Available online at www.sciencedirect.com

SCIENCE DIRECT®

Ecological Modelling 172 (2004) 69–85

ECOLOGICAL MODELLING

www.elsevier.com/locate/ecolmodel

A data assimilation technique applied to estimate parameters for the NEMURO marine ecosystem model

Hiroshi Kuroda[1], Michio J. Kishi*

Graduate School of Fisheries Sciences, Hokkaido University, 3-1-1 Minato-cho, Hakodate, Hokkaido 041-8611, Japan

Received 12 July 2002; received in revised form 29 July 2003; accepted 13 August 2003

Abstract

We have applied a data assimilation technique to determine biological parameters in the PICES (North Pacific Marine Science Organization) proto type lower trophic level model (NEMURO). North Pacific Ecosystem Model for Understanding Regional Oceanography (NEMURO) has about 80 biological parameters and 11 initial values. We used a sensitivity analysis to choose eight parameters which mostly impacted the simulated values of interest. These parameters were selected as control variables for the data assimilation. Using an adjoint method, we assimilated biological and chemical data from Stn.A7 (off Hokkaido, Japan) into the model. Twin experiments were conducted to determine whether the data constrain those eight control variables. Model output, using optimum parameter values determined by the assimilation, agreed with the data better than those obtained with the first guess parameter values. But some problems still remain even in the simulations using the optimum parameters: namely a large discrepancy is seen between the simulation and data for small zooplankton and the simulated bloom of large phytoplankton that is too large.

Keywords: "NEMURO"; Stn.A7; Adjoint method

1. Introduction

Data assimilation techniques have been proven as vital tools to develop quantitatively realistic ecosystem models, by improving their accuracy, efficiency and prognostic ability. These techniques fix unknown parameters in the governing equations and save our time in finding parameter sets that can simulate the data better. Models' structures can also be refined based on the correlations between parameters by using their error covariance matrices (Tziperman and Thacker, 1989; Matear, 1995).

There have been excellent reviews of ecosystem model of Northern Pacific (Frost and Kishi, 1999; Kawamiya, 2002, for instance). As Franks (1995) pointed out, "modelers must be guided by the questions they are asking", and here we have focused on the parameter sensitivity of ecosystem model in Northern Pacific. The state-of-the-art model in the Northern Pacific is North Pacific Ecosystem Model for Understanding Regional Oceanography (NEMURO), which has 11 ecological compartments, which were constructed by the PICES Model Task Team during a workshop held in January 2000, in NEMURO, Hokkaido, Japan (Eslinger et al., 2000). NEMURO

* Corresponding author. Tel.: +81-138-40-8873; fax: +81-138-40-8873.
E-mail addresses: kuroda@dpc.ehime-u.ac.jp (H. Kuroda), kishi@salmon.fish.hokudai.ac.jp (M.J. Kishi).
[1] Present address: Graduate School of Science and Engineering, Ehime University, 2-5 Bunkyo-cho, Matsuyama, Ehime 790-8577, Japan.

0304-3800/$ – see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2003.08.015

consists of the material flows and equations to determine them but do not have fixed parameter values, although the basic values for Station P (inside Alaskan gyre), Stn.A7 (off Hokkaido) and the Bering Sea are listed in the table. This study addresses the important work of determining the parameter values objectively and quantitatively via data assimilation and thereby improving immature NEMURO. Part of this study focused on Stn.A7 (41.5°N, 145.5°E) off Kushiro, Japan (Fig. 1). Stn.A7 is on the outer region of the Oyashio, and it is generally believed that the effects of horizontal advection can be neglected in constructing a basic physical–ecological coupled model (Kishi et al., 2001). At this early stage we have reduced the one-dimensional model to a box model to reduce the computational load for the optimization.

We have designed this study by referring to typical papers written over the past several years. In particular, the fundamental mathematical concept behind the adjoint model was based on the work of Lawson et al. (1995), who applied an adjoint method to a prey–predator model and showed that the data assimilation method was effective for estimating parameter values of ecosystem models. We also referred to the work of Lawson et al. (1996) and Spitz et al. (1998), who tested the ability of the adjoint data assimilation method to recover parameters with twin experiments. The ecosystem model, data sets from Stn.A7 and the data assimilation technique are presented in Section 2, Section 3 and Section 4, respectively. Results of the data assimilation for Stn.A7 are given in Section 5. Discussion and conclusions are presented in Sections 6 and 7, respectively.

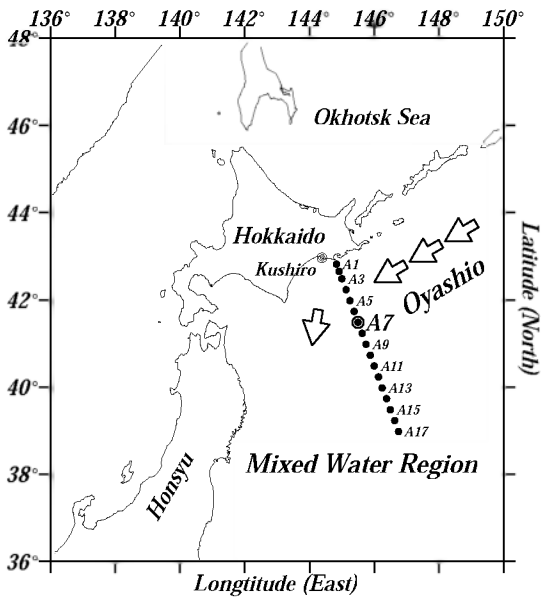

Fig. 1. Location of Stn.A7. Some arrows indicate schematic Oyashio path.

2. Ecosystem model

2.1. Structure

"NEMURO" (Fig. 2) was based on the KKYS model (Kawamiya et al., 1995, 1997), revised by Kishi et al. (2001). It was formulated by adding silicate cycling and predatory (larger) zooplankton to Kishi et al.'s (2001) model, yielding in a model with the following 11 compartments.

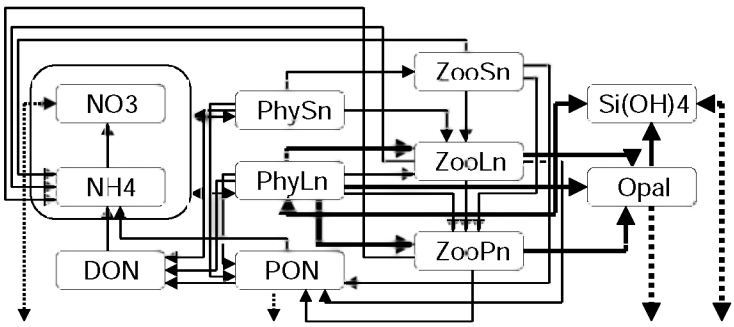

Fig. 2. Diagram of 11-compartment ecosystem model called "NEMURO". Thin arrows and thick arrows indicate nitrogen flows and silicon flows respectively, double-headed dashed arrows represent the exchange of the material with bottom below 30 m depth and downward-pointing dashed arrows denote sinking of material out of the box model.

- Small phytoplankton (PhyS)
- Large phytoplankton (PhyL)
- Small zooplankton (ZooS)
- Large zooplankton (ZooL)
- Predator zooplankton (ZooP)
- Nitrate (NO_3)
- Ammonium (NH_4)
- Particulate organic nitrogen (PON)
- Dissolved organic nitrogen (DON)
- Silicate ($Si(OH)_4$)
- Opal (Opal)

Symbols for each are in parentheses. Phytoplankton are separated into two compartments: PhyS consists of flagellates and other small phytoplankton; and PhyL represents diatoms. Diatoms also use silicon in making their shells, and therefore silicon flows are incorporated into PhyL. Zooplankton is divided into three compartments, ZooS, ZooL, and ZooP. ZooL represents copepods, mainly *Neocalanus* spp., which is a dominant species in the Northeastern Pacific near Japan and has an annual ontogenetic vertical migration (Tsuda et al., 1999). *Neocalanus* spp. migrates in summer into deep water between 250 and 2000 m depth, spawns during winter, hatches, and ascends to the upper layers in spring as nauplius to graze primary production (Miller et al., 1984; Miller and Clemons, 1988). Especially in the northwestern Pacific, the life cycle of copepods was revealed by Tsuda et al. (1999). Based on their research, the most dominant spices off Hokkaido is *Neocalanus plumchrus* which ascends in early May from deeper layers (around 1000 m) and remains in the euphotic layer until the end of August. In this model, ZooL was set to move out of the model scope on 31st of August and 10% of them were set to return on 1st of April every year (Kishi et al., 2001). The equations among nitrogen-based compartments, i.e. PON, DON, NO_3, and NH_4 are exactly the same as in Kishi et al. (2001). Two silicon-based compartments were added to Kishi et al.'s (2001) model; one is $Si(OH)_4$ and the other is particulate matter, Opal. Governing equations, more detailed structure and parameter values for NEMURO are presented in the PICES Science Report (hereafter "PSR"), Eslinger et al. (2000) and in Fujii et al. (2002), which combined NEMURO with a carbon cycle model (which does not affect the ecosystem dynamics because calcium and carbon are always plentiful throughout the ocean and do not limit biological production).

