Nansen Environmental and Remote Sensing Center

A non-profit environmental research center affiliated with the University of Bergen

Thormølensg. 47
N-5006 Bergen, Norway
Tel: +47 55 20 58 00
Fax: +47 55 20 58 01

NERSC technical report 328

Report on error quantified hindcast with data assimilation (GreenSeas)

by

E. Simon [1], A. Samuelsen [1], L. Bertino [1]

1. Mohn-Sverdrup Center (MSC) / NERSC, Bergen.

This work is a deliverable to the EU FP 7th GreenSeas project

November 1, 2012

Nansen Environmental and Remote Sensing Center (NERSC)
Thormølensgate 47
N-5006 Bergen
Norway

phone +47 55205800
fax +47 55205801
email admin@nersc.no

TITLE GreenSeas Report on error quantified hindcast with data assimilation 30 September 2012	**REPORT IDENTIFICATION** 328 Deliverable D5.2
CLIENT EU (Seventh Framework Programme)	**CONTRACT NO. 265294**
CLIENT REFERENCE EU	**AVAILABILITY** Public
AUTHORS Ehouarn Simon Annette Samuelsen Laurent Bertino	**AUTHORIZATION** Bergen, November 1, 2012 Prof. Stein Sandven Director NERSC

Contents

Report on error quantified hindcast with data assimilation (GreenSeas)

E. Simon [1], A. Samuelsen [1], L. Bertino [1]

1. **Mohn-Sverdrup Center (MSC) / NERSC, Bergen.**

1 Introduction

The 4-year hindcast of the North Atlantic and Arctic biology for the period 2007-2010 started as planned and it is currently being continued. Sea Surface Temperature, along-track sea level anomalies, ice concentration and remotely sensed chlorophyll-a are assimilated in two steps with a log-transformed extension of the deterministic ensemble Kalman-filter in the coupled physical-biogeochemical model HYCOM-NORWECOM. The processing of the first three years is almost completed and the full hindcast should be obtained before the end of the year 2012. Ocean Color data were not assimilated in 2007 in order to assess the impact of the corrections of the physical ocean state variables on the ecosystem model dynamics. Observations located in areas shallower than 300m are not assimilated during the first 6 months of 2008 (warm-up of the biogeochemical assimilation). Daily averages of both physical and biological variables are continuously produced in order to quantify the error in the solution.

2 Model description and experimental setup

The physical model used is the HYbrid Coordinate Ocean Model - HYCOM - (Bleck, 2002). This model is coupled to the ecosystem model NORWECOM (Skogen and Søiland, 1998) and has been used in several studies of the North Atlantic and the Norwegian Sea (Hansen and Samuelsen, 2009; Skogen et al., 2007). The model contains three types of nutrients; nitrate, phosphate and silicate, two types of phytoplankton; diatoms and flagellates, three types of detritus (nitrogen detritus, phosphate detritus and biogenic silica), and oxygen. A few modifications have been done to the model in the last couple of years (Samuelsen and Bertino, 2011). The model now includes two species of zooplankton; micro-zooplankton and meso-zooplankton. The feeding parameterizations for the zooplankton follows that of the model ECOHAM (Pätsch and Kühn, 2008). The micro-zooplankton feed on flagellates, diatoms, and detritus, while the meso-zooplankton feed on micro-zooplankton, diatoms and detritus.

2.1 Model setup

A relatively coarse model of the North Atlantic with resolution of about 50km in the Norwegian Sea was used. The atmospheric forcing used was the ERA-Interim reanalysis (Dee et al., 2011). For river input TRIP (Oki and Sud, 1998) was used for the freshwater-input. The nutrients in the river were derived from GlobalNEWS model results (Seitzinger et al., 2010). Both the freshwater- and nutrient-input from the rivers are climatological and thus there is no interannual variability present in the river forcing. The model uses the KPP mixing scheme (Large et al., 1994) Relaxation to climatology is applied to the surface salinity field with a relaxation time scale of 200 days, while no relaxation is applied to the sea surface temperature. Nutrients and oxygen are relaxed back to climatological data (Conkright et al., 1998) at the southern and northern boundary.. The ecosystem model was initialized with climatological values for the three nutrients and oxygen and constant low values for the remaining variables. The sediment layer was initialized with constant values. The model was run up to the end of 2011. In-situ measurements of chlorophyll for the period 1997-2009 were used for comparison.

2.2 Assimilation

The coarse resolution prototype of TOPAZ4 (coarse-TOPAZ4) is derived from the actual TOPAZ4 system (Sakov et al., 2012) that has transitioned from using the traditional perturbed observations EnKF scheme (Burgers et al., 1998) to the deterministic EnKF, or DEnKF, that was developed by Sakov and Oke (2008). The analysis is performed in the model grid space and divided into two steps. In a first step, the physical data (SST, along-track SLA and ice concentration) are assimilated in the physical component (HYCOM) of the coupled model. In this first step, the biological component is not included in the analysis state vector. The instances of negative layer thickness or ice concentration, should they occur, are corrected in a post-processing procedure. The mass balance is also corrected to account for vertical updates of isopycnals during analysis. However, biological tracers being defined as concentrations, changes in layer thickness might lead to changes in the amount of substance, due to changes in volume that are not compensated by changes in concentrations of the species. In order to insure the conservation of the amount of species for each tracer at each horizontal grid point (conservation in the water column), a vertical remapping of the tracer is also performed. This remapping uses the WENO interpolations that are already embedded in HYCOM. In the second step, surface chlorophyll concentrations data are assimilated in the biological component (NORWECOM) of the coupled model. It means that the dynamics of the physical ocean is not corrected by the assimilation of biological data. Biological state variables and parameters (phyto- and zooplankton mortality rates) are log-transformed prior to assimilation in order to prevent issues arising from the positiveness of the variables (Simon and Bertino, 2012).

Several changes in the system occurred during the simulation. Due to the local collapse of some parameter components of the ensemble towards their bounds (mortality rates equal to 0% or 100%) late December 2008, a new set of parameter values was drawn on 1 January 2009 based on the last local analyzed means and variances. In the mean time, new minimum and maximum bounds were defined for the parameters (0.1% and 99.9%) in order to prevent numerical issues. Furthermore, the occurrences of negative values in the phosphate detritus for the forecast steps during the bloom 2009 led to a correction of the formulation of the interactions between the phosphate detritus and the zooplankton. The simulation was restarted on 1 January 2009 with the corrected version of the model and the new set of the parameters

3 Validation against the assimilated satellite surface chlorophyll-a concentration

3.1 Time evolution of the error

The domain-averaged error time series in Fig. 1 show the seasonal peak of error during the summer phytoplankton bloom. The first peak corresponds to the physical data assimilation only in 2007 and the second and third peaks to the assimilation of Ocean Colour data in 2008 and 2009 with the characteristic assimilation jigsaw. It is worth noting that the assimilation of Ocean Colour data has proven effective in 2008: even in a relatively long 7-days window, the forecast errors remain well inferior to the errors obtained when assimilating the physical data only. A peak in the RMS error and standard deviation is observed on 9 July 2008. It corresponds to the extension of the observed domain due to the introduction of observation shallower than 300m in the assimilation system.

However, we note a large growth of the error during the bloom period in 2009. This is mostly due to erroneous large concentrations located in few restricted areas of the domain (around 30 N). In these areas, the parameter estimation leads to a combination of values that encourages the phytoplankton growth (low phytoplankton and large zooplankton mortality rates) due to too low chlorophyll concentrations in winter 2008-2009. However, the spread of the ensemble is locally too low early spring and does not allow large corrections of the parameters values that would be required during the spring bloom 2009 (too large chlorophyll concentration).

