APPENDIX

Table A1. Diatom names and acronyms. The 10 most abundant diatoms are given in boldface.

Diatom	Acronym	Diatom	Acronym
Amphora aequalis	AAEQ	*Gomphonema minutum*	GMIN
Achnanthes sp.	ACH	*Gomphonema olivaceum*	GOLIV
Achnanthidium clevei var.	ACCLB	*Gomphonema parvulum*	GPRV
Achnanthidium clevei	ACCL	*Gomphonema pumilum*	GPUM
Amphora copulata	ACOP	*Gomphonema olivaceoides*	GQDR
Amphora fogediana	AFOG	*Gomphonema sarcophagus*	GSRC
Achnanthes lacunarum	ALAC	*Gomphonema tergestinum*	GTRG
Amphora inariensis	AMIN	*Gyrosigma macedonicum*	GYMAC
Achnanthidium minutissimum	AMSS	*Hannea arcus*	HARC
Amphora ovalis	AOVAL	*Hantzschia amphioxys*	HAYX
Amphora pediculus	**APED**	*Hippodonta rostrata*	HROS
Amphora thumensis	ATHUM	*Luticola mutica*	LMUT
Aulacoseira granulata	AUGR	*Meridion circulare var. constrictum*	MCCC
Amphora veneta	AVEN	*Meridion circulare*	MCRC
Cocconeis disculus	CDIS	*Martyana martyi*	MMRT
Cymatopleura elliptica	CELL	*Melosira varians*	MVAR
Cyclotella juriljii nom. nud.	**CJUR**	*Nitzschia alpina*	NALP
Cyclotella meneghiniana	CMHGN	*Navicula antonii*	NANT
Cocconeis neothumensis	CNTHUM	*Navicula capitatoradiata*	NCPR
Cyclotella ocellata	**COCE**	*Navicula cryptocephala*	NCRPH
Cocconeis placentula	**CPLA**	*Nitzschia dissipata*	NDISS
Cocconeis placentula var. euglypta	CPLE	*Neidium dubium*	NDUB
Cocconeis placentula var. lineata	CPLL	*Navicula gregaria*	NGRG
Caloneis schumaniana	CSCH	*Navicula hasta*	NHAS
Cavinula scutelloides	**CSCU**	*Navicula krsticii*	NKRS
Cymbella affiniformis	CAFF	*Navicula lanceolata*	NLAN
Cymbella lanceolata	CLAN	*Nupela lapidosa*	NLAP
Cymbella neocistula	CYNC	*Nitzschia linearis*	NLIN
Diatoma angusticostata	DANG	*Navicula praetarita*	NPRA
Denticula tenuis	DCNT	***Navicula prespanense***	**NPRE**
Diadesmis gallica var. perpusilla	DGLPS	*Navicula protracta*	NPTR
Diploneis mauleri	**DMAU**	*Nitzschia recta*	NREC
Diatoma mesodon	DMES	*Navicula reinhardtii*	NRERH
Diploneis modica	DMOD	***Navicula rotunda***	**NROT**
Diploneis ovalis	DOVAL	*Navicula subhastatula*	NSHA
Epithemia adnata	EADN	***Navicula subrotundata***	**NSROT**
Encyonema caespitosum	ECAES	*Nitzschia subacicularis*	NSUA
Encyonema minutum	EMIN	*Navicula tripunctata*	NTPT
Encyonopsis microcephala	ENCYM	*Navicula viridulacalcis*	NVCAL
Encyonema silesiacum	ESLS	*Navicula viridula*	NVIR
Epithemia sorex	ESOR	*Orthoseira roseana*	OROS
Fragilaria capucina var.	FCAPV	*Placoneis balcanica*	PBAL
Fragilaria capuccina	FCAPV	*Pinnularia borealis*	PBOR
Fallacia ochridana	FOCH	*Placoneis minor*	PCLM
Fragilaria parasitica	FPAR	*Placoneis elginensis*	PELG
Frustulia vulgaris	FVUL	*Planothidium lanceolatum*	PLLA
Gomphonema clavatum	GCLA	*Planothidium rostratum*	PLLR
Geissleria decussis	GDEC	*Placoneis neoexigua*	PNEO
Gomphonema italicum	GITA	*Pseudostaurosira brevistriata*	PSBR

Table A1 (ctd). Diatom names and acronyms. The 10 most abundant diatoms are given in boldface.

