

# Tailoring palaeolimnological diatom-based transfer functions

Julien M.J. Racca, Irene Gregory-Eaves, Reinhard Pienitz, and Yves T. Prairie

**Abstract:** This paper presents a method designed to build species-tailored diatom–environment models. Using a pruning algorithm of artificial neural networks, powerful species-tailored models constrained to water temperature, water depth, and dissolved organic carbon were developed from a 109-lake training set from northwestern Canada and Alaska. The reasoning behind the approach is that the implementation of a single, gradient-based, organism–environment relationship should only use species that are comprehensively influenced by the variable of interest. By pruning species according to their relevance to each of the three studied variables, the cross-validated performances of all three models were significantly increased, suggesting that nonrelevant species have corrupting influences and need to be removed. The removal of corrupting species also suggests that palaeolimnological transfer functions based on an appropriate subset of useful species are more independent.

**Résumé :** Nous présentons une méthode pour construire des modèles diatomées–environnement basés sur les espèces. À l'aide d'un algorithme d'élagage tiré des réseaux neuraux artificiels, nous avons mis au point des modèles basés sur les espèces avec une contrainte pour la température de l'eau, la profondeur et la concentration de carbone organique dissous à partir d'une série expérimentale de données sur 109 lacs du nord-ouest canadien et de l'Alaska. Le raisonnement qui sous-tend la méthode est que l'établissement d'une relation particulière organisme–environnement basée sur un gradient ne devrait utiliser que des espèces qui sont influencées de façon globale par la variable considérée. Par l'élagage des espèces d'après leur pertinence vis-à-vis chacune des trois variables étudiées, les performances des trois modèles déterminées par validation croisée sont significativement améliorées, ce qui indique que les espèces non pertinentes ont une influence nuisible et doivent être retirées. Le retrait des espèces nuisibles indique aussi que les fonctions de transfert paléolimnologiques basées sur un sous-ensemble approprié d'espèces utiles sont plus indépendantes.

[Traduit par la Rédaction]

## Introduction

Transfer functions that quantify the modern relationships between the composition of diatom assemblages and environmental variables for a set of lakes are routinely used in palaeolimnological studies to infer quantitative environmental changes from past diatom assemblage data. Several methods, based on different algorithm types, have been successfully applied to model the complex relationships between taxon assemblages and environmental variables: weighted averaging regression – calibration based approach (ter Braak and van Dam 1989; Birks et al. 1990), weighted averaging partial least-squares regression (ter Braak and Juggins 1993), maximum likelihood based approach (ter Braak and van Dam 1989; ter Braak et al. 1993), full probability based approach (Bayesian modeling) (Ellison 1996; Toivonen et al. 2001; Vasko et al. 2000), and artificial neural networks based approach (Racca et al. 2001; Köster et al. 2004).

While it is clear that the predictive ability of any of these methods depends ultimately on the degree to which the distribution of the biota assemblages is actually determined by environmental characteristics, it is also affected by the sampling characteristics of the modern data set (distributions and ranges of the environmental variables, number of samples, number of taxa, amount of noise, etc.) (Racca and Prairie 2004). Because relationships between the composition of species assemblages and environmental variables are extracted from a restricted set of lakes, the predictive ability of a particular model is necessarily dependent on the choice of lakes included in the training set. In general, modern training sets are designed either to be as encompassing as possible or to focus on a predetermined environmental gradient. Thus, depending on the subsequent use of a model, lakes in a training set are first chosen to cover a large range of the environmental variable of interest but also to cover a small range of other variables. With such a design, it is expected

Received 13 January 2004. Accepted 11 September 2004. Published on the NRC Research Press Web site at <http://cjfas.nrc.ca> on 21 February 2005.  
J17917

**J.M.J. Racca<sup>1</sup> and R. Pienitz.** Paleolimnology–Paleoecology Laboratory, Centre d'Études Nordiques et Département de Géographie, Université Laval, Québec, QC G1K 7P4, Canada.

**I. Gregory-Eaves.** Department of Biology, University of Ottawa, 150 Louis Pasteur Street, Ottawa, ON K1N 6N5, Canada.

**Y.T. Prairie.** Département des Sciences Biologiques, Université du Québec à Montréal, Case postale 8888 succ. Centre-Ville, Montréal, QC H3C 3P8, Canada.

<sup>1</sup>Corresponding author (e-mail: [racca.julien@courrier.uqam.ca](mailto:racca.julien@courrier.uqam.ca)).

that the variation in assemblage data will be attributed principally to the changes in the environmental variable of interest. Also, it is expected that the effects of other variables will have little impact on the variation in the assemblage data. As a result, diatom transfer functions based on such optimal training sets generally exhibit good predictive power (e.g., Fallu and Pienitz 1999), whereas models based on more limnologically diverse lakes, where variation in species assemblages can be influenced by several environmental gradients, exhibit generally lower predictive power (e.g., Philibert and Prairie 2002).

In this study, we suggest that much more powerful models can be developed from a less optimal training set design. Here, we apply a method to build optimal subtraining sets from an existing nonoptimal full training set. In contrast with the aforementioned design where the training set is constructed based on a choice of lakes, the proposed method deals with the selection of taxa according to their relative contribution in a model. We hypothesize that most of the complexity in a training set, which can be attributed either to the effect of the multiple environmental influences on assemblage data in limnologically diverse lakes or to stochastic variability within the data set, could be better constrained if only species whose distribution is comprehensively dependent on the variable being studied are included in a training set.

To test this idea, we developed three diatom-based training sets constrained to (i) water temperature, (ii) water depth, and (iii) dissolved organic carbon (DOC) from an original training set where the lakes spanned a wide range in all of these three gradients resulting from the combination of two diatom-training sets from northwestern Canada and Alaska (Pienitz et al. 1995; Gregory-Eaves et al. 1999). The main objective of the study was to show that the predictive ability of a model can be increased when it is species tailored to a particular variable (i.e., when only the subset of species whose distribution and abundance are comprehensively related to the environmental variable is used).

## Materials and methods

### Study area

The extended modern training set of 109 lakes used in this study results from the combination of a 58-lake training set from the Yukon and the Northwest Territories (Pienitz et al. 1995) and a 51-lake training set from Alaska (Gregory-Eaves et al. 1999) (Fig. 1). The 58 lakes from the Yukon and the Northwest Territories are located between Whitehorse and Tuktoyaktuk ranging from 60°37'N to 69°35'N and from 132°04'W to 138°22'W. The locations of the 51 Alaskan lakes range from 60°28'N to 69°35'N and from 141°38'W to 150°49'W. Details of the limnological, physiographic, and geological features for each training set are reported in Pienitz et al. (1995, 1997) and Gregory-Eaves et al. (1999). The lakes of each training set were chosen to span a broad north-south climatic gradient, but by combining the two sets, a fairly large east-west gradient is also captured. Combining the two data sets enlarges the ranges of several limnologically and nonlimnologically associated variables. We summarize the ranges covered by lake-related variables for the 109 training set sites in Table 1.

### Sample collection

Sample collection in the field and measurements of related environmental variables (water temperature, water depth, pH, conductivity, and water transparency) were done in the summers of 1990 and 1996 for the Yukon and Alaskan training sets, respectively. Laboratory analyses of nutrients, major ions, and trace metals were performed by the National Water Research Institute (Burlington, Ontario) following standard methods. Full details of field sampling methods, water chemistry, and other analyses are provided in Pienitz et al. (1995, 1997) and Gregory-Eaves et al. (1999, 2000).