The errors predicted by the EnKF (blue line) correlates well with the variations of the actual errors, but underestimate the errors by half in the 2008 and third in the 2009 main bloom period (from May to August)

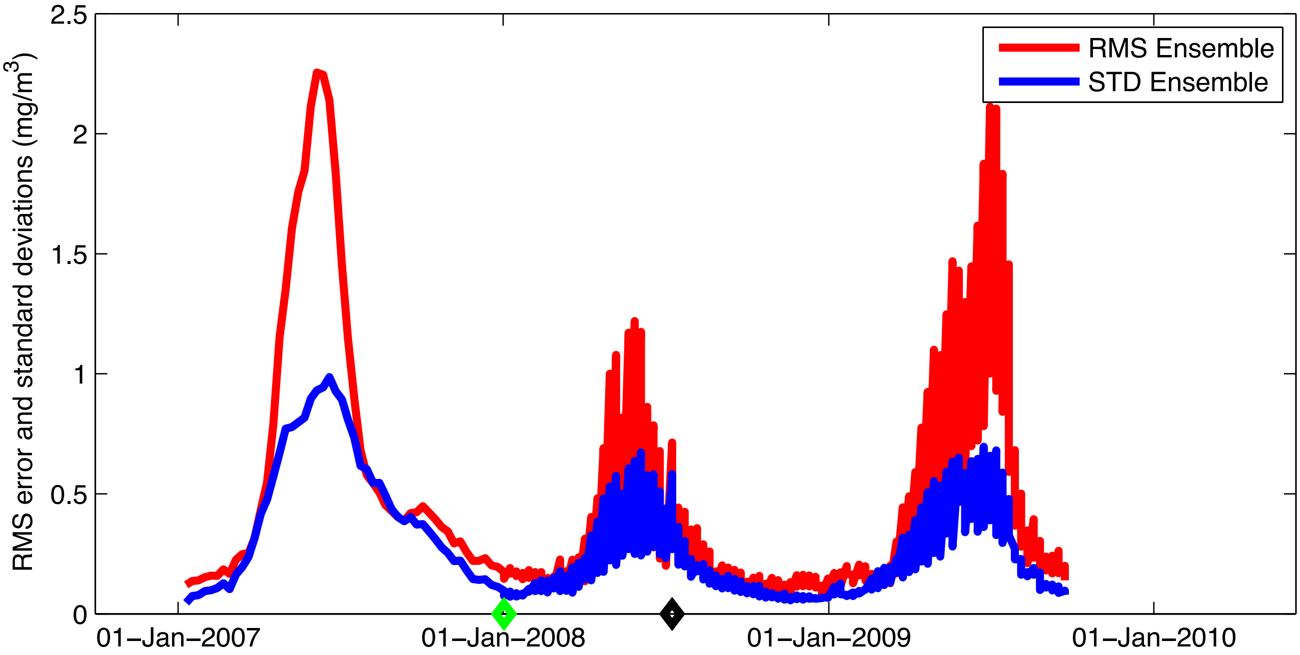

Figure 1: Surface chlorophyll concentration: time evolution of the RMS error and standard deviation computed from the assimilated observation. The green diamond highlights the date of the first biological analysis. The black diamond highlights the date of the first assimilation of Ocean Colour data in areas shallower than 300m.

and slightly in winter (from December to January). It indicates that the model and measurement errors are well tuned for late summer and early winter conditions, but that additional sources of errors could be considered for the summer (for example, the optical properties of the water or the parameterization of the mortality of the different plankton).

3.2 Regional distribution of the errors

The visual comparison of satellite versus modeled surface chlorophyll is shown to highlight the evolution of the spatial distribution of the error during the different seasons and the impact of the assimilation over the years. The modeled surface chlorophyll corresponds to the 7-day forecast ensemble mean, that is to say the surface chlorophyll concentration estimated from the model before assimilation of the observations shown in the figures.

First, we note that the 3D ecosystem model in physical assimilation-only mode (that is in 2007) tends to underestimate the surface chlorophyll concentration in most parts of the domain south of 50N (West Tropical Atlantic, North American Shelves and European Shelves) while too large concentrations are produced off West Africa (see Fig. 2) and near the Southern bound of the domain at the Equator. It is important to note that satellite observations are not present above 50N during that period of the year due to the long winter polar night and weather conditions (clouds). As expected, one assimilation step is not enough to significantly improve the modeled chlorophyll concentrations (9 January 2008). However, the assimilation of observations over one year leads to a correction of a large part of the deficiencies in January 2009. The upwelling area off Senegal has been reduced and we note larger concentrations in the Middle Atlantic Bight and the Gulf of Mexico.

The same deficiencies of the model are still observed in March in physical assimilation-only mode (see Fig. 3) Furthermore, the northward increase of the observed domain as the daytime increases in the North suggests that the chlorophyll concentration is also underestimated in the North part of the domain (e.g. North Sea and Norwegian Sea) during the cold period. However, we already note a slight improvement of the chlorophyll distribution in March 2008 after a 2-month assimilation period. The benefits of the assimilation are more evident in March 2009. The modeled chlorophyll distribution is in agreement with the observation in the South Western part of the domain (West Tropical Atlantic, Caribbean Sea, Gulf of Mexico). However, we note that concentrations are still too low near the sub-Polar Gyre. We note also a slightly overestimated concentration in the North Sea.

During the bloom period in spring (Fig. 4) and summer (Fig. 5) the behavior of the ecosystem model radically changes in the Northern part of the domain. In physical assimilation-only mode, the chlorophyll concentration in the European Shelves, in the sub-Polar regions and in the Arctic Ocean (Bering Strait) is now exaggerated compared to the satellite observations. In the meantime, the model continues to underestimate the surface chlorophyll concentrations in the West Tropical Atlantic. Despite the large reduction of the RMS error noted in Fig. 1, the assimilation cannot prevent the occurrence of a patch of large chlorophyll concentrations in the sub-Polar regions (Irminger Sea) for the bloom peak, early May 2008 (Fig. 4). This large patch of error is not present anymore, early May 2009. However, the deficiencies in the parameter estimation result in patches of too large chlorophyll concentrations all along the Gulf Stream.

Early August 2008 (Fig 5.), we note a significant improvement of the chlorophyll distribution when the primary production peak is over. The errors are mainly localized in the Arctic Ocean near the Bering Strait. Due to the short length of the observation window in that area, the assimilation of observation during several years is required to significantly improve the modeled chlorophyll concentration. So, we note a better representation of the surface chlorophyll concentration in the Arctic Ocean early August 2009, notably near the Bering Strait. However, the previous patches of large concentrations are still present in the central Atlantic at latitude 25 N to the Irminger basin due to weak corrections on some parameters in this area.

Finally, we note in early fall (Fig. 6) that the EnKF-assimilated chlorophyll concentration is in agreement with the observations in 2008 and 2009 while the free-running model overestimates the chlorophyll in the sub-Polar regions due to a strong bloom in spring and continues to provide too large (resp. low) concentrations off Africa (resp. in the Western Tropical Atlantic). Furthermore, the assimilation of physical data - notably the observations of ice concentration - plays a key role in the good representation of the inter-annual variability of the chlorophyll concentration in the Arctic Ocean. The differences in the ice cover between the years 2007 (minimum), 2008 and 2009 are clearly observed from the modeled surface chlorophyll distribution. Unfortunately, there are no Ocean Color observations in the Arctic Ocean during that period of the year. Finally, we note that the too large chlorophyll concentrations present in the central Atlantic during the spring bloom 2009 in the EnKF-assimilated simulation result in a slight overestimation of the surface chlorophyll concentration in the central Atlantic mainly off Portugal. Furthermore, the underestimation of the chlorophyll concentration along the North Brazil Current by the EnKF-assimilated simulation can be partly explained by the influence of the Amazon River and the resulting large uncertainties in the observations in that area and in the nutrient river inputs.

