Assessment of a Physical-Biogeochemical Coupled Model System for Operational Service in the Baltic Sea

Zhenwen Wan[1*], Jun She[1], Marie Maar[2], Lars Jonasson[1], Jesper Baasch-Larsen[3]

1. Centre for Ocean and Ice, Danish Meteorological Institute, Lyngbyvej 100, DK-2100 Copenhagen, Denmark
2. National Environmental Research Institute, Aarhus University, Department of Marine Ecology, Frederiksborgvej 399, PO box 358, DK-4000 Roskilde, Denmark
3. Danish Defence Center for Operational Oceanography, Overgaden Oven Vandet 62B, DK-1023 Copenhagen K, Denmark

*Corresponding author: zw@dmi.dk, phone: 0045 3915 7284, fax: 0045 3915 7300

Key words: ecosystem model, ecological model, biogeochemical model, Baltic Sea, model assessment, model validation, operational oceanography

Abstract

Thanks to the abundant observation data, we are able to deploy the traditional point-to-point comparison and statistical measures in combination with a comprehensive model validation scheme to assess the skills of the biogeochemical model ERGOM in providing an operational service for the Baltic Sea. The model assessment concludes that the operational products can resolve the main observed seasonal features for phytoplankton biomass, dissolved inorganic nitrogen, dissolved inorganic phosphorus and dissolved oxygen in euphotic layers, as well as their vertical profiles. This assessment reflects that the model errors of the operational system at the current stage are mainly caused by insufficient light penetration, excessive organic particle export downward, insufficient regional adaptation and some of improper initialization. This study highlights the importance of applying multiple schemes in order to assess model skills rigidly and identify main causes for major model errors.

1. Introduction

Assessment of an operational model is different from validation of a model targeted at a specific research task. An operational model should serve broader interests than a research model generally does, since the users of the model results can be interested in various subdomains and processes. This is especially true during the early development phase of an operational model to supply biogeochemical information service. During the preliminary phase, there are no specific user needs, simply because user groups have not been well developed yet. Of course, there are general concerns in ecological operational oceanography, e.g. eutrophication, harmful algae blooms and oxygen depletion. Therefore, an operational model should produce sensible results in the entire model domain for all targeted state variables. In fact, the development of ocean models are endless practices where developers always do their best to work towards moving targets. As a goal of this stage, the model is aiming at reproducing the main observed seasonal features for phytoplankton biomass, nutrients concentration and dissolved oxygen concentration in euphotic layers.

Various ecosystem models have been developed for the Baltic Sea (Neumann, 2000; Edelvang et al., 2005; Savchuk et al., 2008; Eilola et al., 2009). The biogeochemical model ERGOM developed by Neumann (2000) and Neumann et al. (2002) has been applied in a number of investigations of the Baltic Sea ecosystem. The model inherited the advances of previous ecological models developed for the Baltic Sea (Stigebrandt and Wulff, 1987; Fennel, 1995; Fennel and Neumann, 1996) and has been further developed. Fennel and Neumann (2003) introduced stage-structured copepod models in order to replace the bulk description of zooplankton and improve the link to higher trophic levels. In the study on eutrophication and shifts in nitrogen fixation, Neumann and Schernewski (2008) introduced iron-phosphate-complex in combination with Dissolved Inorganic Phosphorus (DIP) in order to simulate the mineralization of detritus in the sediment. Kuznetsov et al. (2008) added seven state variables so as to simulate C, N, P cycling separately. Maar et al. (2011) added silicate as one more state variable so as to be able to model the ecosystem in the entire

3

salinity gradient region covering the Baltic Sea and the North Sea. Other examples of ERGOM application studies include the inter-annual variability in cyanobacteria blooms (Janssen et al., 2004), the assessment of two nutrient abatement strategies (Neumann and Schernewski, 2005), and the fate of river-borne nitrogen (Neumann, 2007).

As one part of the EU projects ECOOP (http://www.ecoop.eu) and MyOcean (http://www.myocean.eu.org), the ecosystem model ERGOM (Neumann, 2000; Neumann et al., 2002) is coupled with the circulation model HBM (https://hbmsvn.dmi.dk/) (Berg and Poulsen, 2012) for providing GMES (Global Monitoring of Environment and Security) Marine Service in the Baltic Sea. This paper presents an assessment of the operational model system with focus on its biogeochemical service, through comparing model results and observations comprehensively.

2. Models, data and methods

2.1 Physical model

The physical model is the HIROMB-BOOS ocean circulation model (HBM) (Berg and Poulsen, 2012). The core of the physical model, the circulation model, is based on the primitive geophysical fluid dynamics equations for the conservations of volume, momentum, salt and heat. The circulation model has been coupled to a Hibler-type sea ice model. The wind, air pressure, air temperature, humidity, evaporation/precipitation and cloud cover are taken into account in the parameterizations of surface boundary conditions. Water levels of tides and surges and monthly climatology of temperature and salinity are imposed as outer lateral boundary conditions. River runoff is included as an inner lateral condition. The model setup fully covers both the Baltic Sea and the North Sea with four two-way nested subdomains (Table 1). Our targeted area is the Baltic Sea (Fig. 1)

Table 1. Model grids

Subdomains	Longitude	Latitude	Lon.Res	Lat.Res	Lay.

North Sea	4°07'30"W-11°57'30"E	48°31'30"-65°52'30"N	5'	3'	50
Danish Straits	9°20'25"-14°49'35"E	53°35'15"-57°35'45"N	50"	30"	75
Wadden Sea	6°10'50"-10°29'10"E	53°13'30"-55°41'30"N	1'40"	1'	24
Baltic Sea	14°37'30"-30°17'30"E	53°31'30"-65°52'30"N	5'	3'	109

Abbreviations: Lat.Res for latitude resolution, Lon.Res for longitude resolution, Lay. for number of layers.

Location for Fig. 1

The products by the operational weather model High Resolution Limited Area Model of the Danish Meteorological Institute are used to provide atmospheric forcing drivers for the physical model (She et al., 2007a). The daily river runoffs are provided by the operational hydrological model HBV run by the Swedish Meteorological Hydrological Institute (Bergström, 1976 and 1992) in combination with observations from the German Bundesamt für Seeschifffahrt und Hydrographie and climatology. The previous versions of HBM were validated by She et al. (2007a, b). The current version was validated in the Scientific Calibration Report V2 for WP6 (http://www.myocean.eu.org/).

2.2 Ecosystem model

The applied version of ERGOM is close to the original version by Neumann et al. (2002). ERGOM originally adopted Redfield ratio for the phytoplankton stoichiometry. Wan et al. (2011) documented that a non-Redfield ratio is more suitable in the Baltic Sea than the Redfield ratio. Moreover, Wan et al. (2012) demonstrated that a spatially variable N/P ratio is more close to the real phytoplankton stoichiometry in the Baltic Sea than a fixed non-Redfield ratio does. In the current study, the model setup and configuration are the same as in the MyOcean Scientific Calibration Report V2 for WP6, but the source code is upgraded to implement the spatially variable N/P ratio (Wan et al., 2012).

5

Initial fields for ammonia, nitrate, DIP and Dissolved Oxygen (DO) are set through merging the data from the World Ocean Atlas 2001 (WOA01, Conkright et al., 2002) and the data from the International Council for the Exploration of the Sea (ICES) (http://www.ices.dk/indexfla.asp). Initial fields for the biological state variables have been adjusted through repetitive runs. The open boundary conditions for nitrate, DIP and DO are interpolated from the climatology of WOA01 data while the remaining state variables are set to zero. The bioloadings are from the same data sources for river runoffs mentioned above. The atmospheric nutrient depositions are based on Langner et al. (2009) and Eilola et al. (2009).

2.3 The comprehensive validation scheme

The comprehensive validation scheme makes use of all available in-situ data in order to reflect the model skill overall, rather than only at selected stations or over a part of the spatio-temporal domain. This scheme compares model results with observations along the specified dimension (e.g. temporal evolution, vertical profile or horizontal distribution). For technical details, refer to Wan et al. (2011). In this study, the 4-dimensional spatiotemporal grid to delimit data representation has a horizontal resolution of 0.5°x0.5°, a vertical resolution of 4 m and a temporal resolution of 15 days.

2.4 Statistical measures

To assess the model skills we use the following statistical measures: coefficient of determination (R^2), i.e. square of correlation coefficient, Model Efficiency (ME) (Nash and Sutcliffe, 1970), Cost Function (CF) (OSPAR Commission, 1998) and Percentage of Bias (PB) (Allen et al., 2007). ME is a measure of the ratio of the model error to the data variability,

$$ME = 1 - \frac{\sum (D - M)^2}{\sum (D - \overline{D})^2}, \qquad (1)$$

where D is the data, M is the corresponding model value, while the overbar denotes an averaging operation. ME is cited as a performance indicator: >0.65 excellent, 0.65-0.5 very good, 0.5-0.2 good, <0.2 poor (Maréchal, 2004). CF is a measure of the "goodness of fit" between model and data,

$$CF = \frac{\sum |M - D|}{n\sigma_D}, \qquad (2)$$

where σ_D is the standard deviation of data and *n* is the number of samples in the dataset. CF is cited as a performance indicator: <1 very good, 1-2 good, 2-3 reasonable, >3 poor (Radach and Moll, 2006). |PB| is cited as a performance indicator: <10 excellent, 10-20 very good, 20-40 good, >40 poor (Maréchal, 2004) and PB is given,

$$PB = \frac{\sum (D - M)}{\sum D} * 100. \qquad (3)$$

2.5 Observations

The observations used for model assessment are downloaded from ICES database. We have used the following observation types: temperature, salinity, chlorophyll (Chl) *a*, Dissolved Inorganic Nitrogen (DIN = ammonia + nitrate only), dissolved inorganic phosphorous and DO. The data coverage ranges 10°~30°E and 54°~66°N (Fig. 1) and from January 1, 2007 to December 31, 2008. The total record numbers for temperature, salinity, Chl *a*, DIN, DIP and DO are listed in Table 3. The ICES database is searched for monthly based time-series records. It ends up with 18 stations which have monthly based time-series records for almost all of the targeted state variables during 2007 and 2008. The station locations are shown in Fig. 1.

2.6 Simulation

The simulation is the same as the inter-comparison experiment described in the Scientific Calibration Report V2 for WP6 of the MyOcean project, i.e. a model hindcast for years of 2007 and

7

2008. The only difference to that inter-comparison experiment is using the upgraded source code with a spatially variable N/P ratio (Wan et al., 2012).

3. Results

Although ERGOM includes nine state variables, we present the model-observation comparison for only DIN, DIP, Chl *a* and DO, in consideration of the availability of observations. Temperature and salinity of the model results are also compared with observations in order to supply information on the skills of the circulation model. We examine the temporal dynamics in surface and bottom layers at 18 stations (Fig.s 2-12), the vertical profile at Station I in the Gotland deep (Fig. 13) and the bias distribution along different dimensions (Fig.s 14-16). The surface/global statistical measures are listed in Tables 2 and 3, whose performance scores are listed in Table 4.

Table 2. Statistical measures of model-observation comparison in the surface layer

	NS	$Mean^o$	$Mean^m$	PB	R^2	ME	CF
temperature	2077	9.8	9.7	-1.1	0.94	0.93	0.07
Salinity	2008	9.3	9.2	-1.1	0.96	0.96	0.05
DIN	1548	3.6	1.5	-58	0.10	0.04	19.0
DIP	1551	0.34	0.33	-4.7	0.35	0.33	1.3
Chl *a*	1291	3.5	3.0	-14	0.06	0.03	6.9
DO	1814	352	337	-4.0	0.34	0.21	1.2

Abbreviations: NS for number of samplers, $Mean^o$ for mean value of observations, $Mean^m$ for mean value of model results, PB for percentage of bias, R^2 for square of correlation coefficient, i.e. coefficient of determination, ME for model efficiency, CF for cost function.

Table 3 Statistical measures of model-observation comparison overall

	NS	$Mean^o$	$Mean^m$	PB	R^2	ME	CF
temperature	16534	7.8	7.9	1.2	0.89	0.89	0.11
salinity	16208	11	11	-2.2	0.98	0.98	0.02
DIN	10517	3.1	4.6	26	0.07	-0.18	2.24
DIP	10549	0.90	1.1	-2.2	0.87	0.86	0.22
Chl *a*	5644	2.3	2.7	-14	0.15	0.11	3.09
DO	14070	276	290	4.9	0.80	0.77	0.36

Abbreviations same as in Table 2.

Table 4 Performance scores

	Surface layer			All layers		
	PB	ME	CF	PB	ME	CF
DIN	Poor	Poor	Poor	Good	Poor	Reasonable
DIP	Excellent	Good	Very good	Excellent	Excellent	Very good
Chl *a*	Very good	Poor	Poor	Very good	Poor	Poor
DO	Excellent	Good	Very good	Excellent	Excellent	Very good

Scores are accorded to Nash and Sutcliffe (1970), OSPAR Commission (1998) and Allen et al. (2007).

3.1 Temperature

In the surface layer, the model results fit observations very well at all the 18 stations in terms of seasonal variability (Fig. 2). In details, model matches observation best in the winter months but with more bias in the summer months, which can be up to 2 °C off. Northeastern Baltic sea coastal stations (M, O, R) have larger model errors than others. In statistics using all model-observation pairs in surface layer (far beyond 18 stations), PB is -1.1 only, R^2 is up to 0.94, ME is up to 0.93, and CF is 0.07 (Table 2). It means that the performance scores are either “excellent” or “very good” in the surface layer.