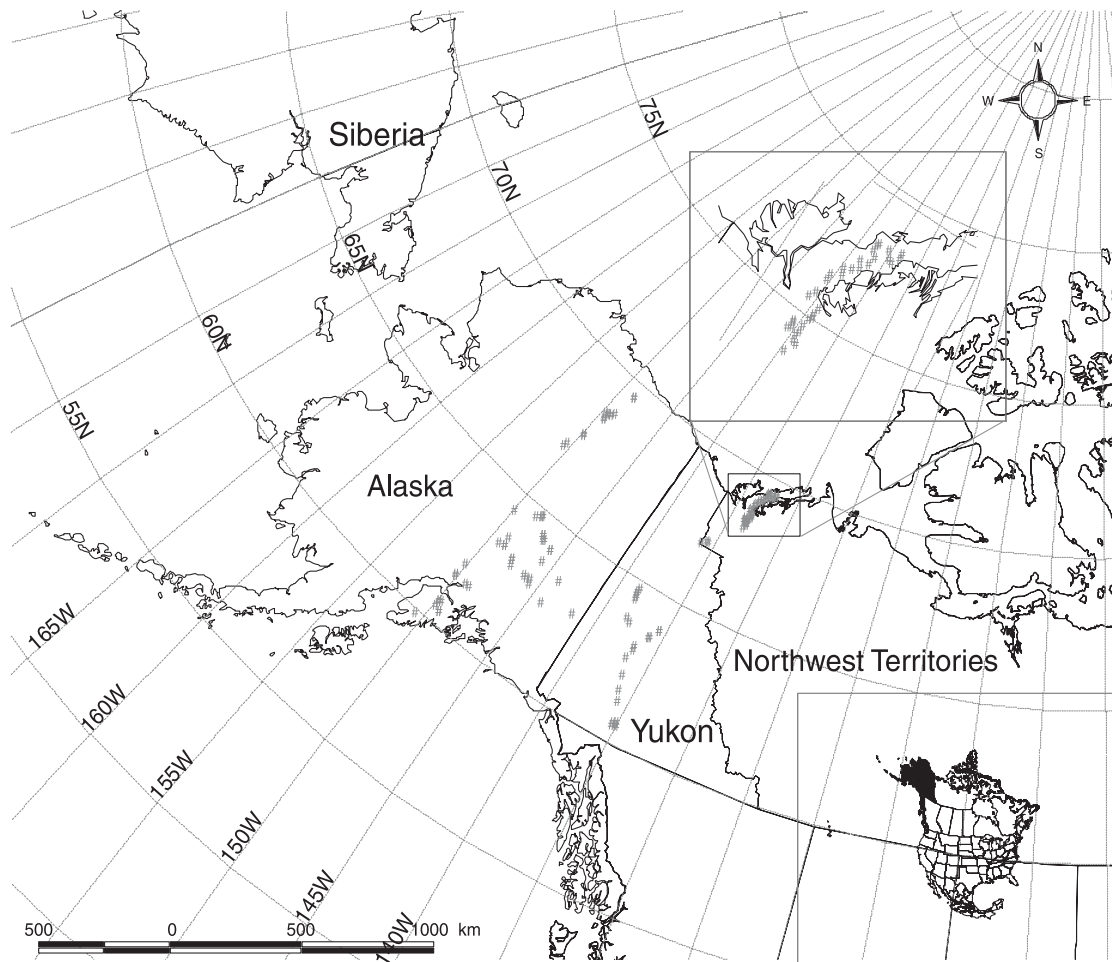
Diatom slides were prepared by treating surface sediment samples of each site using standard methods (Pienitz et al. 1995). Identification and enumeration of diatom valves were done along random transects under oil immersion using light microscopy. For each slide, between 300 and 500 diatom valves were identified to the lowest taxonomic level using primarily the following taxonomic sources: Krammer and Lange-Bertalot (1986–1991), Foged (1981), Patrick and Reimer (1966, 1975), Cumming et al. (1995), Camburn et al. (1984–1986), and Fallu et al. (2000). The harmonization of species identification between the two training sets was made on multiple occasions during several diatom taxonomic workshops (e.g., Arctic and Antarctic Diatom Workshop (5th), 1995, unpublished report, Queen's University, Kingston, Ontario). A total of 545 diatom species were identified in the 109 surface samples but only 259, with a relative abundance of 1% in at least three samples, were used for the analyses. Details regarding these species can be found in Appendix A.

### Numerical analyses

#### *Artificial neural networks transfer functions*

Although the ecological response curve of all species in regard to one environmental gradient is often assumed to be unimodal, a mixture of different response curves is often observed in palaeolimnological training data sets (e.g., unimodal, skewed unimodal, sigmoidal increasing, or sigmoidal decreasing) (Birks 1998). We used artificial neural networks to implement the transfer functions because artificial neural networks are capable of accommodating the full range of species response curves (Leshno et al. 1993). So-called multilayer perceptrons, one type of network architecture trained with a back-propagation algorithm (Rumelhart et al. 1986), have been successfully applied in quantitative palaeolimnology (Racca et al. 2001; Philibert et al. 2003; Köster et al. 2004) and palaeoceanography (Malmgren and Nordlund 1997; Peyron and De Vernal 2001). Here, the same network architecture is used. In this type of network, neurons are arranged in a distinct layered topology: one input layer (representing independent variables (species)), one hidden layer, and one output layer (representing dependent variables (environmental variables)). All neurons from one layer are connected to all neurons in the adjacent layers and all of these connections have a weight that represents the parameters of the network. By back-propagation (iterative process), the weights of the connections are adjusted by feeding a set of input-output pattern pairs many times. As a result of these weight adjustments, internal hidden neurons, which are not part of the input or output, come to represent important

**Fig. 1.** Map of study sites showing the position of the 51 Alaskan and 58 Yukon calibration lakes. Inset shows the location in North America.



features of the task domain and the relationship between input and output is captured by the interactions of these units. This relationship (function) can then be used to predict new output (i.e., values of environmental variables) from new input data (i.e., species assemblages). Background information on neural networks is available in various introductory textbooks such as Bishop (1995) and more details of this methodology as applied to palaeolimnology can be found in Racca et al. (2001).

#### ***Building constrained training set based transfer functions***

To build optimal subtraining sets, we used the skeletonization pruning algorithm of artificial neural networks (Moser and Smolensky 1989). Pruning algorithms (e.g., Reed 1993) are comparable with backward elimination in regression models. Backward elimination starts with all independent variables and sequentially removes the least relevant one and stops if the model performance drops below a given threshold by the removal of any of the remaining independent variables. The skeletonization algorithm was already applied to estimate the functionality of individual species in the Surface Water Acidification Program training data set (Birks et al. 1990) with the objective of removing nonrelevant and redundant species from a pH model (Racca et al. 2003). Using skeletonization, the relevance of one species (i.e., its relative

contribution) is determined as an estimation of the change in the model error (i.e., root mean square error, RMSE) when this species is omitted: the more the model error increases, the more a species is relevant and vice versa. The estimated relevance can, therefore, be viewed as a direct measure of the numerical importance of each species in the model and can be used to remove species according to their relative contribution.

Here, we used the skeletonization method to prune diatoms according to their contribution to water temperature, water depth, and DOC. Pruning routines were applied until optimal constrained sets of species were reached. This involves consecutive and alternative steps of training–pruning simulation, as skeletonization pruning is a dynamic procedure. Details of the skeletonization–pruning algorithm used in this study can be found in Racca et al. (2003).

#### ***Pruning procedures and model validation***

The same pruning parameter setup was applied to each simulation: from the initial full data set, noncontributing species were removed (one by one) according to their numerical importance until the removal of one species increased the model error over a fixed criterion. For each training–pruning simulation, species removal begins when the model has converged (i.e., when the error remains quasi-