NEMURO has capabilities that previously developed ecosystem models lack. First, this model enables us to understand things that simple NPZ models or general nitrogen-based models could never explain, e.g. how the transition of phytoplankton species occurs due to the shift of limitations. Therefore it is useful to apply this model to regions where switching of nitrogen and silicon limitation controls biochemical cycles. Second, this model by design can describe roughly but accurately the material flows for the North Pacific so as to be applicable over the entire basin (with appropriately varying parameter values). Third, it can be coupled with higher trophic ecosystem models (e.g. of Pacific saury), chiefly through the predatory zooplankton.

2.2. Model numerics

The ecosystem model is incorporated into a box model with a depth of 30 m, run with a time step of 6 h. For the assimilation, this box model is run for 2 years to achieve a steady state annual cycle to avoid the influence of initial values. For the twin experiment data sets, the model was also run for 2 years.

3. Data from Stn.A7

Biological, chemical and physical data have been collected since 1990 by the Fisheries Agency off Hokkaido, Japan (Saito et al., 1998; Kasai et al., 2001). Data at Stn.A7 are used in this study because it is located outside Oyashio and the effect of horizontal advection is small (Kishi et al., 2001). For this study, the data sets are separated into two sets, i.e. one used to force the box model and one used to constrain control variables. Chlorophyll, nitrate, zooplankton wet weight, silicate, temperature, salinity data are monthly averaged and also vertically averaged for upper 30 m from 1990 to 1999 (hereafter we call this data set MMV). Whenever more than seven measurements are available for a quantity during each month at each depth, the maximum and minimum values are removed because Warm Core Rings (WCR) sometimes passed through this region changing physical and chemical environments a great deal. For the box

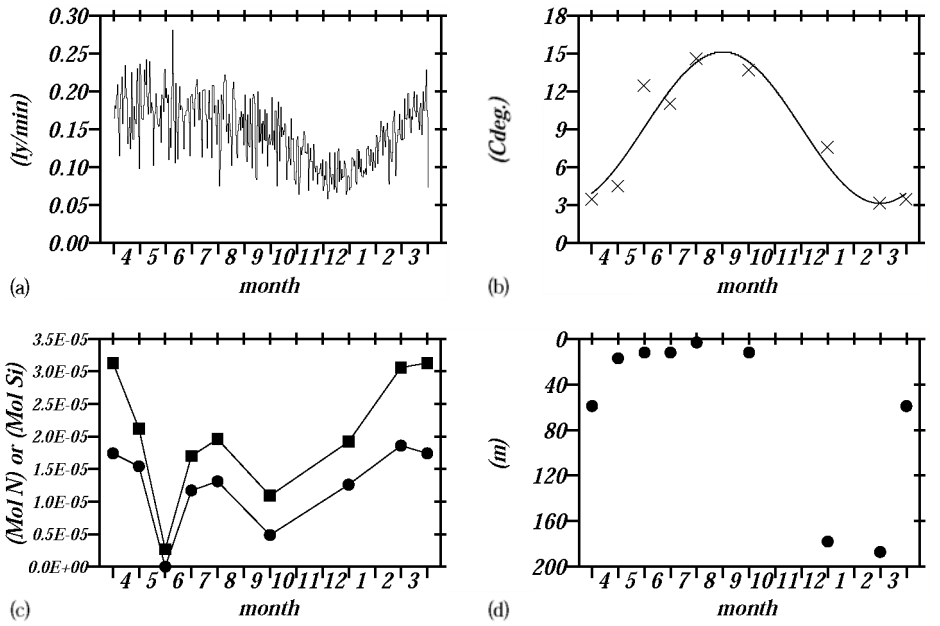

Fig. 3. (a) Light intensity at surface observed at Kushiro. (b) Crosses indicate monthly mean temperature above 30 m and curve denotes a sinusoidal curve fitted to them by a least-square method. (c) Monthly mean concentration of nitrate (closed circle) and silicate (closed square) at 30 m. (d) Mixed layer depth calculated from vertical gradient of monthly mean sigma-*t*.

model, we integrate each MMV vertically for the upper 30 m (hereafter MV).

3.1. Forcing data

The daily maximum light intensity at NEMURO weather station is averaged for the period of 1991–1996 and used as a daily light forcing, assuming the daily variation of light intensity to follow a sine-curbed function (Fig. 3a). Water temperatures from MV are from a fit to a sinusoid by a least-square method (Fig. 3b). Mean monthly nitrate and silicate concentrations at 30 m are taken out of MMV, and linearly interpolated (Fig. 3c). Seawater exchange coefficients that determine the nutrient exchange rate across the bottom of the model were fixed as follows:

(*Case I*) the shallow mixed layer season (from April to October); 0.03% per day,
(*Case II*) the period when mixed layer is rapidly deepened (in November and December); 3.0% per day,
(*Case III*) the deep mixed layer season (from January to March); 1.5% per day.

These values were calculated based on Sverdrup transport (0.3% per day) at 100 m depth. The Si/N uptake ratio by PhyL (in NEMURO PhyL is supposed to represent diatoms which take up silicate as well as nitrate) is fixed as follows; we assume that there are some relations between the uptake Si/N ratio by phytoplankton and that of surrounding water, and that the Si/N ratio of PhyL is equal to the ratios of the temporal changes in nitrate and in silicate in the surrounding water. This value is calculated from observed values (Fig. 4a and b) and 1.172 is assigned to Si/N ratio.

3.2. Data to constrain control variables

Data for nutrients such as nitrate and silicate processed by MV is used to constrain control variables (Fig. 4a and b). Though the rest of assimilated data are chl-*a* concentration reflecting total phytoplankton quantity and wet weight of two sizes of zooplankton, these data cannot be applied directly and must be processed to correspond to ecosystem model compartments. So we divided the chl-*a* concentration into two categories, i.e. PhyS, PhyL based on Saito et al. (1998) and Kasai et al. (2001), and converted from

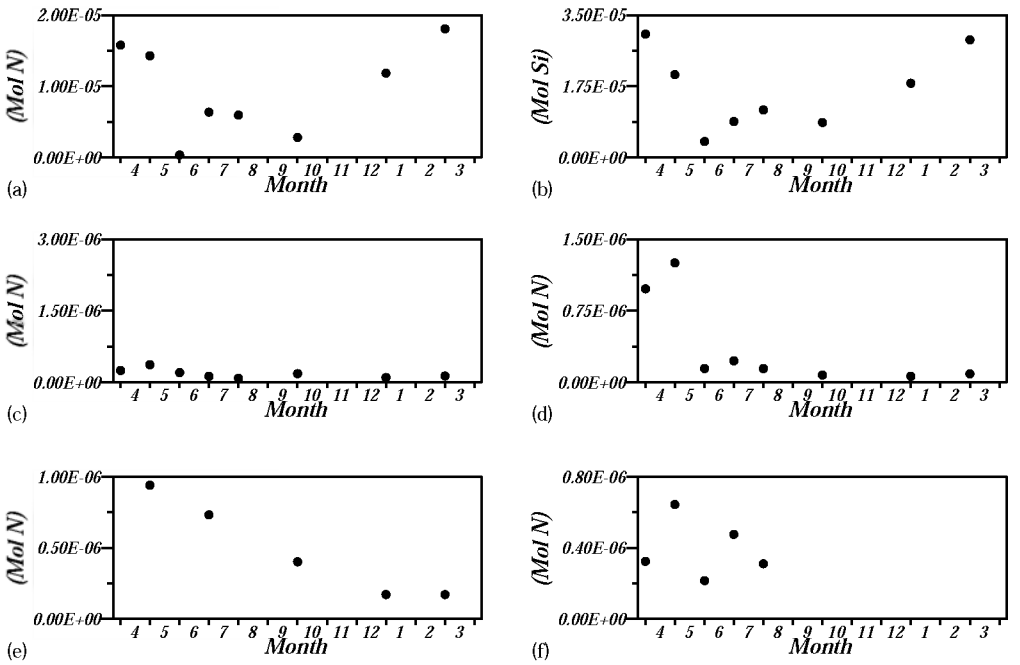

Fig. 4. Monthly mean biological and chemical data above 30 m: (a) nitrate, (b) silicate, (c) PhyS, (d) PhyL, (e) ZooS, and (f) ZooL, respectively.

chl-*a* to nitrogen using the ratio 1.59 g chl-*a*/mol N (Fig. 4c and d). We also convert from carbon to nitrogen using a C/N ratio of 5.63 (Fig. 4e and f) because data for zooplanktons corresponding to ZooS (Shinada et al., 2000) and ZooL, was recorded in milligrams of carbon. Thus, six types of data are available for constraints.