9

In the bottom layer, the seasonal cycle is less visible at water depth deeper than 50 m. The model catches the observed seasonal pattern for the shallow stations in Kattegat, Western Baltic Sea, Bothnian Sea and Bothnian Bay (C, D, N, P, Q and R) and the deep stations in Central and North Baltic Proper (F-K), but are rather off for stations A, E, L (Fig. 3). The temporal evolution of vertical profiles of the model (Fig. 13a) matches well that of observations in general (Fig. 13g). There are however some minor errors. For example, the model temperature at depth 90-120 m is persistently higher than observations, and there exists downward temperature gradient in November and December above 40 m in model results but not in observations which indicates that the model has less vertical mixing. The spatial mean of observations is caught well by the corresponding mean of model results (Fig. 14a). The mean of observations at one depth plane is also well reproduced by the corresponding model results (Fig. 15a), but the model errors are larger in layers below 100 m than above, up to 0.5 °C. The percentage bias of model to observation is mostly smaller than ± 10% (Fig. 16a). The global statistical measures PB, R^2, ME and CF are 1.2, 0.89, 0.89 and 0.11, respectively (Table 3). It means that the performance scores are also either “excellent” or “very good” in the bottom layer.

Location for Fig. 2

Location for Fig. 3

3.2 Salinity

In the surface layer, the model results reproduce the observed seasonal variability in south of 59°N, i.e. stations A-K, where salinity is higher than 6.0 psu (Fig. 4). No salinity observations are available at stations L and M. At stations N-R , the mean values of model results are close to those of observations, but the model cannot reproduce the fine seasonal dynamics which is mostly smaller than 1.0 psu. The surface statistical measures PB, R^2, ME and CF are -1.1, 0.96, 0.96 and 0.05, respectively (Table 2).

In the bottom layer, seasonal cycle is not visible (Fig. 5). The fit between model results and observations is quite similar as in surface layer (Fig. 5). The temporal profile of model results (Fig. 13b) matches that of observations in general (Fig. 13h). The observed halocline depth is around 60 m, while the modeled one varies between 40 m and 80 m. The spatial mean of the salinity observations is caught perfectly by the model (Fig. 14d). The mean of the observations at one depth plane is also well reproduced (Fig. 15d). Regarding the spatial distribution of the model errors, the percentage bias of the model to observation is mostly smaller than ± 5% (Fig. 16 d). The model generally has positive biases in coastal regions, but negative biases in offshore regions. The model bias can be larger than ± 10% in the Bothnian Bay. The global statistical measures PB, R^2, ME and CF are -2.2, 0.98, 0.98 and 0.02, respectively (Table 3).

Location for Fig. 4

Location for Fig. 5

3.3 DIN

In the surface layer, the model results at all the 18 stations reproduce the observed seasonal variability, high values during winter and low values during summer (Fig. 6). For winter nutrients, the model underestimates the surface DIN in the western Baltic Sea (stations A-D) and Gulf of Finland (stations L-O) but with a fine match in the central Baltic Sea (stations E-K), Bothnian Sea and Bothnian Bay (stations P-R). Notably, the underestimation of DIN decreases from Eastern Skagerrek to Kategatte and Arkona Basin (stations A-D). The timing of abrupt DIN consumption in model results is consistent with that in observations at the deep water stations G-K, but later than that of observations in coastal stations A-F, M-P and R. The surface statistical measures PB, R^2, ME

11

and CF are -58, 0.10, 0.04 and 19, respectively (Table 2). The performance indicators, however, show the model quality of surface DIN quality is “poor” (Table 4) although as shown above, the modeled surface DIN does reproduce many important measured features at the 18 stations.

In the bottom layer, seasonal pattern of DIN varies between stations. Clear pattern is found in the stations north of 59N (L-R), with high values in winter and low values in summer. No clear seasonal change patterns can be identified in stations A-K. The model results are close to the observed seasonal variations at the shallow water stations C, D, M, O, P and Q, and reproduce the basic seasonal pattern at stations B, L, N and R, but are rather off at deep stations A and F-K (Fig. 7). It is noted that the overestimation of the bottom DIN is only found in the central Baltic Sea (stations G-K). At the shallower stations, the model estimates mean DIN well except for a underestimation of the winter DIN in Golf of Finland (stations L-N). The temporal evolution of the vertical profile at station I shows that the model can reflect the observed seasonal variations only in the upper 20 m. Model results for DIN (Fig. 13c) are much higher than observations in layers 80 m below (Fig. 13i). The seasonal variation is less than that of observations (Fig. 14e). The model generally underpredicts DIN above 30 m, but overpredicts below 60 m (Fig. 15e). The model bias has a clear horizontal pattern (Fig. 16e). Negative model bias mainly appears in the Danish Straits, the Polish coasts, the Gulf of Finland and the Finland coasts, while large positive model bias appears in the western Baltic proper and the western Bothnian Sea. The global statistical measures PB, R^2, ME and CF are 26, 0.07, -0.18 and 2.24, respectively (Table 3)), which is “poor” for ME, “reasonable” for CF and “good” for PB (Table 4).

Location for Fig. 6

Location for Fig. 7

3.5 DIP

12

In the surface layer, the model reproduces the basic seasonal variation pattern, with high values during winter and low values during summer at all the 18 stations (Fig. 8). The model results match observations at offshore stations E-K, and can only follow the basic seasonal pattern but not resolve the detailed variations at the coastal stations M-P. The model errors of the surface DIP are similar to that of the surface DIN. The winter DIP peak values are underestimated in coastal stations A-D and N-O. The surface statistical measures PB, R^2, ME and CF are -4.7, 0.35, 0.33 and 1.3, respectively (Table 2), which implies that the model quality is “good” to “excellent” for the surface DIP in terms of the performance indicators in Table 4.

In the bottom layer, the model results are close to observations and can reproduce the observed seasonal variability at most of stations, except coastal stations A, J, L and R (Fig. 9). The temporal evolution of vertical profile shows that the model can reproduce the observed seasonal variability in upper 20 m (Fig. 13d, j) and the model results are close to observations in layers 80 m below. The seasonal pattern of model results mostly follows that of observations, except that the model underpredicts DIP during winter (Fig. 14c). The model results match well the observations in vertical profiles (Fig. 15c). The horizontal distribution of model bias is featured with large positive values in the Bothnian Sea and the Bothnian Bay (Fig. 16c). The highest PB is up to 100 and even higher. The global statistical measures PB, R^2, ME and CF are -2.2, 0.87, 0.86 and 0.22, respectively (Table 3). This indicates that overall performance of the model in simulating DIP is “excellent” (Table 4).

Location for Fig. 8

Location for Fig. 9

3.6 Chl *a*

In the surface layer, the model reproduces the basic seasonal variation pattern with 2 or 3 bloom peaks during April to October and a recession during November to February (Fig. 10). The model's bloom peak values are generally larger than 3 mg m^{-3} and the recession values are smaller than 1 mg m^{-3}, which are close to those of observations. The surface statistical measures PB, R^2, ME and CF are -14, 0.06, 0.03 and 6.9, respectively (Table 2), which gives a "good" performance in terms of PB and "poor" in ME and CF (Table 4).

The model results show that Chl *a* mostly appear in the upper layer 30 m above (Fig. 13e), in agreement with observations (Fig. 13k). The temporal evolution of the vertical profile of observations is quite complex, which the model fails to reproduce. The spatial means show that the general seasonal evolution of model results is close to that of observations, but the model underpredicts spring bloom peak, especially in year 2008 (Fig. 14b). The overall vertical profile of model results is quite consistent with that of observations (Fig. 15b). The model results have positive biases in the Danish Straits, the Gulf of Finland and the Bothnian Bay, and negative bias in the Baltic proper (Fig. 16b). As Chl *a* appears mainly in the upper layers 20 m above, the global statistical measures are close to the surface statistical measures. The global statistical measures PB, R^2, ME and CF are -14, 0.15, 0.11 and 3.09, respectively (Table 3), which means "very good" in terms of PB, but "poor" in ME and CF.

Location for Fig. 10

3.7 DO

In the surface layer, model results are generally consistent with observations at all the 18 stations in terms of seasonal variability (Fig. 11). The consistency seems to decrease with salinity. The model has one month advance of the timing of the seasonal maxima during spring. The surface statistical measures PB, R^2, ME and CF are -4.0, 0.34, 0.21 and 1.2, respectively (Table 2), with performance scores ranging from "very good" to "excellent" (Table 4).

14

In the bottom layer, the model reproduces seasonal variations at shallow water stations, but is rather off at the deep water stations E-K (Fig. 12). The temporal evolution of the vertical profile shows that the model (Fig. 13f) can reproduce the seasonal variation of observations (Fig. 13l) in the upper 60 m, but diverges in layers 60-120 m. The observed minima within euphotic layers appear subsurface during summer, but the corresponding modeled minima appear at the surface. The modeled summer values (June-October) are generally higher than observations (Fig. 14f). The general vertical profile of model results is close to that of observations, but the biases increase downward below 60 m (Fig 15f). The model errors are mostly smaller than ± 20% (Fig. 16f). Relative large model errors exist in the western Baltic proper and the western Bothnian Sea. The global statistical measures PB, R^2, ME and CF are 4.9, 0.80, 0.77 and 0.36, respectively (Table 3), with performance scores ranging from “very good” to “excellent” (Table 4).

Location for Fig. 11

Location for Fig. 12

Location for Fig. 13

Location for Fig. 14

Location for Fig. 15

Location for Fig. 16

4. Discussions

4.1 Model validity

The comprehensive comparison presented above includes the model-observation pairs in the order of 10^4 for almost every targeted state variable, thanks to the relatively abundant observation network in the Baltic Sea. Though the model-observation comparison is comprehensive, it is not obvious which aspects of model results are valid as the products of operational oceanography. Literally, model validation is a general phrase which might generate confusions sometimes and needs clarifications specifically (Rykiel, 1996; Radach and Moll, 2006). There are no written criteria to judge whether a model is valid for operational oceanography. While we are developing and improving our operational model system, we follow two criteria: that the quantitative model skills should be among the right order of this type of models, and that the model should be able to reproduce major observed features for interested scales.

As values of ecological parameters can differ a lot across systems, various statistical measures have been adopted in assessing model skills in previous studies. The statistical measures CF, ME and PB are applied in the ecological model validation studies nearby the Baltic Sea (Radach and Moll, 2006; Allen et al., 2007; Neumann and Schernewski, 2008; Lewis and Allen, 2009). According to these three statistical criteria (Maréchal, 2004; Radach and Moll, 2006) and the results (Table 3), the model skills for temperature, salinity, DIP and DO are scored either "excellent" or "very good". The model skill for Chl *a* is only scored "very good" of PB criterion, but "poor" according to both CF and ME criteria. The model skill for DIN is scored "good" of PB criterion, "reasonable" of CF criterion, but "poor" according to ME criterion. Although same "scores" do not always mean same level of model performances, the statistical measures provide a possibility to inter-compare skills across models applied in different regions. In comparison with other models in the Baltic Sea and nearby regions, the overall skills of this model system are at the same level of this type of models (Edelvang et al., 2005; Lacroix et al., 2007; Lewis and Allen, 2009, Almroth and Skogen 2010).

4.1.1 Model validity of seasonal variability in surface

Observations show spring blooms start in March and last to late April or early May. The system is featured with abrupt nutrient consumption for both DIN and DIP and a similar abrupt increase of phytoplankton biomass. The model captures these features (Fig. 6, 8, 10), although there is some timing delay at stations outside of the Baltic proper. After spring blooms until late October or early November, surface DIN remains depleted at most of stations, surface DIP however is only depleted for a rather short duration at the shallow water stations, but continuously decreases and then gradually recovers from July at the deep water stations E-K. In autumn, the system is featured with abrupt nutrient recovery by wind mixing and autumn blooms of phytoplankton. During winter, nutrient concentrations remain high and phytoplankton biomass remains low. These features are mostly captured by the model (Fig. 6, 8, 10).

The model-observation biases of Chl a in surface layer seems unusually high in summer at Stations O and P, meanwhile the observed Chl a is unusually low (Fig. 10). The satellite detected Chl a (http://marcoast.dmi.dk/chlorophyll.php) is used as another reference. The modeled Chl a is compared with the satellite detected Chl a (Fig. 17). Both the modeled and satellite detected Chl a are mostly higher 4 mg m^{-3} in June and July at those two stations, but the observational Chl a is lower than 2 mg m^{-3}, which is unusual in summer. We think the observations at those two stations might be problematic. The additional comparison provides also a reference for stations where in-situ observations are missed, e.g. at Station Q, and Station E in 2007. All in all, the modeled Chl a is quite consistent with the satellite detected Chl a, except for winter months. In winter, the satellite detected Chl a is generally poor and much discrepancy with observations.

Location for Fig. 17.

4.1.2 Model validity of vertical profile

The model generally reproduces the observed vertical profiles except poorly for DIN (Fig. 15). The temporal evolution of vertical profiles at the Gotland Deep station I shows that the model's vertical profiles are close to the observed ones, although there is a lot of fine difference (Fig. 13). For example, the maximum vertical gradient appears at depth of 60 m for observations (Fig. 13b, c, d, f), but the corresponding model position is at depth of 80 m (Fig. 13h, i, j, l). It means the vertical profiles of model at a specific station are not always consistent with observations, however, the overall pattern of vertical profiles are generally good. We think that the model errors at different horizontal locations probably cancel out greatly.

