as other ensemble assimilation filters, since the integration of a realistic model is by far more computationally expensive than the assimilation scheme; ii) the GHOSH assimilation algorithm improved the agreement between forecasts and observations without producing unrealistic effects on the non-assimilated variables. Moreover, a sensitivity analysis revealed correlation between higher order and a key nutrient (nitrate) estimation improvement, highlighting the importance of the GHOSH enhanced order of approximation in boosting DA performances on non-assimilated variables.

1 Introduction

The Gauss-Hermite High-Order Sampling Hybrid (GHOSH) filter, introduced in Part 1 (Spada et al., 2024), presents a promising advancement in data assimilation (DA) methodologies, particularly for addressing the challenges posed by non-linearities

https://doi.org/10.5194/egusphere-2026-3182
Preprint. Discussion started: 3 September 2026

and model complexity in geophysical systems. Compared to a traditional second-order method, the GHOSH filter improves the accuracy of the state estimation in systems governed by complex dynamic.

In Part 1, the reliability of the new filter has been proven in a large set of twin experiments based on an idealized model commonly used to test DA methodologies (Lorenz96 Lorenz, 1996). In the present work, we extend the application of the GHOSH filter to the much more complex and realistic marine biogeochemical application (the OGSTM-BFM) currently used in the Copernicus Marine Service (Salon et al., 2019), featuring assimilation of satellite observations.

Biogeochemical data assimilation (DA) is relatively recent both in terms of methods and assimilated variables (Fennel et al., 2019). Surface chlorophyll concentration, which is estimated by satellite ocean colour observations, is the most commonly assimilated biogeochemical variable since its relatively high coverage and its near real-time availability make it suitable also for operational applications (e.g., Song et al., 2016; Tsiaras et al., 2017; Santana-Falcón et al., 2020; Pradhan et al., 2019; Teruzzi et al., 2018b; Ciavatta et al., 2016). The high non-linearity of the biogeochemical processes makes it difficult to properly estimate the cross-correlations between variables, thus degradation of non-assimilated variables is often a problematic issue (see e.g., Santana-Falcón et al., 2020; Ford, 2020; Fowler et al., 2023; Mamnun et al., 2022; Wang et al., 2020).
Hence, given its complexity and relevance in the operational framework, the OGSTM-BFM application serves as a critical testbed for evaluating the GHOSH filter's performance in a real-world scenario.

In this work, the GHOSH skill is assessed qualitatively, looking at the dynamics of chlorophyll, nutrients and their correlations, and quantitatively, in terms of improved agreement with observations without degradation of non-assimilated variables. Moreover, a sensitivity analysis has been designed to highlight the importance of the GHOSH enhanced approximation order by computing its correlation with the root mean square deviation (RMSD) metric. Finally, the feasibility of a real implementation of the GHOSH filter is supported also by an analysis of the computational resources needed for the filter solution with respect to the overall cost of the application.

Section 2 presents the experimental design and introduces the OGSTM-BFM model, the data used for assimilation and performance evaluation, and a summary of the relevant GHOSH key elements. Section 3 shows the results of the experiments, that are discussed in Section 4. Finally, Appendix A contains some specific details of the GHOSH implementation.

2 Material and methods

2.1 The OGSTM-BFM model

The biogeochemical model used in this study is the coupled transport-biogeochemical model OGSTM-BFM, which is the biogeochemical component of the Mediterranean Copernicus Marine Service (Salon et al., 2019). The OGSTM transport model resolves the advection, the diffusion and the sinking terms of the biogeochemical variables (Lazzari et al., 2010).
The biogeochemical BFM model describes the biogeochemical cycles of carbon, nitrogen, phosphorus and silicon through the dissolved inorganic, particulate living organic and non-living organic compartments, and includes nine plankton functional types, as well as dissolved oxygen, pH and the carbonate system (for a total of 51 tracers as state variables). The OGSTM-BFM model has been applied to study the chlorophyll and primary production (Cossarini et al., 2021; Teruzzi et al., 2021, 2018b;

https://doi.org/10.5194/egusphere-2026-3182
Preprint. Discussion started: 3 September 2026

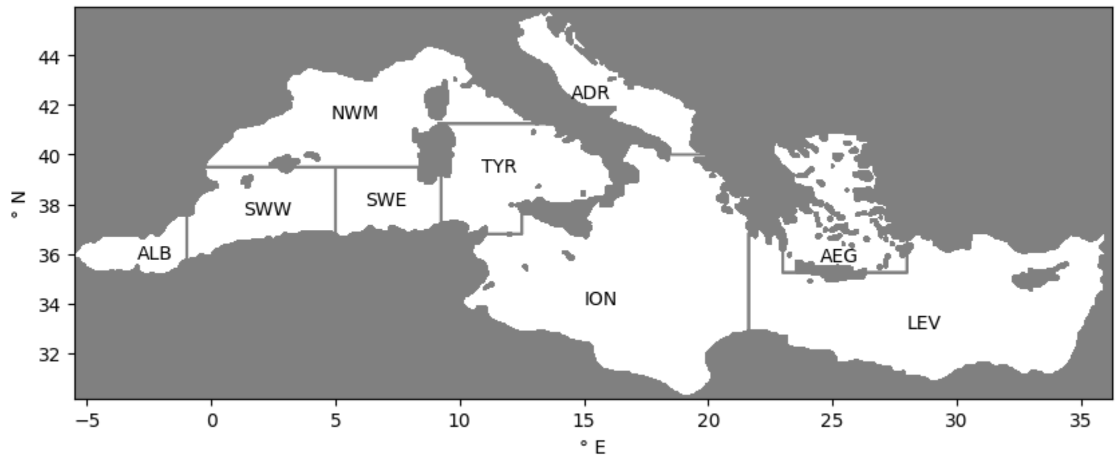


Figure 1. Sub-basins partition of the Mediterranean Sea.

Lazzari et al., 2012), nutrient cycles (Lazzari et al., 2016), alkalinity, carbon cycle and carbon fluxes (Cossarini et al., 2015; Melaku Canu et al., 2015) and oxygen dynamics (Di Biagio et al., 2022).

The OGSTM-BFM is forced by the output of an ocean general circulation model (i.e., NEMO, from the Copernicus Marine Mediterranean system in the present application; Salon et al., 2019) that provides fields of physical quantities (i.e., velocity, temperature, salinity, diffusivity, and solar radiation and wind at the surface from a physical reanalysis) driving the transport processes, the kinetic rates of chemical reactions, and energy and matter exchanges at the air-sea interface.

The present setup (terrestrial and atmospheric inputs, boundary conditions) is described in Salon et al. (2019).

2.2 Observations

The assimilated observations are satellite chlorophyll maps obtained from the CMEMS multisensor product OCEANCOLOUR_MED_CHL_L3_REP_OBSERVATIONS_009_073 (Volpe et al., 2017). The original data at 1 km resolution were checked for spikes, excluding values exceeding by two standard deviations the climatological mean and remapped on the model grid resolution. Moreover, in order to reduce the effect of cloud coverage, 7 daily satellite observations are averaged on a weekly basis.

Additionally, in situ climatological datasets for nitrate and phosphate, to be used for validation, are computed from a dataset that integrates the in situ EMODnet data collections (Buga et al., 2018) and the data sources listed in Lazzari et al. (2016). The climatology is calculated for the 0-30 m layer on 9 sub-basins represented in Fig. 1 (Salon et al., 2019) and for 4 different seasons (January-March; April-June; July-September; October-December).

https://doi.org/10.5194/egusphere-2026-3182
Preprint. Discussion started: 3 September 2026

2.3 The GHOSH filter

The GHOSH filter, described in details in (Spada et al., 2024), is an ensemble DA filter that uses $r+1$ ensemble members to represent the state estimation and its uncertainty. r is the dimension of the uncertainty subspace, i.e., the subspace of the state space spanned by the ensemble members. The GHOSH filter features an enhanced order of approximation (named h) in the s-dimensional subspace (included in the uncertainty subspace) where the approximation error may be maximum. This is realized by partially parameterizing (in the moment tensor μ) the prior uncertainty pdf (i.e., probability density function) and by using a principal component analysis (PCA) to chose the subspace of the best approximation order. Before the PCA computation, a scaling matrix $\mathbf{\Lambda}$ is used to rescale the state variables, i.e. to define which variables are of more interest.