4. Data assimilation method

4.1. Sensitivity analysis

As a rule in a data assimilation for estimating parameter values, the balance between the number of control variables and quantity of data used to constrain them is very important. We do not have enough data to treat all of NEMURO's approximately 80 ecological parameters and 11 initial values as control variables. Therefore, we performed a sensitivity analysis based on the Monte-Carlo method (as in Kishi, 1994) to select the ecological parameters to be constrained. We thus excluded following parameters: 11 initial values, the Si/N ratio of PhyL and parameters for temperature dependence in biological process or decomposition, which double all rates for each 10 °C increase in temperature. The ecological parts of the model are run to calculate the biomass of each compartment until steady state solutions are obtained. Values of each compartment constrained by data and ecological parameters are stored to be used as the base line values (same values as "PSR") around which random perturbations are generated and put into the Monte-Carlo error analysis. We performed a Monte-Carlo error analysis with 1000 individual calculations with input parameters perturbed independently over random error distribution with limit of ±50% of base line values. Principal component analysis is used to reduce the 1000 sets of output of biological parameters values. The PCA indicates four factors, which together explains 18% of the variance in data space (Table 1). The reason why this variance is so small is that "NEMURO" is made of and is controlled by the interaction of many compartments, meaning that many parameters must be significant for "NEMURO". The first and fourth factors are clearly related to photo-

Table 1
Factor pattern 1st principal component to 4th principal component

	P.C.1	P.C.2	P.C.3	P.C.4
IoptS	-0.036	0.004	0.021	-0.023
IoptL	0.021	-0.081	-0.068	-0.021
VmaxS	0.326	0.132	0.188	0.198
KNO3S	-0.175	-0.103	-0.053	-0.167
KNH4S	-0.080	-0.079	-0.128	0.131
PusaiS	-0.014	-0.008	-0.035	0.048
VmaxL	-0.232	-0.062	0.064	0.277
KNO3L	0.112	-0.005	0.080	-0.243
KNH4L	0.069	0.045	-0.062	-0.179
KSiL	-0.018	-0.008	0.035	0.069
PusaiL	-0.026	-0.020	-0.070	-0.018
ResPS0	-0.071	-0.087	-0.085	0.008
ResPL0	0.043	-0.026	0.025	-0.124
MorPS0	-0.180	-0.019	0.376	-0.313
MorPL0	-0.025	-0.031	-0.090	-0.076
GammaS	-0.089	-0.052	-0.121	-0.005
GammaL	0.004	0.036	0.008	0.015
GRmaxS	-0.021	0.071	0.188	-0.003
LamS	0.023	0.043	0.024	-0.034
GRmaxLps	-0.064	0.089	-0.149	-0.097
GRmaxLpl	0.002	0.028	0.096	-0.049
GRmaxLzs	0.026	-0.043	-0.014	-0.044
LamL	-0.025	0.091	-0.165	0.023
GRmaxPpl	0.177	-0.168	-0.164	-0.185
GRmaxPzs	-0.045	-0.169	-0.071	0.012
GRmaxPzl	0.042	-0.101	0.212	0.152
LamP	0.140	-0.386	0.089	-0.144
PusaiPL	-0.025	0.117	0.057	0.009
PusaiZS	-0.006	-0.038	-0.083	0.080
AlphaZS	-0.004	0.029	-0.007	0.147
BetaZS	0.022	0.024	0.043	-0.037
AlphaZL	0.001	0.106	-0.088	-0.106
BetaZL	-0.058	0.080	-0.263	-0.063
AlphaZP	-0.016	0.034	-0.128	0.074
BetaZP	-0.023	-0.256	-0.001	0.027
MorZS0	-0.005	-0.053	-0.029	0.035
MorZL0	-0.035	-0.024	-0.006	0.068
MorZP0	-0.024	0.172	0.023	-0.315
VP2N0	0.034	0.009	-0.083	-0.022
VP2D0	0.021	0.026	-0.131	0.075
VD2N0	-0.014	0.091	0.186	-0.090
VO2S0	0.004	0.047	0.038	-0.049
Nit0	0.015	-0.035	-0.033	-0.105
PhySn	0.397	-0.153	-0.319	0.219
PhyLn	-0.395	0.341	0.063	0.145
ZooSn	0.204	0.411	0.165	-0.099
ZooLn	0.048	0.439	-0.468	-0.128
NO3	-0.465	-0.049	-0.093	0.263
SiOH4	-0.299	-0.219	-0.198	-0.423
Eigen value	3.357	2.092	1.803	1.665
Proportion	6.852	4.270	3.679	3.397

Relatively larger elements of these Eigen vector are hatched, as for biological parameters, more than 0.1 and for the output, more than 0.2.

synthesis, while the second and the third are related to grazing of ZooL. We focused on eight parameters with relatively large vector values in several factors and chose them as control variables: maximum photosynthetic rate at 0 °C by PhyS, VmaxS; maximum photosynthetic rate at 0 °C by PhyL, VmaxL; Ivlev constant of ZooP, LamP; mortality rate of ZooP at 0 °C, MorZP0; mortality rate of PhyS at 0 °C, MorPS0; half saturation coefficient for NO_3 by PhyS, KNO_3S; maximum rate of grazing at 0 °C by PhyL to ZooP, GRmaxPpl; and maximum rate of grazing at 0 °C by ZooL to ZooP, GRmaxPzl.

4.2. Cost function

The assimilation minimizes a cost function that measures the differences between data and the corresponding model output. In this study, the cost function is defined by assuming that data is 100% reliable and is written in a least-square manner,

$$F = \sum_{j=1}^{6} \sum_{i=\mathrm{Start}}^{\mathrm{End}} M_j (C_j(i) - \mathrm{data}_j(i))^2 \quad (1)$$

where $\mathrm{data}_j(i)$ denotes observed values whose observed time is equivalent to model time step i and whose type is j, and $C_j(i)$ represents model output corresponding to the $\mathrm{data}_j(i)$. Each j from 1 to 6 is an assigned number to briefly express six types of data. *Start* and *End* also indicates the beginning and the last time step of an assimilated period, respectively. In addition M_j is a important weight and if it were not for M_j in (1), it would be impossible to equally evaluate six types of square errors and there M_j is defined as follows:

$$M_j(i) = \begin{cases} \left(\dfrac{\max(\overline{\mathrm{data}_k})}{\overline{\mathrm{data}_j}} \right)^2, & \text{if } \mathrm{data}_j(i) \text{ exists} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where $\overline{\mathrm{data}_j}$ is the annual average of the data type j and $\max(\overline{\mathrm{data}_k})$ is the maximum of the annual average of the assimilated data. This weight is the same as Lawson et al. (1996), except for the second power. By use of this second power weight, terms in (2) concerning small-order data type are less underestimated than by Lawson et al.'s and after assimilation best results are obtained. Details about the influence of selecting weights on the assimilation will be discussed in Section 6.

Start and *End* should be selected carefully in order to focus on the model output without the influence of initial values. Here *Start* and *End* were respectively 120 (at 1st May in the first year) and 2880 (at 1st April in the third year) time steps. Consequently, we assimilate data into the model for about 2 years and discuss the model output in the second year by using optimum parameter values after the data assimilation.

4.3. Optimization

Our optimization routine consists of the conjugate gradient method (FR method) and the quadratic interpolation method. The merit of this type of optimization is computationally fast convergence of control variables. On the other hand, it often finds local, not global, minima. Using metaheuristics such as genetic algorithms, simulated annealing, and greedy randomized adaptive search procedures will yield better results, though the computational load is much heavier. At this stage, however, computational speed is given priority because NEMURO exists in the earliest stages of development and we expect this data assimilation to greatly improve NEMURO. Therefore we will implement data assimilation by use of a metaheuristic after further development of NEMURO.

Next for our optimization to finish successfully without numerical instability, control variables need to be mapped from unbounded space into bounded space. The eight control variables have different orders and empirical values, and all of them are defined as values greater than zero. The mapping technique excludes undesirable (e.g. negative in this case) values for some control variables. Consequently, the sin-squared function (Box, 1966; Vallino, 2000) is introduced as follows:

$$\begin{aligned} &\vec{a} = (a(1), a(2), a(3), \ldots, a(8))^T \\ &\vec{a\_s} = (a\_s(1), a\_s(2), a\_s(3), \ldots, a\_s(8))^T \\ &a(i) = aL(i) + (aU(i) - aL(i))\sin^2(a\_s(i)) \\ &aL(i) \le a(i) \le aU(i) \quad (1 \le i \le 8) \end{aligned} \tag{3}$$

where $\vec{a}$ vector mapped into $\vec{a\_s}$ vector is composed of eight control variables and a control variable $a(i)$ has empirically known to take values ranging from $aL(i)$ to $aU(i)$. And next in order to obtain $a\_s(i)$, the inverse transform to $a(i)$ is applied.

$$a\_s(i) = \sin^{-1}\left(\sqrt{\frac{a(i) - aL(i)}{aU(i) - aL(i)}}\right) \tag{4}$$

In all our optimization routine, this $\vec{a\_s}$ vector is utilized instead of the $\vec{a}$ vector and to calculate conjugate directions, $a(i)$'s gradient computed by use of adjoint method (as will be mentioned in Section 4.4) should be also transformed to the gradient with respect to $a\_s(i)$ as follows:

$$\begin{aligned} \frac{\partial F}{\partial a\_s(i)} &= \frac{\mathrm{d}(a(i))}{\mathrm{d}(a\_s(i))}\frac{\partial F}{\partial a(i)} \\ &= 2(aU(i) - aL(i))\sin(a\_s(i)) \\ &\quad \times \cos(a\_s(i))\frac{\partial F}{\partial a(i)} \end{aligned} \tag{5}$$

where F represents the cost function in Section 4.2.