4.2 Model errors and likely causes

4.2.1 Insufficient light penetration

The model underestimates the amplitudes of seasonal variations for Chl *a*, DIN, DIP and DO (Fig. 14b, c, e, f). In details, the model underestimates the seasonal maxima for Chl *a*, DIN, DIP, but overestimates the seasonal minimum for DO. We think the insufficient light penetration is the main cause. The observed DIN is depleted down to 40-60 m (Fig. 13c), but the model results show DIN depletion is only down to 30 m and the duration of DIN depletion is shorter. The insufficient light penetration leads to underestimation of nutrient uptake and phytoplankton biomass. It means the primary production is underestimated, thus the maximum DO concentration during spring blooms is underpredicted (Fig. 11).

4.2.2 Bottom layer vulnerability in deep water areas

The model results reflect a model vulnerability in bottom layer in deep water areas, i.e. in the Gotland deep. The first, the modeled bottom salinity are continuously decreasing at Stations I and J, but there are no clear decreasing trends in observations (Fig. 5, i, j). The second, the observed bottom DIN at the Gotland deep (Station I) has an obvious increasing trend from May of 2007 to July of 2008, however, the corresponding model results show a decreasing trend (Fig. 7, i). The likewise model-observation discrepancy occurs as to DIP (Fig. 9, i). The third, the observed bottom DO shows a decreasing trend, however, the corresponding model results show an increasing trend (Fig. 12, i). The negative DO gets larger and larger, meaning hydrogen sulphide was taking place.

18

The main cause for this model vulnerability is due to the improper vertical grid. Although the model has 109 vertical layers fro the Baltic Sea (Table 1), they are arranged: 2 m for the surface layer, 1 m for each of the following 98 layers, and 3 m, 6 m, 8 m, 16 m, 25 m for the 100-104th layer respectively, and 50 m for each of the rest 5 layers. The thickness of bottom layer at both Stations I and J are 50 m. At first, the too thick bottom layer introduced errors in the initialization, as we see the initial bottom DO was set positive due to grid interpolation (Fig. 12, i). Actually, the initial bottom nitrate was also wrongly set much higher than observation for same reason (not presented). The model results in the bottom layer at Station I reflect that the dead organic detritus was remineralized first through consuming the positive DO and then through oxidizing the wrongly initialized high nitrate. In fact, the real remineralization was occurring through oxidizing sulphide, as the negative DO increased. The second, the too thick bottom layer diluted the effects of water-sediment flux on the bottom water. That's why the modeled dynamics in the bottom layer is slow, not comparable to the observed dynamics. The third, too thick bottom might not accurately reproduce the hydrodynamics, as we see the model-observation discrepancy for salinity (Fig. 5, i, j). Inaccurate hydrodynamics could also exacerbate the model biases.

If the initialization errors are negligible and the real variations are not dramatic, the model can follow observations in the bottom layer in deep water areas, as we see at Stations J and K (Fig.s 7, 9, 12). It means the model does not include fundamental errors. This supports the speculation that the model vulnerability failed to recaptured the observed biogeochemical dynamics at the Gotland deep was mainly caused by the improperly coarse vertical grid. On the other hand, there might exist another possibility: the remineralization rate under anoxic condition might also be slower than the reality.

4.2.3 Insufficient regional adaptation

Although the horizontally variable N/P ratio improves the model adaptation for different regions (Wan et al., 2012), the model shows better performance in offshore regions than in coastal regions, and better in the Baltic proper than outside (Fig. 16). The model shows the best performance for the deep water stations (F-K). This might be caused by the parameter values being tuned for the Baltic proper (Neumann, 2000; Neumann et al., 2002). The model's regional adaptation can be further improved by allowing more parameters to vary regionally and refining the boundary inputs, like river loadings. Modeled spring blooms at stations outside of the Baltic proper occur later than observed. Suspended particles are reported influential for the timing of spring blooms (Tian et al., 2010).

4.2.4 Uncertainties in forcing and initialization

One of the major model errors in DIN and DIP occur in coastal regions influenced by the river runoff (station A-E, L-O in Figs. 6, 8 and 16). The river nutrient loading used in this study is based on mainly the HBV model output. Due to lack of observations, a detailed validation of river loading may not be feasible. Moreover, only big rivers are included. Recent study found that small rivers may have a significant contribution to the total river nutrient loading to the Baltic Sea (unpublished). For ecological modeling, including nutrient loads from smaller rivers will improve not only the total amount of nutrient inputs to the Baltic Sea but also the locations of the riverine nutrient sources.

Some impacts from improper initial conditions may last for quite a long period, even for the whole simulation duration, especially in deep areas and near bottom. For example, the large initial errors for bottom DIN and DO at stations G, J, K last for quite a long period (Fig. 7 and 12). The comparison between vertical profiles of model results and those of observations reflects obvious differences for DIN and DO at the beginning of simulation. The initial model errors only decay slowly (Fig 13c, f, i, l). The strong permanent stratification of salinity of observations is located at the depth of 60 m, while the corresponding stratification of model results is at the depth of 80 m, none of them even changes at all during two years of simulation (Fig. 13b, h). This might reflect that insufficient vertical mixing slows down the initial errors decaying.

20

4.3 Assessment schemes

Statistical measures and point-to-point comparison are the common schemes to assess model skills (Lacroix et al., 2007; Lewis and Allen, 2009; Ruzicka, 2011). Statistical measures can use all available data and avoid subjective involvement in selecting observed data. However, there are two caveats that we must be aware of. First, statistical measures cannot ensure a proper representation for each observed data. For example, the statistical measures show the model-observation fit is rather poor for DIN in surface (Table 2), however, the point-to-point comparison shows that model results can reflect the basic seasonal variability (Fig. 6). This inconsistency is caused by extreme outliers in data set, like the data from estuaries. In some other cases, equal representation of each data is not reasonable. For example, two observations respectively from densely and sparsely sampled areas (in time or space) should not equally contribute to the spatial mean. Second, statistical measures are usually used to show the overall model skill, rather than describe model skills along different dimensions. The point-to-point comparison is very effective to analyze the model performance at the selected station, especially to evaluate model robustness to reproduce a certain dynamic process, provided time-series of observed data. The shortcoming of the point-to-point comparison includes the following four aspects. First, the point-to-point comparison has a limited representation, as the ecological properties can differ a lot in various sub-regions. Second, the point-to-point comparison is limited to the stations with time-series of data, but other data, e.g. those from cruises will not be used. Third, it is inevitable to have subjective involvement in selecting stations and layers, which is necessary for model developer’s sake of good representation to analyze model performance, but not appreciable for users/customers who are interested in an objective assessment of the quality of the operational products. Finally, it is inconvenient to implement a point-to-point comparison at too many stations.

The comprehensive comparison scheme (Wan et al., 2011) uses all available observations in the entire model domain. This scheme deploys a grid in the spatial-temporal domain to properly

21

distribute data representations. The gridded data from all resources makes it possible to analyze the model skills along different dimensions (Fig. 14, 15, 16). There is no subjective involvement in selecting data. Thus, the comprehensive validation scheme can provide a relatively rigorous and throughout assessment of model skills along different dimensions. However, the comprehensive validation scheme will only be effective for systems with abundant observations. Thus, the comprehensive validation cannot replace the point-to-point comparison. It is important to deploy the traditional point-to-point comparison and statistical measures along with the comprehensive validation in order to assess model skills quantitatively.

5. Summary

Following the inter-comparison experiments of the MyOcean project, the model system with the latest feature (Wan et al., 2012) is assessed for its skills in providing biogeochemical information service. The abundant observation data in the Baltic Sea allow us to implement a comprehensive model validation scheme, which makes use of all available observation data to assess model skills along each dimension. The comprehensive model validation scheme combined with the traditional point-to-point comparison and statistical measures makes it possible to provide a relatively rigorous assessment of model skills and to identify the major model errors and the main causes behind. According to criteria used in the Baltic Sea and nearby regions (Maréchal, 2004; Radach and Moll, 2006), model skills for temperature, salinity, DIP and DO is scored either “excellent” or “very good”. The model skill for Chl *a* is only scored “very good” on the PB criterion, but “poor” according to both CF and ME criteria. The model skill for DIN would be scored “good” on the PB criterion, “reasonable” on the CF criterion, but “poor” according to the ME criterion.

This assessment reflects that the model errors are mainly caused by insufficient light penetration, excessive organic particle export downward, insufficient regional adaptation and uncertainties in riverine nutrient loading, physical forcing and initial fields. This study highlights the importance to apply multiple schemes (the comprehensive validation scheme, the point-to-point comparison and

22

the statistical measures) in order to assess model skills rigidly and to identify main causes for major model errors effectively.

Acknowledgements

We would like to thank Per Berg for technical assistance with the HBM model code and setups, and to thank the Swedish Meteorological and Hydrological Institute and the Bundesamt für Seeschifffahrt und Hydrographie in Hamburg, Germany for providing the river data. This work was supported by European Commission FP6 and FP7 projects ECOOP (Contract No. 036355), MYCOEAN (Contract No. 218812) and MEECE (Contract No. DK18159104).

23

References

Allen, J. I., Holt, T. J., Blackford, J., and Proctor, R.: Error quantification of a high-resolution coupled hydrodynamic-ecosystem coastal-ocean model: Part 2. Chlorophyll-a, nutrients and SPM, J. Marine Syst., 68, 381–404, 2007.

Almroth, E. and Skogen, M.D.: A North Sea and Baltic Sea model ensemble eutrophication assessment, Ambio. 39, 59-69, 2010.

Berg, P. and Poulsen, J. W.: Implementation details for HBM. DMI Technical Report No. 12-11, ISSN: 1399-1388, Copenhagen, 2012.

Bergström, S.: Development and application of a conceptual runoff model for Scandinavian catchments, Ph.D. thesis., SMHI Reports RHO, No. 7, Norrköping, 1976.

Bergström, S.: The HBV model – its structure and applications, SMHI Reports RH, No. 4, Norrköping, 1992.

Conkright, M. E., Locarnini, R., Garcia, H., O'Brien, T., Boyer, T. P., Stephens, C., and Antonov, J.: World ocean atlas 2001, objective analyses, data statistics and figures, CDROM documentation, National Oceanographic Data Center, Silver Spring, MD, 2002.

Edelvang, K., Kaas, H., Erichsen, A. C., Alvarez-Berastegui1, D., Bundgaard, K., and Jørgensen, P. V.: Numerical modeling of phytoplankton biomass in coastal waters, J. Marine Sys., 57, 13–29, 2005.

Eilola, K., Meier, H. E. M., and Almroth, E.: On the dynamics of oxygen, phosphorus and cyanobacteria in the Baltic Sea: a model study, J. Marine Sys., 75, 163–184, 2009.

Fennel, W.: Model of the yearly cycle of nutrients and plankton in the Baltic Sea, J. Mar. Syst. 6, 313-329, 1995.

Fennel, W. and Neumann, T.: The mesoscale variability of nutrients and plankton as seen in a coupled model, Ger. J. Hydrogr. 48, 49-71, 1996.

Fennel, W. and Neumann, T.: Variability of copepods as seen in a coupled physical biological model of the Baltic Sea, ICES Marine Science Symposia 219, 208-219, 2003.

Janssen, F., Neumann, T., and Schmidt, M.: Inter-annual variability in cyanobacteria blooms in the Baltic Sea controlled by wintertime hydrographic conditions, Mar. Ecol. Prog. Ser. 275, 59-68, 2004.

Kuznetsov, I., Neumann, T., and Burchard, H.: Model study on the ecosystem effect of a variable C:N/ P ratio for cyanobacteria in the Baltic Proper, Ecol. Model., 219, 107–114, 2008.

Lacroix, G., Ruddick, K., Park, Y., Gypens, N., and Lancelot, C.: Validation of the 3D biogeochemical model MIRO&CO with field nutrient and phytoplankton data and MERIS-derived surface chlorophyll images, J. Mar. Syst. 64, 66-88, 2007.

Langner, J., Andersson, C., and Engardt, M.: Atmospheric input of nitrogen to the Baltic Sea basin: present situation, variability due to meteorology and effect of climate change, Boreal Environ. Res., 14, 226–237, 2009.

Lewis, K. and Allen, J. I.: Validation of a hydrodynamic-ecosystem model simulation with time-series data collected in the western English Channel, J. Marine Syst., 77, 296–311, 2009.

Maar, M., Møller, E. F., Larsen, J., Kristine, S. M., Wan. Z., She, J., Jonasson, L., and Neumann, T.: Ecosystem modeling across a salinity gradient from the North Sea to the Baltic Sea, Ecol. Model., 222, 1696–1711, 2011.

Maréchal, D.: A Soil-Based Approach to Rainfall-Runoff Modelling in Ungaugaed Catchments for England and Wales, PhD Thesis, Cranfield University. 157 pp, 2004.

Neumann, T.: Towards a 3D-ecosystem model of the Baltic Sea, J. Marine Sys., 25, 405–419, 2000.

Neumann, T.: The fate of river-borne nitrogen in the Baltic Sea: An example for the River Oder. Estuar., Coast. and Shelf S. 73, 1-7, 2007.