In this implementation, μ describes (up to order h) a Gaussian distribution, and the scaling matrix $\mathbf{\Lambda}$ is diagonal. Being $\mathbf{P}=\mathbf{L}\mathbf{L}^T$ the uncertainty covariance matrix, $\mathbf{\Lambda}$ contains the uncertainty variance (i.e., the diagonal of $\mathbf{P}$), which is computed as the squared Euclidean norms of the rows of $\mathbf{L}$. In the present application, the $\mathbf{\Lambda}$ matrix has been chosen valuing simplicity and generality over skill performance. The implemented scaling matrix is probably a suitable choice for many systems, since it implies that the PCA is computed on the Pearson correlation matrix and that it is not affected by the different amplitudes of the variable standard deviations.

During the forecast and assimilation steps, the uncertainty expressed by the ensemble member's covariance matrix is inflated by a factor $\frac{1}{\rho}$ (ρ is called forgetting factor, see e.g., Pham et al., 1998) and combined with a covariance matrix $\mathbf{Q}$ representing the parametric uncertainty that is non-dependent by the ensemble (as typically happens in hybrid filters, see e.g., Carrassi et al., 2018).

The filter is implemented in its localized version, which takes into account only correlations within a certain maximum horizontal distance (the localization radius).

2.4 Experimental design

Implemented in the OGSTM-BFM biogeochemical model system of the Mediterranean Sea at $\frac{1}{4}^o$ resolution, the local GHOSH filter uses 16 ensemble members (i.e., an uncertainty subspace of dimension $r=15$) to perform a weekly assimilation of surface chlorophyll from satellite ocean color (as in Teruzzi et al., 2018a) during the period 1 January 2013 - 1 January 2014. Consistently, an OGSTM-BFM model run without assimilation (control run) is used as a reference to investigate the GHOSH filter performances in some different settings.

Given the non-Gaussianity of biogeochemical tracer concentrations, the GHOSH filter is applied to the log-transformed biogeochemical variables (Ciavatta et al., 2016; Pradhan et al., 2019; Song et al., 2016; Ford, 2020). To avoid instabilities due to large negative values of logarithms of very small concentrations, a cut-off is applied to state variables, removing values lower than 10^{-4}. This solution dynamically removes unwanted correlations that can typically occur at deeper layers, where some tracers (such as, for example, phytoplankton biomass and chlorophyll) tend to very low and negligible concentrations.

The control run initial condition and the starting ensemble were produced based on a 10-year historical simulation of the OGSTM-BFM model without assimilation. For each year, 31 days centred on the 1st of January were log-transformed after

https://doi.org/10.5194/egusphere-2026-3182
Preprint. Discussion started: 3 September 2026

applying a 10^{-4} cutoff on concentrations, for a total of 310 simulated multivariate states. The mean of the multivariate states was used as control initial condition and mean of the starting ensemble, while a PCA of the ensemble of $\mathbf{\Lambda}$-scaled states provided the 15 (i.e., the dimension of the uncertainty subspace r) principal components used in the GHOSH filter's high-order sampling method (i.e., the GHOSH sampling method described in Spada et al., 2024) to sample the initial ensemble.

The covariance matrix of the parametric uncertainty $\mathbf{Q}$ has been assumed diagonal and invariant in time, with variances inversely proportional to the height of the cells, in order to represent the lower variability when averaging in a bigger cell volume. Given that the resolution of the present model setup is not high enough to adequately resolve coastal processes, assimilating observations in coastal areas could lead to instabilities. Therefore, the $\mathbf{Q}$ uncertainty variance in the coastal areas was reduced by applying a fourth-degree polynomial smoothing function and setting the covariance values at zero in the grid cells closest to the coast. More precisely, the uncertainty starts to decrease where sea depth is 200 metres and reaches zero in areas shallower than 50 metres.

Also the covariance matrix $\mathbf{R}$ of the observation uncertainty has been assumed diagonal and constant in time. Using the same fourth-degree polynomial smoothing applied in the $\mathbf{Q}$ case, the inverse of the observation uncertainty variances is smoothed to zero in the coastal areas (i.e., the $\mathbf{R}$ variances increase in the coastal areas oppositely to $\mathbf{Q}$). Furthermore, the fourth-degree polynomial smoothing was used to weight the local version of $\mathbf{R}^{-1}$ to gently remove assimilation of observations beyond the localization radius of 90 km.

In the assimilation experiment, the GHSOH filter is configured to achieve an approximation order $h = 5$ in a subspace of dimension $s = 3$. A forecast, which includes inflation ($\rho = 0.9$), hybridization with $\mathbf{Q}$ and resampling, is performed with daily frequency. Since a tuning of the parametric uncertainty (represented in $\mathbf{Q}$) and observation uncertainty ($\mathbf{R}$) may be necessary to produce an effective assimilation when system uncertainties are not fully known (e.g., Cossarini et al., 2019; Ford and Barciela, 2017; Janjić et al., 2018; Mattern et al., 2017; Tandeo et al., 2020), the forgetting factor and the uncertainty variances have been chosen (after many short tests not included in this description) in order to provide effective assimilation increments without introducing degradation on biogeochemical variables.

As a sensitivity analysis, 17 additional tests (summarized in tab. 1) are carried out varying the polynomial order h, the forgetting factor ρ and the forecast frequency. In particular, the polynomial order h (with the corresponding number s of principal components approximated with order h) varies as follows:

- $h = 2$, i.e., a second order of approximation in the whole uncertainty subspace ($s = 15$);
- $h = 3$, i.e., third order of approximation in the top 8 principal components ($s = 8$) and second order in the remaining 7 components;
- $h = 5$, i.e., fifth order of approximation in the top 3 principal components ($s = 3$), and second order in the rest (12 components).

The forecast frequency refers to how often the GHOSH filter's forecast phase and the corresponding high-order resampling phase (see part 1, Spada et al., 2024) are performed. Since at every forecast step the forgetting factor is applied and the

https://doi.org/10.5194/egusphere-2026-3182
Preprint. Discussion started: 3 September 2026

Test name	**Order** (h)	**Forgetting factor** (ρ)	**Forecast frequency**
ctrl	Free model		
T_h2_rho1_d	2	1	daily
T_h3_rho1_d	3		
T_h5_rho1_d	5		
T_h2_rho0.9_d	2	0.9	
T_h3_rho0.9_d	3		
T_h5_rho0.9_d	**5**		
T_h2_rho0.8_d	2	0.8	
T_h3_rho0.8_d	3		
T_h5_rho0.8_d	5		
T_h2_rho1_w	2	1	weekly
T_h3_rho1_w	3		
T_h5_rho1_w	5		
T_h2_rho0.7_w	2	0.7	
T_h3_rho0.7_w	3		
T_h5_rho0.7_w	5		
T_h2_rho0.5_w	2	0.5	
T_h3_rho0.5_w	3		
T_h5_rho0.5_w	5		

Table 1. Experimental setup: control run, assimilation experiment (in bold) and sensitivity analysis tests.

covariance matrix $\mathbf{Q}$ of the parametric uncertainty is added to the forecast covariance, a daily forecast differs substantially from a weekly forecast, even if the observation (and assimilation) frequency remains weekly.
Appendix A reports additional technical details specific of the GHSOH filter settings.

The tests are carried out in the GALILEO cluster at the Italian HPC centre CINECA, mounting 2 Intel Broadwell CPU (Intel(R) Xeon(TM) E5-2697 v4 @2.3GHz, 18 cores each) per each node. The parallel Fortran 2003 implementation of the GHOSH algorithm ran on 46 nodes (corresponding to 1656 cores).

https://doi.org/10.5194/egusphere-2026-3182
Preprint. Discussion started: 3 September 2026

3 Results

In the assimilation experiment, the GHOSH performances have been evaluated by considering several metrics: the forecast skill of the assimilated variable (chlorophyll, Section 3.1) and of the non-assimilated variables (nutrients, Section 3.2), and the quality of the system state uncertainty estimation (Section 3.3). The results are discussed in detail through a qualitative and quantitative comparison with the control run (i.e., a free run with no assimilation). The RMSD between simulation (before assimilation) and observations (Section 2.2) is used as a quantitative performance indicator. More in detail, when an observation is available, the square distance between the available observations and the forecast before assimilating those observations, is computed. Then, the observations are assimilated, and the model is integrated in time, up to the next available observations. The procedure is repeated during the whole experiment length, with all the computed square distances concurring to the final RMSD score.

The approximation order impact on DA performances has been studied comparing the scores of 18 tests in different configurations (sensitivity analysis, Section 3.4).

Finally, the GHOSH applicability in a real complex marine biogeochemical application with assimilation of ocean color observation has been assessed by measuring the impact of the assimilation routines on the total computational cost.

3.1 Surface chlorophyll

The surface chlorophyll estimated by the control run and the GHOSH filter has been compared with satellite data from January to December 2013. A detailed analysis of a specific relevant event (winter bloom, end of March - beginning of April 2013) is shown in Fig. 2. The RMSD time series is shown in Fig. 3.