**Table 1.** Summary of the ranges of environmental variables for both training sets and for the combined training set.

	Yukon training set				Alaska training set				Combined training set			
	Min.	Max.	Mean	SD	Min.	Max.	Mean	SD	Min.	Max.	Mean	SD
ALT (m)	15.0	1387.0	361.6	122.0	35.0	1028.7	429.4	320.0	15.0	1387.0	393.4	361.5
DEPTH (m)	1.0	49.0	7.4	4.0	0.8	54.0	10.5	5.0	0.8	54.0	8.8	10.3
pH (unity)	5.9	9.3	7.9	8.1	7.0	9.3	7.9	7.9	5.9	9.3	7.9	8.0
TEMP (°C)	12.0	23.0	18.2	18.0	9.8	22.9	15.5	16.1	9.8	23.0	16.9	17.0
COND (µS·cm <sup>-1</sup> )	24.0	1500.0	160.7	112.0	6.0	434.0	100.7	82.0	6.0	1500.0	132.7	168.4
TPU (µg·L <sup>-1</sup> )	3.0	55.1	15.6	12.3	1.3	475.8	24.9	11.2	1.3	475.8	20.0	46.1
DOC (mg·L <sup>-1</sup> )	3.1	35.1	12.3	10.8	2.7	257.0	66.5	44.4	2.7	257.0	37.7	54.1
DIC (mg·L <sup>-1</sup> )	0.3	134.2	17.9	13.1	1.1	59.5	13.6	12.0	0.3	134.2	15.9	17.9
TKN (µg·L <sup>-1</sup> )	72.0	1293.0	440.8	368.0	50.0	2400.0	511.4	346.0	50.0	2400.0	473.8	373.9
SiO <sub>2</sub> (mg·L <sup>-1</sup> )	0.1	12.5	2.2	1.0	0.1	10.8	2.2	1.2	0.1	12.5	2.2	2.8
SO <sub>4</sub> (mg·L <sup>-1</sup> )	0.5	1242.0	41.3	8.9	0.2	43.1	7.3	5.1	0.2	1242.0	25.4	121.8
Ca (mg·L <sup>-1</sup> )	3.5	50.3	19.6	18.6	0.3	50.0	14.4	13.2	0.3	50.3	17.2	12.2
Na (mg·L <sup>-1</sup> )	0.2	187.0	8.9	3.9	0.2	11.6	1.9	1.0	0.2	187.0	5.6	18.6
K (mg·L <sup>-1</sup> )	0.1	29.9	1.9	1.2	0.2	15.6	1.1	0.6	0.1	29.9	1.5	3.3
Cl (mg·L <sup>-1</sup> )	0.2	75.2	7.3	2.0	0.1	6.0	1.4	1.1	0.1	75.2	4.5	10.2
Fe (µg·L <sup>-1</sup> )	5.9	1660.0	213.3	86.6	10.0	1920.0	180.2	57.0	5.9	1920.0	197.8	342.3
Mn (µg·L <sup>-1</sup> )	2.0	160.0	23.7	15.0	1.0	58.7	12.7	8.8	1.0	160.0	18.5	23.8

**Note:** ALT, altitude; DEPTH, water depth; TEMP, water temperature; COND, conductivity; TPU, total phosphorus unfiltered; DOC, dissolved organic carbon; DIC, dissolved inorganic carbon; TKN, total Kjeldahl nitrogen.



constant even after iterations are added) and species removal is stopped when the error increases by 10% over the error at the starting point of pruning. This procedure was repeated as long as the training–pruning simulations were possible (i.e., a new training–pruning simulation was applied to reduced data sets until no retraining iteration after a species deletion could lower the error below the threshold necessary to end pruning). We kept this threshold quite low (10% over the RMSE at the starting point) so as to ensure that the removal steps were small (i.e., where few species are removed at each step instead of only one or two training–pruning simulations where many species are removed). Moreover, the number of iterations for the retraining phase of each training–pruning simulation was kept low (20) for the same reason (i.e., the longer is the retraining phase, the better the model error converges and more species are removed at each training–pruning simulation). Although several pruning parameter configurations could be used, our experience is that the order of species removal is not dramatically changed. Ultimately, only the number of pruned species can be modified by the pruning parameter configuration. The pruning procedures were performed using Stuttgart Neural Network Simulator v4.2 (Zell et al. 1996).

However, because skeletonization pruning of species is based on the change of the apparent error function (i.e., apparent RMSE), a validation of the pruned model was made using a standard back-propagation model with cross-validation based on leave-one-out jackknifing. It is this cross-validated error term that we ultimately seek to improve by tailoring the transfer function (i.e., by pruning out the species judged irrelevant for the particular variable modeled). For this purpose, the same methodology as proposed in Racca et al. (2001, 2003) was applied using a cross-validation routine (CROSVL) (R. Racca, Département des Sciences, Université de Nouvelle Calédonie, BP 4477, 98847 Noumea CEDEX, Nouvelle-Calédonie, unpublished program) of YANNS (Boné et al. 1998).

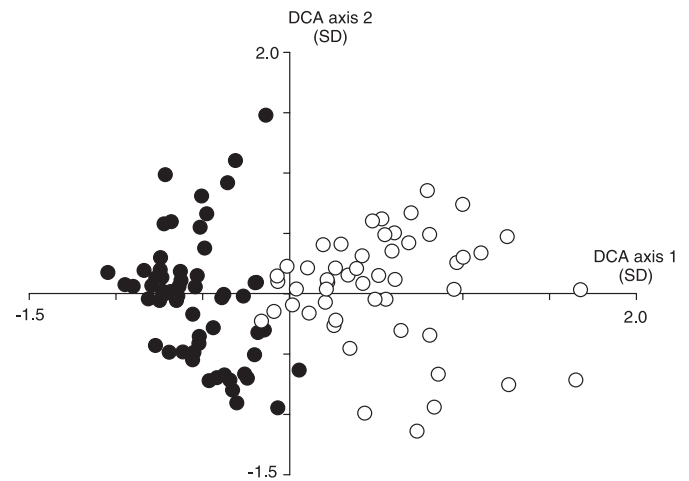
## Results

### Data set characteristics

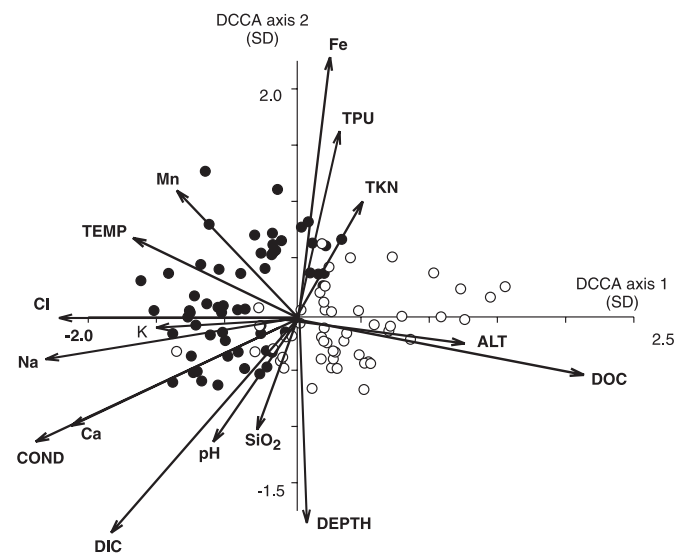
Because of the strong latitudinal gradient covered by both training sets, several distinctive changes in diatom assemblage composition were apparent between the boreal forest sites in the south and the arctic tundra sites in the north (Pienitz et al. 1995; Gregory-Eaves et al. 1999). Here, to explore the similarities or dissimilarities in the diatom assemblages between the two training sets, a detrended correspondence analysis grouping all (259) species was carried out. The detrended correspondence analysis shows a clear distinction between species assemblages from the 58 lakes in the Yukon data set and the 51 lakes in the Alaskan data set (Fig. 2). The percentage of cumulative variance captured by the first two axes is 11.4% of the species data. The first two axes of the detrended correspondence analysis are significant according to Monte Carlo permutation tests (with 199 unrestricted permutations,  $p \leq 0.05$ ).

A detrended canonical correspondence analysis was carried out (Fig. 3) to test if the dissimilarity between the diatom assemblages can be explained by environmental differences between the two regions. Only if the variables that

**Fig. 2.** Detrended correspondence analysis (DCA) plot showing diatom assemblage dissimilarities between Alaskan and Yukon lakes. Solid circles represent lakes from the Yukon and Northwest Territories, and open circles represent Alaskan lakes.



**Fig. 3.** Detrended canonical correspondence analysis (DCCA) plot of the 109-lake data set. Solid circles represent lakes from the Yukon and Northwest Territories, and open circles represent Alaskan lakes. TEMP, water temperature; DEPTH, water depth; TPU, total phosphorus unfiltered; TKN, total Kjeldahl nitrogen; COND, conductivity; DOC, dissolved organic carbon; DIC, dissolved inorganic carbon.



cover different ranges in the two regions explain a great proportion of species variation we can ensure that assemblage dissimilarity is due to these environmental variables. The percentage of cumulative variance in the species–environment relationship captured by the first two detrended canonical correspondence analysis axes is 35.45%, with eigenvalues of 0.284 for axis 1 and 0.157 for axis 2. All canonical axes are significant according to Monte Carlo permutation tests (with 199 unrestricted permutations,  $p = 0.018$ ). In the detrended canonical correspondence analysis plot, ordination of the lakes on the first axis shows a separa-

tion between Yukon and Alaskan lakes. From the relative position of the lakes in the environmental space, it is clear that species assemblages in Alaska are generally associated with more DOC-rich and more high-altitude conditions than species assemblages in the Yukon and Northwest Territories.