4 Validation against independent in-situ data

The reanalysis results were compared to in-situ data, for chlorophyll, nutrients, temperature and salinity. The model values used were those of the grid cell containing the measurement point interpolated to the depth of the measurement. The observations were provided by Institute of Marine Research in Norway through the MyOcean2 project and are thus located in the North Sea, Norwegian Sea, Barents Sea and around Svalbard. The only persistent time-series in the data set is that from Ocean Weather Station M. Certain areas have extensive cruises at specific time of the year, for example in the Barents Sea there are generally only observations available in September. There is data available in 2007, 2008, and 2009, but the 2009 dataset is quite limited, although there is still a regular time series at Station M available (Fig. 7).

Time series of surface chlorophyll from satiation M show that the initial increase of chlorophyll in the spring is earlier in the assimilation run also when only physics is assimilated. This may be du to a more realistic seasonal shoaling of the mixed layer. However in 2007 the overestimation of chlorophyll is about the same in the reanalysis and the control run, while in 2008 the bias is reduced in the reanalysis relative to the control run. Note that the chlorophyll values observed in 2007 were generally higher than those observed in 2008. The maximum measured values in 2007 being close 6 Chl mg/m3, while the maximum measured values in 2008 were about half that. The model, on the other hand tends to produce similar maximum values every year and therefore the bias was overall larger in 2008 for both runs.

One particular point of interest for the in-situ comparison was the effect the assimilation of chlorophyll would have on the concentrations of nutrients. The reanalysis as well as a free running control run were thus compared to these values. From the surface nutrients (Fig. 8 and 9) we see that that the reanalysis tend to be slower in mixing the nutrients back to the surface during winter, otherwise the two runs are fairly similar. At depth (Fig. 10 and 11) the values diverge with time in 2007 when only physics is assimilated. In 2008 when chlorophyll is also added to the assimilation the difference between the two runs seem to stabilize, but it is necessary to also evaluate subsequent years to be sure that there is no drift induced by the assimilation. Below 200 m the model generally has lower nutrient values that the observations, but while at 500 meters the assimilation run is closer to the observation, at 1000 meters and below (not shown) it is the control run that is closest to the observations.

In summary while the assimilation of satellite chlorophyll clearly reduces the bias in chlorophyll, the affect on the nutrients less clear. We se that in the assimilation run the nutrients are closer to the observation at 500 meters, but at 1000 meters it is the control run that has values closest to the observations. The nutrient response at depth is a slow process, hence no definite conclusion can be drawn until after the end of the reanalysis and a longer in-situ time-series has been obtained.

5 Conclusion

A combined state parameter estimation simulation is currently running at NERSC for the purpose of the reanalysis of the North Atlantic and Arctic Oceans biology during the period 2007-2010. Observations of sea surface temperature (SST), along-track sea level anomaly (TSLA), ice concentration and satellite-derived chlorophyll concentration are assimilated in the coupled model HYCOM-NORWECOM in two steps with the deterministic ensemble Kalman filter (DEnKF) every week. The biological corrections are applied to the state variables and the phyto- and zooplankton mortality rates of the model NORWECOM.

The evaluation of the time evolution of the RMS error computed from the assimilated surface Ocean Colour observations shown a relatively good control of the error growth during the first year (2008) compared to the solution in 2007 where no Ocean Colour data are assimilated. The improvement of the surface chlorophyll distribution observed in 2008 is confirmed with the in-situ data present in the North Sea (station M). Furthermore, we do not note degradation in the nutrients concentrations (surface and at 500m) that could occur due to the data assimilation when compared to in-situ data in this particular station.

However, we note a large growth of the error during the bloom period in 2009 mostly associated with erroneous large concentrations located in few restricted areas of the domain (central Atlantic and Irminger Sea). In these areas, local collapses of the parameter component of the ensemble in winter lead to sets of values of the mortality rates encouraging the phytoplankton growth and so, extremely large chlorophyll concentrations during the spring bloom. Unfortunately, a comparison to in-situ data could not be done because of the lack of data.

The simulation is now approaching the winter period 2009-2010. We plan to investigate the regional distribution of the different mortality rates as well as their seasonal evolution over the period 2008-2009 in order to apply regional time-evolving parameters for the last year of the hindcast

Acknowledgments

The project acknowledges the Norwegian supercomputing project NOTUR2 for a grant of computer time on the national High Performance Computing facilities (project `nn2993k`).

References

Bleck R.: An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates, *Ocean Modelling*, 4(1), 55-88, 2002.

Burgers G., van Leeeuwen P.J. and Evensen G.: Analysis scheme in the Ensemble Kalman Filter. *Mon. Weather Rev.*, 126, 1719-1724, 1998.

Canuto, V.M. et al.: Ocean turbulence, III: New GISS vertical mixing scheme. *Ocean Modelling*, 34(3-4), 70-91,2010.

Conkright, M.E. et al.: World Ocean Database 1998 Documentation and quality control, National Oceanographic Data Center, Silver Spring, MD, 1998.

Dee, D.P. et al.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656), 553-597, 2011.

Hansen, C. and Samuelsen, A.: Influence of horizontal model grid resolution on the simulated primary production in an embedded primary production model in the Norwegian Sea. *Journal of Marine Systems*, 75(1-2), 236-244, 2009.

Large, W.G., McWilliams, J.C. and Doney, S.C.: Oceanic vertical mixing - a review and a model with a nonlocal boundary-layer parameterization. *Reviews of Geophysics*, 32(4), 363-403, 1994.

Oki, T. and Sud, Y.C.: Design of Total Runoff Integrating Pathways (TRIP) A global river channel network. *Earth Interact*, 2, 1-37, 1998.

Pätsch, J. and Kühn, W.: Nitrogen and carbon cycling in the North Sea and exchange with the North Atlantic - a model study, Part I. Nitrogen budget and fluxes. *Continental Shelf Research*, 28(6), 767-787, 2008.

Sakov P. and Oke P.R.: A deterministic formulation of the ensemble Kalman filter: an alternative to ensemble square root filters. *Tellus*, 60A, 361-371, 2008.

Sakov P., Counillon F., Bertino L., Lister K.A., Oke P.R. and Korablev A.:TOPAZ4: an ocean-sea ice data assimilation system for the North Atlantic and Arctic. *Ocean Sci.*, 8, 633-656, 2012.

Samuelsen, A. and Bertino, L.: Arctic Ocean ecosystem modeling in MyOcean. *Mercator Ocean Quarterly Newsletter*, 40, 36-40, 2011.

Seitzinger, S.P. et al.: Global river nutrient export: A scenario analysis of past and future trends. *Global Biogeochemical Cycles*, 24(null), GB0A08, 2010.

Simon E. and Bertino L. : Gaussian anamorphosis extension of the DEnKF for combined state and parameter estimation: application to a 1D ecosystem model, *Journal of Marine Systems*, 89, 1, 1-18, 2012.