Lastly, it is also very important what values are selected for $aL(i)$ and $aU(i)$ from the empirical value range. Each ecosystem model parameter (control variable) has a range that increases with the parameter's uncertainty. On the other hand, for a fixed same value of $a\_s(i)$, the larger difference between $aL(i)$ and $aU(i)$ is, the larger the scaled gradient computed by (5) is. The magnitude of this gradient impacts the speed of convergence for each control variable, such that a larger gradient for a given control variable results in faster convergence for that variable. Because the uncertainties for our control variables are not well known, we set all $aL(i)$ and $aU(i)$ by the same ratio relative to $a(i)$ in order to equally evaluate all control variables in optimization, therefore we give $aL(i)$ and $aU(i)$ one-tenth and ten times the values employed in "PSR", respectively. Though Vallino (2000) used almost the same mapping technique (similar standardization), $aU(i)$ (or $aL(i)$) he assigned to each control variable had different ratio relative to $a(i)$, individually. He may have intended to implicitly adjust individual convergence speeds for control variables by combining each parameter's uncertainty with the mapping technique.

4.4. Adjoint construction

We construct the adjoint model using a Lagragian function that allows us to treat dependent variables like independent ones and to compose the program with ease. Detailed explanations about these are seen in Lawson et al. (1995). For our adjoint construction, however, one point should be noted. In NEMURO, to reproduce annual ontogeneric vertical migration of copepods (ZooL) as mentioned in Section 2.1, a quantity of ZooL on 1st of April are obtained by multiplying the quantity on 31st of August of the previous year by the returning ratio (10%). Because ZooL is

Table 2
Parameter values in twin experiment

	Base value	First guess value (ratio)	Upper value	Lower value
VmaxS (day^{-1})	1.00E+00	1.400E+00 (+40%)	1.000E−01	1.00E+01
VmaxL (day^{-1})	1.00E+00	0.600E+00 (−40%)	1.000E−01	1.00E+01
LamP (mol^{-1} N)	1.40E+06	0.800E+06 (−43%)	1.40E+05	1.40E+07
MorZP0 (day^{-1})	5.85E+04	3.850E+04 (−34%)	5.85E+03	5.85E+05
MorPS0 (day^{-1})	5.85E+04	7.850E+04 (+34%)	5.85E+03	5.85E+05
KNO_3S (mol N)	3.000E−06	1.000E−06 (−67%)	3.000E−07	3.000E−05
GrmaxPpl (day^{-1})	4.000E−01	6.000E−01 (+50%)	4.000E−02	4.00E+00
GrmaxPzl (day^{-1})	4.000E−01	2.000E−01 (−50%)	4.000E−02	4.00E+00

not always integrated forward every one time step, Lagragian multipliers at a time step also need to be computed backward by using not only one-step-future multipliers but also other time step's multipliers.

4.5. Twin experiments

For the last step of data assimilation into the ecosystem model, we conducted a twin experiment to investigate whether or not the quantity of data is sufficient for eight control variables. We also check by this experiment that the adjoint model has been correctly constructed. Data for the twin experiment were sub-sampled during the period from the second to third years of a third-year run of the box model using biological and physical parameter values used in "PSR". We used the values of all 11 compartments on 1st of April in the second year as initial values for the box model in the data assimilation.

Next we independently perturb control variables over random error distributions from 35 to 65% of base values, and employ these as first guess parameter values (Tables 2 and 3). Results indicated that quantity of data and data type available was enough for control variables to converge and that the adjoint model was also correctly constructed, because the optimum parameter values obtained from this assimilation were equal to their base values (Fig. 5).

Table 3
Parameter values before and after assimilation

Control variable	First guess parameter value	Optimum parameter value	Unit
VmaxS	1.00E+00	1.20E−01	day^{-1}
VmaxL	1.00E+00	4.63E−01	day^{-1}
LamP	1.40E+06	7.65E+05	mol^{-1} N
MorZP0	5.85E+04	6.08E+04	day^{-1}
MorPS0	5.85E+04	4.80E+05	day^{-1}
KNO_3S	3.00E−06	4.73E−06	mol N
GrmaxPpl	4.00E−01	8.07E−01	day^{-1}
GrmaxPzl	4.00E−01	4.02E−02	day^{-1}

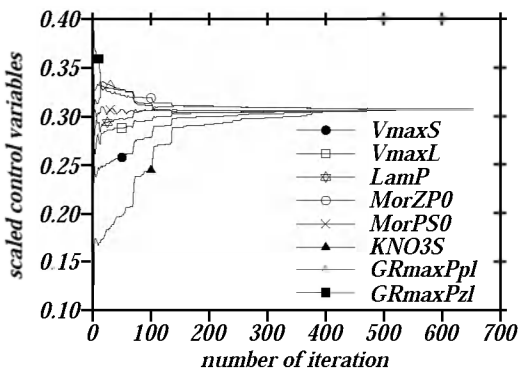

Fig. 5. Variations of scaled eight control variables in a twin experiment. Number of iterations states what times the conjugate gradient with respect to the control variables are calculated.

5. Results from Stn.A7 data

Base values in the twin experiments are used for the first guess parameter values and initial values of 11 compartments were selected as follows. Five compartments (nitrate, silicate, PhyS, PhyL, and ZooL) hold observed values on April which are utilized as their initial values (see Fig. 4), while for the rest of the six compartments, steady values on 1st of April computed in the twin experiment are employed.

Box model output computed using the first guess parameter values are shown in Fig. 6 and monthly mean observed values shown in Fig. 4 are compared in Fig. 6a–f, which indicates that the error between data