### Initial training set based transfer functions

From the initial training set of 259 species and 109 sites, diatom – artificial neural networks based models for water temperature, water depth, and DOC were built. We present plots of observed values versus predicted values (Figs. 4a–4c). The predictive power of each model is low as expressed here as the relative measure of relationship strength ( $r^2$  jackknife) between the predicted and observed values. The absolute measures of uncertainty associated with the predictions (i.e., the RMSE of prediction (RMSEP)) of each model are approximately twice those generally obtained in other studies of diatom-based water temperature, water depth, and DOC models (e.g., Pienitz et al. 1995; Fallu and Pienitz 1999; Gregory-Eaves et al. 1999). Moreover, the relationships between observed and predicted values are not linear for either the water temperature model or the DOC model, suggesting that these models are strongly biased.

### Constrained training set based transfer functions

More optimal training sets were obtained by pruning the initial full species data set (259) according to the contribution of the diatom species to the performance of either the water temperature, water depth, or DOC model. By using the skeletonization algorithm until the “apparent” performances (RMSE) of the models increased over 10% from the minimum, the species-tailoring procedure reduced the initial data set by 65.4% (260 to 90 species) for the water temperature based model, 59.6% (259 to 105) for the water depth based model, and 49.2% (259 to 132) for the DOC-based model. Species composition of the three subtraining sets is shown in Appendix A. Only 24 species (9.2% of the initial full training set) are common to all three tailored sets. Forty-seven species (18.1%) are common to the water temperature and water depth training sets, 50 species (19.2%) to the water temperature and DOC sets, and 55 species (21.1%) to the water depth and DOC models. That the pruning algorithm selects different groups of species indicates that the relative importance of several species to the apparent statistics of a model is critically dependent on the environmental variable considered. Nevertheless, the question remains as to what extent each group of species can be used to improve the predictions of the environmental variable for which they were selected. In other words, how does the exclusion of species improve the cross-validated predictive performances of the models? To answer this question, we built cross-validated (leave-one-out jackknifing) models for each of the three environmental variables using their corresponding subtraining set (plots of observed versus predicted values are presented in Figs. 4d–4f). The predictive power of each pruned training set based model is improved when compared with the corresponding initial training set based models (Figs. 4d–4f). The strength ( $r^2$  jackknife) of the relationship between the predicted and observed values increased from 0.34 to 0.68 for water temperature, from 0.60 to 0.77 for water depth, and from 0.33 to 0.67 for DOC. This leads to a decrease in

the absolute measure of uncertainty associated with the predictions of each model. The improvement in predictive power of each constrained model is statistically significant ( $F = 1.92, 1.7, \text{ and } 1.79; p = 0.0074$ ) for temperature, depth, and DOC, respectively.

## Discussion

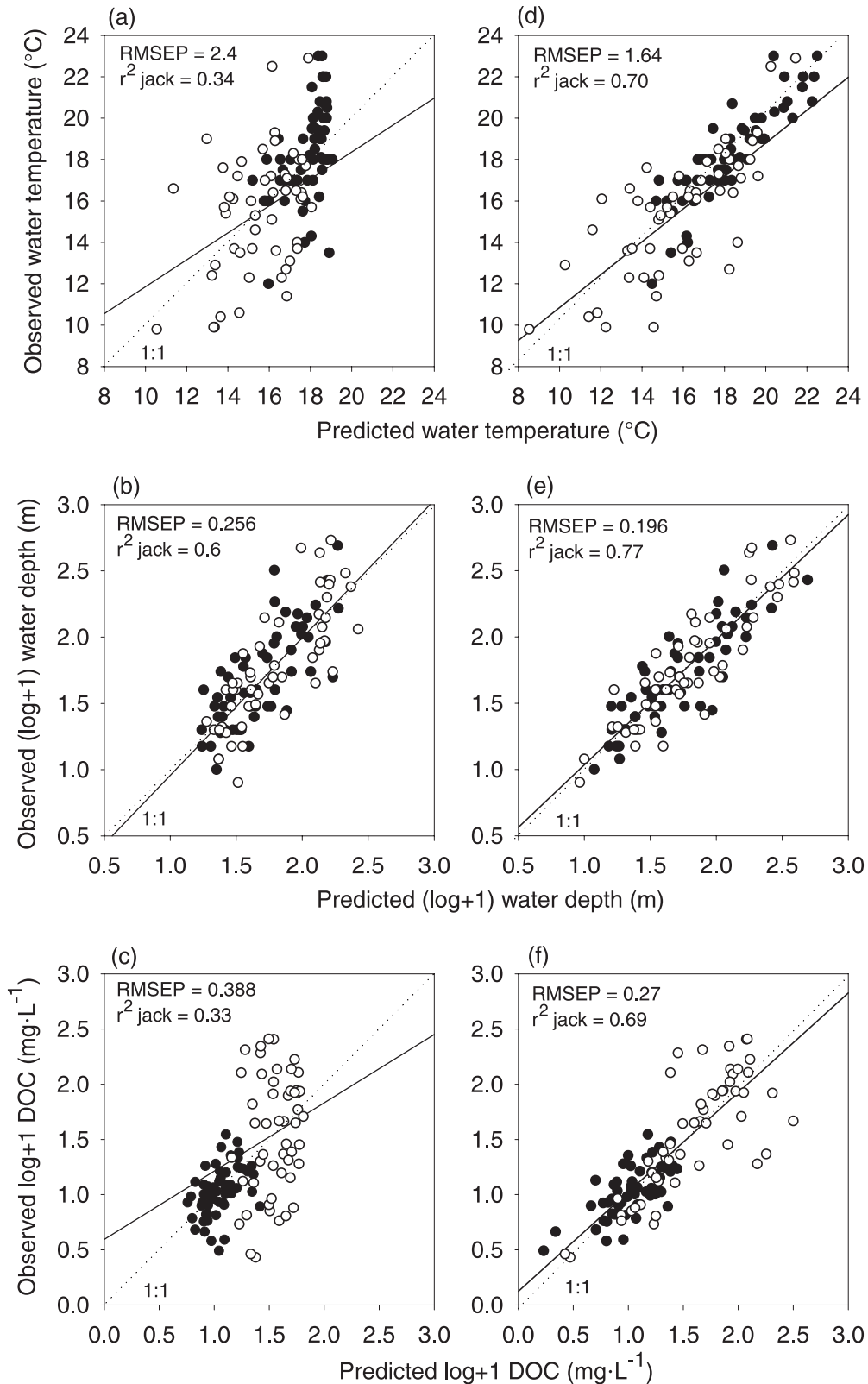
### Species selection and model improvement

The proposed environmentally dependent pruning method used here allowed us to build tailored training sets for diatom-based water temperature, water depth, and DOC models. The constrained training sets were built separately from the initial training set according to the relative contribution of each species to each of the three variables studied. By removing noncontributing species, the predictive power of the models increased significantly in all three cases, suggesting that pruning is an efficient method for improving model performance. Indirectly, these results also imply that the species that were removed by the pruning method were in fact corrupting the models based on all species. Moreover, the species removal also improved the prediction characteristics of the models. For example, when all species were included in the water temperature model, the predictions never exceeded 20 °C, while the predictions were very close to the observations (up to 24 °C) when only the species that seemed to be useful to model water temperature were included. Clearly, the removal of noncontributing species is beneficial, both to improve prediction power and to reduce model bias.

Our results clearly demonstrate that the assumption that all species are ecologically relevant and therefore contribute to the accuracy of the prediction is questionable. Nevertheless, the question of how the noncontributing species affect the model remains difficult: are noncontributing species simply a source of random noise or is there a more complex coupling between the species and their environment? By analogy with simple modeling techniques such as multiple regression, the inclusion of species that carry no information about their environment should not negatively affect performance of a model: the modeling procedure should normally simply ignore them by assigning them very little weight. However, because species removal actually improved model performance, our results suggest that these species had a genuinely corrupting influence. In our view, this is most likely conceivable for species that are multiply determined (i.e., for species that are strongly influenced by more than one environmental gradient). Unless these influencing gradients are always correlated to the same extent and in the same way, no modeling technique can reasonably cope with possibly conflicting environmental signals.