Skogen M. and Søiland H. : A user's guide to NORWECOM v2.0. The NORWegian Ecological Model system, *Technical Report Fisken og Havet 18*, Institute of Marine Research, Norway,1998.

Skogen, M.D., Budgell, W.P. and Rey, F.: Interannual variability in Nordic seas primary production.*ICES Journal of Marine Science*, 64(5), 889-898, 2007.

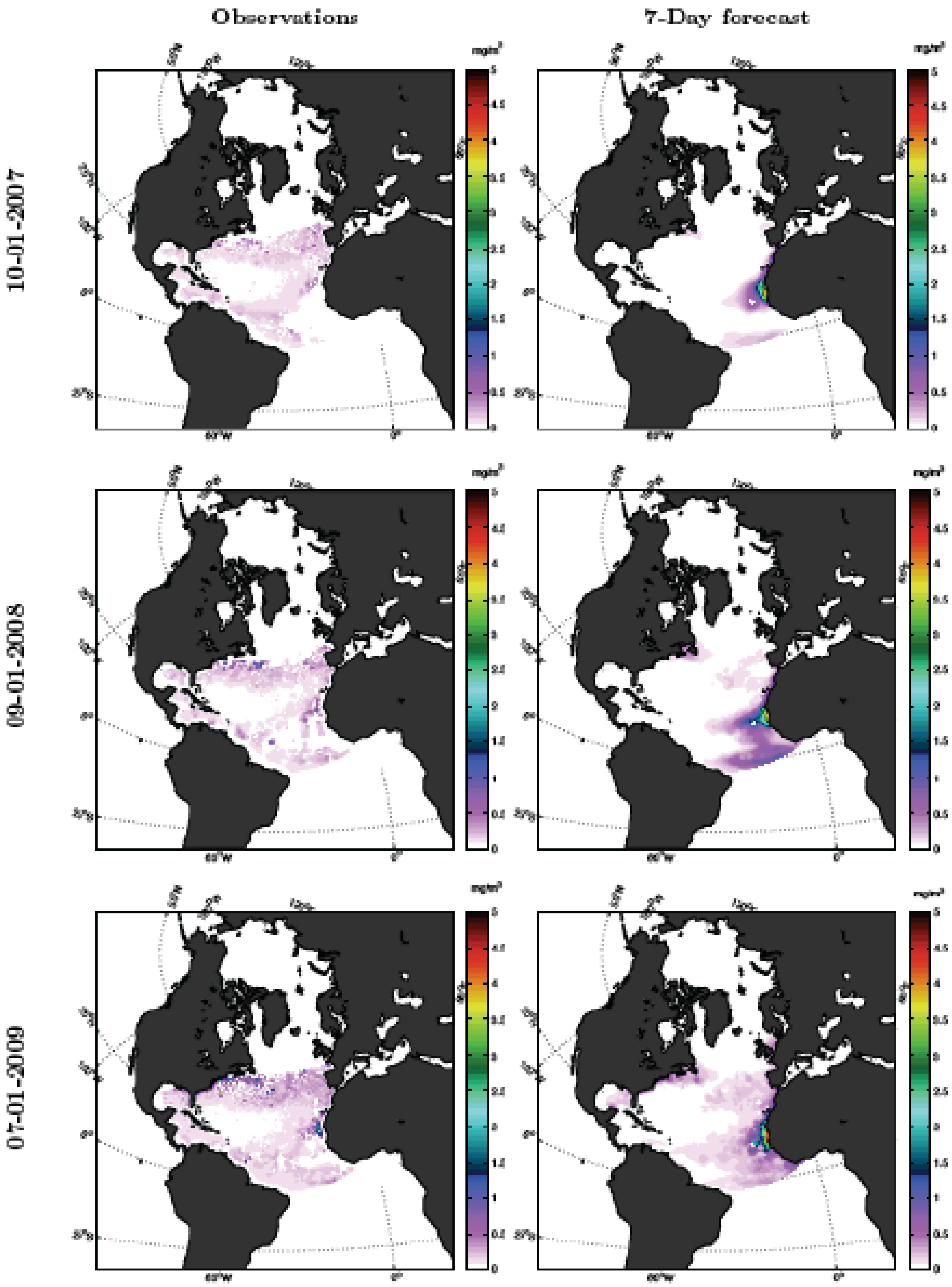

Figure 2: Comparison of observed versus free-running modeled (2007) and EnKF-assimilated (2008, 2009) surface concentrations of Chla early January

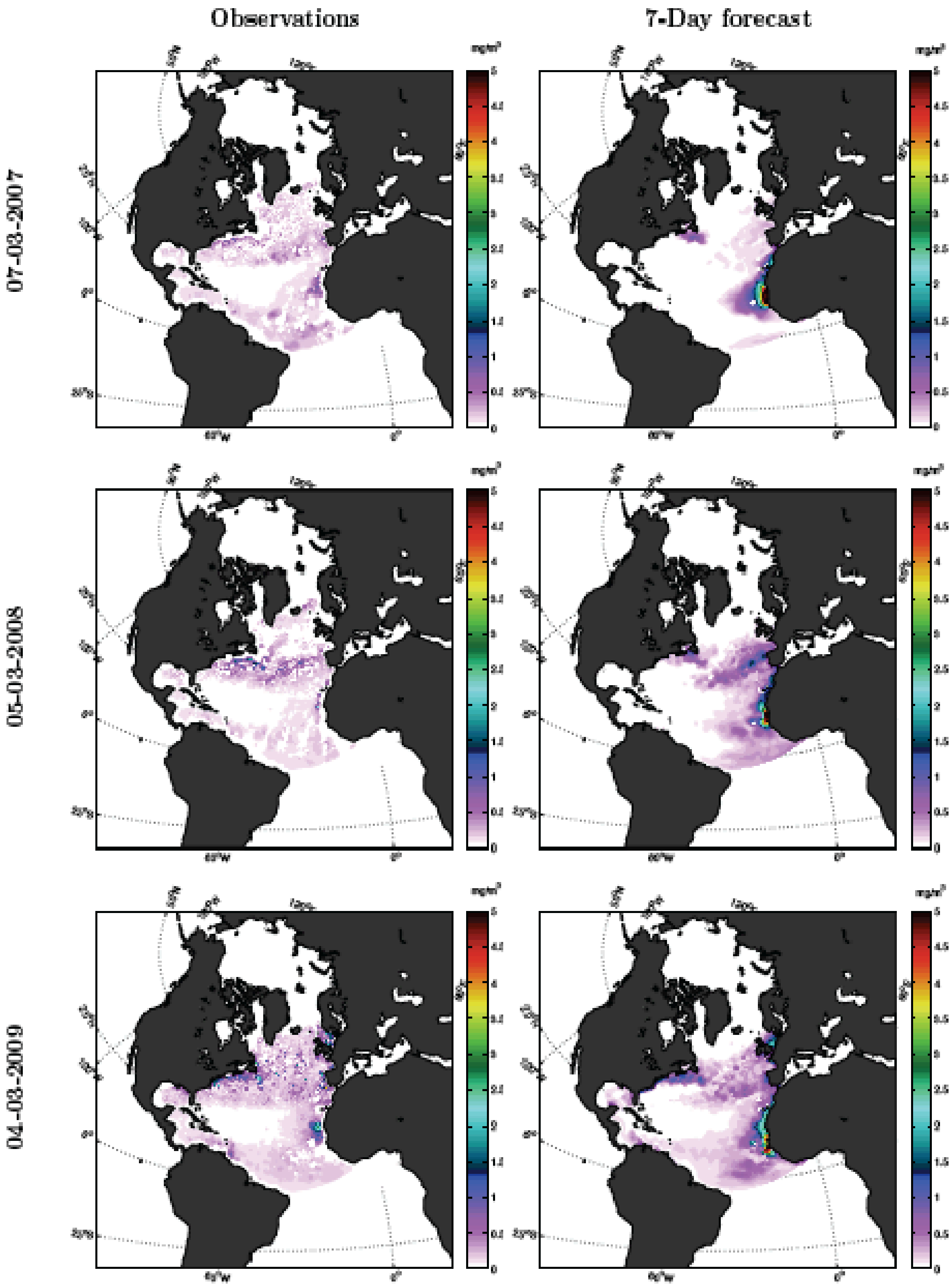

Figure 3: Comparison of observed versus free-running modeled (2007) and EnKF-assimilated (2008, 2009) surface concentrations of Chla early March.