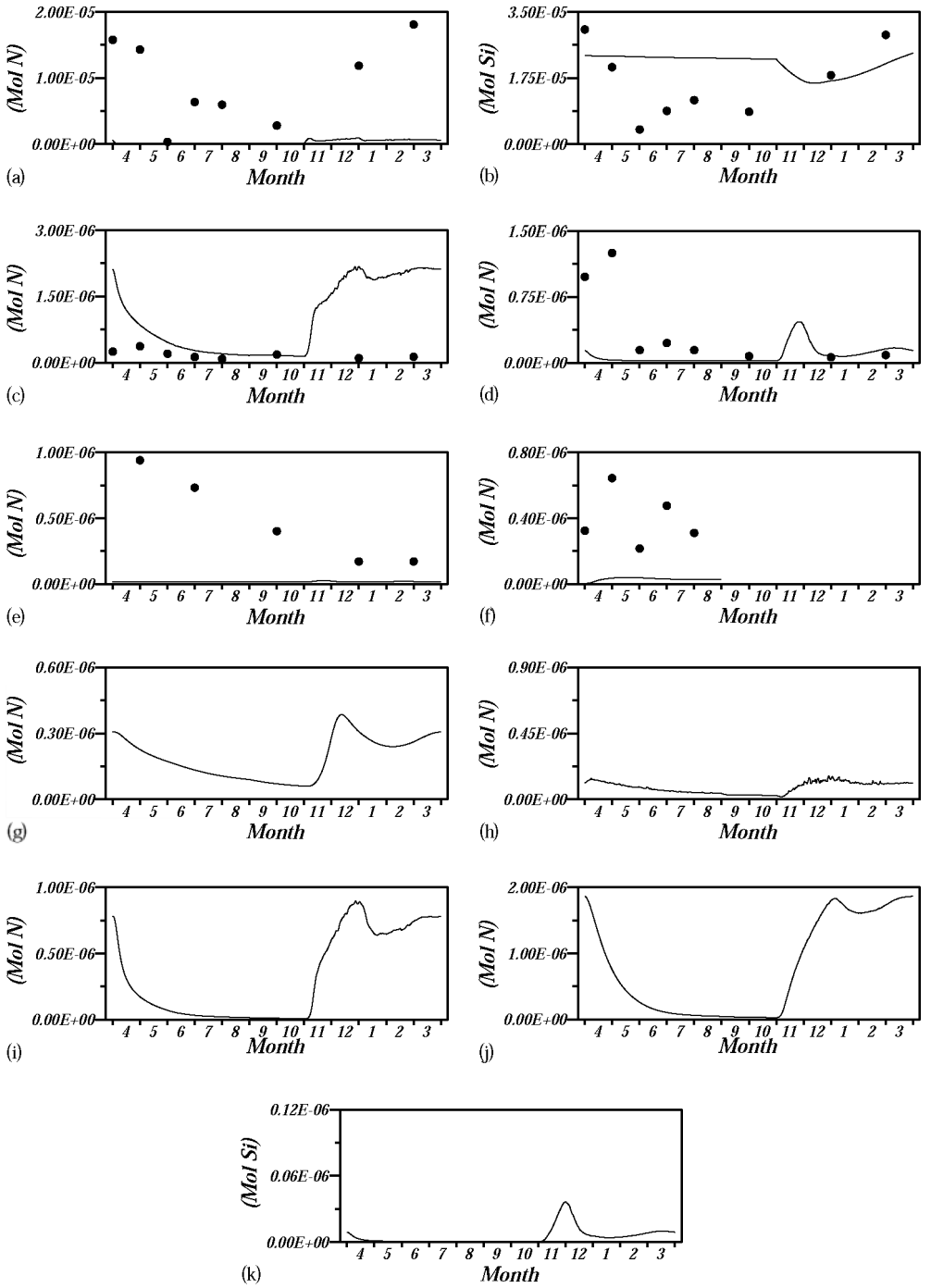

Fig. 6. Data and model output by using first guess parameter values described in Table 3. Closed circles indicate data shown in Fig. 5 and curves are model results: (a) NO_3, (b) $Si(OH)_4$, (c) PhyS, (d) PhyL, (e) ZooS, (f) ZooL, (g) ZooP, (h) NH_4, (i) PON, (j) DON, and (k) Opal.

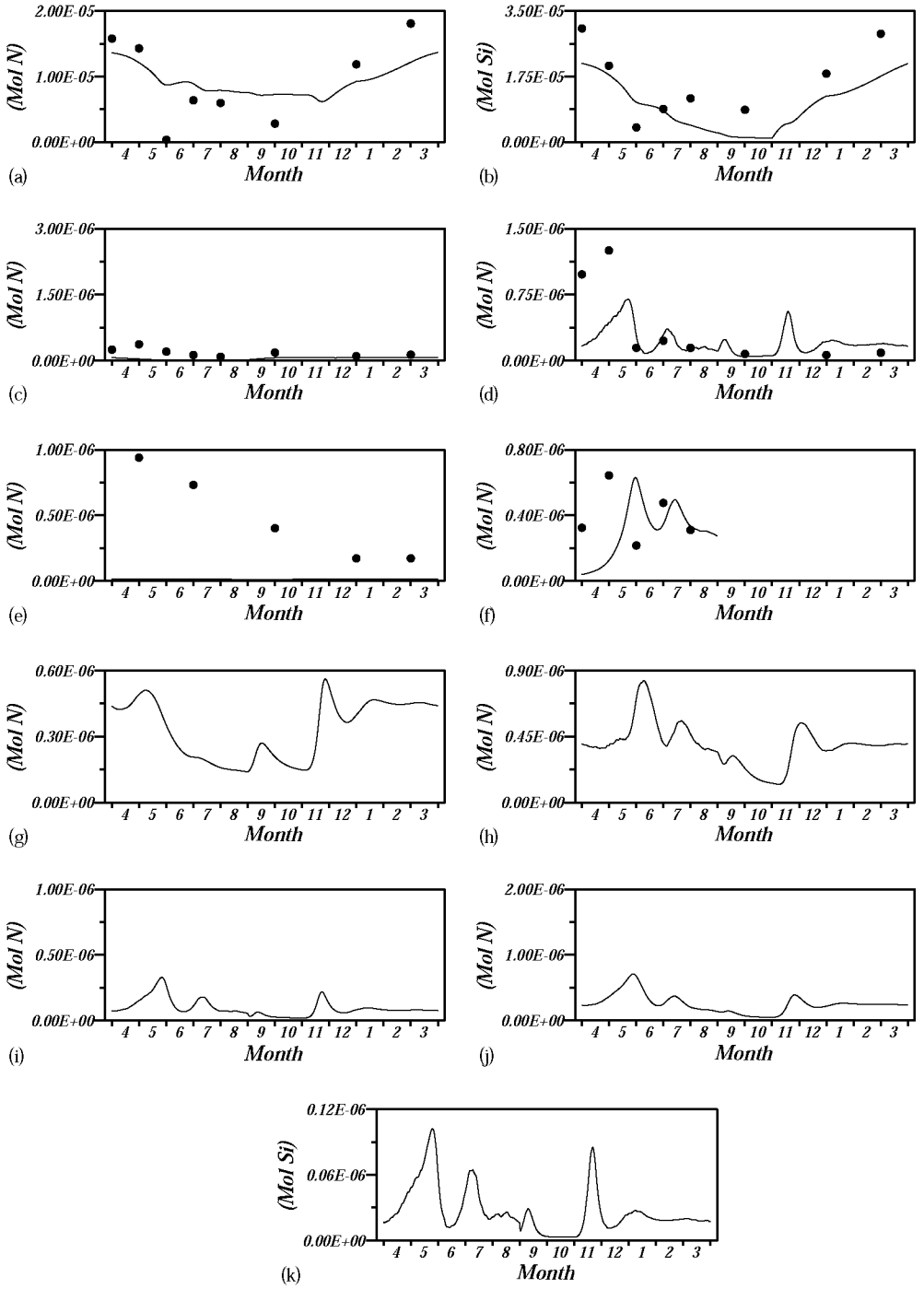

Fig. 7. Same as Fig. 6, except for using optimum parameter values obtained by data assimilation.

and the output are very large and that parameter values in "PSR" are ineffective for this box model forced by real temperature, light, and so on. It is therefore vital to change them to reproduce biochemical characteristics shown in observed data. Fig. 7 is the same as Fig. 6, except for using the optimum parameter values obtained by the data assimilation. When comparing Fig. 6a–f with Fig. 7a–f, the output shown in Fig. 7 seems to agree better with the data. In other words, this demonstrates that our assimilation improved model output effectively and objectively. However, some misfits between data and the output remain (e.g. the misfit concerning ZooS). In the following section we will discuss the reasons for such misfits.

6. Discussion

6.1. Misfit between data and output

We focus on the following points:

(1) The change with time in domination of PhyS and PhyL in the simulation using optimum values does not seem realistic and the simulated annual cycle for PhyS is too weak.
(2) As mentioned in Section 5, the data assimilation did not significantly improve the simulations of ZooS (compared to data). Therefore, the assimilation was ineffective for ZooS.
(3) The data assimilation seems to have improved the simulations of nitrate and silicate, except that the model fails to simulate the rapid decrease observed in May–June.
(4) Differences between the model and data corresponding to ZooL are relatively large, though the time dependence of the output is consistent with the data (i.e. increases in April–May, followed by sudden decreases, then increases again and decreases in July–August).
(5) Results for PhyL are desirable, except for the fourth big bloom occurring from November to December that is never observed in the region around Stn.A7.

Generally, the causes of phenomena like (1)–(3) are often attributed to inadequacies in the model's structure. However, in this study we used only a subset of the biological parameters as control variables. Those misfits occurred because not all the biological parameters included in the model were treated as control variables. Though the data available was sufficient to constrain only eight control variables, addition of other parameters to control variables made the assimilation unstable (results not shown). Thus these results are the best that we can do with this much data. It is fruitful, however, to discuss the causes of the problems from (1) to (3) and to investigate what parameters should be added as control variables after obtaining other biological and chemical data or what parameter values were misestimated in "PSR".

And so we discuss the relationship between the assimilation process and the structure of the box model. The assimilation is more efficient for PhyL than for PhyS because PhyL utilizes both nitrate and silicate for its growth while PhyS uses only nitrate. Therefore, control of concentration of PhyL in the assimilation process results in control of nitrate and silicate at the same time. This means once approximate annual cycles of these three compartments are determined, that of PhyS is determined by the quantity of nitrate PhyS can use. In our case PhyS need to suppress the increase in biomass since quantity of nitrate left for PhyS is thought to be a little.

This is confirmed by interpreting the optimum parameter values of VmaxS, KNO_3S and MorPS0; these were changed by factors of about one-eighth, one-and-a-half, and one-tenth of their first guesses, respectively. The assimilation tends to favor lower concentrations of PhyS, so that PhyS takes up fewer nitrates, and therefore the optimum rate of mortality for PhyS is much higher than its initial guess values.

As for (2), one might imagine that the decrease in PhyS makes ZooS decrease since ZooS can increase only by grazing PhyS in the model. But this is not true of the case. Comparing Fig. 6a with Fig. 6c, it is shown that though PhyS is abundant in the period from November to April, ZooS does not increase so much. In other words, increasing in ZooS is suppressed not by depletion of feed but by too high grazing pressure on ZooS. And it is necessary to reset PSR's parameter values concerning the grazing on ZooS by both ZooL and ZooP.