Neumann, T., Fennel, W., and Kremp, C.: Experimental simulations with an ecosystem model of the Baltic Sea: a nutrient load reduction experiment, Global Biogeochem. Cy., 16, 7(1)–7(19), 2002.

Neumann, T., Schernewski, G.: An ecological model assessment of two nutrient abatement strategies for the Baltic Sea, J. Mar. Syst. 56, 195-206, 2005.

Neumann, T. and Schernewski, G.: Eutrophication in the Baltic Sea and shifts in nitrogen fixation analyzed with a 3-D ecosystem model, J. Marine Sys., 74, 592–602, 2008.

OSPAR, Villars, M., Vries, I.D., Bokhorst, M., Ferreira, J., Gellers-Barkman, S., Kelly-Gerreyn, B., Lancelot, C., Ménesguen, A., Moll, A., Pätsch, J., Radach, G., Skogen, M., Soiland, H., Svendsen, E., Vested, H.J., 1998. Report of the ASMO modeling workshop on eutrophication issues, 5-8 November 1996. The Hague, The Netherlands. OSPAR Commission Report, RIKZ, The Hague: 102.

Radach, G. and Moll, A.: Review of the Three-Dimensional Ecological Modelling Related to the North Sea Shelf System – part 2: Model Validation and Data Needs. Oceanography and Marine Biology – an Annual Review, vol. 44, pp. 1-60, 2006.

Ruzicka, J.J., Wainwright, T.C., and Peterson, W.T.: A simple plankton model for the Oregon upwelling ecosystem: Sensitivity and validation against time-series ocean data, Ecological Modelling, 222 (6), 1222-1235, 2011.

Rykiel, E.J., Jr.: Testing ecological models: the meaning of validation, Ecol. model. 90, 229-244, 1996.

Savchuk, O. P., Wulff , F., Hille, S., Humborg, C., and Pollehne, F.: The Baltic Sea a century ago – a reconstruction from model simulations, verified by observations, J. Marine Sys., 74, 485–494, 2008.

She, J., Berg, P., and Berg, J.: Bathymetry eff ects on water exchange modeling through the Danish Straits, J. Marine Sys., 65, 450– 459, 2007a.

She, J., Høyer, J., and Larsen, J.: Assessment of sea surface temperature observational networks in the Baltic Sea and North Sea, J. Marine Sys., 65, 314–335, 2007b.

Stigebrandt, A. and Wulff, F.: A model for the dynamics of nutrients and oxygen in the Baltic proper, J. Mar. Res. 45, 729-759, 1987.

Tian, T., Merico, a., Su J., Staneva, J., Wiltshire, K., and Wirtz, K.: Importance of resuspended sediment dynamics for the phytoplankton spring bloom in a coastal marine ecosystem, J. Sea Res., 62, 214–228, 2009.

Wan, Z., Jonasson, L., and Bi, H.: N/ P ratio of nutrient uptake in the Baltic Sea, Ocean Sci., 7, 693–704, 2011.

Wan, Z., Bi, H., She, J., Maar, M., and Jonasson, L.: Model study on horizontal variability of nutrient N/P ratio in the Baltic Sea and its impacts on primary production, nitrogen fixation and nutrient limitation, Ocean Sci. Discuss., 9, 385-419, doi:10.5194/osd-9-385-2012, 2012.

Figure legends

Fig. 1 Topography of the Baltic Sea (unit: m) and location of time-series observational stations A-R (marked with *).

Fig. 2 Seasonal variability of temperature in surface layer

Red solid curve (black dashed cycles) for model results (observations). Unit: °C. Panels A-R for Stations A-R (Fig. 1), respectively.

Fig. 3 Seasonal variability of temperature in bottom layer

Notations same as in Fig. 2.

Fig. 4 Seasonal variability of salinity in surface layer

Red solid curve (black dashed cycles) for model results (observations). Unit: PSU. Panels A-R for Stations A-R (Fig. 1), respectively.

Fig. 5 Seasonal variability of salinity in bottom layer

Notations same as in Fig. 4.

Fig. 6 Seasonal variability of DIN in surface layer

Red solid curve (black dashed cycles) for model results (observations). Unit: mmol m^{-3}. Panels A-R for Stations A-R (Fig. 1), respectively.

Fig. 7 Seasonal variability of DIN in bottom layer

Notations same as in Fig. 6.

Fig. 8 Seasonal variability of DIP in surface layer

Notations same as in Fig. 6.

Fig. 9 Seasonal variability of DIP in bottom layer

Notations same as in Fig. 6.

Fig. 10 Seasonal variability of Chl *a* in surface layer

Red solid curve (black dashed cycles) for model results (observations). Unit: mg m^{-3}. Panels A-R for Stations A-R (Fig. 1), respectively.

Fig. 11 Seasonal variability of DO in surface layer

Notations same as in Fig. 6.

Fig. 12 Seasonal variability of DO in bottom layer

Notations same as in Fig. 6.

Fig. 13 Temporal evolutions of vertical profile in the Gotland deep at station I

Panels A-F for observations of temperature, salinity, DIN, DIP, Chl *a*, DO, respectively; Panels G-R for model results of them. Units: temperature -- °C; Chl *a* – mg m^{-3}; DIN, DIP, DO – mmol m^{-3}.

Fig. 14 Overall pattern of seasonal variability

Red solid curve (black dashed cycles) for model results (observations). Panels A-F for, temperature, Chl *a*, DIP, salinity, DIN, DO, respectively. Units same as in Fig. 13.

Fig. 15 Overall pattern of vertical profile

Notations same as in Fig. 14.

Fig. 16 Horizontal pattern of model's percentage errors

Panels A-F for, temperature, Chl *a*, DIP, salinity, DIN, DO, respectively. Units %.

Fig. 17 Inter-comparison among modeled, satellite detected and in-situ observed Chl *a* in surface layer

Blue solid curve for satellite detected results. Other notations same as in Fig. 10.

Fig. 1

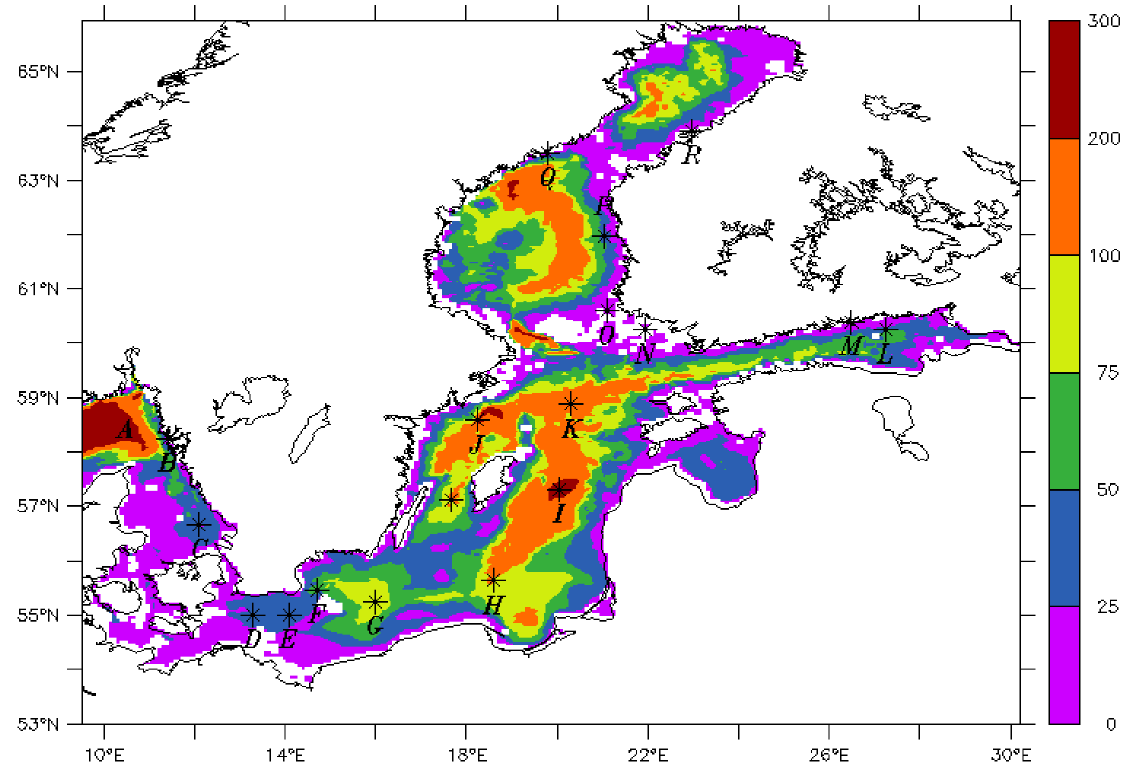

65°N
63°N
61°N
59°N
57°N
55°N
53°N
10°E
14°E
18°E
22°E
26°E
30°E
300
200
100
75
50
25
0
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R