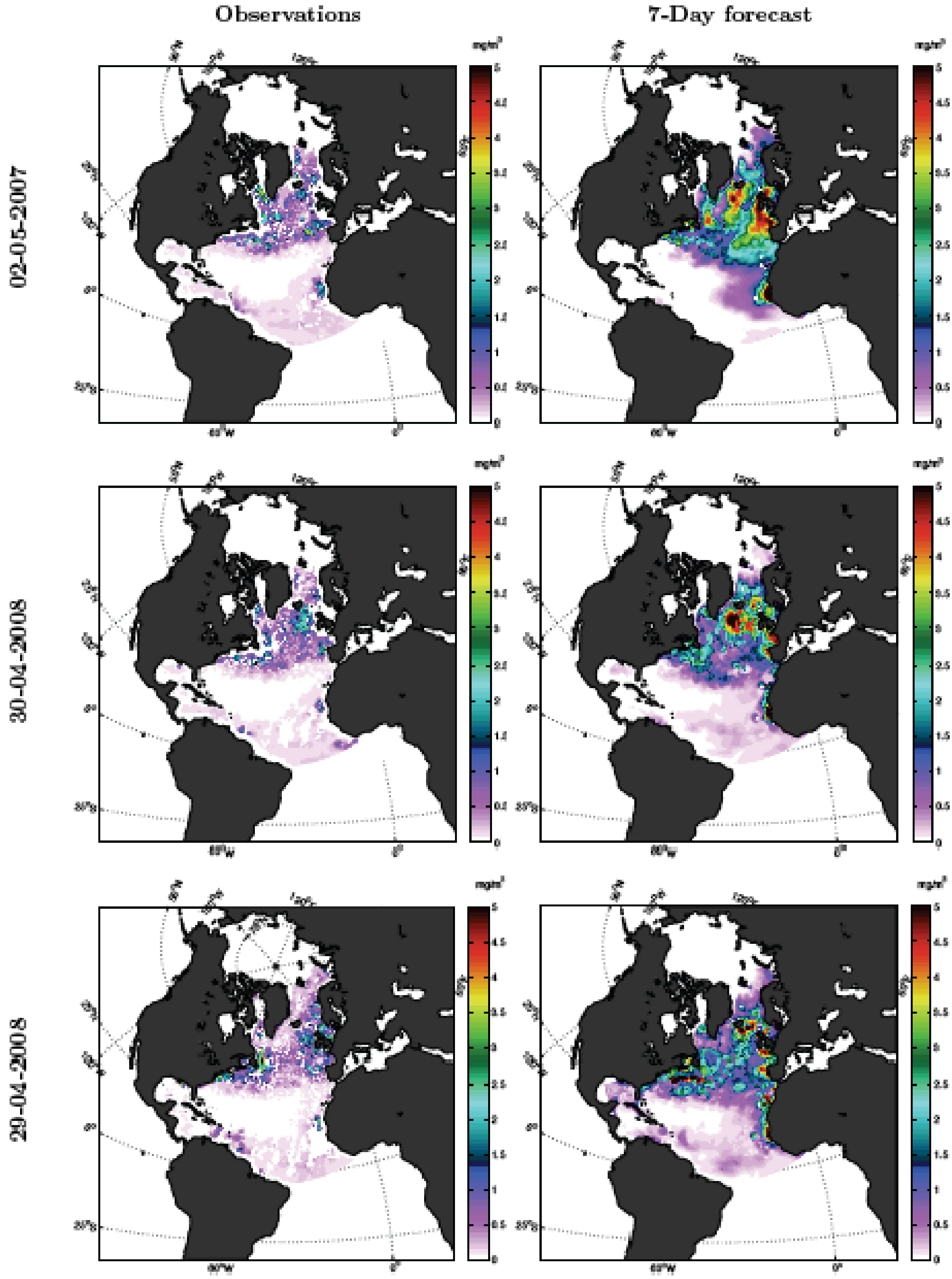

Figure 4: Comparison of observed versus free-running modeled (2007) and EnKF-assimilated (2008, 2009) surface concentrations of Chla early May