The sequence of ocean color maps (**a** row of Fig. 2) describes a typical winter bloom in the western Mediterranean, with the northernmost part still under the convective phase (i.e., blue hole surrounded by high chlorophyll concentration patches during the first phase (maps of 5 and 19 March, Fig. 2**a**), the start of the surface bloom in the whole area of the northern western Mediterranean (map of 26 March) and its continuation in the following weeks (maps of 2 and 9 April in Fig. 2). Additionally, high chlorophyll levels are present along the Algerian Current from the beginning of March as a result of the active mesoscale dynamics. The high concentration of chlorophyll along the Algerian Current decreases in the second half of March. On the other hand, with the exception of few coastal river-influenced areas, the eastern Mediterranean Sea has a much lower chlorophyll concentration due to the lower intensity of the winter vertical mixing (Fig. 2**a**).

The control run anticipates the start of the bloom in the north-western region by 2-3 weeks, with the peak of chlorophyll concentration being reached in the map of 19 March. The evolution of chlorophyll in the Algerian current is quite well reproduced in the winter period, even if it is somewhat overestimated in the first weeks. Finally, the control run tends to slightly overestimate the chlorophyll concentration in the entire eastern basin in March (Fig. 2**b**). The maps of the pre-assimilation for the March 5 and 12 dates of the GHOSH run (Fig. 2, panels **c1** and **c2**) show that the BFM model tends to predict surface bloom formation in the northwestern region as in the control run (i.e. too early bloom start). However, the assimilation on March 12 corrects this tendency, allowing the model to maintain a low concentration in the following forecast periods until the correct

https://doi.org/10.5194/egusphere-2026-3182
Preprint. Discussion started: 3 September 2026

Test name	Order (h)	Chlorophyll RMSD (mg/m^3)	Phosphate RMSD ($mmol/m^3$)	Nitrate RMSD ($mmol/m^3$)
ctrl	no	0.084	0.036	0.99
T_h2_rho1_d	2	0.068	0.035	0.93
T_h3_rho1_d	3	0.068	0.035	0.85
T_h5_rho1_d	5	0.068	0.035	0.83
T_h2_rho0.9_d	2	0.059	0.033	1.50
T_h3_rho0.9_d	3	0.060	0.033	1.14
T_h5_rho0.9_d	**5**	**0.059**	**0.033**	**0.83**
T_h2_rho0.8_d	2	0.059	0.033	1.96
T_h3_rho0.8_d	3	0.058	0.032	2.15
T_h5_rho0.8_d	5	0.058	0.032	1.76
T_h2_rho1_w	2	0.071	0.035	0.94
T_h3_rho1_w	3	0.071	0.035	0.89
T_h5_rho1_w	5	0.071	0.035	0.91
T_h2_rho0.7_w	2	0.063	0.034	0.94
T_h3_rho0.7_w	3	0.063	0.035	0.93
T_h5_rho0.7_w	5	0.063	0.034	0.91
T_h2_rho0.5_w	2	0.060	0.033	1.35
T_h3_rho0.5_w	3	0.060	0.032	1.02
T_h5_rho0.5_w	5	0.060	0.033	0.91

Table 2. Summary of the root mean square deviation (RMSD) indicators. Results for control run, assimilation experiment (in bold) and sensitivity analysis tests.

onset of bloom on March 26 (Fig. 2, panels **d2** to **d4**). In fact, the assimilation on March 12 has changed the nutrient conditions in the surface layer so that the condition remains unfavorable for rapid phytoplankton growth and no major correction takes place in the following week (e.g. small differences between the map before and after the assimilation on March 19). During the bloom period (after March 26), assimilation corrects some spatial mismatch of the bloom pattern, and essentially the phenology of the event is well captured. While the corrections to chlorophyll dynamics along the Algerian current appear marginal (i.e. difference between before and after maps), the GHOSH corrects the state of phytoplankton and nutrients in the euphotic zone in the eastern Mediterranean as early as the first week of March and no further significant corrections are necessary thereafter. In fact, the before assimilation maps show correctly low concentrations for the entire period.

Time series of RMSD computed between satellite and model before the assimilation are plotted in Fig. 3. Since the values of surface chlorophyll can vary considerably during the year, the relative root mean square deviation (RRMSD, i.e., the RMSD divided by tracer concentration) is also provided, as a normalized proxy of the skill. Figure 3 shows that the GHOSH

https://doi.org/10.5194/egusphere-2026-3182
Preprint. Discussion started: 3 September 2026

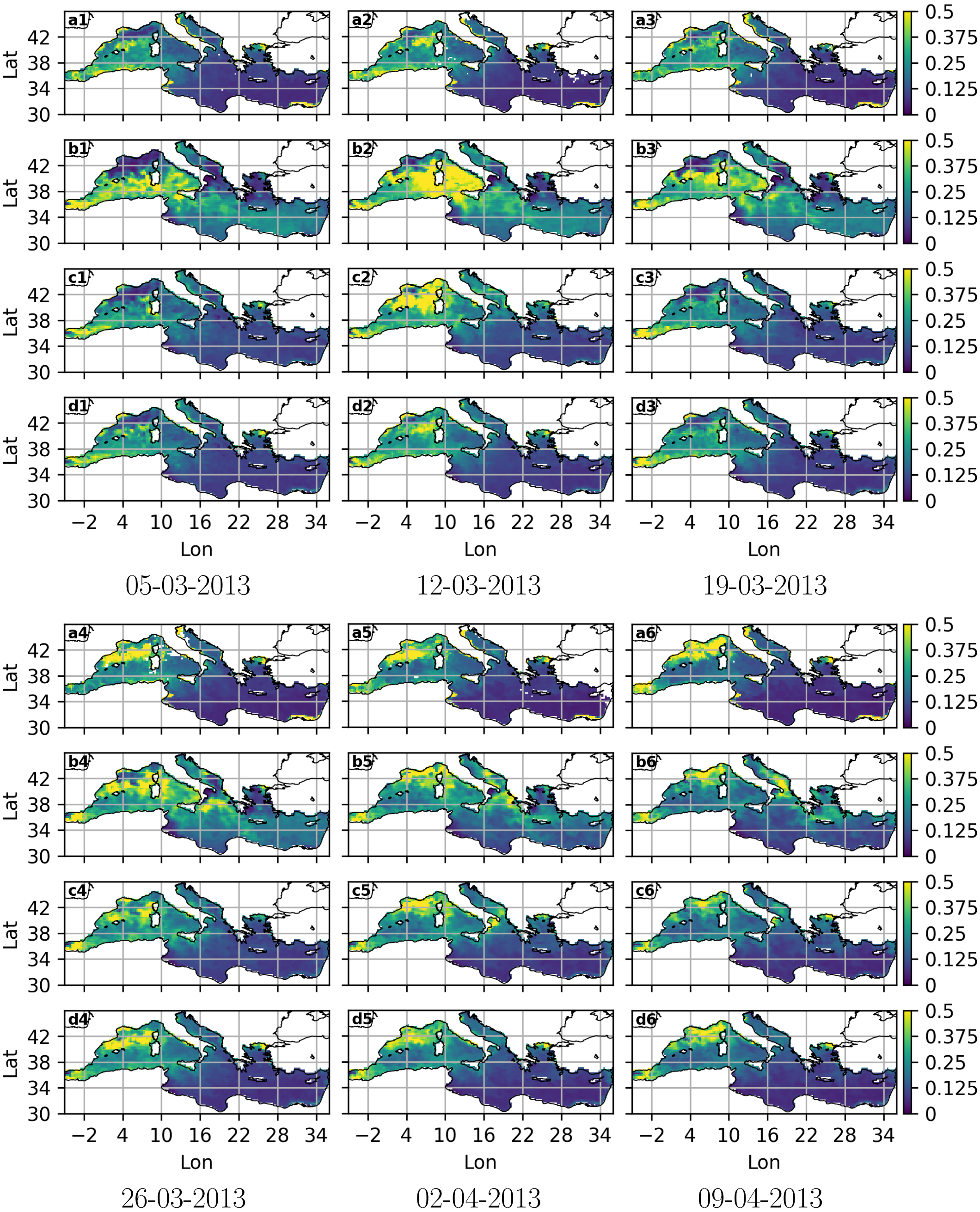


Figure 2. Surface chlorophyll (mg/m^3) comparison: satellite (**a**), control run (**b**) and data assimilation (before and after assimilation, **c** and **d** respectively) during a winter bloom event (19.03.2013 - 09.04.2013).

https://doi.org/10.5194/egusphere-2026-3182
Preprint. Discussion started: 3 September 2026