Fig. 2

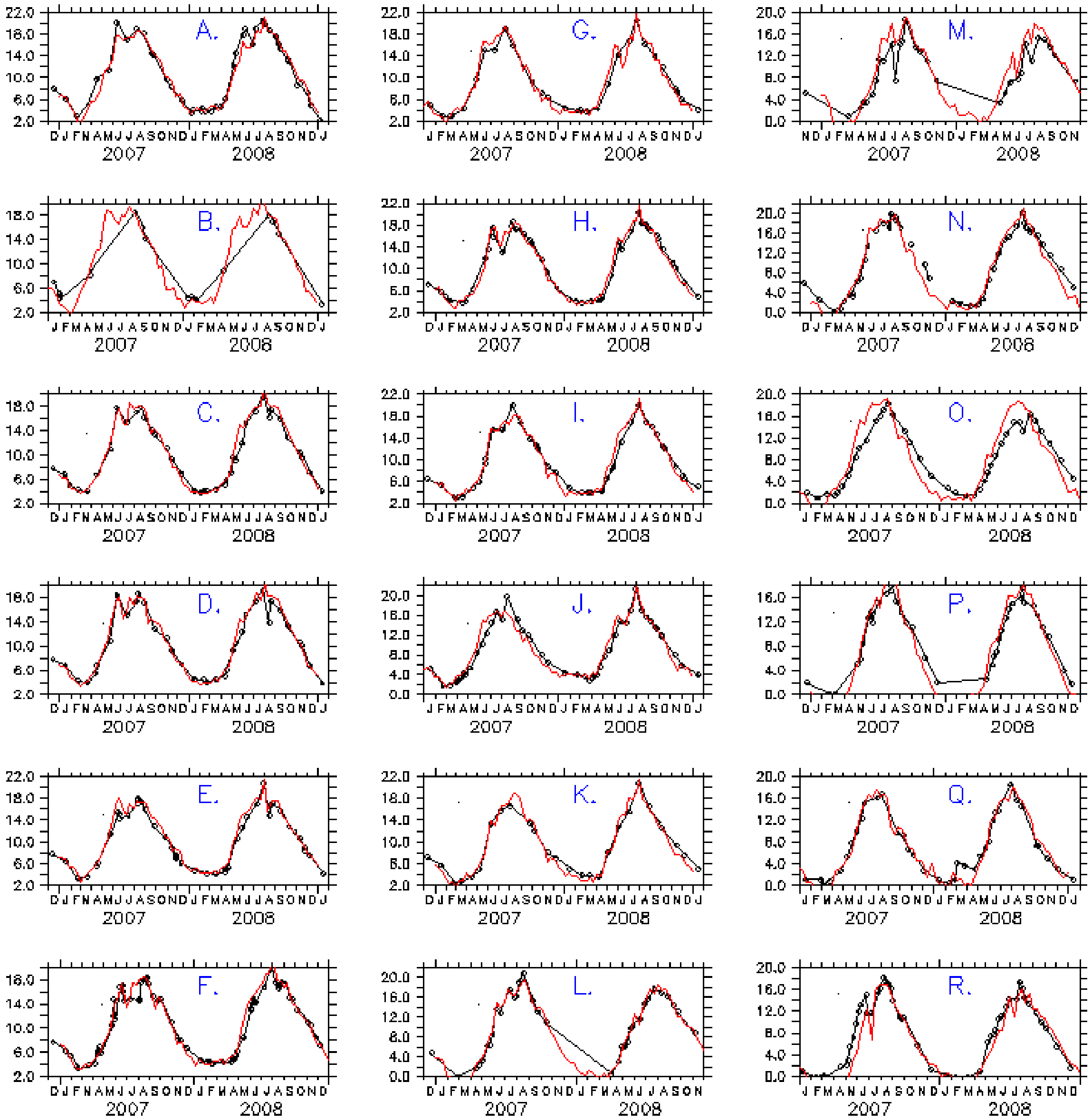


Fig. 3

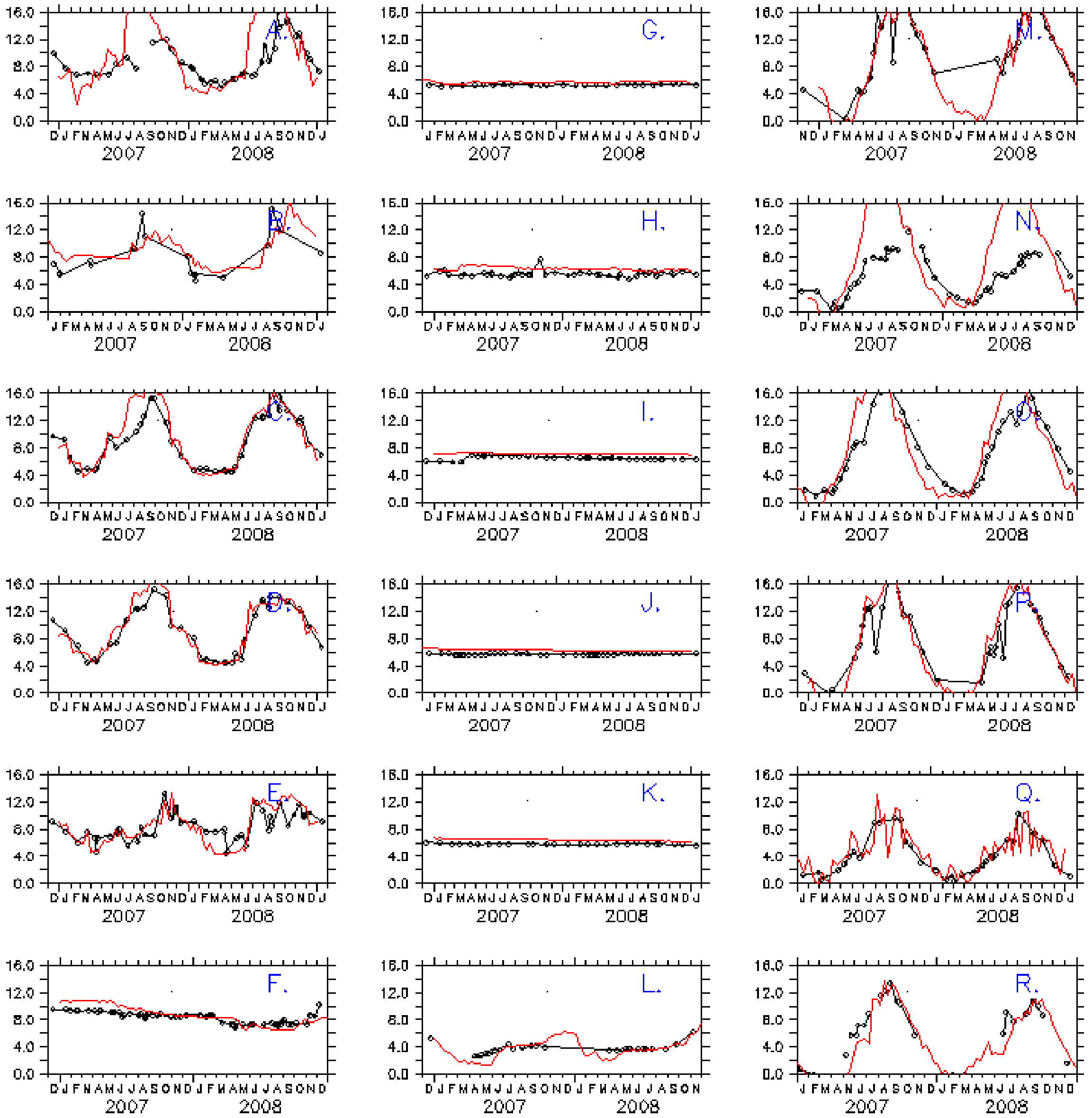


Fig. 4

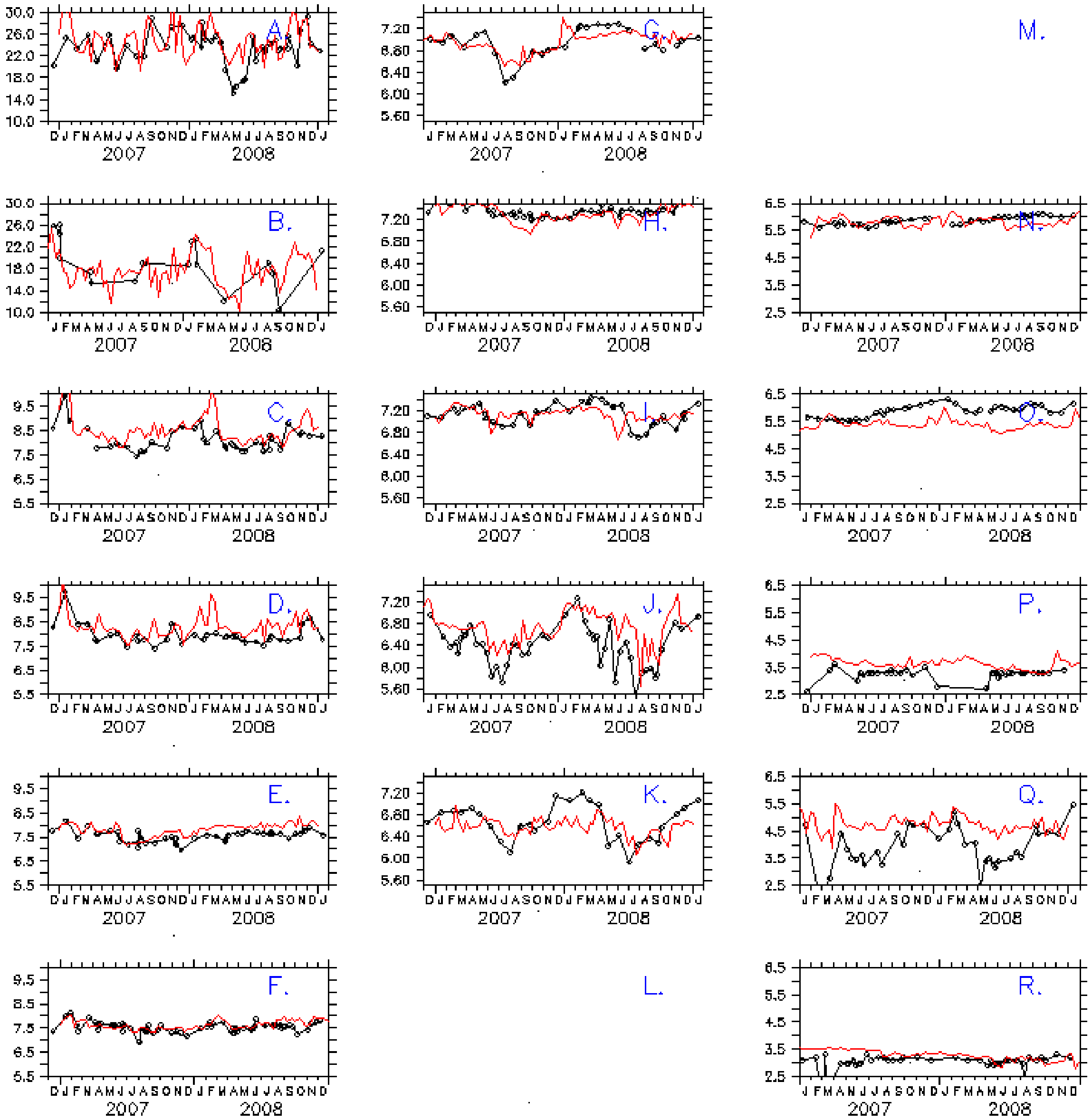


Fig. 5

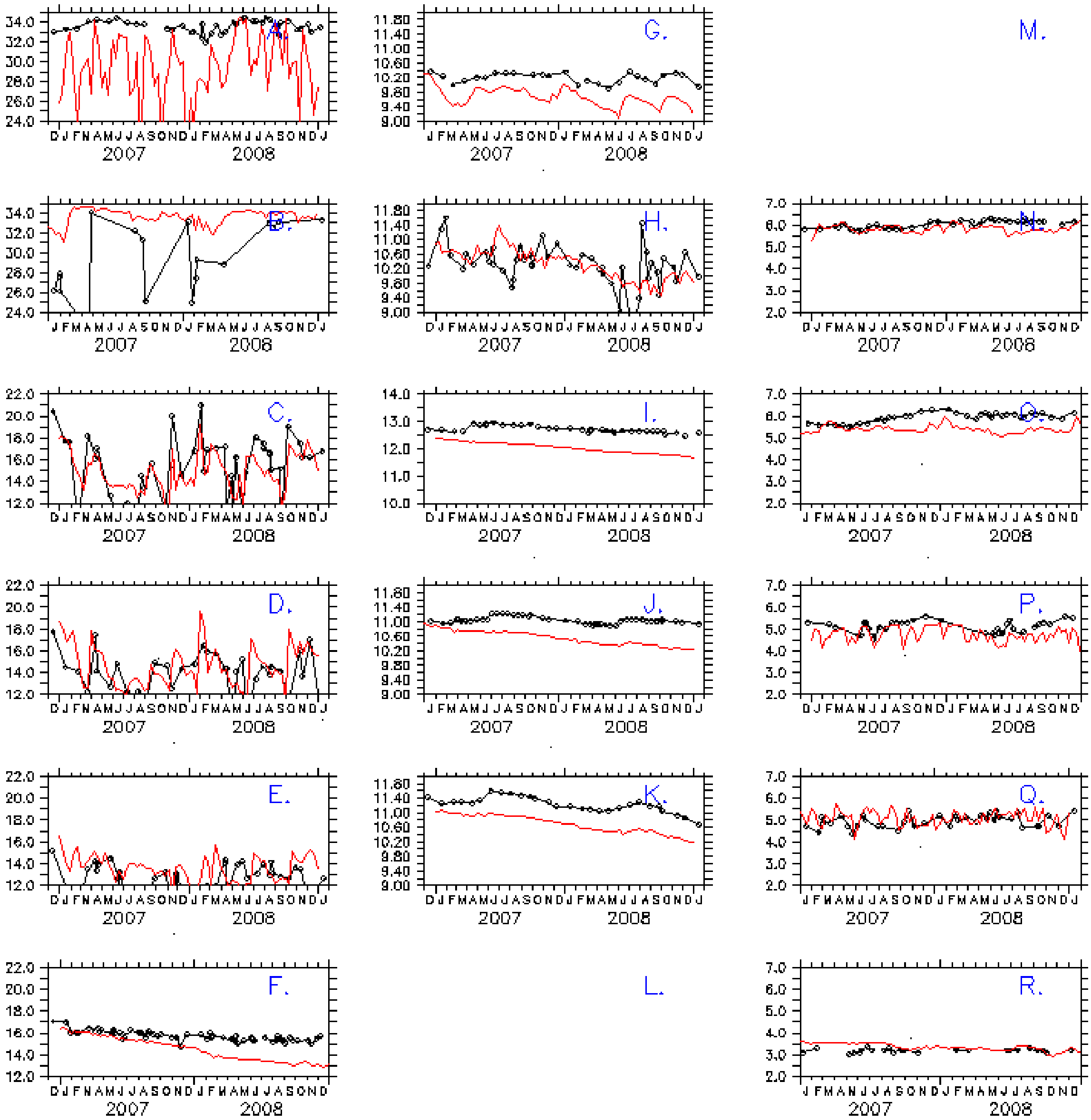


Fig. 6

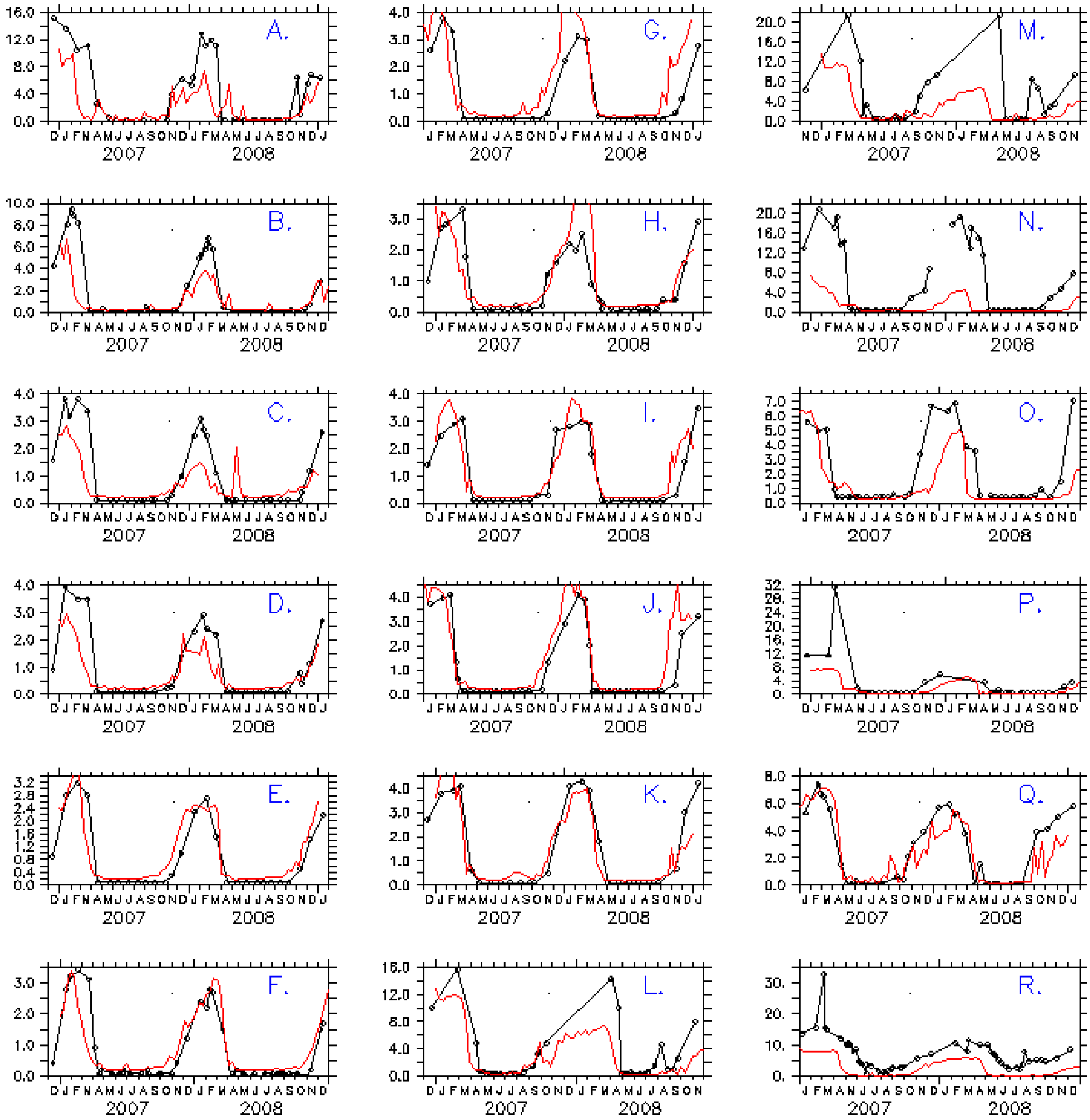


Fig. 7

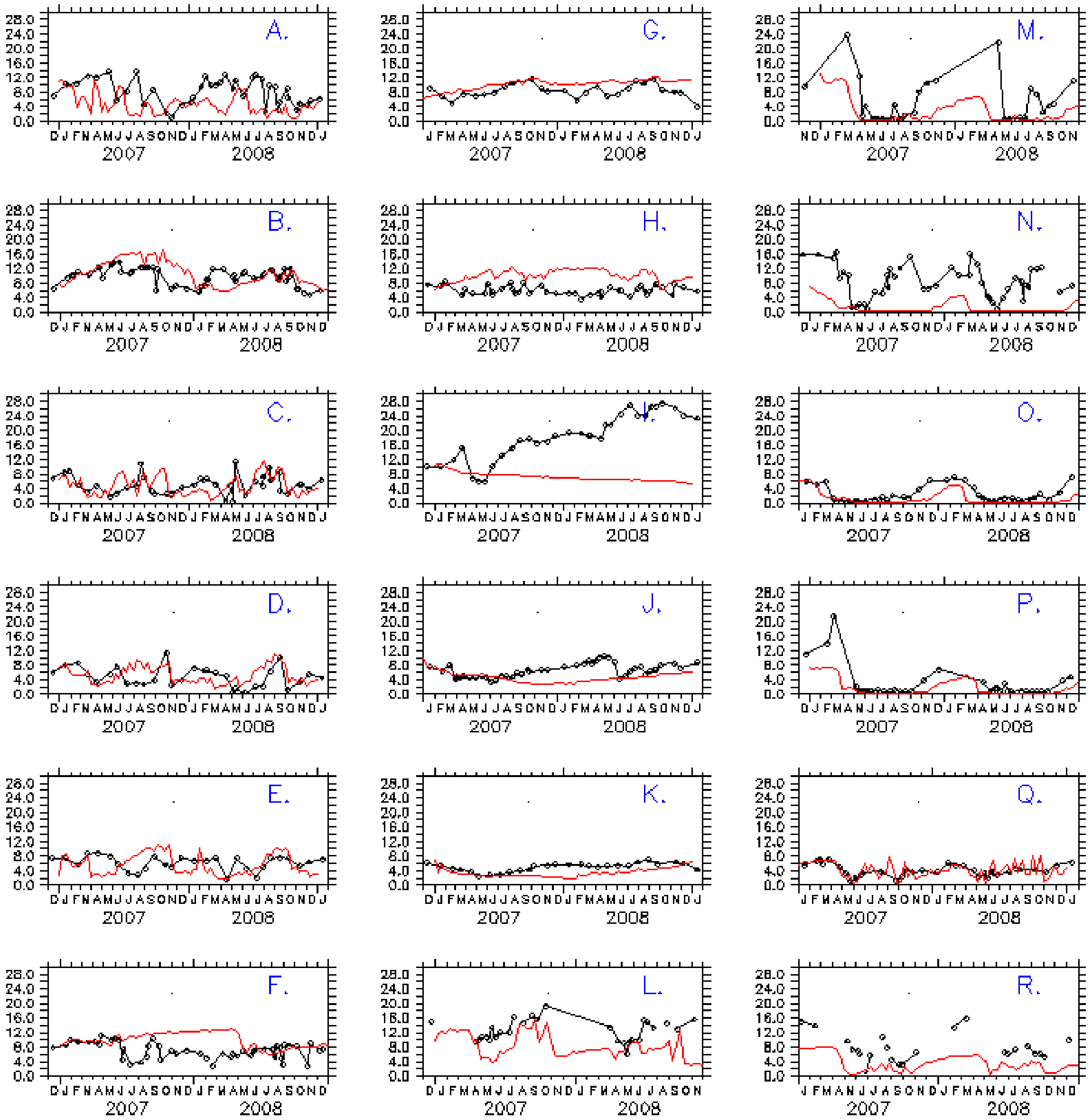


Fig. 8

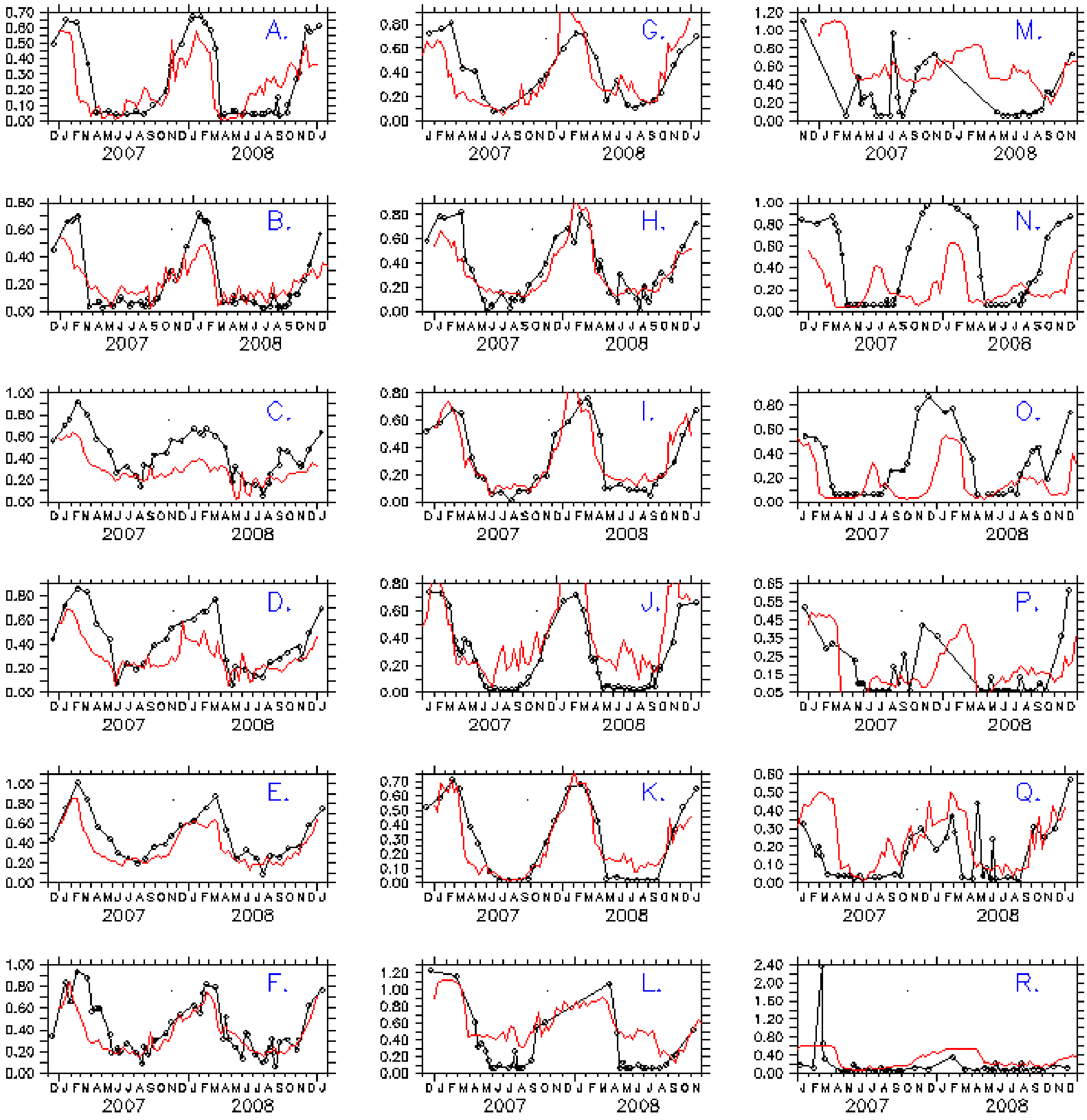


Fig. 9

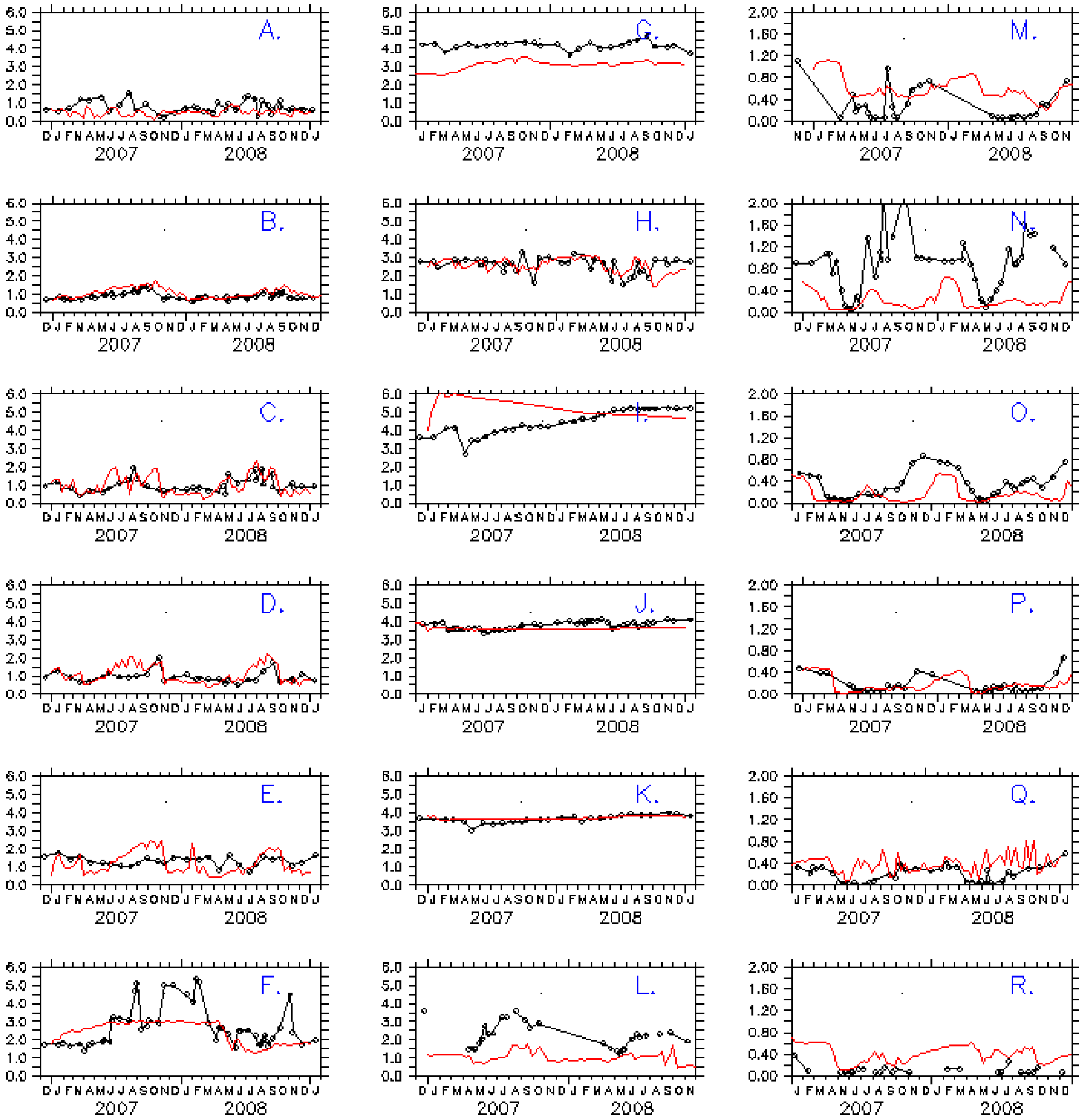


Fig. 10

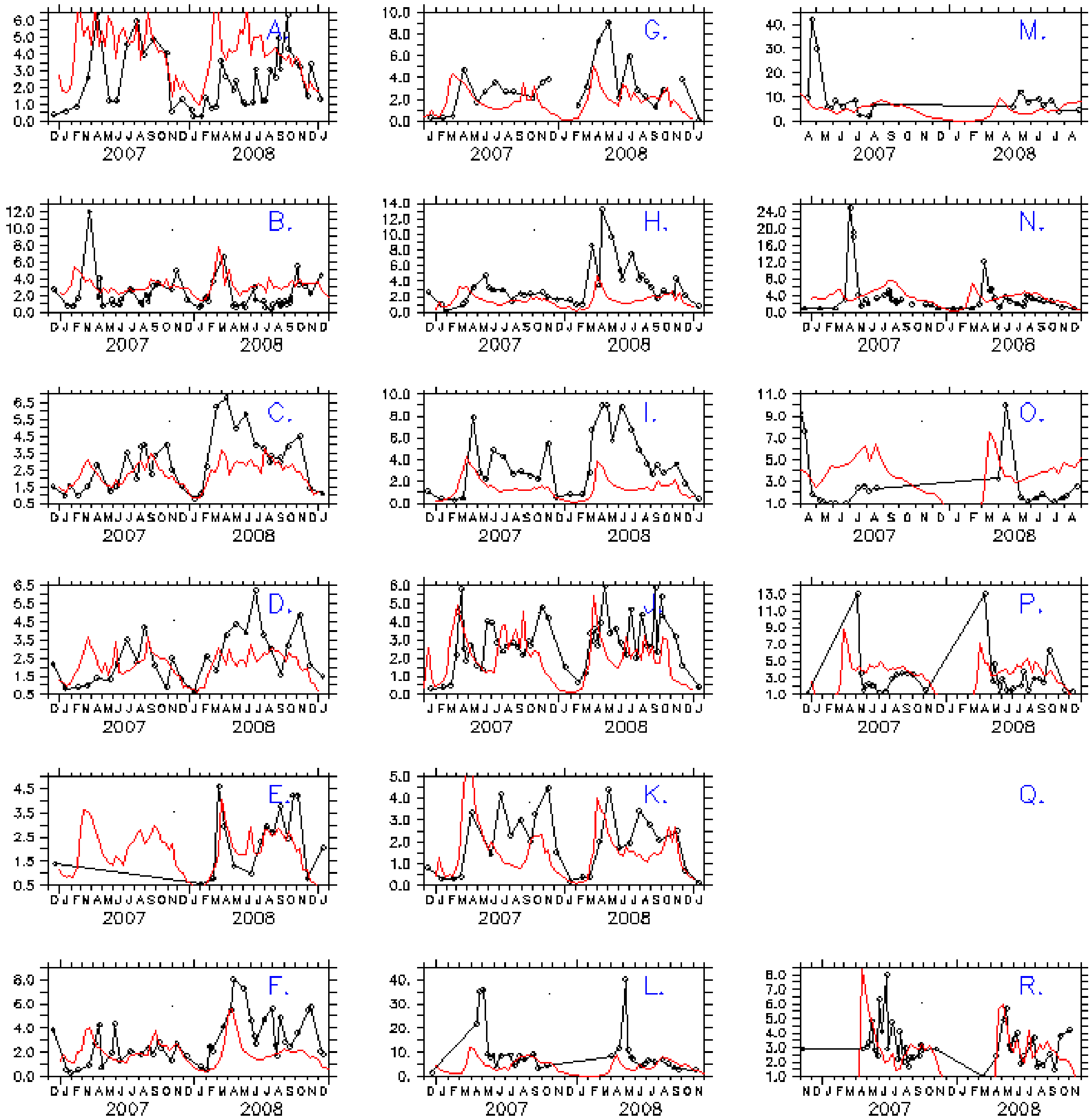


Fig. 11

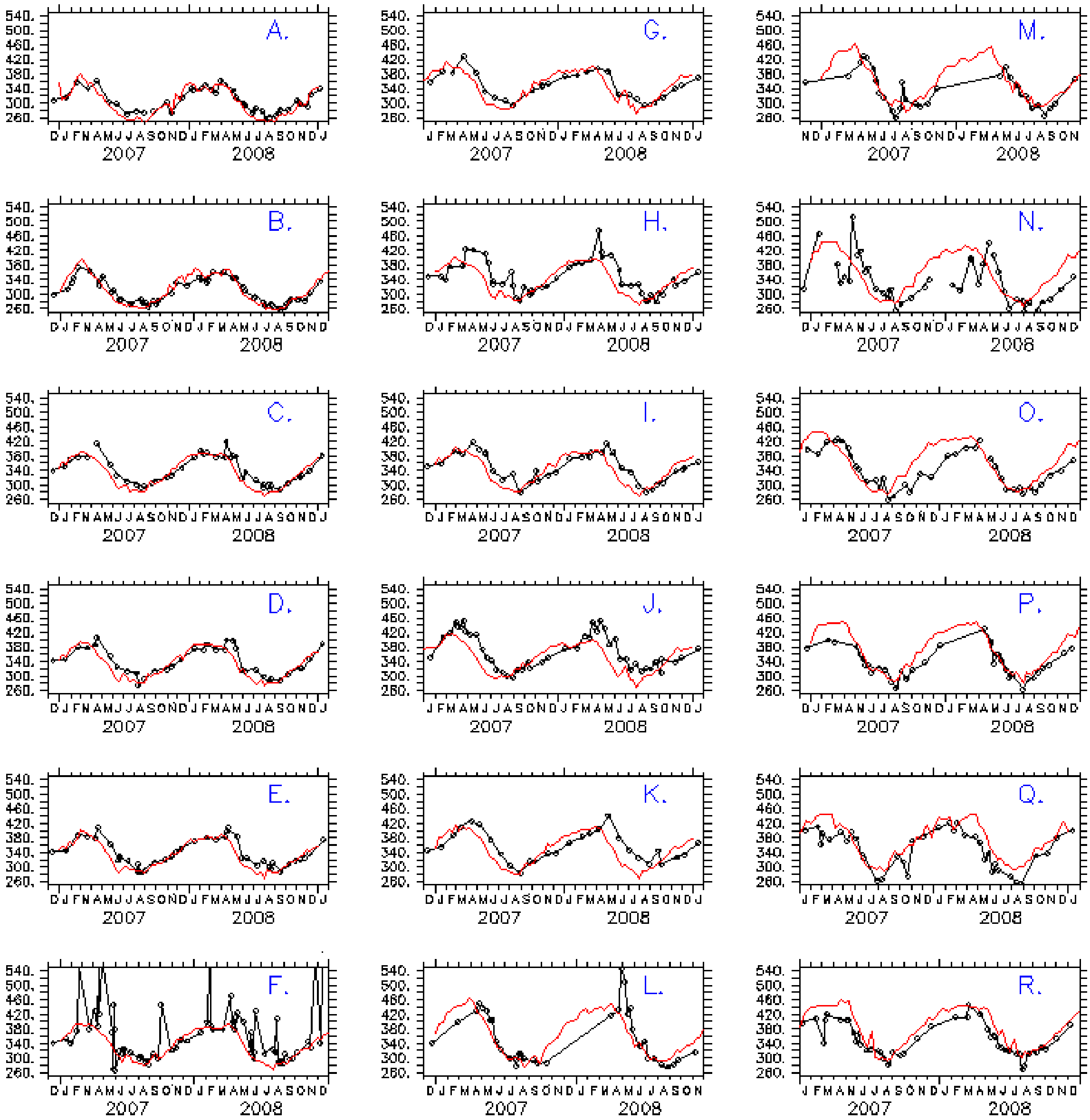

A.
B.
C.
D.
E.
F.
G.
H.
I.
J.
K.
L.
M.
N.
O.
P.
Q.
R.