While these multiply determined species are probably an important source of model corruption, there may also be other ways by which variations in the abundance of certain species actually confound a model. For example, there may be several alternative stable species assemblages for a given lake driven by interspecific relationships among the diatoms. As such, these types of relationships are never considered in model building, although they are likely to occur in nature. However, we know of no method that is able to assess the relative importance of these confounding influences. To this extent, the question of how noncontributing species affect

**Fig. 4.** Plots of observed against jackknife-predicted values for the water temperature model, water depth model, and DOC model (*a–c*) when all species are used and (*d–f*) when pruned training sets are used. The jackknife-predicted values are based on a validation set (leave-one-out). Fitted lines are based on model I regression. Solid circles represent lakes from the Yukon and Northwest Territories, and open circles represent Alaskan lakes.



the predictive capacities of a model remains open. Until such a question can be addressed, we argue that the best palaeolimnological models will be those tailored only with the appropriate species (i.e., those that are useful in a predictive sense). In this context, pruned models for water temperature, water depth, and DOC do not necessarily contain species that are exclusively influenced by only one of these variables. Indeed, a species that is multiply determined can still be useful if at least some of the information that it carries can, in some sense, be generalized. This appears to be the case for all species that are common to all tailored training sets (9.2% of the species in our case). Conversely, the species that are multiply influenced but for which no generalization can be achieved will necessarily be eliminated by the pruning algorithm.

### Toward independent transfer functions

That noncontributing species can corrupt the empirical predictive power of the models suggests that any changes in the distribution and abundance of these species in the past could affect their reconstruction capabilities. This is an important lacunae of unpruned models. Because the distribution and abundance of these noncontributing and possibly corrupting species may be strongly influenced by several variables (physical, chemical, and (or) biological) characteristic of the lake system, changes in any of these environmental parameters will alter their abundance. Thus, we could wrongfully infer changes in the variable of interest even if it remained unchanged. Because of this, we suggest that the removal of noncontributing taxa can potentially reduce the effects of other environmental influences. We contend that tailored models are probably less sensitive than others and more independent because they are specifically designed to quantify the changes of one environmental variable using only species that respond to this variable in a way that can be generalized.

Until now, the effects of multiple influence and interaction of environmental gradients (correlated or not) on species assemblages were only partially addressed in the design phase of palaeolimnological studies by selecting lakes to be included in the training set (S. Hausmann and F. Kienast, Paleolimnology–Paleoecology Laboratory, Centre d'Études Nordiques, Département de Géographie, Université Laval, Québec, QC G1K 7P4, Canada, unpublished data). However, while a preselection is often possible in certain regions, mainly for those where information on lakes is easy to obtain before sampling, sampling is limited and logistically difficult and expensive, for example, in remote northern regions. In these regions, controlling the number of influencing variables by reducing the number of sites in an existing training set is one alternative way to constrain the multiple environmental influences on species. In this case, a sub-training set of selected sites is defined in which the environmental variable of interest has the largest range possible but in which the ranges of secondary variables are also kept narrow (S. Hausmann and F. Kienast, unpublished data). However, this a posteriori selection could be problematic for at least two reasons. First, the number of sites, often an important parameter in the success of the modeling approach, could end up being too small if many secondary variables are detected in the original training set. Second, and more

importantly, a model based on such a “site-selected” training set where few situations of interaction and (or) multiple influences are structuring the species data would be incapable of implementing these situations. As a result, such models will perform poorly when applied to down-core species data if interactions and (or) multiple influences occurred in the past. This second problem is also relevant to models based on modern training sets where an a priori selection of sites is made to avoid the effects of secondary gradients.

Ideally, the implementation of every organism–environment relationship should be based on modern training sets where situations of multiple influences and (or) interactions structure species assemblages: only a model that has the possibility of “learning” from multivariate patterns will have the potential to give realistic inferences when applied to multiply induced past assemblage data. Thus, the more examples of similar situations of multiple influences or interactions occur in a training set, the more a model will be capable of implementing these situations. However, if a model is not able to learn from some situations because too few examples of these occur in a given training set, then these situations need to be avoided. Therefore, we suggest that more effort should be made toward the development of efficient calibration models in which only nongeneralized situations of multiple influences and interactions are avoided (i.e., like our pruning algorithm do) rather than toward the development of calibration models in which all situations of multiple influence or interaction (i.e., generalizable and not) are avoided (i.e., like in methods based on a priori or a posteriori selection of sites). In other words, more attention should be given to build efficient univariate-based models from a multivariate organism–environment relationships training set rather than attempting to build univariate models from a pseudo-univariate organism–environment relationships training set. To be efficient, an appropriate univariate model (based on multivariate relationships) should be capable of reaching the two following goals. First, the model should have the ability to implement only the generalizable relationships between assemblage data and each structuring environmental variable (i.e., where species that suffer from nongeneralizable multivariate interaction are excluded). Second, it should have the capacity to make independent predictions. The method proposed in this study is designed to reach these two goals: the selection of species is made to create an optimal model for a given variable by removing species whose distributions are independent of the variable of interest. In addition, by making an environmentally dependent selection of species to be included in a particular training set, a transfer function based on these species will be quasi-independent (a certain dependence will occur only in cases where species are common to several subtraining sets).

We believe that these observations to be important, as independent transfer functions are required in situations where any reconstructed environmental variable may be confounded by the influence of other factors. For example, few researchers have attempted to model water depth because changes in nutrient concentration and (or) light quality may or may not covary with lake level fluctuations (Wolin and Duthie 1999). Similarly, reconstruction of lake depth is problematic because changes in lake level could be the con-



sequence of changes in temperature and (or) the result of changes in relative humidity. The application of our three quasi-independent diatom transfer functions for reconstruction of past changes in lake depth, lake water temperature, and DOC concentration could provide substantial insight into the magnitude of past climatic and environmental changes in northwestern Canada and Alaska.

In conclusion, in this study, we have applied a method that is designed to build tailored palaeolimnological models in situations where several important environmental variables structure species data in a training set. In contrast with the idea of a priori or a posteriori selection of lakes to reduce secondary gradients, the proposed method deals with the selection of a subset of numerically useful species. The reasoning behind the approach is that the implementation of a single gradient-based organism–environment relationship should use only species that are comprehensively influenced by the variable of interest. Such an approach based on taxon selection appears to be attractive for two reasons. First, the selection of species is made to create an optimal model for a given variable by removing taxa with distributions that are independent of the variable of interest. The resulting tailored training set can then be used to develop more powerful models. Second, several quasi-independent models of species–environment relationships could be developed from the same original training set because each model will be based on different subsets of relevant species. Once validated using other data sets, this method could prove a very useful tool for developing several tailored transfer functions from the same modern training set and (or) from training sets where several environmental variables are important in structuring species assemblage data.

## Acknowledgements

This paper is a contribution to the Natural Sciences and Engineering Research Council of Canada (NSERC)-sponsored collaborative research opportunity (CRO) project on “Late Pleistocene paleoclimates of eastern Beringia”. It is also a contribution to groupe de recherche interuniversitaire en limnologie (GRIL)-UQAM. This research has been supported by NSERC grants to R. Pienitz and Y.T. Prairie. Logistic support by Centre d’Études Nordiques is gratefully acknowledged. We thank the reviewers and Matthew Wild for their constructive comments on the manuscript.