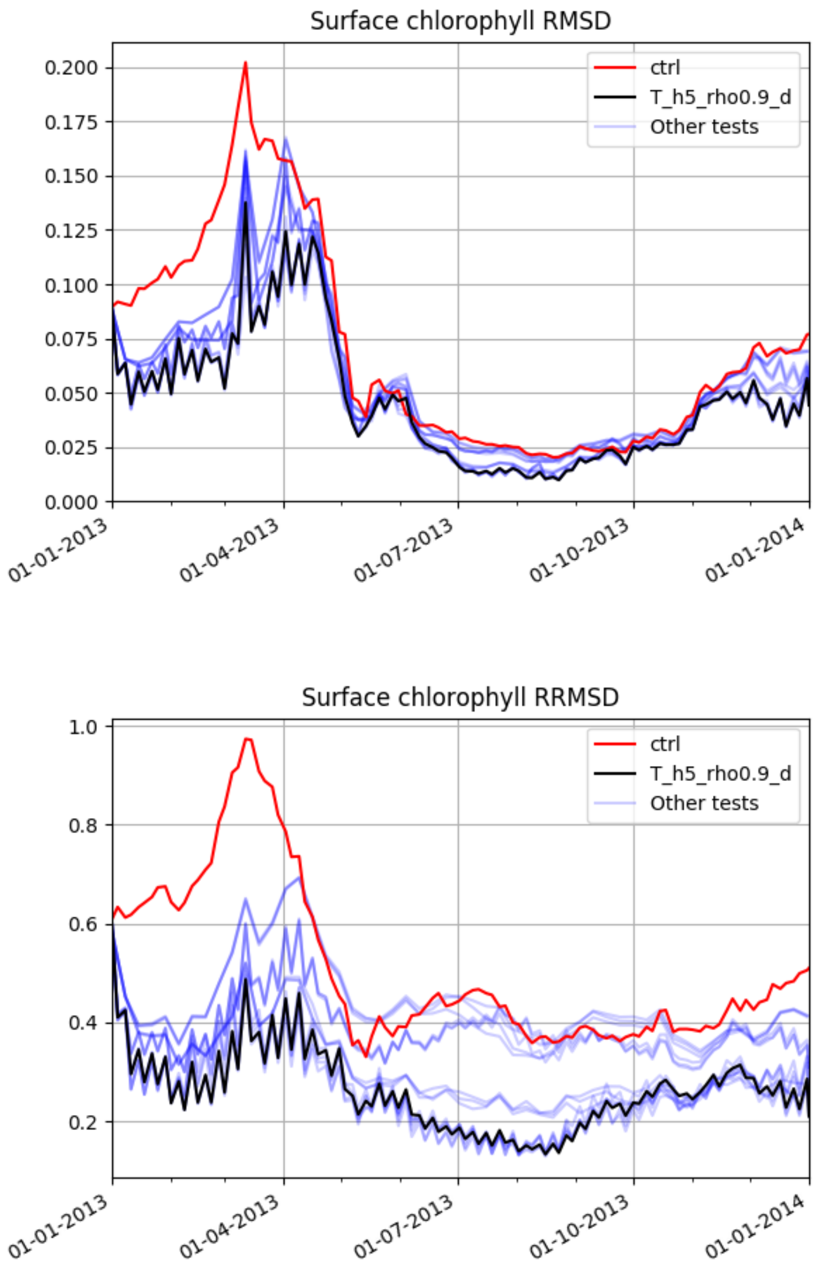


Figure 3. Surface chlorophyll RMSD (mg/m^3) and RRMSD over time.

assimilation produces an improvement in the assimilated variable (i.e., surface chlorophyll) with respect to the control run.
190 The RMSD of the control run follows a seasonal cycle, with maximum values in late winter-early spring, corresponding to the

https://doi.org/10.5194/egusphere-2026-3182
Preprint. Discussion started: 3 September 2026

bloom period, and minimum values during summer when the surface layer is mostly nutrient depleted and chlorophyll is at its lowest level.

The winter-spring RMSD is significantly reduced by the GHOSH filter whereas the RMSD reduction results fairly small during summer, due to a general smaller values of modelled surface chlorophyll in that season. In fact, when the RRMSD is considered, the summer improvement due to assimilation is more detectable.

As a proxy of the overall performance of each 1-year experiment, the RMSD between the simulation and observation, averaged in space and time, is summarized in Tab. 2: the chlorophyll RMSD for *ctrl* (control run) is 0.084, while it is 0.059 for the GHOSH test, resulting in a 30% reduction.

3.2 Nutrients

The effect of DA on non-assimilated variables has been assessed using independent observations to validate phosphate and nitrate (Tab. 2), which can both act as limiting nutrient for phytoplankton growth in the Mediterranean Sea (Lazzari et al., 2016).

The Hovmöller diagrams of phosphate and nitrate in the DA and the control runs show the effects of the assimilation on the two nutrient compartments (Fig. 4). Both nitrate and phosphate show lower concentrations in the 150–400 m layer from April in the simulation with satellite assimilation, with highest impact at the beginning of April. At the surface (0–100 m layer), the assimilation slightly enhances the phosphate concentrations, particularly between April and July, while it reduces the nitrate concentrations. However, these effects do not degrade the typical vertical distribution of nutrients in the Mediterranean Sea, and the seasonal sequence of late-spring/winter mixed conditions and of summer stratification is preserved.

Interestingly, the Hovmöller diagrams of Fig. 4 show that the DA is able to correct the profiles of the nutrients down to 300–400 m by changing the shape of the nutricline (i.e., deepening and creating a gentler shape) during the month of the deep chlorophyll maximum formation (from April to June) and by re-setting the nutricline to a shallower level after summer.

3.3 Uncertainty estimation

In the GHOSH filter, a weighted ensemble is used to estimate the uncertainty of the forecast and analysis states. The standard deviation of the weighted ensemble is used to quantify the uncertainty, whereas a comparison of the standard deviations before and after assimilation is calculated to evaluate the uncertainty reduction produced by the assimilation. Moreover, the ensemble correlation matrix is analyzed to investigate the relationships between variables. The standard deviation and correlation are shown for the assimilated variable and one of the nutrients during a bloom event in winter.

During the winter bloom (March 26), the spatial patterns of chlorophyll and phosphate standard deviations at the surface (second row of Fig. 5) reflect those of their concentration maps (first row of Fig. 5): areas with high surface concentrations in the western Mediterranean Sea are associated with the highest values of the ensemble standard deviation given the high variability of the surface bloom dynamics.

As an effect of the assimilation, the areas with the highest standard deviation for both variables in the western Mediterranean Sea decrease, proving the effectiveness of the surface assimilation in constraining the surface phytoplankton bloom dynamics.

https://doi.org/10.5194/egusphere-2026-3182
Preprint. Discussion started: 3 September 2026