Diatom	Acronym		Diatom	Acronym
Pinnularia subcapitata	PSCP		*Surirella angusta*	SANG
Rhoicosphenia abbreviata	RABB		*Surirella minuta*	SMIN
Rhopalodia gibba	RHGB		*Sellaphora perbacilloides*	SPBA
Reimeria sinuata	RSIN		*Sellaphora pupula*	SPUP
Surirella angusta	SANG		*Stauroneis gracilis*	SRGR
Surirella minuta	SMIN		*Staurosira construens var. binodis*	STCB
Sellaphora perbacilloides	SPBA		*Staurosira construens*	STCO
Sellaphora pupula	SPUP		*Staurosira construens var. venter*	STCV
Placoneis neoexigua	PNEO		*Stauroneis phoenicenteron*	STPHN
Pseudostaurosira brevistriata	PSBR		***Staurosirella pinnata***	**STPNN**
Pinnularia subcapitata	PSCP		*Stauroneis smithii*	STSM
Rhoicosphenia abbreviata	RABB		*Tryblionella angustata*	TANG
Rhopalodia gibba	RHGB		*Tabellaria flocculosa*	TFLOC
Reimeria sinuata	RSIN		*Ulnaria ulna*	UULN

Table A2. Performance (Correlation coefficient and RMSE) of the ensembles of regression trees (Bagging and Random Forest) on training data and estimated with 10-fold cross validation.

	Bagging					Random Forest			
	CC		RMSE			CC		RMSE	
	Train	Xval	Train	Xval		Train	Xval	Train	Xval
APED	0.96	0.28	1.20	2.73		0.96	0.29	1.24	2.71
CJUR	0.94	0.29	3.71	7.11		0.92	0.27	3.95	7.17
COCE	0.97	0.54	7.37	18.05		0.97	0.54	8.08	18.13
CPLA	0.95	0.16	2.33	5.15		0.96	0.35	2.42	4.72
CSCU	0.96	0.29	3.44	8.50		0.96	0.37	3.70	8.16
DMAU	0.96	0.30	1.12	2.56		0.96	0.33	1.18	2.52
NPRE	0.94	0.40	1.24	2.66		0.93	0.30	1.35	2.68
NROT	0.95	0.23	1.61	3.50		0.95	0.29	1.58	3.36
NSROT	0.95	0.09	2.19	4.89		0.95	0.13	2.18	4.69
STPNN	0.93	0.16	1.48	3.11		0.95	0.22	1.48	2.94

Table A3. Performance (Correlation coefficient and RMSE) of the ensembles of multi-target regression trees (Bagging and Random Forest) on training data and estimated with 10-fold cross validation.

	Bagging					Random Forest			
	CC		RMSE			CC		RMSE	
	Train	Xval	Train	Xval		Train	Xval	Train	Xval
APED	0.95	0.30	1.41	2.68		0.95	0.30	1.41	2.68
CJUR	0.95	0.30	3.79	7.10		0.95	0.30	3.79	7.10
COCE	0.96	0.51	8.68	18.52		0.96	0.51	8.68	18.52
CPLA	0.96	0.36	2.18	4.71		0.96	0.36	2.18	4.71
CSCU	0.95	0.36	4.06	8.17		0.95	0.36	4.06	8.17
DMAU	0.96	0.36	1.24	2.48		0.96	0.36	1.24	2.48
NPRE	0.94	0.28	1.42	2.74		0.94	0.28	1.42	2.74
NROT	0.95	0.26	1.73	3.41		0.95	0.26	1.73	3.41
NSROT	0.95	0.18	2.28	4.64		0.95	0.18	2.28	4.64
STPNN	0.95	0.16	1.53	2.99		0.95	0.16	1.53	2.99