## References

- Birks, H.J.B. 1998. D.g. Frey & E.s. Deevey Review #1 — numerical tools in palaeolimnology — progress, potentialities, and problems. *J. Paleolimnol.* **20**: 307–332.
- Birks, H.J.B., Line, J.M., Juggins, S., Stevenson, A.C., and ter Braak, C.J.F. 1990. Diatoms and pH reconstruction. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **327**: 263–278.
- Bishop, C.M. 1995. *Neural networks for pattern recognition*. Oxford Clarendon Press, Oxford, UK.
- Boné, R., Crucianu, M., and Asselin de Beauville J.P. 1998. Yet another neural network simulator (YANNS). Proceedings of the Conference, NEURAl Networks and their Applications (NEURAP’98), Marseille, France. pp. 421–424.
- Camburn, K.E., Kingston, J.C., and Charles, D.F. 1984–1986. Paleocological investigation of recent lake acidification. PIRLA Diatom Iconograph. PIRLA Unpubl. Rep. Ser. No. 3, Indiana University, Bloomington.
- Cumming, B.F., Wilson, S.E., Hall, R.I., and Smol, J.P. 1995. Diatoms from British Columbia and their relationship to salinity, nutrients and other limnological variables. *Bibliotheca Diatomologica* 31. J. Cramer Verlag, Stuttgart, Berlin.
- Ellison, A.M. 1996. An introduction to Bayesian inference for ecological research and environmental decision-making. *Ecol. Appl.* **6**: 1036–1046.
- Fallu, M.-A., and Pienitz, R. 1999. Diatomées lacustres de Jamésie–Hudsonie (Québec) et model de reconstitution des concentrations de carbone organique dissous. *Écoscience*, **4**: 603–620.
- Fallu, M.-A., Allaire, N., and Pienitz, R. 2000. Freshwater diatoms from northern Québec and Labrador (Canada). *Bibliotheca Diatomologica*, 45. J. Cramer Verlag, Stuttgart, Berlin.
- Foged, N. 1981. *Diatoms in Alaska*. Bibliotheca Phycologia 53. J. Cramer Verlag, Vaduz, Berlin.
- Gregory-Eaves, I., Smol, J.P., Finney, B.P., and Edwards, M.E. 1999. Diatom-based transfer functions for inferring past climatic and environmental changes in Alaska, USA. *Arct. Antarct. Alp. Res.* **31**: 353–365.
- Gregory-Eaves, I., Smol, J.P., Finney, B.P., Lean, D.R.S., and Edwards, M.E. 2000. Characteristics and variation in lakes along a north–south transect in Alaska. *Arch. Hydrobiol.* **147**: 193–223.
- Köster, D., Racca, J.M.J., and Pienitz, R. 2004. Diatom-based inference models and reconstructions revisited: methods and transformations. *J. Paleolimnol.* **32**: 233–245.
- Krammer, K., and Lange-Bertalot, H. 1986–1991. *Bacillariophyceae*. In *Süßwasserflora von Mitteleuropa*, Band 2(1–4). Edited by H. Ettl, J. Gerloff, H. Heynig, and D. Mollenhauer. Gustav Fischer Verlag, Stuttgart and Jena, Berlin.
- Leshno, M., Lin, V.Y., Pinkus, A., and Schocken, S. 1993. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, **6**: 861–867.
- Malmgren, B.A., and Nordlund, U. 1997. Application of artificial neural networks to paleoceanographic data. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **136**: 359–373.
- Moser, M., and Smolensky, P. 1989. Skeletonization: a technique for trimming the fat from a network via relevance assessment. *Adv. Neural Information Processing Syst.* **1**: 107–115.
- Patrick, R., and Reimer, C. 1966. *The diatoms of the United States*. Vol. 1. Fragilariaceae, Eunotiaceae, Achnantheaceae, Naviculaceae. Academy of Natural Sciences of Philadelphia, Philadelphia, Pa.
- Patrick, R., and Reimer, C. 1975. *The diatoms of the United States*. Vol. 2. Entomoneidaceae, Cymbellaceae, Gomphonemaceae, Epithemiaceae. Academy of Natural Sciences of Philadelphia, Philadelphia, Pa.
- Peyron, O., and De Vernal, A. 2001. Application of artificial neural networks (ANN) to high-latitude dinocyst assemblages for the reconstruction of past sea-surface conditions in arctic and subarctic seas. *J. Quat. Sci.* **16**: 699–709.
- Philibert, A., and Prairie, Y.T. 2002. Diatom-based transfer function for western Quebec lakes (Abitibi and Haute-Mauricie): the possible role of epilimnetic CO<sub>2</sub> concentration in influencing diatom assemblages. *J. Paleolimnol.* **27**: 465–480.
- Philibert, A., Prairie, Y.T., and Carcaillet, C. 2003. 1200 years of fire impact on biogeochemistry as inferred from high-resolution diatom analysis in a kettle lake from the *Picea mariana* – moss domain (Quebec, Canada). *J. Paleolimnol.* **30**: 167–181.
- Pienitz, R., Smol, J.P., and Birks, H.J.B. 1995. Assessment of freshwater diatoms as quantitative indicators of past climatic

change in the Yukon and Northwest Territories, Canada. *J. Paleolimnol.* **13**: 21–49.

Pienitz, R., Smol, J.P., and Lean, D.R.S. 1997. Physical and chemical limnology of 59 lakes located between the southern Yukon and the Tuktoyaktuk Peninsula, Northwest Territories (Canada). *Can. J. Fish. Aquat. Sci.* **54**: 330–346.

Racca, J.M.J., and Prairie, Y.T. 2004. Apparent and real bias in numerical transfer functions in palaeolimnology. *J. Paleolimnol.* **31**: 117–124.

Racca, J.M.J., Philibert, A., Racca, R., and Prairie, Y.T. 2001. A comparison between diatom-based pH inference models using artificial neural networks (ANN), weighted averaging (WA) and weighted averaging partial least squares (WA–PLS) regressions. *J. Paleolimnol.* **26**: 411–422.

Racca, J.M.J., Wild, M., Birks, H.J.B., and Prairie, Y.T. 2003. Separating wheat from chaff: diatom taxon selection using an artificial neural network pruning algorithm. *J. Paleolimnol.* **29**: 123–133.

Reed, R. 1993. Pruning algorithms — a survey. *IEEE (Inst. Electr. Electron Eng.) Trans. Neural Networks*, **4**: 740–747.

Rumelhart, D.E., Hinton, G.E., and Williams, R.J. 1986. Learning representations by back-propagating errors. *Nature (Lond.)*, **323**: 533–536.

ter Braak, C.J.F., and Juggins, S. 1993. Weighted averaging partial least-squares regression (WA–PLS) — an improved method for reconstructing environmental variables from species assemblages. *Hydrobiologia*, **269**: 485–502.

ter Braak, C.J.F., and van Dam, H. 1989. Inferring pH from diatoms — a comparison of old and new calibration methods. *Hydrobiologia*, **178**: 209–223.

ter Braak, C.J.F., Juggins, S., Birks, H.J.B., and van der Voet, H. 1993. Weighted averaging partial least squares regression (WA–PLS): definition and comparison with other methods for species–environment calibration. *In* *Multivariate environmental Hstatistics. Edited by G.P. Patil and C.R. Rao.* Elsevier Science Publishers, Amsterdam, Netherlands. pp. 525–560.

Toivonen, H.T.T., Mannila, H., Korhola, A., and Olander, H. 2001. Applying Bayesian statistics to organism-based environmental reconstruction. *Ecol. Appl.* **11**: 618–630.

Vasko, K., Toivonen, H.T.T., and Korhola, A. 2000. A bayesian multinomial gaussian response model for organism-based environmental reconstruction. *J. Paleolimnol.* **24**: 243–250.

Wolin, J.A., and Duthie, H.C. 1999. Diatoms as indicators of water level change in freshwater lakes. *In* *The diatoms: applications for the environmental and earth sciences. Edited by E.F. Stoermer and J.P. Smol.* Cambridge University Press, Cambridge, UK. pp. 183–204.

Zell, A., Mamier, G., Vogt, M., Mache, N., Hübner, R., Döring, S., Herrmann, K., Soyez, T., Schmalzl, M., Sommer, T., Hatzigeorgiou, A., Posselt, D., Schreiner, T., Kett, B., and Clemente, G. 1996. Stuttgart Neural Network Simulator v4.2. Stuttgart Institute for Parallel and Distributed High Performance Systems. University of Stuttgart, Stuttgart, Berlin [ftp/informatik/uni-stuttgart/de/in pub SNNS SNNSv4.1].

## Appendix A

**Table A1.** Selected diatom species for the water temperature (TEMP), water depth (DEPTH), and dissolved organic carbon (DOC) models.