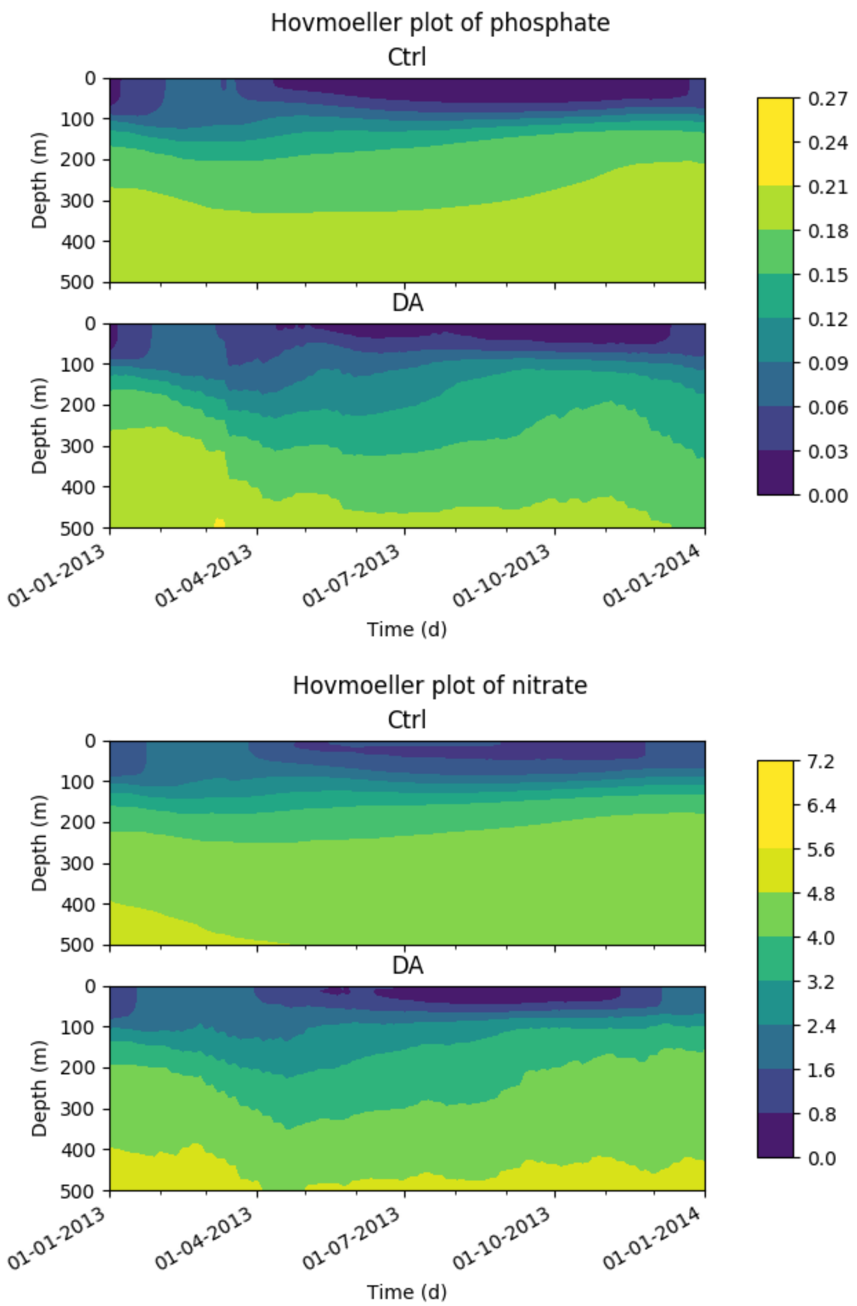


Figure 4. Hovmöller diagrams of nutrients (phosphate and nitrate, $mmol/m^3$): comparison between the control run *ctrl* (up) and the assimilation experiment (bottom)

https://doi.org/10.5194/egusphere-2026-3182
Preprint. Discussion started: 3 September 2026

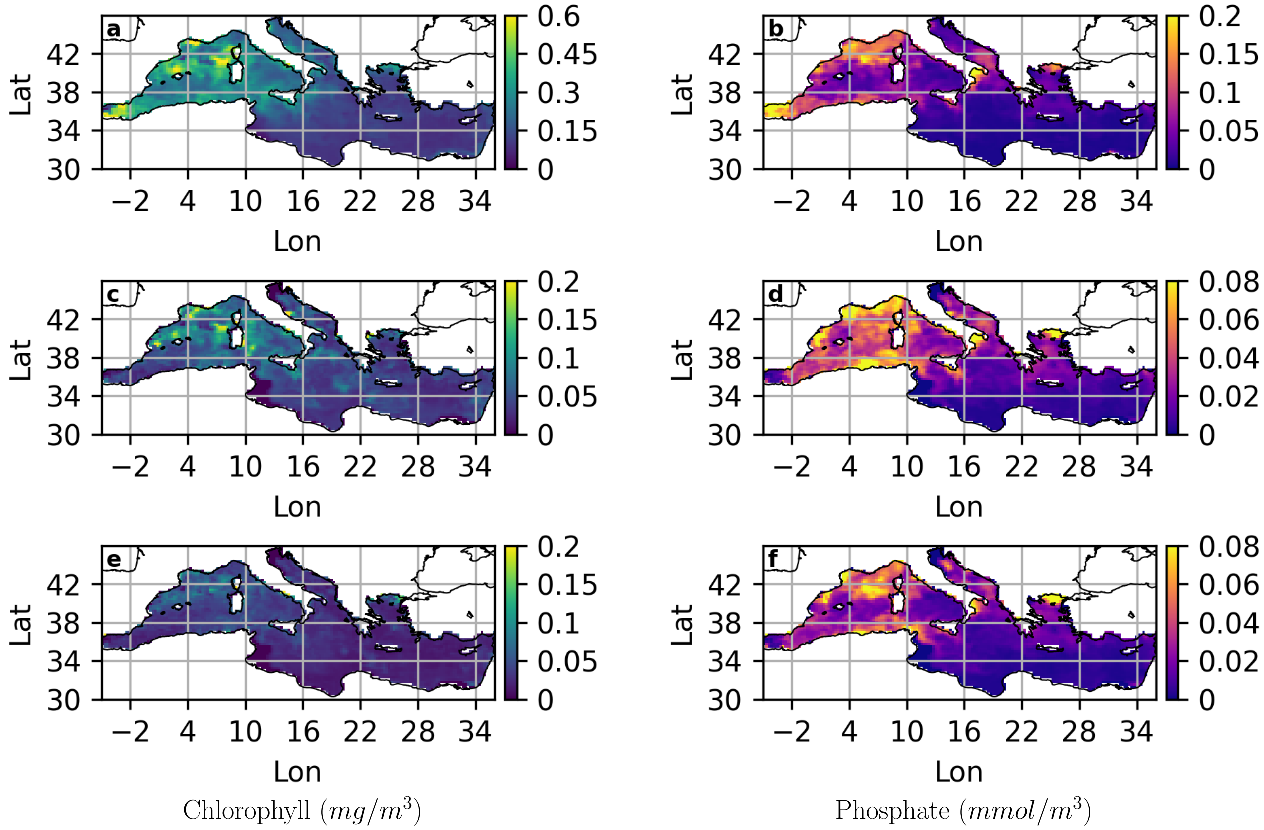


Figure 5. Maps of chlorophyll and phosphate (a and b) and maps of their uncertainty (ensemble standard deviation) before assimilation (c and d) and after assimilation (e and f) for the day 26.03.2013.

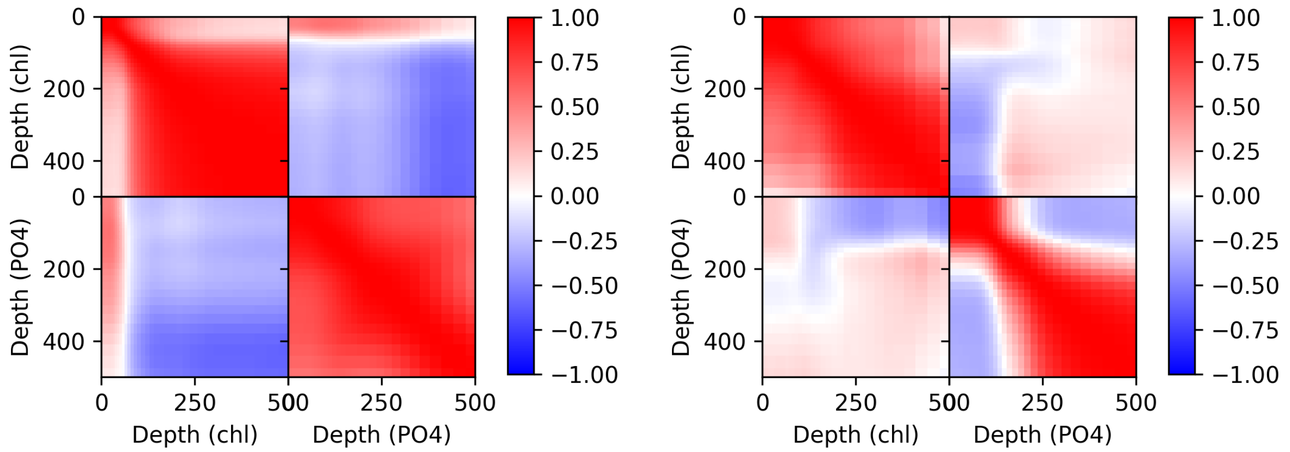
Moreover, a relevant reduction in the standard deviation, in relative terms, is also observed in the less productive eastern Mediterranean Sea.

On average, the assimilation reduces the standard deviation at the surface by 33% and 22% for chlorophyll and phosphate, respectively.

A few spots of high standard deviations remain only in some coastal areas in the western Mediterranean Sea for both variables. The system has been tuned to focus on open sea and the assimilation in shallow waters is less effective due to imposed higher observation uncertainties in the coastal areas (see Section 2.4 for details).

A multivariate correlation matrix (Figure 6) of the north western region shows that the assimilation of satellite chlorophyll influences mostly the chlorophyll in the upper 50 m layer and can impact the phosphate profile to a depth up to 400 m, with

https://doi.org/10.5194/egusphere-2026-3182
Preprint. Discussion started: 3 September 2026



(a) Western Mediterranean

(b) Eastern Mediterranean

Figure 6. Post-assimilation multivariate (chlorophyll and phosphate) correlation during winter (19.03.2013). Horizontally averaged vertical profiles in the western (left) and eastern (right) Mediterranean Sea.

correlation values between the surface chlorophyll and phosphate decreasing with depth. In the more oligotrophic conditions of the Eastern Mediterranean Sea, chlorophyll and phosphate are less correlated than in the western Mediterranean, whereas the impact of surface assimilation on chlorophyll is deeper (i.e., correlations with surface chlorophyll is high down to nearly 400 m).