The reason for misfit (4) comes from the timing of ZooL's ascent back into the model box. In this study the ZooL returns to the model box on 1st of April, and the survival ratio during winter, Rrec, is

set to be 10%. But the calculated values of ZooL on 1st of April and on 1st of August are both about 3.0E−07 mol. If the calculated value of ZooL on 1st of April fits to the observed data, the concentration of ZooL on 30th of August should be approximately equal to 3.0E−06 mol. In short, within just a month of August, ZooL should increase by 10 times. Therefore it is natural that not the amplitude but only the tendency of the fluctuations in ZooL is consistent with that of the observed data. This problem will be solved by advancing the date when ZooL returns.

Lastly the cause of misfit (5) lies in the box model itself. In the annual cycle of PhyL, four times blooms appear, which we call Peak-1 (May–June), Peak-2 (early July), Peak-3 (early September), and Peak-4 (November–December). Peak-1 is regarded as a first spring bloom, Peak-2 is caused by reduction of the grazing pressure by ZooL and ZooP, which temporarily increase around Peak-1 with increase in PhyL but then decrease because of a decline in PhyL concentration. Peak-3 also is caused by drop in the grazing pressure by ZooL, which disappears out of the model box due to annual vertical migration. Next the cause of Peak-4 needs to be considered using the governing equation of ecosystem. The following series of equations represent photosynthesis by PhyL and by integrating $\mathrm{d(PhyLn)}/\mathrm{d}t = \mathrm{GppPLn}$, we can compute the increase with time in PhyL that photosynthesis contribute to.

$$\mathrm{GppPLn} = \mathrm{VmaxL} \times \min\left(\frac{\mathrm{NO_3}}{\mathrm{NO_3} + \mathrm{KNO_3L}}\exp(-\Psi_\mathrm{L} \times \mathrm{NH_4}) + \frac{\mathrm{NH_4}}{\mathrm{NH_4} + \mathrm{KNH_4L}}\frac{\mathrm{Si(OH)_4}}{\mathrm{Si(OH)_4} + \mathrm{KSiL}}\right) \quad (6a)$$

$$\times \exp(\mathrm{kGpp_L} \times \mathrm{TMP}) \quad (6b)$$

$$\times \frac{1}{\mathrm{Depth}}\int_{-\mathrm{Depth}}^{0}\left(\frac{I}{\mathrm{IoptL}}\exp\left(1 - \frac{I}{\mathrm{IoptL}}\right)\right)\mathrm{d}Z \quad (6c)$$

$$\times \mathrm{PhyLn}$$

$$I = I_0 \exp(-\kappa|Z|)$$

$$\kappa = \alpha_1 + \alpha_2(\mathrm{PhySn} + \mathrm{PhyLn})$$

$\mathrm{NO_3}$, $\mathrm{NH_4}$, $\mathrm{Si(OH)_4}$, PhySn, PhyLn indicates concentration of nitrate, ammonium, silicate, PhyS, and PhyL, respectively and details about these equations were described in Fujii et al. (2002).

The photosynthesis by PhyL is controlled by three limiting factors: nutrients (6a), temperature (6b), and light intensity (6c). Fig. 8 shows the annual fluctuations of three terms (i.e. 6a–c) in second year. In Fig. 8 four points on each curve indicate the moment corresponding to Peak-1, -2, -3, and -4, respectively. To investigate the cause of Peak-4, we compare the magnitude of the limiting factors at Peak-1 and -4. At the former, (a), (b), and (c) is found 0.75, 1.75, and 0.9, respectively, while at the latter, 0.75, 2.0, and 0.9, respectively. Though contribution of temperature at Peak-4 is slightly larger, at both peaks three limiting factors impact the photosynthesis by PhyL to almost the same degree.

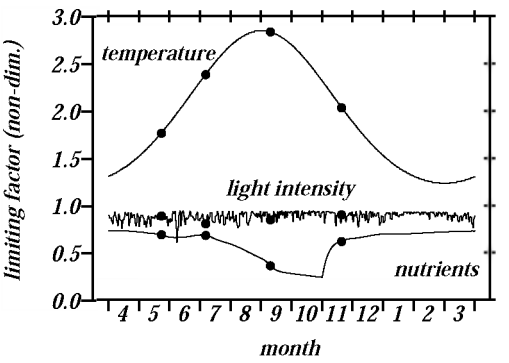

Fig. 8. Annual fluctuations of three limiting factor which control photosynthesis by PhyL; nutrients, temperature, and light intensity. Details are seen in the text. Four closed circle marked on each line indicate the moment when Peak-1, -2, -3, and -4 occurs, respectively.

As for temperature, Peak-1 exists in the season when temperature is increasing by radiative heating while Peak-4 appears when temperature is decreasing due to cooling in winter, so that the effect of temperature has nearly equal impact on the two peaks. As for nutrients especially silicate, Peak-1 exists in the period when nutrients is decreasing because of consumption during the first bloom while Peak-4 occurs when it is increasing because the mixed layer is rapidly deepening from November to December. Thus the influence of nutrients on photosynthesis is also about the same by these two peaks.

Lastly for light, we should take up Steele's (1962) light equation, which is applied in the ecosystem model to express the dependence of photosynthesis

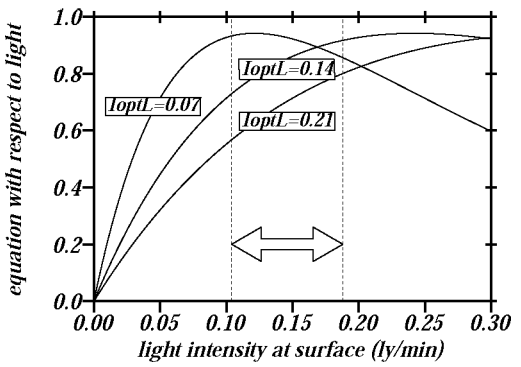

Fig. 9. $L(I_0)$ (that mentioned in the text) as a function of light intensity at surface when assigning 0.07, 0.14, and 0.21 to IoptL, respectively. The range which a double-headed arrow indicates is average ±S.D. of light intensity at surface.

on light as described in Eq. (6c). In the absence of blooms for two kinds of phytoplanktons, the concentrations of PhyS and PhyL are roughly estimated as 10×10^{-7} mol N from Fig. 7c and d, and the summation of them become to be approximately 2.0×10^{-7} mol N. With these concentrations assigned to both PhyS and PhyL, κ turns to be constant and we can also integrate Eq. (6c) vertically. Consequently, the vertically integrated Eq. (6c) turns out to be a function of only light intensity at surface I_0 and will be described as $L(I_0)$ hereafter. Fig. 9 shows $L(I_0)$ as a light intensity at surface I_0 using three different types of optimum light intensity IoptL such as 0.07, 0.14, and 0.21 light year/min, respectively (Parsons et al., 1984) indicated the parameter range of IoptL is approximately between 0.03 and 0.20 light year/min. And a horizontal double-headed arrow also indicates

"annual mean ± S.D. (0.146 ± 0.042)"

of observed light intensity at surface shown in Fig. 3a.

Pay attention to the curve for IoptL = 0.07 which value is used in the model, and it is noted that within the range the doubled-headed arrow indicates (i) $L(I_0)$ is greater than 0.8, (ii) it has relatively small gradient, and (iii) the sign of gradient inverts at $I_0 = 0.11$, which value is nearly close to the lowest light intensity in the annual variation. In other words, light intensity at the surface significantly contributes to photosynthesis by PhyL through the annual cycle but cannot reproduce remarkable annual cycle because light limitation can hardly fluctuate. Moreover, photoinhibition occurs through the annual cycle so that in winter the contribution of light to photosynthesis by PhyL exceeds that in summer though light intensity in winter is weaker. Correspondingly, when focusing on the curve of IoptL = 0.21 within the arrow range, $L(I_0)$ is smaller than 0.8 so that light intensity contributes less to photosynthesis by PhyL than using IoptL = 0.07, but the steeper gradient with positive value will probably bring about reproduction of a remarkable annual cycle. On the other hand, photoinhibition can never happen in this range.

Therefore, to incorporate the effect of annual light fluctuations into the ecosystem model and to suppress the pseudo-bloom of PhyL at Peak-4, IoptL needs to be assigned a value greater than 0.07.