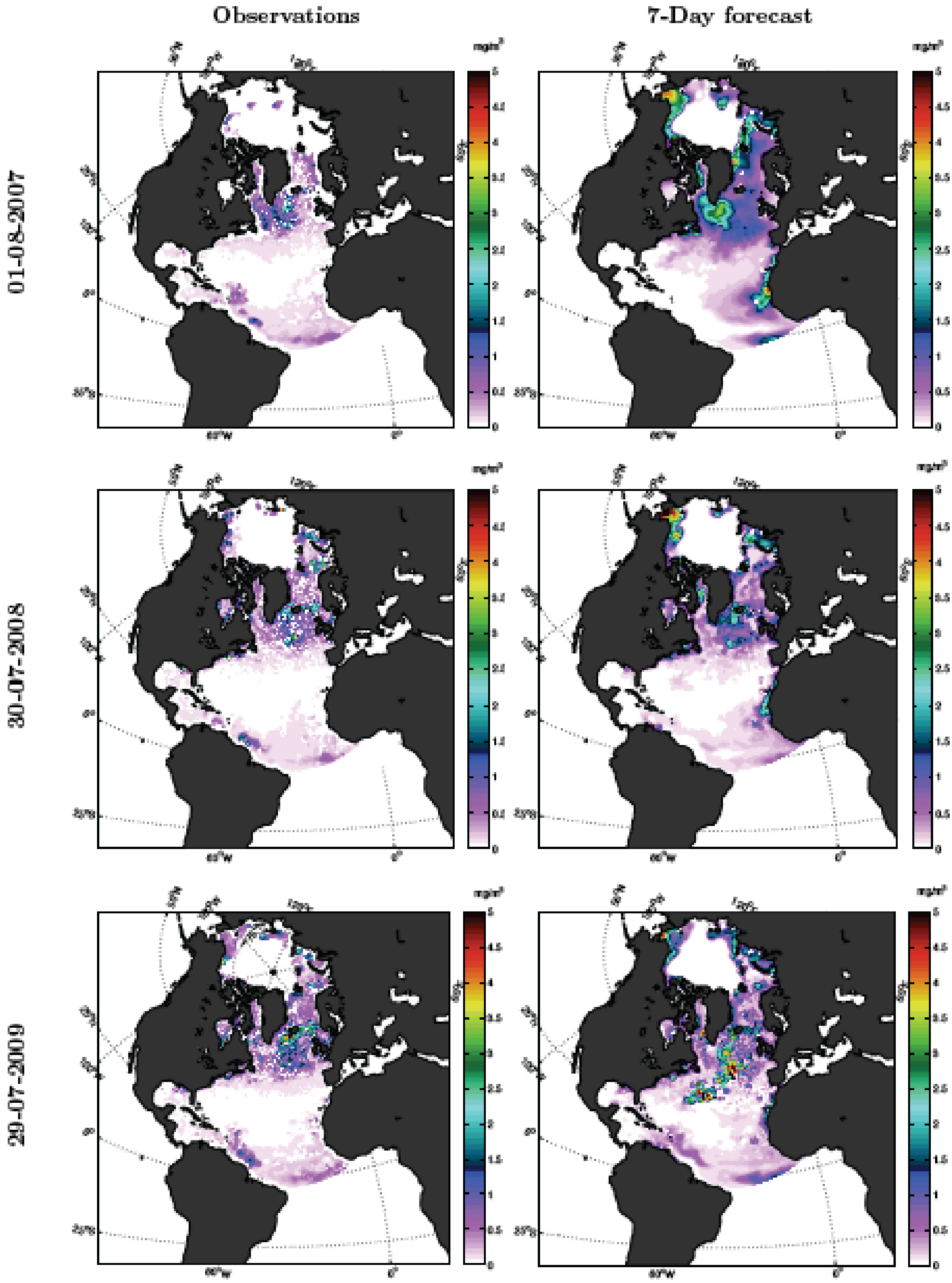

Figure 5: Comparison of observed versus free-running modeled (2007) and EnKF-assimilated (2008, 2009) surface concentrations of Chla early August

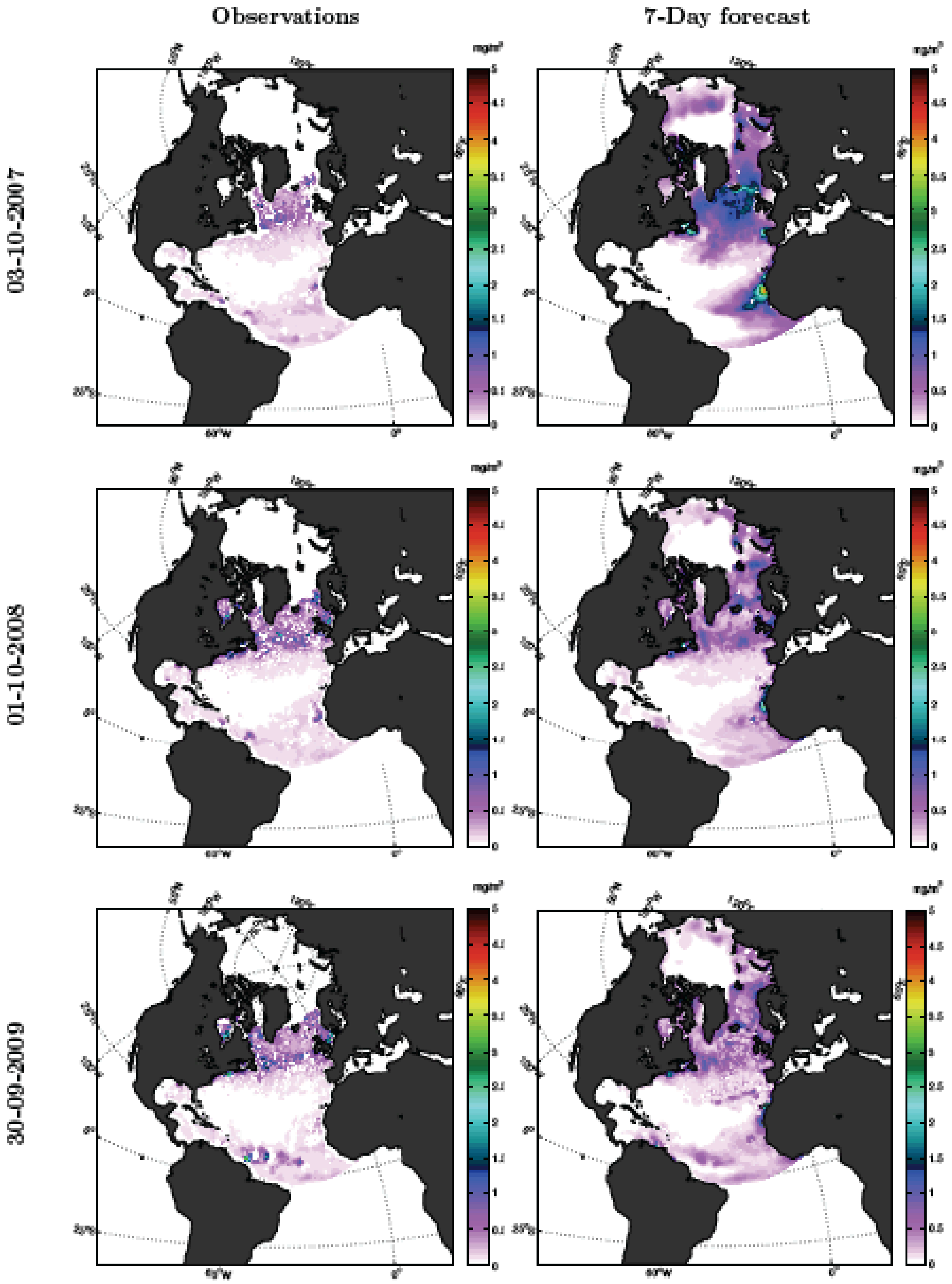

Figure 6: Comparison of observed versus free-running modeled (2007) and EnKF-assimilated (2008) surface concentrations of Chla early October.

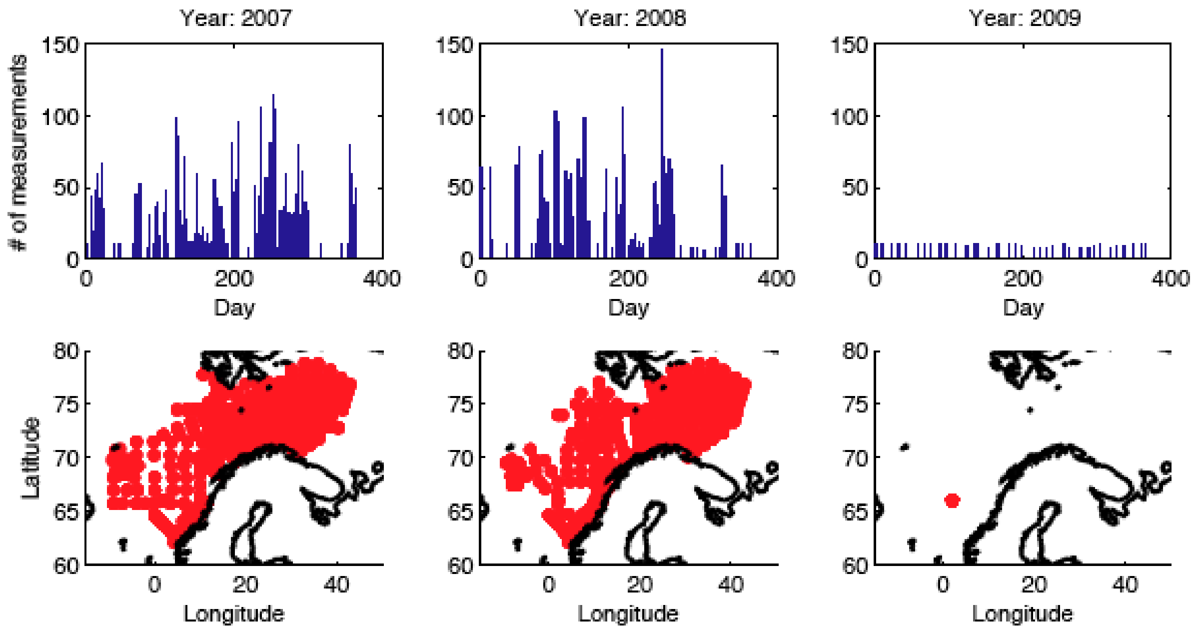

Figure 7: : The upper panels show the number of chlorophyll measurements as a function of day of year and the lower panel show the geographical coverage if the data set used.