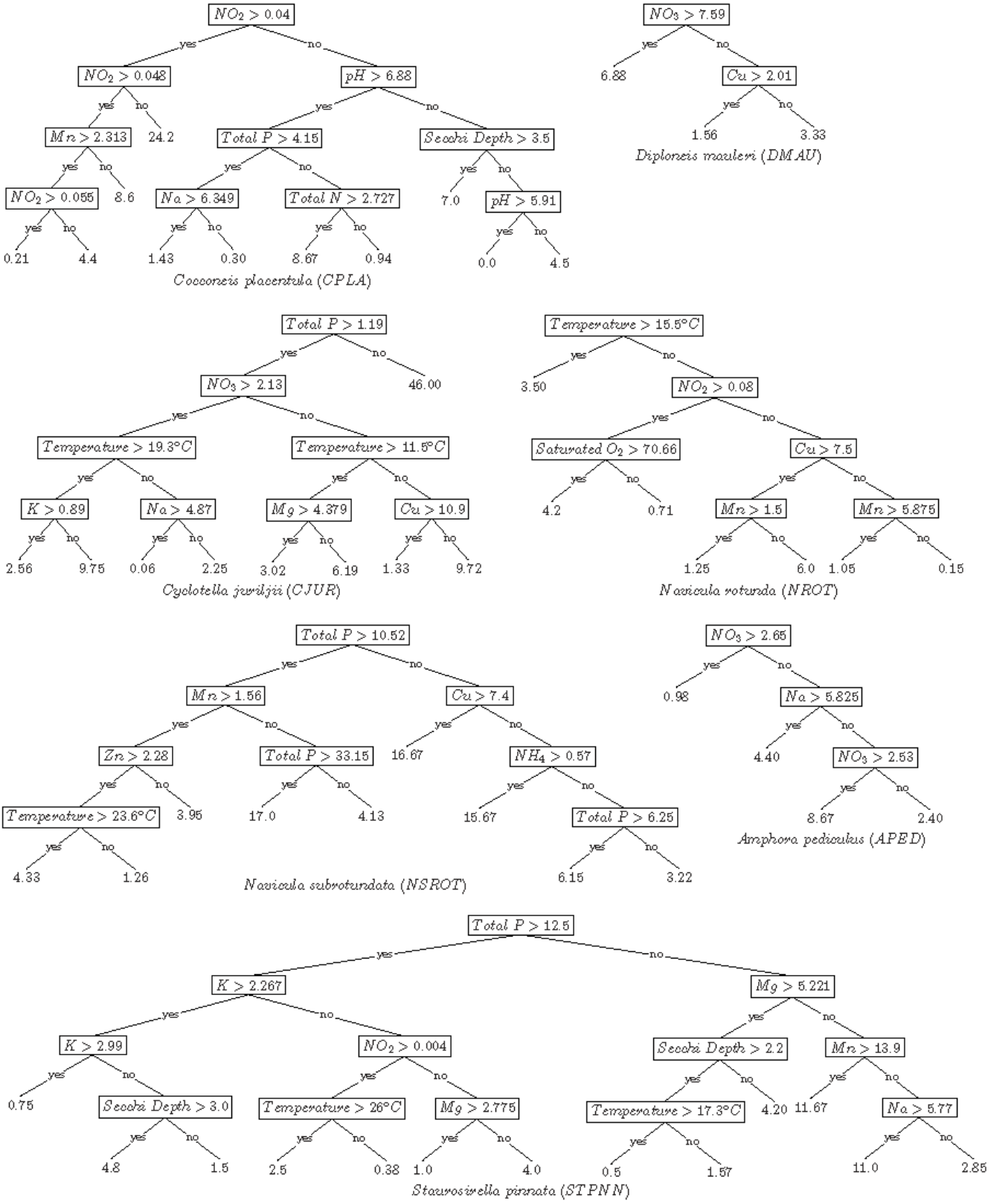

Figure A1. Regression trees for the predicting the abundance of 7 of the top 10 most abundant diatoms (the trees for the remaining diatoms are given in Figures 4,5 and 6)