3.4 Sensitivity analysis

Performance sensitivity to the approximation order h has been evaluated by looking at RMSD scores in the tested configurations. Table 2 shows that the performance of the assimilated variable is substantially unaffected by the approximation order, whilst h demonstrates a relevant impact on a non-assimilated variable, namely the nitrate. In particular, when the order h is 5, nitrate RMSD between GHOSH and independent observations improves up to 45% with respect to order $h = 2$ (scored in the set of tests with $\rho = 0.9$ and daily forecasts). Every $h = 5$ test has a smaller RMSD than every corresponding $h = 2$ and, in general, data clearly show that increasing h implies an improvement in RMSD. In fact, the nitrate's RMSD values, centred and normalized over their standard deviation, present a strong negative correlation with h (Pearson's correlation coefficient is -0.80).

3.5 Computational cost

Each 1-year assimilation test, consisting of 16 ensemble members, needed approximately 4 hours to be computed in the CINECA's GALILEO super computer, running in parallel on 46 computational nodes (1656 cores).

https://doi.org/10.5194/egusphere-2026-3182
Preprint. Discussion started: 3 September 2026

The computational cost of the GHOSH filter itself is marginal compared with the model integration time and the I/O operations, being less than 1% of the total computational time.

4 Discussion

In the realistic 3D high-dimensional parallel implementation with assimilation of real satellite data, the GHOSH filter proved (in Section 3) to be a reliable data assimilation scheme with 30% RMSD reduction for the assimilated variable compared to the control run. Moreover, the GHOSH assimilation did not degrade non-assimilated variables, actually improving their RMSDs with respect to independent observations (Fig. 3 and Fig. 4). Consistently, the GHOSH assimilation results allowed catching some relevant system dynamics (e.g. formation of the winter bloom in the North Western region) and spatial and inter-variable covariances (Fig. 5 and Fig. 6). Further, a sensitivity analysis underlines the role of the high-order sampling in improving performances in a non-assimilated variable (Section 3.4).

The results obtained from coupling the GHOSH filter to the OGSTM-BFM model show a major improvement in the nitrate performance when increasing the approximation order h. However, chlorophyll and phosphate RMSDs are not substantially affected by h. Consistently with the twin experiments results in Part 1 (Spada et al., 2024), this behaviour may depend on the relatively low number of ensemble members (16, $r = 15$), which imply that only a limited number of principal components ($s = 3$) takes advantage of the higher order $h = 5$. Indeed, with the normalization rules chosen in this experiment (i.e., $\mathbf{\Lambda}$ in Section 2.3), the marginal effects of the higher approximation order on chlorophyll and phosphate could be motivated by the fact that they possibly slightly contribute to the first 3 principal components of the state uncertainty. A larger number of ensemble members (i.e., bigger r) would allow a larger s-subspace, probably extending the high-order effects on further state variables. This view is consistent with the generally observed benefits of a larger ensemble in data assimilation based on EnKF (ensemble kalman filter) (see e.g., Edwards et al., 2015; Nerger et al., 2005; Sacher and Bartello, 2008).

The scaling matrix $\mathbf{\Lambda}$ used in the sampling procedure plays an important role in the filter, since it defines which variables are of more interest in the computation of the most relevant components of the state uncertainty. When implementing GHOSH in the realistic OGSTM-BFM application, we used a normalization matrix to represent $\mathbf{\Lambda}$ (Section 2.3) in order to properly scale the different biogeochemical variables. However, this strategy implies that the principal components are composed by variables that have a high correlation with as many other variables as possible, including processes that can be less relevant for an analysis that is more interested in the photic layer (e.g., processes/dynamics occurring in deep layers). Alternative options can be investigated to optimize the assimilation among the model variables. For instance, similarly to scaling strategies applied in sampling methods proposed in literature (Dovera and Della Rossa, 2012), $\mathbf{\Lambda}$ could be designed to focus on the most relevant processes. In particular, $\mathbf{\Lambda}$ can be a sparse matrix with non-zero coordinates corresponding to variables and regions of interest (as the photic layer) based on information derived by a climatology or a multi-annual simulation.

As with $\mathbf{\Lambda}$, a preliminary approach is applied for the definition of the parametric uncertainty covariance matrix $\mathbf{Q}$ (Section 2.3). The diagonal matrix $\mathbf{Q}$ can not be very representative in the relatively complex OGSTM-BFM application. A more advanced parameterization of $\mathbf{Q}$ deserves to be further investigated for future applications, e.g., using a non-diagonal low rank

https://doi.org/10.5194/egusphere-2026-3182
Preprint. Discussion started: 3 September 2026

matrix as in 3D-variational methods (Teruzzi et al., 2014), or applying methods already used in numerical weather prediction (e.g., Gustafsson et al., 2018, and the works cited therein).

However, even in its basic implementation, the adopted hybrid approach (together with the GHOSH's lower need of inflation proved in Spada et al., 2024) prevented the "collapse into the mean" effect (not shown), which can affect ensemble algorithms (e.g., Anderson, 2007). Indeed, the hybridization adopted in each *forecast* phase has the beneficial effect of maintaining a reasonable spread, even in tests without forgetting factor (i.e., $\rho = 1$).

The implemented statistical moments μ (Section 2.3) describe a Gaussian uncertainty pdf (see also Discussion in Spada et al., 2024). However, the log-transformation of the tracers before applying the assimilation filter implies a log-Gaussian uncertainty pdf on concentrations, which is well-suited for non-negative biogechemical variables (see e.g., Ciavatta et al., 2016; Pradhan et al., 2019; Song et al., 2016; Ford, 2020).

On the other hand, the transformation has the side effect of changing also the observation operator (that is represented by the sum of the chlorophyll components of different phytoplankton species), which becomes non-linear. In this context, the resampling after forecast featured by the GHOSH filter is beneficial, since it exploits the high-order sampling's enhanced capability of managing non-linearity (showed in Part 1) not only with the model operator but also with the non-linear observation operator.

From a computational point of view, the negligible (less than 1%) amount of time spent on DA operations in our tests, is compliant with literature results (see e.g., Nerger et al., 2005). Indeed, the main source of computational cost is the integration of the model scaled by the ensemble size. This behaviour is also in line with the GHOSH asymptotic computational complexity that is the same as in EnKF second-order schemes (see Part 1 Spada et al., 2024).

5 Conclusions

GHOSH is a novel DA ensemble filter that was introduced in Part 1, and features an enhanced approximation order. In this work, it has been tested in a 3D parallel implementation of a state-of-the-art Mediterranean physical-biogeochemical model with assimilation of ocean color observations. Testing a GHOSH implementation in a computationally demanding realistic application was a mandatory step to assess the suitability of the novel filter in the geosciences context. Moreover, the challenge of tracking inter-variables covariances in the complex processes characterizing biogeochemical applications represents an interesting benchmark.

Our results show that, from a computational point of view, the cost of the GHOSH filter is comparable to other second-order deterministic filters (e.g., SEIK, ETKF), since the computational cost chiefly depends on the ensemble size that proportionally expands the model integration cost.

Furthermore, GHOSH improved the agreement between the forecast and observations without producing unrealistic effects on the non-assimilated variables. Considering relevant marine ecosystem dynamics, GHOSH filter successfully constrained the uncertainty of the phytoplankton winter bloom events while assimilating ocean color. Moreover, a sensitivity analysis indicates that the use of a higher order has positive impacts on the performance of a non-assimilated variable (e.g., nutrients).

https://doi.org/10.5194/egusphere-2026-3182
Preprint. Discussion started: 3 September 2026

Code availability. A GHOSH python implementation is available from the GitHub page: https://github.com/Sword-Code/PythonDA under the licence GNU GPLv3. The FORTRAN implementation of the GHOSH filter is based on the OGSTM model (available at https://github.com/inogs/ogstm) coupled with the BFM model (available at https://bfm-community.github.io/www.bfm-community.eu/). The exact version of the OGSTM-BFM-GHOSH system used to produce results and plots described in Section 3 is archived on Zenodo (https://doi.org/10.5281/zenodo.12819521).