6.2. Assimilation model structure

For understanding the causes of misfit between model output and data, we should also evaluate the structure of the assimilation model, especially its cost function. As shown in Section 4.2, we constructed the cost function by assuming that observed data are 100% reliable, which allowed us to ignore error correlation between compartments and to introduce the quadratic weights calculated from annual means of observed data (Case A). This weight accounts for the relative differences of biomass among compartments but not for seasonal changes in each compartment. For instance, the quantity of phytoplankton at Stn.A7 changes by more than a factor of 10 over a year (as shown in Fig. 4d), so in the cost function the difference between minimum and maximum concentration is quadratically magnified (in the above case, more than 100 times). To incorporate this influence into the assimilation model and to evaluate the importance in selecting weights, another two types of weights are tested for the cost function: one is an original weight, the other is the weight used by Lawson et al. (1996). Our original formulation (Case B) is:

$$F = \sum_{j=1}^{6} \sum_{i=\mathrm{Start}}^{\mathrm{End}} M_j(i)(C_j(i) - \mathrm{data}_j(i))^2$$

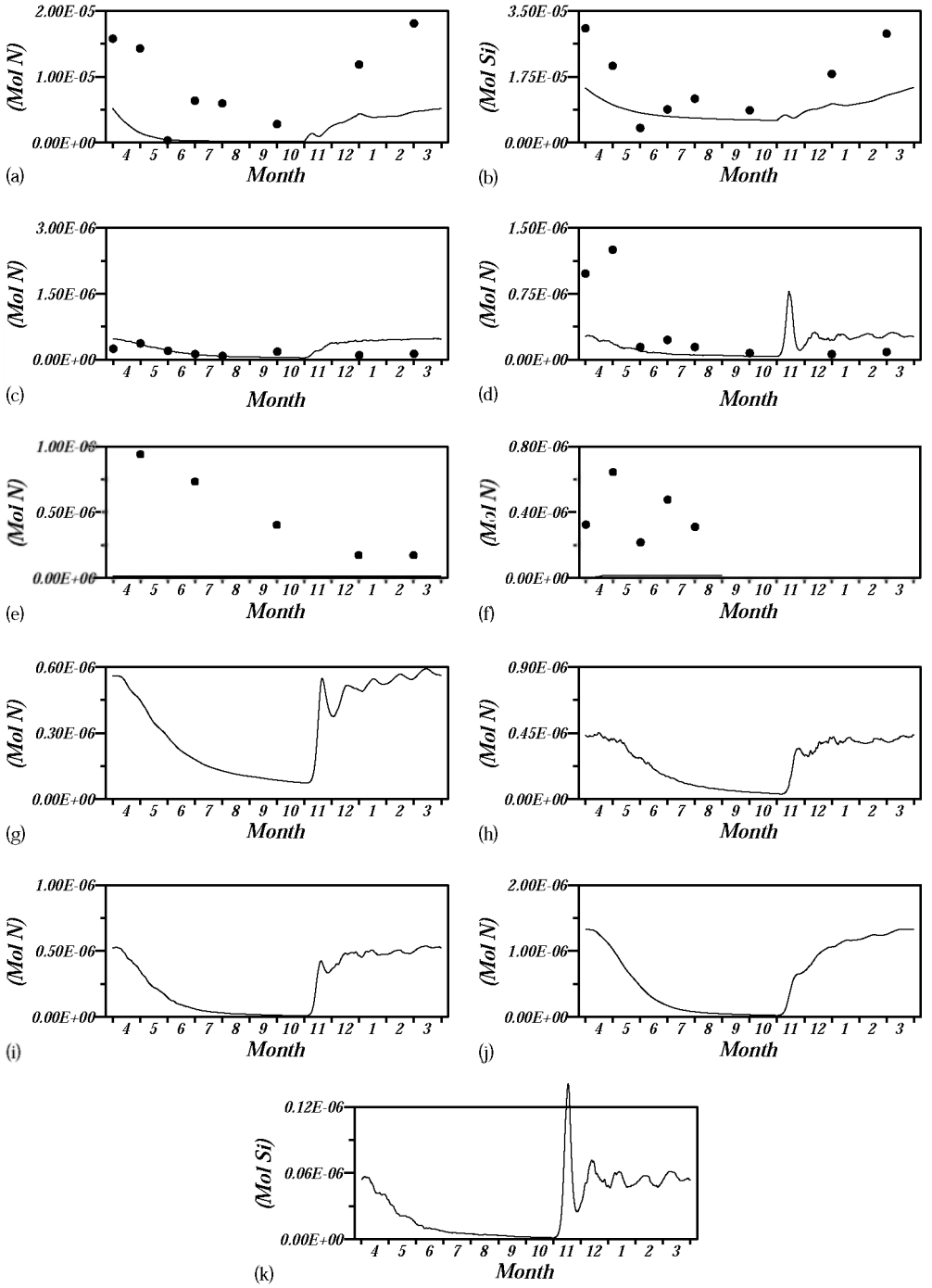

Fig. 10. Same as Fig. 7, except for using Case B's weight in the cost function.

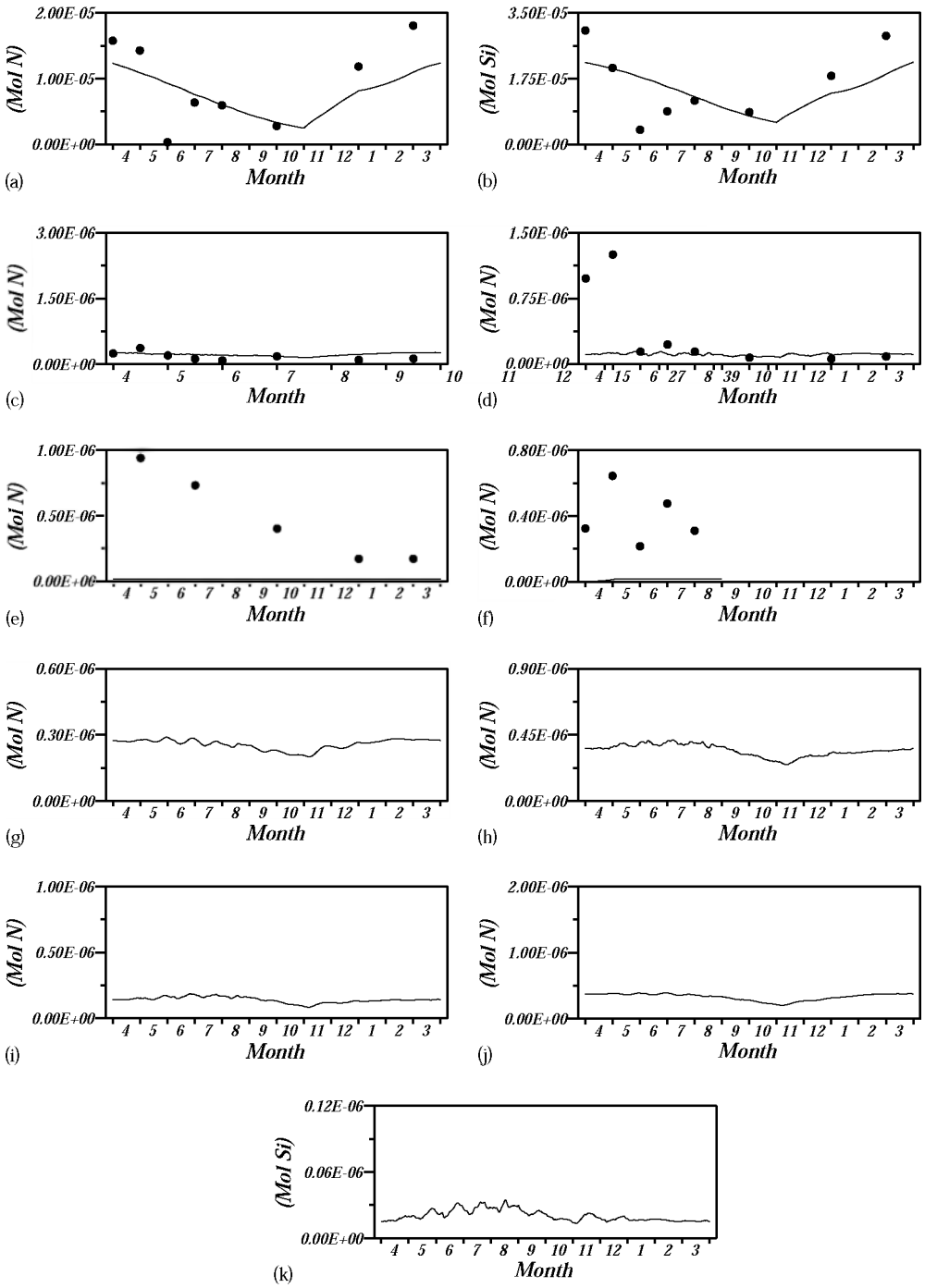

Fig. 11. Same as Fig. 7 or Fig. 10, except for using Lawson et al.'s (1996) weight in the cost function.

$$M_j(i) = \begin{cases} \left(\dfrac{\max(\overline{\text{data}_k})}{\overline{\text{data}_j}}\right)^2 \times \left(\dfrac{\overline{\text{data}_j}}{\text{data}_j(i)}\right)^2 = \left(\dfrac{\max(\overline{\text{data}_k})}{\text{data}_j(i)}\right)^2, & \text{if } \text{data}_j(i) \text{ exists} \\ 0, & \text{otherwise} \end{cases}$$

where the notation is the same as in Section 4.2. The latter (Case C) is formulated as:

$$M_j(i) = \begin{cases} \left(\dfrac{\max(\overline{\text{data}_k})}{\overline{\text{data}_j}}\right), & \text{if } \text{data}_j(i) \text{ exists} \\ 0, & \text{otherwise} \end{cases}$$

These two assimilation models are implemented, optimal parameter values are obtained, and second-year model results from these two parameter sets are shown in Figs. 10 and 11. Firstly we compare Case A (Fig. 7a–f) with Case B (FFig. 10a–f). In Case B, the errors are larger than in Case A on the whole; the quantity of nitrate becomes less than that of the data; in spring the simulated concentration of PhyS fits the data better than in Case A but higher concentrations are maintained in winter due to the ineffective light limitation as described in Section 6.1; silicate varies less over the annual cycle because PhyL remains inactive.