Species	Occurrence	TEMP	DEPTH	DOC
<i>Fragilaria pinnata</i>	99	Selected	Selected	Selected
<i>Achnanthes minutissima</i> (tribe)	105	Selected	Selected	Selected
<i>Fragilaria brevistriata</i>	81		Selected	Selected
<i>Cyclotella stelligera</i>	73	Selected		Selected
<i>Fragilaria construens</i> var. <i>venter</i>	62	Selected	Selected	Selected
<i>Fragilaria pinnata</i> (coarse form)	51	Selected	Selected	
<i>Navicula minima</i>	70			
<i>Navicula cryptotenella</i>	71	Selected	Selected	Selected
<i>Navicula seminulum</i>	61	Selected	Selected	
<i>Nitzschia fonticola</i>	62	Selected	Selected	Selected
<i>Fragilaria brevistriata</i> var. <i>papillosa/inflata</i>	66	Selected		Selected
<i>Navicula pupula</i>	72			Selected
<i>Amphora pediculus</i>	52			Selected
<i>Achnanthes pusilla</i>	55	Selected		Selected
<i>Cymbella microcephala</i>	52	Selected	Selected	Selected
<i>Cyclotella pseudostelligera</i>	34	Selected		
<i>Fragilaria construens</i>	39		Selected	Selected
<i>Brachysira vitreal</i> <i>Anomoeoneis vitrea</i>	53	Selected		Selected
<i>Fragilaria pinnata</i> var. <i>intercedens</i>	41			Selected
<i>Asterionella formosa</i>	33	Selected	Selected	Selected
<i>Fragilaria construens</i> var. <i>pumila</i>	28	Selected		Selected
<i>Navicula vitiosa</i>	42	Selected	Selected	Selected
<i>Fragilaria capucina</i> var. <i>gracilis</i>	36	Selected	Selected	Selected
<i>Achnanthes subatomoides</i>	48	Selected		
<i>Cyclotella rossii</i>	25		Selected	Selected

Table A1 (continued).

Species	Occurrence	TEMP	DEPTH	DOC
<i>Fragilaria pseudoconstruens</i>	32	Selected	Selected	Selected
<i>Cyclotella tripartita</i>	20	Selected	Selected	Selected
<i>Achnanthes conspicua</i>	29			
<i>Cymbella silesiaca</i>	53	Selected	Selected	
<i>Fragilaria tenera</i>	26	Selected	Selected	
<i>Navicula capitata</i> var. <i>hungarica</i>	33		Selected	Selected
<i>Fragilaria nanana</i>	22	Selected		Selected
<i>Achnanthes suchlandtii</i>	35		Selected	
<i>Amphora libyca</i>	43			Selected
<i>Gomphonema parvulum</i>	39	Selected		Selected
<i>Navicula disjuncta</i>	32			Selected
<i>Cyclotella michiganiana</i>	20	Selected	Selected	Selected
<i>Cyclotella</i> cf. <i>ocellata</i>	18	Selected		
<i>Fragilaria famelica</i>	21	Selected		
<i>Navicula cryptocephala</i>	28	Selected		
<i>Cymbella</i> cf. <i>angustata</i>	19			Selected
<i>Cymbella gracilis</i>	27	Selected	Selected	Selected
<i>Nitzschia perminuta</i>	28		Selected	
<i>Aulacoseira subarctica</i>	13	Selected	Selected	
<i>Cocconeis placentula</i> var. <i>euglypta</i>	24			
<i>Achnanthes lanceolata</i> aff. sp. <i>lanceolata</i>	25		Selected	Selected
<i>Navicula cryptotenella</i> fo. PISCES	16	Selected	Selected	
<i>Navicula radiosa</i>	38	Selected		Selected
<i>Caloneis bacillum</i>	24	Selected	Selected	
<i>Pinnularia interrupta</i> /P. <i>biceps</i>	37	Selected		
<i>Fragilaria virescens</i> var. <i>exigua</i>	15		Selected	
<i>Cocconeis placentula</i>	12			
<i>Navicula menisculus</i>	19	Selected	Selected	Selected
<i>Cyclotella delicatissima</i>	9	Selected	Selected	Selected
<i>Diploneis oculata</i>	34		Selected	
<i>Stauroneis anceps</i>	30			Selected
<i>Stauroneis anceps</i> var. <i>gracilis</i>	17	Selected		
<i>Aulacoseira alpigena</i>	10		Selected	Selected
<i>Cyclotella bodanica</i>	15	Selected	Selected	
<i>Cyclotella bodanica</i> var. <i>lemanica</i>	12	Selected	Selected	Selected
<i>Achnanthes lanceolata</i> ssp. <i>frequentissima</i>	14			Selected
<i>Navicula laevis</i>	34			Selected
<i>Navicula submuralis</i>	16			Selected
<i>Amphora inariensis</i>	22	Selected	Selected	
<i>Nitzschia amphibia</i>	16		Selected	Selected
<i>Nitzschia palea</i>	24	Selected	Selected	Selected
<i>Synedra radians</i>	12	Selected		
<i>Stauroneis phoenicenteron</i>	20			
<i>Navicula absoluta</i>	15		Selected	
<i>Navicula digitulus</i>	17			Selected
<i>Navicula schmassmannii</i>	12		Selected	Selected
<i>Nitzschia acicularis</i>	20			
<i>Aulacoseira distans</i> var. <i>humilis</i>	11		Selected	Selected
<i>Aulacoseira distans</i> var. <i>nivalis</i>	6			Selected
<i>Cyclotella comensis</i>	12			Selected
<i>Stephanodiscus alpinus</i>	8	Selected	Selected	
<i>Tabellaria flocculosa</i> (strain IV)	25			
<i>Diatoma tenue</i> var. <i>elongatum</i>	26	Selected		
<i>Fragilaria capucina</i>	10	Selected	Selected	Selected
<i>Fragilaria lapponica</i>	7		Selected	
<i>Eunotia incisa</i>	13	Selected		Selected
<i>Achnanthes altaica</i>	9			

**Table A1** (continued).

Species	Occurrence	TEMP	DEPTH	DOC
<i>Achnanthes curtissima</i>	15		Selected	
<i>Achnanthes impexiformis/impexa</i>	14			Selected
<i>Achnanthes laterostrata</i>	18	Selected		
<i>Achnanthes marginulata</i>	14	Selected	Selected	Selected
<i>Gyrosigma spenceri</i>	10			
<i>Amphipleura kriegieriana</i>	16			Selected
<i>Navicula disjuncta</i> fo. short (<15 µm)	9			Selected
<i>Navicula elginensis</i>	10		Selected	Selected
<i>Navicula indifferens</i>	7			Selected
<i>Navicula mediocris</i>	11			
<i>Navicula rhynchocephala</i>	18			Selected
<i>Gomphonema gracile</i>	12			
<i>Nitzschia graciliformis</i>	16	Selected		Selected
<i>Nitzschia gracilis</i>	20	Selected	Selected	
<i>Cyclotella atomus</i>	6		Selected	
<i>Cyclotella comensis</i> fo. AK 1	10		Selected	
<i>Stephanodiscus</i> cf. <i>medius</i>	10			Selected
<i>Stephanodiscus hantzschii</i>	6		Selected	Selected
<i>Fragilaria capucina</i> var. <i>rumpens</i>	8			
<i>Fragilaria crotonensis</i>	5		Selected	
<i>Fragilaria pinnata</i> var. <i>lancettula</i>	8		Selected	Selected
<i>Fragilaria robusta</i> /F. <i>pseudoconstruens</i> var. <i>bigibba</i>	13	Selected		
<i>Fragilaria vaucheriae</i>	37	Selected	Selected	Selected
<i>Eunotia intermedia</i> /E. <i>vanheurckii</i>	8	Selected		
<i>Achnanthes chlidanos</i>	15			
<i>Achnanthes gracillima</i>	6	Selected	Selected	Selected
<i>Achnanthes levanderi</i>	17	Selected		Selected
<i>Diploneis oblongella</i>	13			
<i>Neidium ampliatus</i>	19			Selected
<i>Navicula minuscula</i>	16		Selected	
<i>Navicula oblonga</i>	5			
<i>Navicula pseudoscutiformis</i>	37			
<i>Cymbella hebridica</i>	10	Selected	Selected	
<i>Cymbella incerta</i>	13			Selected
<i>Cymbella perpusilla</i>	8			Selected
<i>Amphora thumensis</i>	8			Selected
<i>Gomphonema minutum</i>	4	Selected	Selected	
<i>Nitzschia amphibioides</i>	10		Selected	Selected
<i>Rhoicosphenia abbreviata</i>	7		Selected	Selected
<i>Aulacoseira ambigua</i>	9			
<i>Cyclotella rossii</i> fo. AK 1 (tripartita)	8			
<i>Stephanodiscus parvus</i>	4		Selected	Selected
<i>Tabellaria flocculosa</i> (strain III)	10		Selected	Selected
<i>Fragilaria construens</i> var. <i>binodis</i>	10			
<i>Eunotia bilunaris</i>	17			
<i>Achnanthes clevei</i>	10			
<i>Achnanthes microcephala</i>	9	Selected		Selected
<i>Achnanthes rosenstockii</i>	5		Selected	
<i>Achnanthes scotica</i>	11	Selected		Selected
<i>Achnanthes ventralis</i>	17	Selected	Selected	
<i>Frustulia rhomboides</i> var. <i>crassinervia</i>	7		Selected	
<i>Diploneis marginestriata</i>	8			Selected
<i>Stauroneis smithii</i> var. <i>minima</i>	4			Selected
<i>Brachysira brebissonii</i> /Anomoeoneis <i>brachysira</i>	9	Selected	Selected	
<i>Navicula absoluta</i> fo. 1	12			Selected
<i>Navicula bryophila</i>	15			
<i>Navicula cocconeiformis</i>	23			