Appendix A: High-order sampling solutions

Given a certain maximum approximation order h and the dimension s of the subspace in which the h^{th}-order approximation is achieved, the GHOSH filter (during its initialization) needs a solution of a specific non-linear system to encode the moments information (μ) in the GHOSH sampling matrix $\boldsymbol{\Omega}^h$ and in the vector $\boldsymbol{v}$ of the square roots of the ensemble weights (Spada et al., 2024). Most of the time, the non-linear system admits more than one solution, hence different $\boldsymbol{\Omega}^h$ matrices (and corresponding $\boldsymbol{v}$ vectors) are possible. Since $\boldsymbol{\Omega}^h$ and $\boldsymbol{v}$ have a central role in the GHOSH sampling strategy, this appendix reports their values used in the GHOSH configurations explored in this work (Section 2.4).

In the case with $h = 2$ and $s = 15$, the high-order sampling method reduces to the traditional second-order-exact sampling (Pham, 2001). The solution is easily obtained starting from a constant vector $\boldsymbol{v}$ and adding orthonormal columns in order to form an orthogonal matrix.

In the case with $h = 3$ and $s = 8$, a solution with constant weights ($v_m = \frac{1}{4}$ for each $m \in \{1, \ldots, 16\}$) is chosen, namely,

$$\boldsymbol{\Omega}^{h=3} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}^T . \tag{A1}$$

In the case with $h = 5$ and $s = 3$, the chosen solution is

$$\boldsymbol{v} = \Big(\frac{1}{\sqrt{5}} \quad \frac{1}{\sqrt{5}} \quad \frac{1}{2\sqrt{5}} \quad \frac{1}{2\sqrt{6}} \quad \frac{1}{2\sqrt{6}} \quad \frac{1}{2\sqrt{6}} \quad \frac{1}{2\sqrt{6}} \quad \frac{1}{2\sqrt{6}} \quad \frac{1}{2\sqrt{6}} \quad \cdots \\ \cdots \quad \frac{1}{2\sqrt{5}} \quad \frac{1}{2\sqrt{6}} \quad \frac{1}{2\sqrt{6}} \quad \frac{1}{2\sqrt{6}} \quad \frac{1}{2\sqrt{6}} \quad \frac{1}{2\sqrt{6}} \quad \frac{1}{2\sqrt{6}} \Big)^T \tag{A2}$$

and

https://doi.org/10.5194/egusphere-2026-3182
Preprint. Discussion started: 3 September 2026

$$\boldsymbol{\Omega}^{h=5} = \begin{pmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2\sqrt{6}} & \frac{1}{2\sqrt{6}} & \frac{1}{2\sqrt{6}} & \frac{1}{2\sqrt{6}} & \frac{1}{2\sqrt{6}} & \frac{1}{2\sqrt{6}} & \cdots & -\frac{1}{2} & -\frac{1}{2\sqrt{6}} & -\frac{1}{2\sqrt{6}} & -\frac{1}{2\sqrt{6}} & -\frac{1}{2\sqrt{6}} & -\frac{1}{2\sqrt{6}} & -\frac{1}{2\sqrt{6}} \\ 0 & 0 & 0 & \frac{1}{\sqrt{6}} & \frac{1}{2\sqrt{6}} & \frac{1}{2\sqrt{6}} & -\frac{1}{\sqrt{6}} & -\frac{1}{2\sqrt{6}} & -\frac{1}{2\sqrt{6}} & \cdots & 0 & \frac{1}{\sqrt{6}} & \frac{1}{2\sqrt{6}} & \frac{1}{2\sqrt{6}} & -\frac{1}{\sqrt{6}} & -\frac{1}{2\sqrt{6}} & -\frac{1}{2\sqrt{6}} \\ 0 & 0 & 0 & 0 & \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & 0 & \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & \cdots & 0 & 0 & \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & 0 & \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} \end{pmatrix}^{T} . \quad \text{(A3)}$$

Author contributions. GC, S Salon and S Spada designed the study. S Spada developed the algorithms, supervised by SM. S Spada, AT and GC designed the experiments. S Spada implemented the code and performed the simulations collaborating with AT. S Spada wrote the manuscript, with the contribution of AT, GC and CS. All the authors participated in acquiring the funding for the project.

Competing interests. The authors declare no competing interests.

Acknowledgements. We acknowledge the CINECA award under the ISCRA initiative and the HPC-TRES program, for the availability of high-performance computing resources.
This study has been conducted using E.U. Copernicus Marine Service data.
The research reported in this work was supported by OGS and by the SEAMLESS project (https://www.seamlessproject.org/).
This work was partly funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004032 (project SEAMLESS).

https://doi.org/10.5194/egusphere-2026-3182
Preprint. Discussion started: 3 September 2026

References

Anderson, J. L.: An adaptive covariance inflation error correction algorithm for ensemble filters, Tellus A: Dynamic Meteorology and Oceanography, https://doi.org/10.1111/j.1600-0870.2006.00216.x, 2007.

Buga, L., Sarbu, G., Fryberg, L., Magnus, W., Wesslander, K., Gatti, J., Leroy, D., Iona, S., Larsen, M., Koefoed Rømer, J., Østrem, A., Lipizer, M., and Giorgietti, A.: EMODnet Chemistry Eutrophication and Acidity aggregated datasets v2018, https://doi.org/https://doi.org/10.6092/EC8207EF-ED81-4EE5-BF48-E26FF16BF02E, 2018.

Carrassi, A., Bocquet, M., Bertino, L., and Evensen, G.: Data assimilation in the geosciences: An overview of methods, issues, and perspectives, Wires climate change, 9, 2018.

Ciavatta, S., Kay, S., Saux-Picart, S., Butenschön, M., and Allen, J. I.: Decadal reanalysis of biogeochemical indicators and fluxes in the North West European shelf-sea ecosystem, Journal of Geophysical Research: Oceans, 121, 1824–1845, https://doi.org/https://doi.org/10.1002/2015JC011496, 2016.

Cossarini, G., Lazzari, P., and Solidoro, C.: Spatiotemporal variability of alkalinity in the Mediterranean Sea, Biogeosciences, 12(6), 1647–1658, 2015.

Cossarini, G., Mariotti, L., Feudale, L., Mignot, A., Salon, S., Taillandier, V., Teruzzi, A., and D'Ortenzio, F.: Towards operational 3D-Var assimilation of chlorophyll Biogeochemical-Argo float data into a biogeochemical model of the Mediterranean Sea, Ocean Modelling, 133, 112–128, https://doi.org/10.1016/j.ocemod.2018.11.005, 2019.

Cossarini, G., Feudale, L., Teruzzi, A., Bolzon, G., Coidessa, G., Solidoro, C., Di Biagio, V., Amadio, C., Lazzari, P., Brosich, A., and Salon, S.: High-Resolution Reanalysis of the Mediterranean Sea Biogeochemistry (1999–2019), Frontiers in Marine Science, 8, https://doi.org/10.3389/fmars.2021.741486, 2021.

Di Biagio, V., Salon, S., Feudale, L., and Cossarini, G.: Subsurface oxygen maximum in oligotrophic marine ecosystems: mapping the interaction between physical and biogeochemical processes, Biogeosciences, 19, 5553–5574, https://doi.org/10.5194/bg-19-5553-2022, 2022.

Dovera, L. and Della Rossa, E.: Improved initial ensemble generation coupled with ensemble square root filters and inflation to estimate uncertainty, Computational Geosciences, 16, 437–453, https://doi.org/10.1007/s10596-011-9243-5, 2012.

Edwards, C. A., Moore, A. M., Hoteit, I., and Cornuelle, B. D.: Regional Ocean Data Assimilation, Annual Review of Marine Science, 7, 21–42, https://doi.org/10.1146/annurev-marine-010814-015821, pMID: 25103331, 2015.

Fennel, K., Gehlen, M., Brasseur, P., Brown, C. W., Ciavatta, S., Cossarini, G., Crise, A., Edwards, C. A., Ford, D., Friedrichs, M. A. M., Gregoire, M., Jones, E., Kim, H.-C., Lamouroux, J., Murtugudde, R., Perruche, C., , t. G. O. M. E. A., and Team, P. T.: Advancing Marine Biogeochemical and Ecosystem Reanalyses and Forecasts as Tools for Monitoring and Managing Ecosystem Health, Frontiers in Marine Science, 6, https://doi.org/10.3389/fmars.2019.00089, 2019.

Ford, D. and Barciela, R.: Global marine biogeochemical reanalyses assimilating two different sets of merged ocean colour products, Remote Sensing of Environment, 203, https://doi.org/10.1016/j.rse.2017.03.040, 2017.

Ford, D. A.: Assessing the role and consistency of satellite observation products in global physical–biogeochemical ocean reanalysis, Ocean Science, 16, 875–893, https://doi.org/https://doi.org/10.5194/os-16-875-2020, publisher: Copernicus GmbH, 2020.