Again we focus on PhyL for perspective. In Case B, PhyL are not found rapidly increasing or decreasing in the annual cycle, as seen at Peak-1 in Case A. This inactivity of PhyL results from the structure of the cost function. In Case B, the cost function equally evaluates the high concentrations of PhyL in spring and low concentrations in other seasons, while in Case A's cost function the higher concentration in spring is given more weight. Ideally it is desirable that the different weights yield the same optimum parameter values. But, here Case A's data assimilation emphasizing active biological events such as the spring bloom, yields the better result. It is important to comprehend what biological event strongly affects the annual biological cycle and to understand how that event is evaluated through weights in a cost function. Besides this, the results from Case A and Case B suggest that data should be intensively collected in the period of March–June to better constrain models.

Fig. 11 shows results from Case C. Only the result of nitrate and silicate seem to be relatively desirable, while those for the other compartments do not show the seasonal variation at all. This is because, as mentioned in Section 4.2, Lawson et al.'s (1996) weight underestimates compartments with relatively smaller order concentration, or in other words overestimates compartments with relatively larger order.

We should not forget about the assumption that data is reliable. Therefore, in the next assimilation we need to evaluate data variance and error, which result from the precision and accuracy of measurement (due to mechanical and artificial factors) or observed location and time (due to spatial and temporal variations of biological distribution), through more complex weights in a cost function.

7. Conclusions

Ecosystem models with more compartments require that more biological parameters be determined. Almost all of these values depend on measurements and laboratory experiments. In analyzing or improving the model, we often have trouble deciding on which parameters to focus. Our method is effective because it is objective and all parameters need not be changed. This assimilation used only control variables that were expected to strongly affect the model output, and it still yielded more than just the optimum values of control variables. It also revealed what parameter values were misestimated and should be changed.

"NEMURO" still remains immature. In the earliest strategy, even our incomplete data assimilation is useful and successful. It also tells us what parameters excluding control variables should be paid attention to; for example, IoptL and the grazing rate on ZooS by ZooL or ZooP. In the next step, first of all, we need to improve these values systematically and biogeochemically, and then we will quantitatively assimilate data into the one-dimensional ecosystem model to estimate parameter values or discover problems in the model. There is currently an implementation of "NEMURO" coupled to a vertical one-dimensional physical model. A data assimilation should be carried for the physical environment simulated by the turbulent closure model. But in this study we simplified "NEMURO" to the box model. This study has been done to develop our methodology using a simpler model so that

we can use the information gained here in subsequent data assimilations combined with the one-dimensional ecosystem model.

Acknowledgements

We would like to thank Dr. A. Tsuda of Ocean Research Institute of Tokyo University and Dr. H. Saito of Tohoku National Fisheries Research Institution, who provided us the data of Stn.A7 and fruitful discussions on biological interpretation. And thanks are also given to Dr. Y. Yamanaka of Graduate School of Environmental Earth Science of Hokkaido University for constructing the box model. We also wish to express our thanks to Dr. S.L. Smith of the Frontier Research System of Global Change for his help in proof-reading the manuscript. Lastly, the detailed comments provided by the chief-editor and a reviewer are greatly appreciated.

References

Box, M.J., 1966. A comparison of several current optimization methods, and the use of transformations in constrained problems. Compt. J. 9, 67–77.

Eslinger, D.V., Kashiwai, M.B., Kishi, M.J., Megrey, B.A., Ware, D.M., Werner, F.E., 2000. Report of the 2000 MODEL Workshop on lower trophic level modeling. PICES Sci. Rep. 15, 1–77.

Franks, P.J.S., 1995. Coupled physical-biological models in oceanography. Rev. Geophys. (Suppl.), 1177–1187.

Frost, B.W., Kishi, M.J., 1999. Ecosystem dynamics in the eastern and western gyres of the Subarctic Pacific—a review of lower trophic level modelling. Prog. Oceanogr. 43, 317–333.

Fujii, M., Nojiri, Y., Yamanaka, Y., Kishi, M.J., 2002. A one-dimensional ecosystem model applied to time series Station KNOT. Deep-Sea Res. II 49, 5441–5461.

Kasai, H., Saito, H., Kashiwai, M., Taneda, T., Kusaka, A., Kawasaki, Y., Kono, T., Taguchi, S., Tsuda, A., 2001. Seasonal and interannual variations in nutrients and plankton in the Oyashio region: a summary of a 10-years observation along A-line. Bull. Hokkaido Natl. Fish. Res. Inst. 65, 55–134.

Kawamiya, M., 2002. Numerical model approaches to address recent problems on pelagic ecosystems. J. Oceanogr. 58, 365–378.

Kawamiya, M., Kishi, M.J., Yamanaka, Y., Suginohara, N., 1995. An ecological–physical coupled model applied to Station Papa. J. Oceanogr. 51, 635–664.

Kawamiya, M., Kishi, M.J., Yamanaka, Y., Suginohara, N., 1997. Obtaining reasonable results in different oceanic regimes with the same ecological-physical coupled model. J. Oceanogr. 53, 397–402.

Kishi, M.J., 1994. Numerical model for mariculture reflecting upon oxygen consumption by bottom mud. Bull. Coastal Oceanogr. 32, 43–53 (in Japanese with English abstract).

Kishi, M.J., Motono, H., Kashiwai, M., Tsuda, A., 2001. An ecological–physical coupled model with ontogenetic vertical migration of zooplankton in the northwestern Pacific. J. Oceanogr. 57, 499–507.

Lawson, L.M., Spitz, Y.H., Hofmann, E.E., Long, R.L., 1995. A data assimilation technique applied to a predator–prey model. Bull. Math. Biol. 57, 593–617.

Lawson, L.M., Hofmann, E.E., Spitz, Y.H., 1996. Time series sampling and data assimilation in a simple marine ecosystem model. Deep-See Res. 43, 625–651.

Matear, R.J., 1995. Parameter optimization and analysis of ecosystem models using simulated annealing: a case study at Station P. J. Mar. Res. 53, 571–607.

Miller, C.B., Clemons, M.J., 1988. Revised life-history analysis for large grazing copepods in the Subarctic Pacific Ocean. Prog. Oceanogr. 20, 293–313.

Miller, C.B., Frost, B.W., Batchelder, H.P., Clemons, M.J., Conway, R.E., 1984. Life histories of large, grazing copepods in a subarctic ocean gyre: *Neocalanus plumchrus, Neocalanus cristatus,* and *Eucalanus bungii* in the northeast Pacific. Prog. Oceanogr. 13, 201–243.

Parsons, T.R., Takahashi, M., Hargrave, B., 1984. In: Parsons, T.R., Takahashi, M., Hargrave, B. (Eds.), Biological Oceanographic Process, 3rd ed. Pergamon Press, Oxford, New York, 330 pp.

Saito, H., Kasai, H., Kashiwai, M., Kawasaki, Y., Kono, T., Taguchi, S., Tsuda, A., 1998. General description of seasonal variations in nutrients chlorophyll *a* and netplankton biomass along the A-line transect western subarctic Pacific from 1990 to 1994. Bull. Hokkaido Natl. Fish. Res. Inst. 62, 1–62.

Shinada, A., ban, S., Ikeda, T., Tusda, A., Saitoh, H., 2000. Seasonal features of planktonic food chains in the Oyashio region. Bull. Plankton Soc. Jpn. 47, 119–124 (in Japanese).

Spitz, Y.H., Moisan, J.R., Abbott, M.R., Richman, J.G., 1998. Data assimilation and a peragic ecosystem model: parameterization using time series observation. J. Mar. Syst. 16, 51–68.

Steele, J.H., 1962. Environmental control of photosynthesis in sea. Limnol. Oceanogr. 7, 137–150.

Tsuda, A., Saito, H., Kasai, H., 1999. Life histories of *Neocalanus flemingeri* and *N. plumchrus* (Calanoida, Copepoda) in the western subarctic Pacific. Mar. Biol. 135, 533–544.

Tziperman, E., Thacker, W.C., 1989. An Optimal–Control/adjoint—equations approach to studying the oceanic general circulation. J. Phys. Oceanogr. 19, 1471–1485.

Vallino, J.J., 2000. Improving marine ecosystem models: use of data assimilation and mesocosm experiments. J. Mar. Res. 58, 117–164.