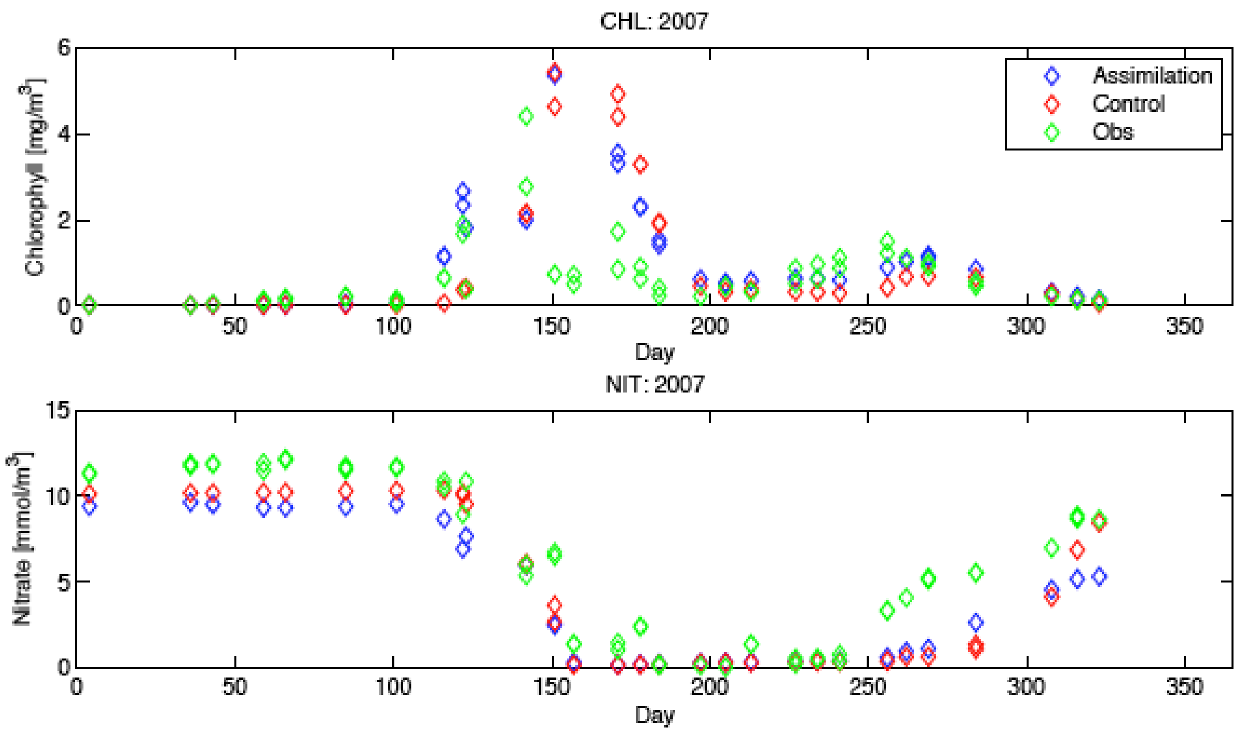

Figure 8: Time series surface of chlorophyll and nitrate at Station M for the reanalysis (blue), control run (red) and observations (green) in 2007 when only physics is assimilated.

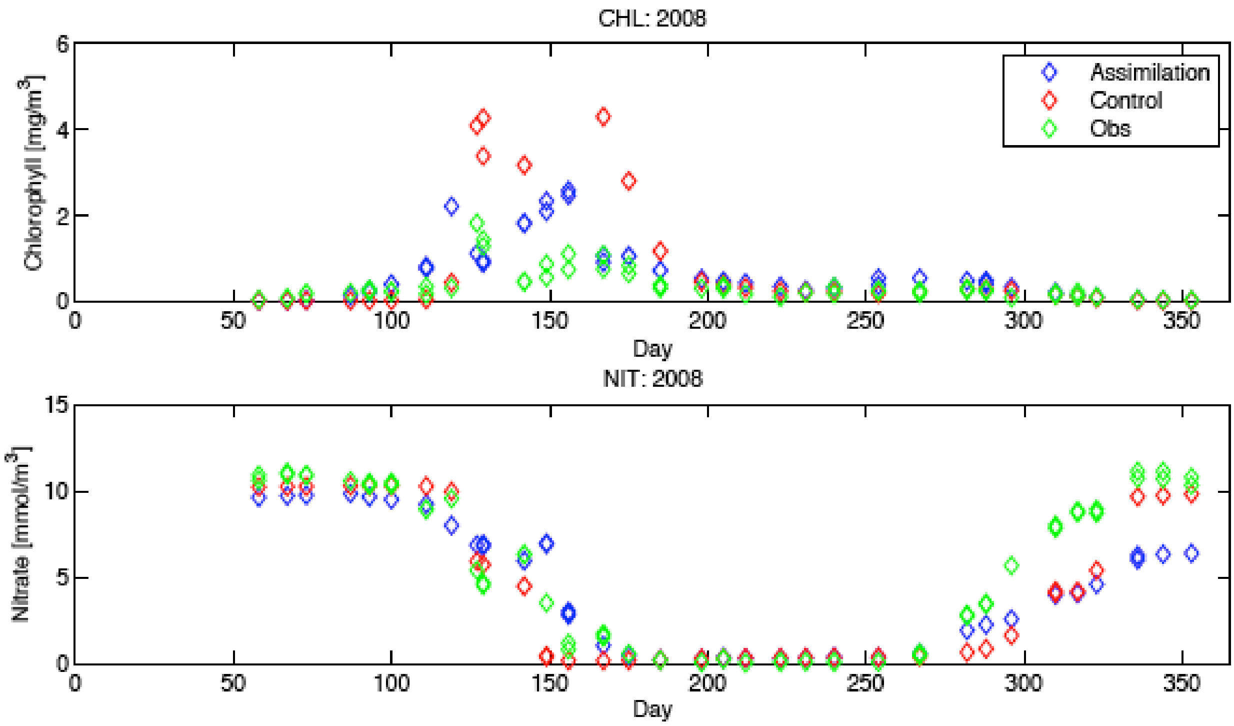

Figure 9: Time series of surface chlorophyll and nitrate at Station M for the reanalysis (blue), control run (red) and observations (green) in 2008 when both physics and chlorophyll is assimilated.