Fowler, A. M., Skákala, J., and Ford, D.: Validating and improving the uncertainty assumptions for the assimilation of ocean-colour-derived chlorophyll a into a marine biogeochemistry model of the Northwest European Shelf Seas, Quarterly Journal of the Royal Meteorological Society, 149, 300–324, https://doi.org/https://doi.org/10.1002/qj.4408, 2023.

Gustafsson, N., Janjić, T., Schraff, C., Leuenberger, D., Weissmann, M., Reich, H., Brousseau, P., Montmerle, T., Wattrelot, E., Bučánek, A., Mile, M., Hamdi, R., Lindskog, M., Barkmeijer, J., Dahlbom, M., Macpherson, B., Ballard, S., Inverarity, G., Carley, J., Alexander, C., Dowell, D., Liu, S., Ikuta, Y., and Fujita, T.: Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres, Quarterly Journal of the Royal Meteorological Society, 144, 1218–1256, https://doi.org/https://doi.org/10.1002/qj.3179, 2018.

Janjić, T., Bormann, N., Bocquet, M., Carton, J. A., Cohn, S. E., Dance, S. L., Losa, S. N., Nichols, N. K., Potthast, R., Waller, J. A., and Weston, P.: On the representation error in data assimilation, Quarterly Journal of the Royal Meteorological Society, 144, 1257–1278, https://doi.org/https://doi.org/10.1002/qj.3130, 2018.

Lazzari, P., Teruzzi, A., Salon, S., Campagna, S., Calonaci, C., Colella, S., Tonani, M., and Crise, A.: Pre-operational short-term forecasts for the Mediterranean Sea biogeochemistry, Ocean Science, 6, 25–39, 2010.

Lazzari, P., Solidoro, C., Ibello, V., Salon, S., Teruzzi, A., Béranger, K., Colella, S., and Crise, A.: Seasonal and inter-annual variability of plankton chlorophyll and primary production in the Mediterranean Sea: a modelling approach, Biogeosciences, 9, 217–233, 2012.

Lazzari, P., Solidoro, C., Salon, S., and Bolzon, G.: Spatial variability of phosphate and nitrate in the Mediterranean Sea: A modeling approach., Deep Sea Research Part I: Oceanographic Research, 108, 39 – 52, https://doi.org/https://doi.org/10.1016/j.dsr.2015.12.006, 2016.

Lorenz, E. N.: Predictability: A problem partly solved, in: Proc. Seminar on predictability, vol. 1, Reading, 1996.

Mamnun, N., Völker, C., Vrekoussis, M., and Nerger, L.: Uncertainties in ocean biogeochemical simulations: Application of ensemble data assimilation to a one-dimensional model, Frontiers in Marine Science, 9, https://doi.org/10.3389/fmars.2022.984236, 2022.

Mattern, J. P., Edwards, C. A., and Moore, A. M.: Improving Variational Data Assimilation through Background and Observation Error Adjustments, Monthly Weather Review, 146, 485–501, https://doi.org/10.1175/MWR-D-17-0263.1, 2017.

Melaku Canu, D., Ghermandi, A., Nunes, P., Lazzari, P., Cossarini, G., and Solidoro, C.: Estimating the value of carbon sequestration ecosystem services in the Mediterranean Sea: An ecological economics approach, Global Environmental Change, 32, 87–95, 2015.

Nerger, L., Hiller, W., and Schröter, J.: A comparison of error subspace Kalman filters, Tellus A, 57, 715–735, https://doi.org/https://doi.org/10.1111/j.1600-0870.2005.00141.x, 2005.

Pham, D. T.: Stochastic Methods for Sequential Data Assimilation in Strongly Nonlinear Systems, Monthly Weather Review, 129, 1194 – 1207, https://doi.org/10.1175/1520-0493(2001)129<1194:SMFSDA>2.0.CO;2, 2001.

Pham, D. T., Verron, J., and Roubaud, M. C.: A Singular Evolutive Extended Kalman filter for data assimilation in Oceanography, Journal of Marine Systems, 16, 323–340, 1998.

Pradhan, H. K., Völker, C., Losa, S. N., Bracher, A., and Nerger, L.: Assimilation of Global Total Chlorophyll OC-CCI Data and Its Impact on Individual Phytoplankton Fields, Journal of Geophysical Research: Oceans, 124, 470–490, https://doi.org/10.1029/2018JC014329, 2019.

Sacher, W. and Bartello, P.: Sampling Errors in Ensemble Kalman Filtering. Part I: Theory, Monthly Weather Review, 136, 3035 – 3049, https://doi.org/https://doi.org/10.1175/2007MWR2323.1, 2008.

Salon, S., Cossarini, G., Bolzon, G., Feudale, L., Lazzari, P., Teruzzi, A., Solidoro, C., and Crise, A.: Novel metrics based on Biogeochemical Argo data to improve the model uncertainty evaluation of the CMEMS Mediterranean marine ecosystem forecasts, Ocean Science, 15, 997–1022, https://doi.org/https://doi.org/10.5194/os-15-997-2019, publisher: Copernicus GmbH, 2019.

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 Science, 16, 1297–1315, https://doi.org/https://doi.org/10.5194/os-16-1297-2020, publisher: Copernicus GmbH, 2020.

https://doi.org/10.5194/egusphere-2026-3182
Preprint. Discussion started: 3 September 2026

Song, H., Edwards, C. A., Moore, A. M., and Fiechter, J.: Data assimilation in a coupled physical-biogeochemical model of the California current system using an incremental lognormal 4-dimensional variational approach: Part 3—Assimilation in a realistic context using satellite and in situ observations, Ocean Modelling, 106, 159–172, https://doi.org/10.1016/j.ocemod.2016.06.005, 2016.

Spada, S., Teruzzi, A., Maset, S., Salon, S., Solidoro, C., and Cossarini, G.: GHOSH v1.0.0: a novel Gauss-Hermite High-Order Sampling Hybrid ensemble filter for computationally efficient data assimilation in geosciences - Part 1, https://doi.org/10.5194/gmd-2023-170, preprint version, 2024.

Tandeo, P., Ailliot, P., Bocquet, M., Carrassi, A., Miyoshi, T., Pulido, M., and Zhen, Y.: A Review of Innovation-Based Methods to Jointly Estimate Model and Observation Error Covariance Matrices in Ensemble Data Assimilation, Monthly Weather Review, 148, 3973–3994, https://doi.org/10.1175/mwr-d-19-0240.1, 2020.

Teruzzi, A., Dobricic, S., Solidoro, C., and Cossarini, G.: A 3-D variational assimilation scheme in coupled transport-biogeochemical models: Forecast of Mediterranean biogeochemical properties, J. Geophys. Res. Oceans, 119, 200–217, 2014.

Teruzzi, A., Bolzon, G., Salon, S., Lazzari, P., Solidoro, C., and Cossarini, G.: Assimilation of coastal and open sea biogeochemical data to improve phytoplankton simulation in the Mediterranean Sea, Ocean Modelling, 132, 46 – 60, https://doi.org/https://doi.org/10.1016/j.ocemod.2018.09.007, 2018a.

Teruzzi, A., Bolzon, G., Salon, S., Lazzari, P., Solidoro, C., and Cossarini, G.: Assimilation of coastal and open sea biogeochemical data to improve phytoplankton simulation in the Mediterranean Sea, Ocean Modelling, 132, 46–60, https://doi.org/10.1016/j.ocemod.2018.09.007, 2018b.

Teruzzi, A., Bolzon, G., Feudale, L., and Cossarini, G.: Deep chlorophyll maximum and nutricline in the Mediterranean Sea: emerging properties from a multi-platform assimilated biogeochemical model experiment, Biogeosciences Discussions, 2021, 1–29, https://doi.org/10.5194/bg-2021-97, 2021.

Tsiaras, K., Hoteit, I., Kalaroni, S., Petihakis, G., and Triantafyllou, G.: A hybrid ensemble-OI Kalman filter for efficient data assimilation into a 3-D biogeochemical model of the Mediterranean, Ocean Dynamics, 67, 673–690, https://doi.org/10.1007/s10236-017-1050-7, 2017.

Volpe, G., Pitarch, J., Colella, S., Brando, V., Forneris, V., Bracaglia, M., and Benincasa, M.: OCEAN COLOUR PRODUCTION CENTRE, QUality Information Document of Ocean Colour Mediterranean and Black Sea Observation Product v1.5. Technical Report., 2017.

Wang, B., Fennel, K., Yu, L., and Gordon, C.: Assessing the value of biogeochemical Argo profiles versus ocean color observations for biogeochemical model optimization in the Gulf of Mexico, Biogeosciences, 17, 4059–4074, https://doi.org/10.5194/bg-17-4059-2020, 2020.