**Table A1** (continued).

Species	Occurrence	TEMP	DEPTH	DOC
<i>Navicula halophila</i>	8			Selected
<i>Navicula jaernefeltii</i>	18		Selected	Selected
<i>Navicula modica</i>	10	Selected		Selected
<i>Navicula</i> cf. <i>oppugnata</i>	3			Selected
<i>Navicula vitabunda</i>	20			Selected
<i>Caloneis silicula</i>	19	Selected		Selected
<i>Pinnularia microstauron</i>	19	Selected	Selected	
<i>Cymbella naviculiformis</i>	12			
<i>Amphora fogediana</i>	15		Selected	
<i>Nitzschia dissipata</i>	8			
<i>Nitzschia radicularia</i>	7			Selected
<i>Nitzschia recta</i>	18	Selected		Selected
<i>Epithemia adnata</i>	6		Selected	
cf. <i>Achnanthes ricula</i>	8			
<i>Aulacoseira italica</i>	3			Selected
<i>Aulacoseira valida</i>	5			
<i>Stephanodiscus minutulus</i>	3		Selected	Selected
<i>Tabellaria flocculosa</i> (strain I)	19	Selected	Selected	Selected
<i>Fragilaria cyclopus/Hannaea arcus</i>	12			
<i>Fragilaria ulna</i> /S. <i>ulna</i>	11		Selected	Selected
<i>Fragilaria delicatissima</i>	7			Selected
<i>Eunotia praerupta</i>	13			
<i>Eunotia faba</i>	6			
<i>Eunotia rhynchocephala</i>	6	Selected		Selected
<i>Achnanthes carissima</i>	8			
<i>Achnanthes exigua</i> var. <i>heterovalva</i>	4			Selected
<i>Achnanthes flexella</i> var. <i>alpestris</i>	11			Selected
<i>Achnanthes oestrupii</i>	11			
<i>Achnanthes saccula</i>	9			
<i>Diploneis elliptica</i>	11		Selected	
<i>Stauroneis kriegerii</i>	5	Selected		
<i>Navicula arvensis</i>	6		Selected	Selected
<i>Navicula jaagii</i>	4	Selected		Selected
<i>Navicula lenzii</i>	5		Selected	
<i>Navicula leptostriata</i>	7		Selected	
<i>Navicula pseudanglica</i>	4	Selected		
<i>Navicula pseudoventralis</i>	7	Selected		
<i>Navicula seminuloides</i>	4			
<i>Navicula subhamulata</i>	4			
<i>Navicula subrotundata</i>	11		Selected	Selected
<i>Navicula trivialis</i>	8		Selected	Selected
<i>Caloneis tenuis</i>	3			Selected
<i>Cymbella amphicephala</i>	14	Selected		Selected
<i>Cymbella</i> cf. <i>cesatii</i>	7			
<i>Gomphonema angustatum</i>	7		Selected	Selected
<i>Gomphonema pumilum</i>	6			Selected
<i>Nitzschia liebtruhii</i>	5		Selected	
<i>Nitzschia pura</i>	17			
<i>Nitzschia valdestriata</i>	8			Selected
<i>Simonsenia delognei</i>	5		Selected	Selected
<i>Denticula tenuis</i>	3			
<i>Stenopterobia curvula</i>	3			Selected
cf. <i>Nitzschia bacillum</i>	5		Selected	
<i>Tabellaria fenestrata</i>	7	Selected	Selected	
<i>Diatoma mesodon</i>	4			
<i>Fragilaria capucina</i> var. <i>mesolepta</i>	7		Selected	Selected
<i>Fragilaria leptostauron</i>	6	Selected	Selected	Selected

**Table A1** (continued).

Species	Occurrence	TEMP	DEPTH	DOC
<i>Fragilaria parasitica</i>	27	Selected	Selected	
<i>Fragilaria neoproducta</i>	4	Selected		Selected
<i>Eunotia flexuosa</i>	5			Selected
<i>Eunotia monodon</i>	4			Selected
<i>Eunotia paludosa</i>	3			
<i>Eunotia circumborealis</i>	4			
<i>Cocconeis cf. diminuta</i>	3		Selected	Selected
<i>Cocconeis neothumensis</i>	7			
<i>Cocconeis placentula</i> var. <i>lineata</i>	6			
<i>Achnanthes helvetica</i>	3		Selected	Selected
<i>Achnanthes didyma</i>	14		Selected	Selected
<i>Achnanthes flexella</i>	6		Selected	Selected
<i>Achnanthes lacus-vulcani</i>	8			Selected
<i>Achnanthes lineariz</i>	13			
<i>Achnanthes petersenii</i>	16		Selected	Selected
<i>Achnanthes peragalli</i>	6		Selected	
<i>Achnanthes zieglerei</i>	4			Selected
<i>Gyrosigma acuminatum</i>	3			Selected
<i>Amphipleura pellucida</i>	8		Selected	Selected
<i>Frustulia rhomboides</i>	3			Selected
<i>Frustulia rhomboides</i> var. <i>saxonica</i>	6		Selected	
<i>Diploneis parmalsubovalis</i>	11			
<i>Stauroneis producta</i>	3			
<i>Stauroneis smithii</i>	12			
<i>Brachysira zellensis</i> / <i>Anomoeoneis brachysira</i> var. <i>zellensis</i>	5			
<i>Brachysira minor</i>	5	Selected	Selected	
<i>Navicula bacillum</i>	3			
<i>Navicula difficillimalarvensis</i>	3		Selected	
<i>Navicula explanata</i>	13	Selected	Selected	
<i>Navicula gerloffii</i>	5			
<i>Navicula ignota</i> var. <i>palustris</i>	5	Selected	Selected	Selected
<i>Navicula laevissima</i> var. <i>perhibita</i>	3	Selected		Selected
<i>Navicula libonensis</i>	3		Selected	
<i>Navicula medioconvexa</i>	6	Selected		Selected
<i>Navicula menisculus</i> fo. AK 1	3			Selected
<i>Navicula pseudolanceolata</i>	5			
<i>Navicula similis</i>	5			
<i>Navicula soehrensensis</i> var. <i>hassiacae</i>	9			
<i>Navicula soehrensensis</i>	6	Selected	Selected	
<i>Navicula subtilissima</i>	4	Selected		Selected
<i>Navicula salinarum</i>	4			
<i>Navicula tuscula</i>	4			
<i>Navicula veneta</i>	5	Selected		
<i>Pinnularia balfouriana</i>	10			
<i>Pinnularia maior</i>	7			
<i>Pinnularia nodosa</i>	8			Selected
<i>Pinnularia viridis</i>	9		Selected	
<i>Cymbella cystula</i>	9			Selected
<i>Cymbella cymbiformis</i>	6			Selected
<i>Cymbella falaisensis</i>	5	Selected	Selected	
<i>Cymbella hustedtii</i>	3			Selected
<i>Cymbella incerta</i> var. <i>crassipunctata</i>	5			Selected
<i>Cymbella lapponica</i> fo. short	6			Selected
<i>Cymbella minuta</i>	13	Selected		
<i>Gomphonema acuminatum</i>	8	Selected		Selected
<i>Gomphonema lateripunctatum</i>	3	Selected		Selected
<i>Gomphonema olivaceum</i>	4			

**Table A1** (concluded).

Species	Occurrence	TEMP	DEPTH	DOC
<i>Nitzschia alpina</i>	9	Selected		Selected
<i>Nitzschia rectiformis</i>	8			
<i>Nitzschia solita</i>	3	Selected	Selected	
<i>Nitzschia supralitorea</i>	3			
<i>Epithemia sorex</i>	5		Selected	
cf. <i>Navicula trivialis</i>	4		Selected	