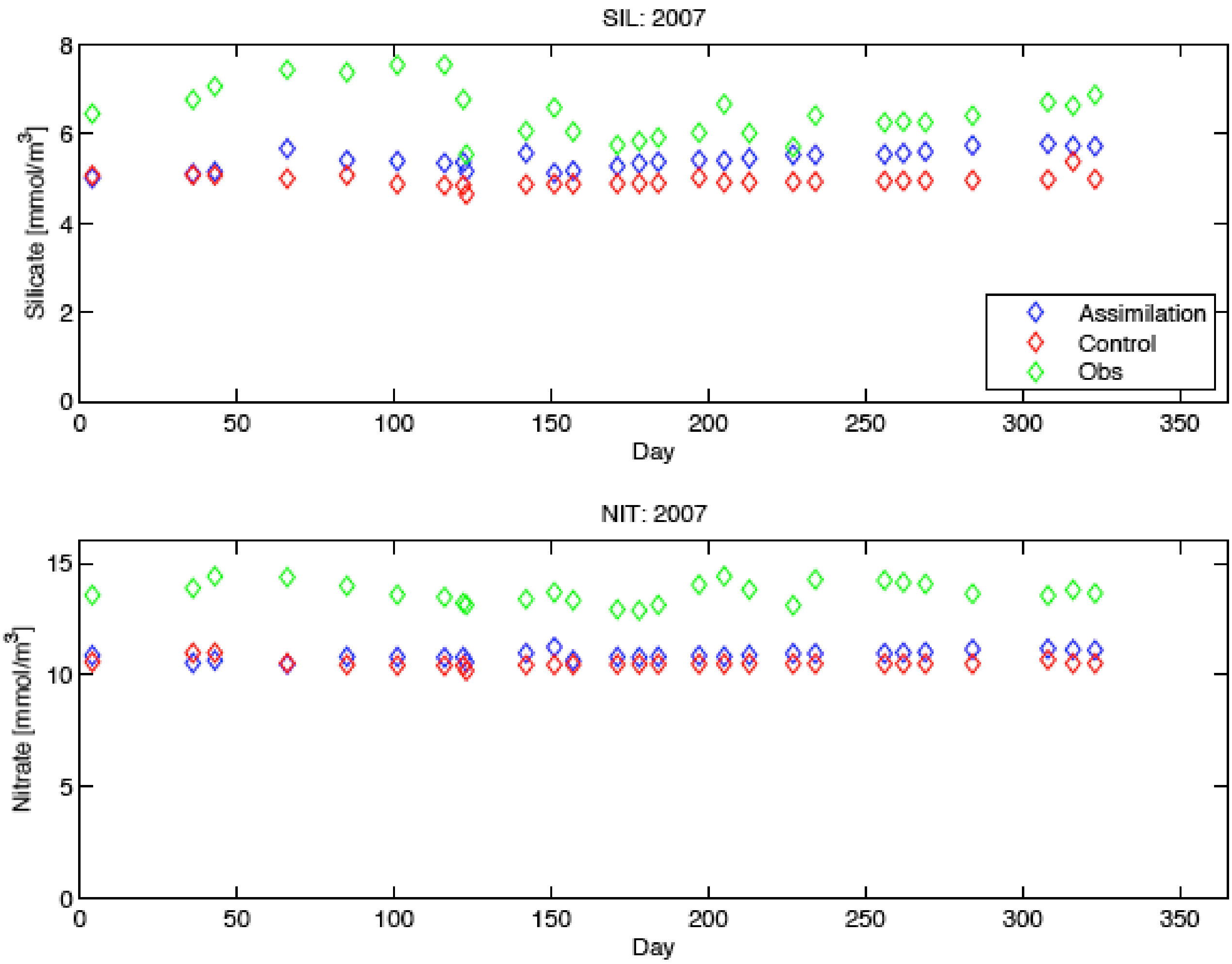

Figure 10: Time series of silicate and nitrate at 500 m at Station M for the reanalysis (blue), control run (red) and observations (green) in 2007 when only physics is assimilated.

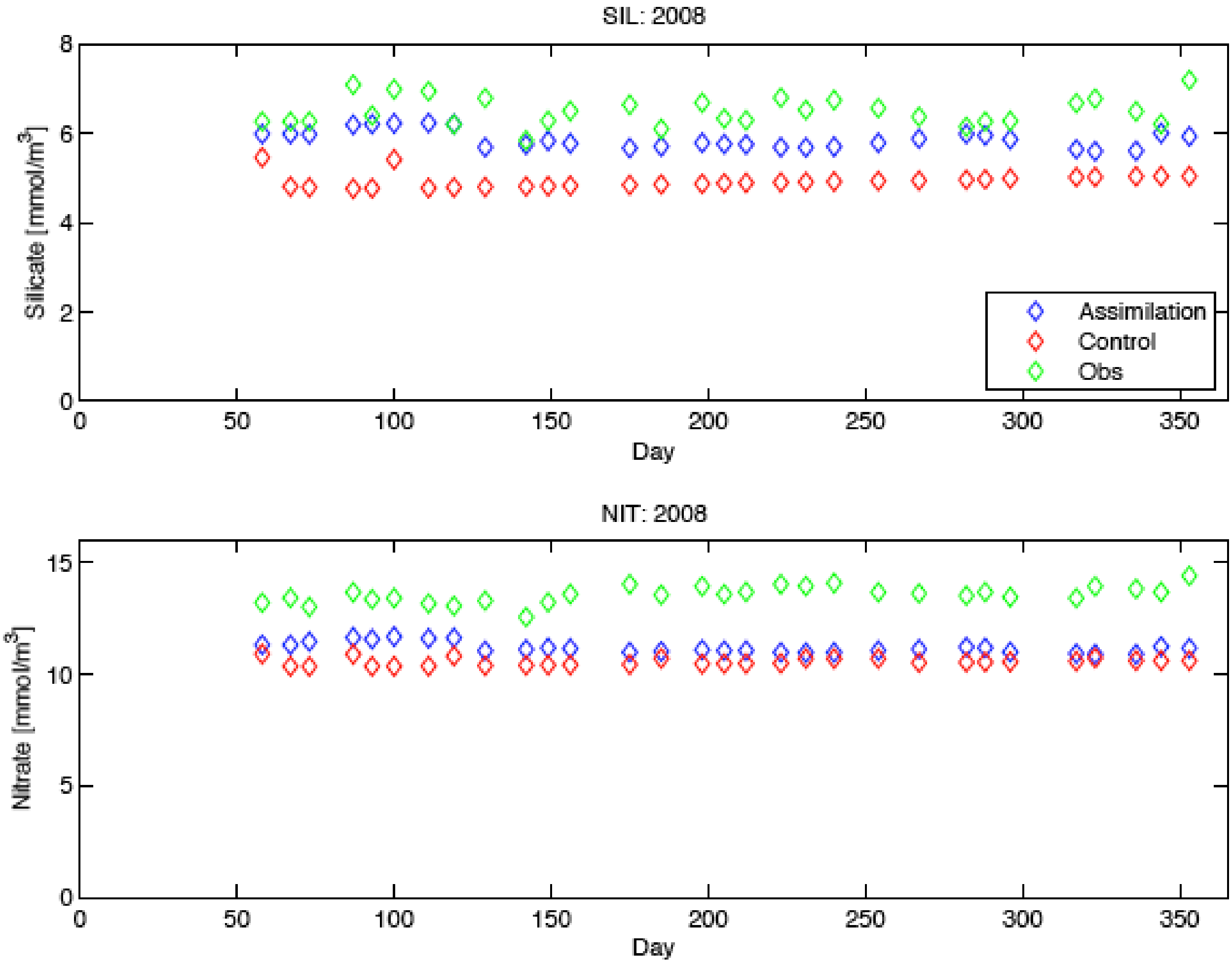

Figure 11: Time series of silicate and nitrate at 500 m at Station M for the reanalysis (blue), control run (red) and observations (green) in 2008 when both physics and chlorophyll is assimilated.