540.
500.
460.
420.
380.
340.
300.
260.
2007
2008

Fig. 12

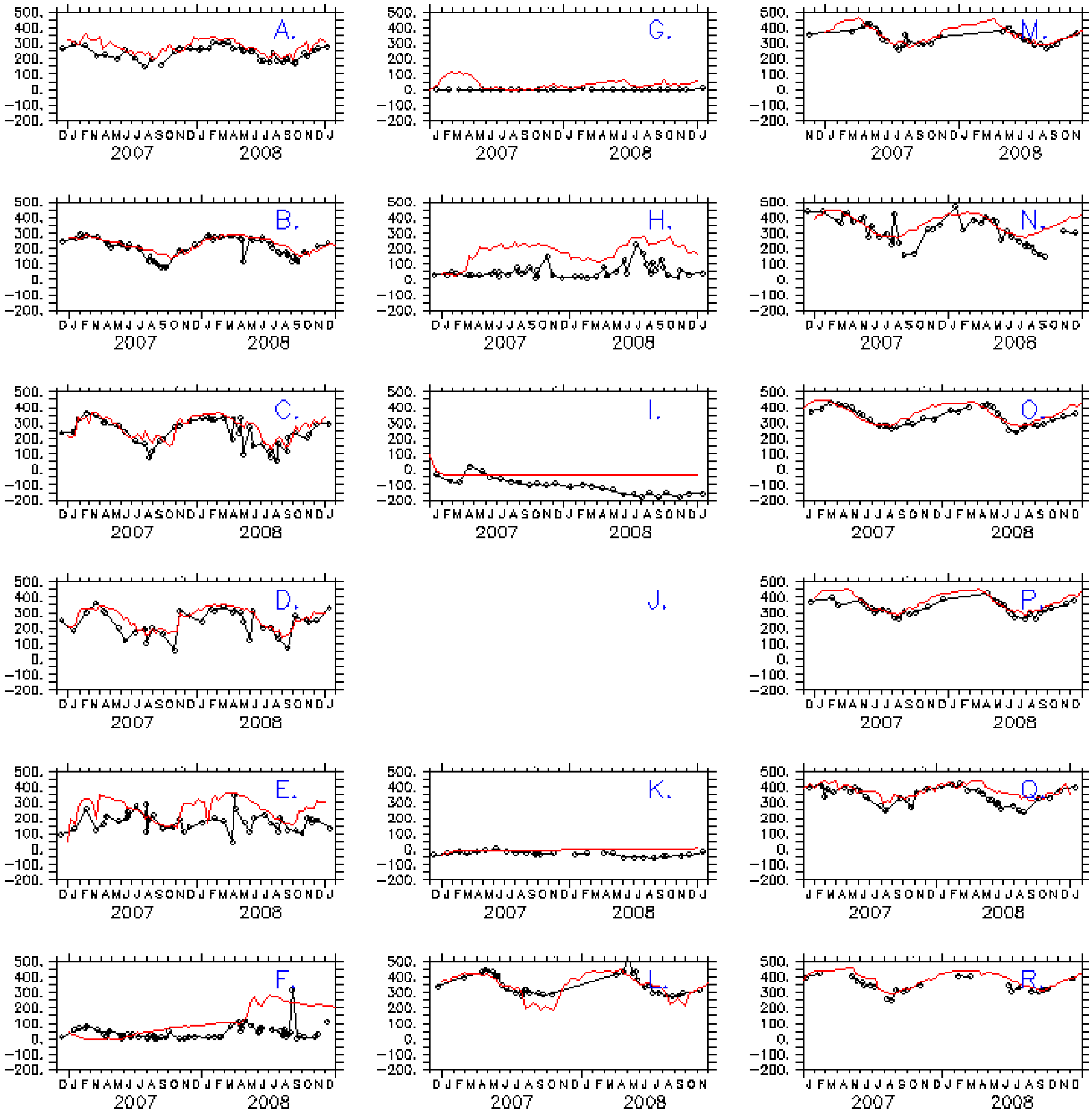


Fig. 13

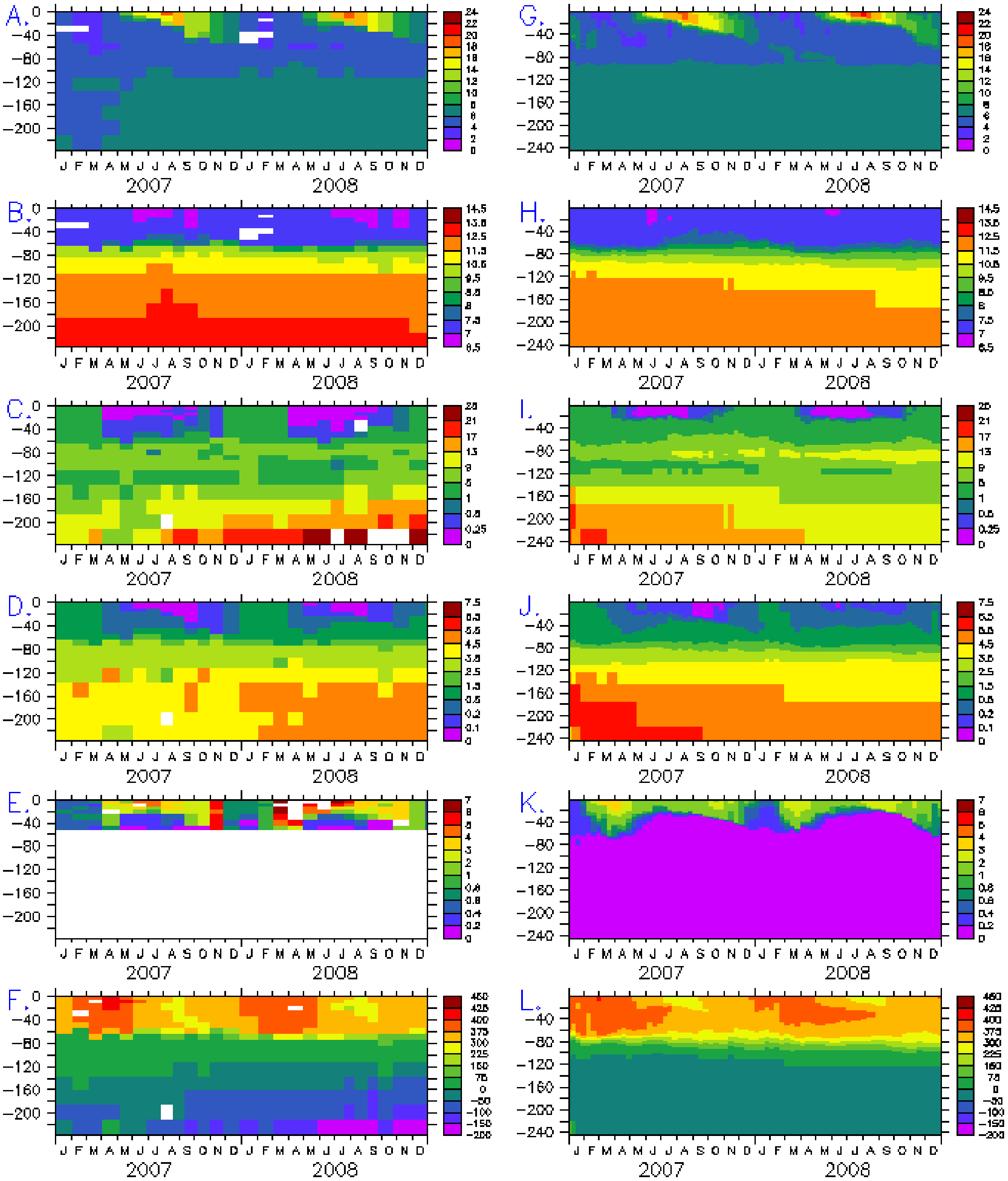
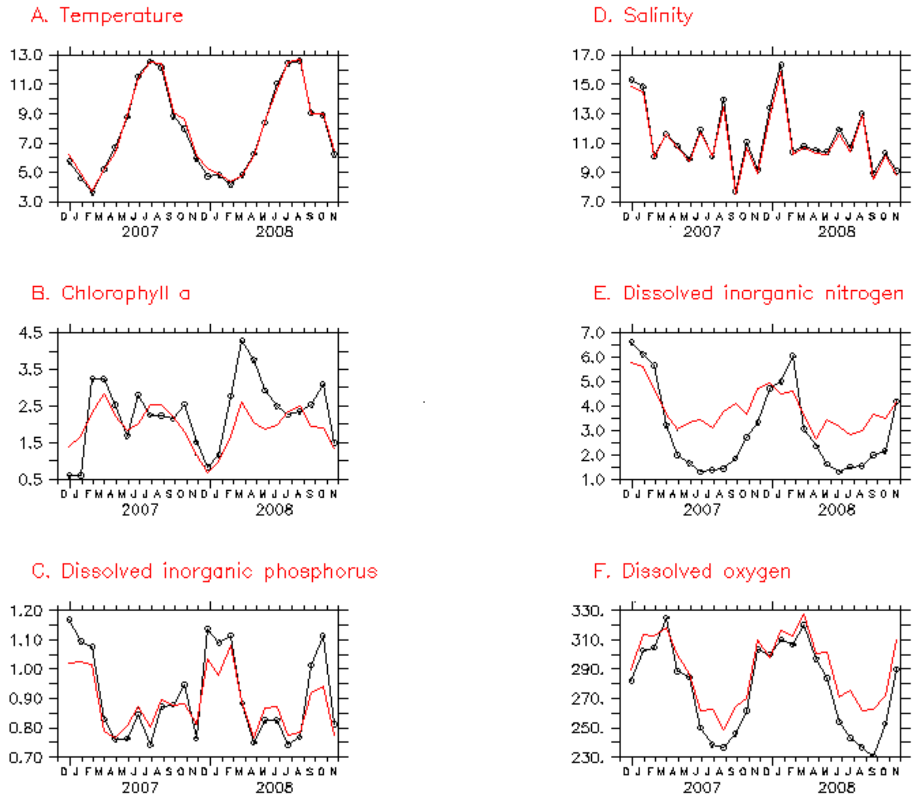

A.
B.
C.
D.
E.
F.
G.
H.
I.
J.
K.
L.
0
−40
−80
−120
−160
−200
−240
J F M A M J J A S O N D J F M A M J J A S O N D
2007
2008

Fig. 14



A. Temperature
B. Chlorophyll a
C. Dissolved inorganic phosphorus
D. Salinity
E. Dissolved inorganic nitrogen
F. Dissolved oxygen
D J F M A M J J A S O N D J F M A M J J A S O N
2007
2008

Fig. 15

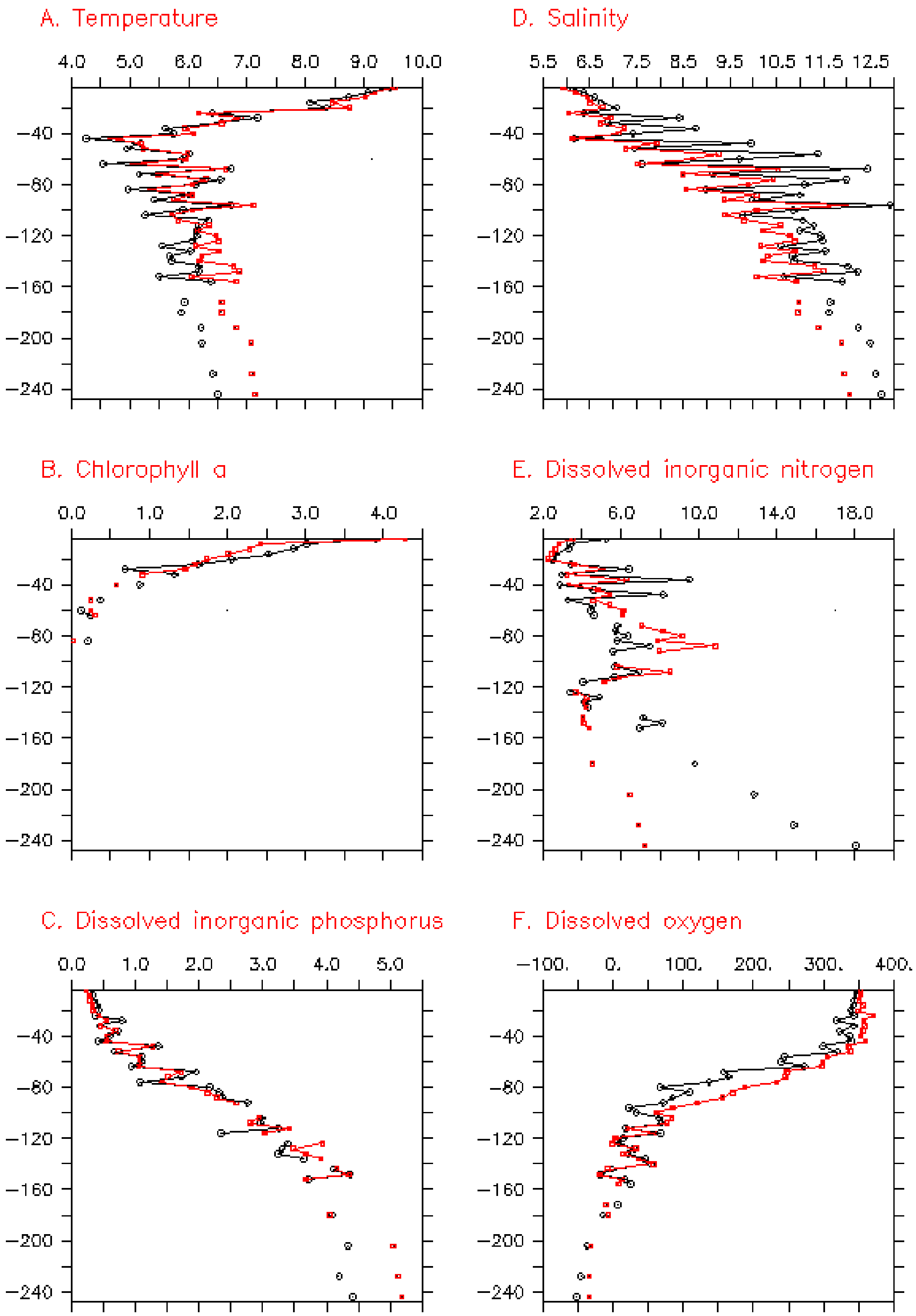

A. Temperature
4.0 5.0 6.0 7.0 8.0 9.0 10.0
−40 −80 −120 −160 −200 −240
D. Salinity
5.5 6.5 7.5 8.5 9.5 10.5 11.5 12.5
−40 −80 −120 −160 −200 −240
B. Chlorophyll a
0.0 1.0 2.0 3.0 4.0
−40 −80 −120 −160 −200 −240
E. Dissolved inorganic nitrogen
2.0 6.0 10.0 14.0 18.0
−40 −80 −120 −160 −200 −240
C. Dissolved inorganic phosphorus
0.0 1.0 2.0 3.0 4.0 5.0
−40 −80 −120 −160 −200 −240
F. Dissolved oxygen
−100. 0. 100. 200. 300. 400.
−40 −80 −120 −160 −200 −240

Fig 16

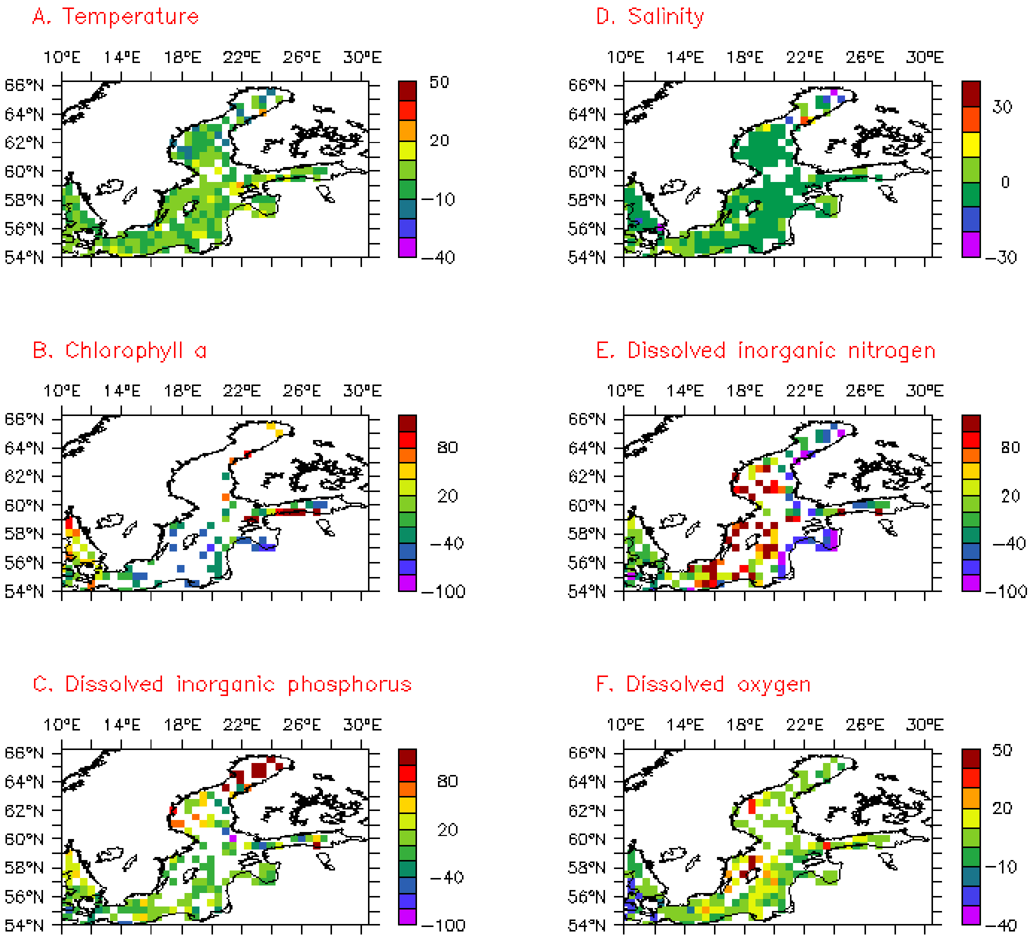

A. Temperature
10°E 14°E 18°E 22°E 26°E 30°E
66°N 64°N 62°N 60°N 58°N 56°N 54°N
50 20 −10 −40
D. Salinity
10°E 14°E 18°E 22°E 26°E 30°E
66°N 64°N 62°N 60°N 58°N 56°N 54°N
30 0 −30
B. Chlorophyll a
10°E 14°E 18°E 22°E 26°E 30°E
66°N 64°N 62°N 60°N 58°N 56°N 54°N
80 20 −40 −100
E. Dissolved inorganic nitrogen
10°E 14°E 18°E 22°E 26°E 30°E
66°N 64°N 62°N 60°N 58°N 56°N 54°N
80 20 −40 −100
C. Dissolved inorganic phosphorus
10°E 14°E 18°E 22°E 26°E 30°E
66°N 64°N 62°N 60°N 58°N 56°N 54°N
80 20 −40 −100
F. Dissolved oxygen
10°E 14°E 18°E 22°E 26°E 30°E
66°N 64°N 62°N 60°N 58°N 56°N 54°N
50 20 −10 −40

Fig. 17