Reading and processing data for use in the LTER-LIFE digital twins: RWS biogeochemical data

Karline Soetaert

first version: 01-10-2023; current version: 30 January 2024

This document prepares the biogeochemical data from Rijkswaterstaat that are to be used in the LTER life dtWad package.

# Rijkswaterstaat metadata

There are many Rijkswaterstaat monitoring stations. The stations from the waddensea are manually inputted:

stns <- c("huisdnbsd", "marsdnd", "helsdr", "malzn", "doovbwt",   
"vliesm", "westmp", "blauwsot", "harlghvmwt", "harlgvhvn",   
"oostmp", "boschgwt", "borndp", "dantzgnd", "dantzgt", "holwdbg",   
"zoutkplzgt", "zoutkplg", "lauwohvmd", "eildbg", "lauws",   
"zuidolwnot", "zuidolwot", "noordpdzl", "ra" )  
dgN <- c( 52.96, 52.98, 52.96, 52.99,  
 53.05, 53.31, 53.29, 53.22,  
 53.18, 53.17, 53.31, 53.40, 53.44,  
 53.40, 53.40, 53.45, 53.48,  
 53.43, 53.41, 53.47, 53.52, 53.48, 53.45, 53.45, 53.48)  
   
dgE <- c( 4.73, 4.75, 4.75, 4.90, 5.03, 5.16, 5.23, 5.28,  
 5.40, 5.41, 5.41, 5.50, 5.60, 5.72, 5.73, 5.96,  
 6.08, 6.13, 6.20, 6.34, 6.44, 6.45, 6.51, 6.60, 6.66)   
  
RWSstations <- data.frame(station=stns, latitude=dgN, longitude=dgE)  
save(file="../processed\_data/RWSstations.rda", RWSstations)

# Rijkswaterstaat biogeochemical data

Biogeochemical data from stations in (and in the vicinity) of the Wadden sea were downloaded from the RWS website <https://waterinfo.rws.nl/>.

The data are in ASCII format, and in Dutch; they can be read with function *readRWS*.

This converts the data in english format - but it does not work for all variables.

Several stations were only sampled in the 1980s. We select the stations that were also sampled in 2000’s.

For some variables only a few measurements were available; these were removed.

Clorinity was estimated at high resolution, for one station (EEMSHVSMPL) only. This parameter was removed.

files <- c("20231019\_010.csv", "20231019\_035.csv", "20231019\_037.csv")  
  
# Read all data files and merge them into wide format (takes a while)  
RWSbiogeo <- readRWS(dir = "../raw\_data/rws\_biogeo/",   
 file = files,   
 format = "wide") # output format

## Warning in reshapeWide(data, idvar = idvar, timevar = timevar, varying =  
## varying, : multiple rows match for variable=Ntot: first taken

## Warning in reshapeWide(data, idvar = idvar, timevar = timevar, varying =  
## varying, : multiple rows match for variable=Ptot: first taken

## Warning in reshapeWide(data, idvar = idvar, timevar = timevar, varying =  
## varying, : multiple rows match for variable=unknown: first taken

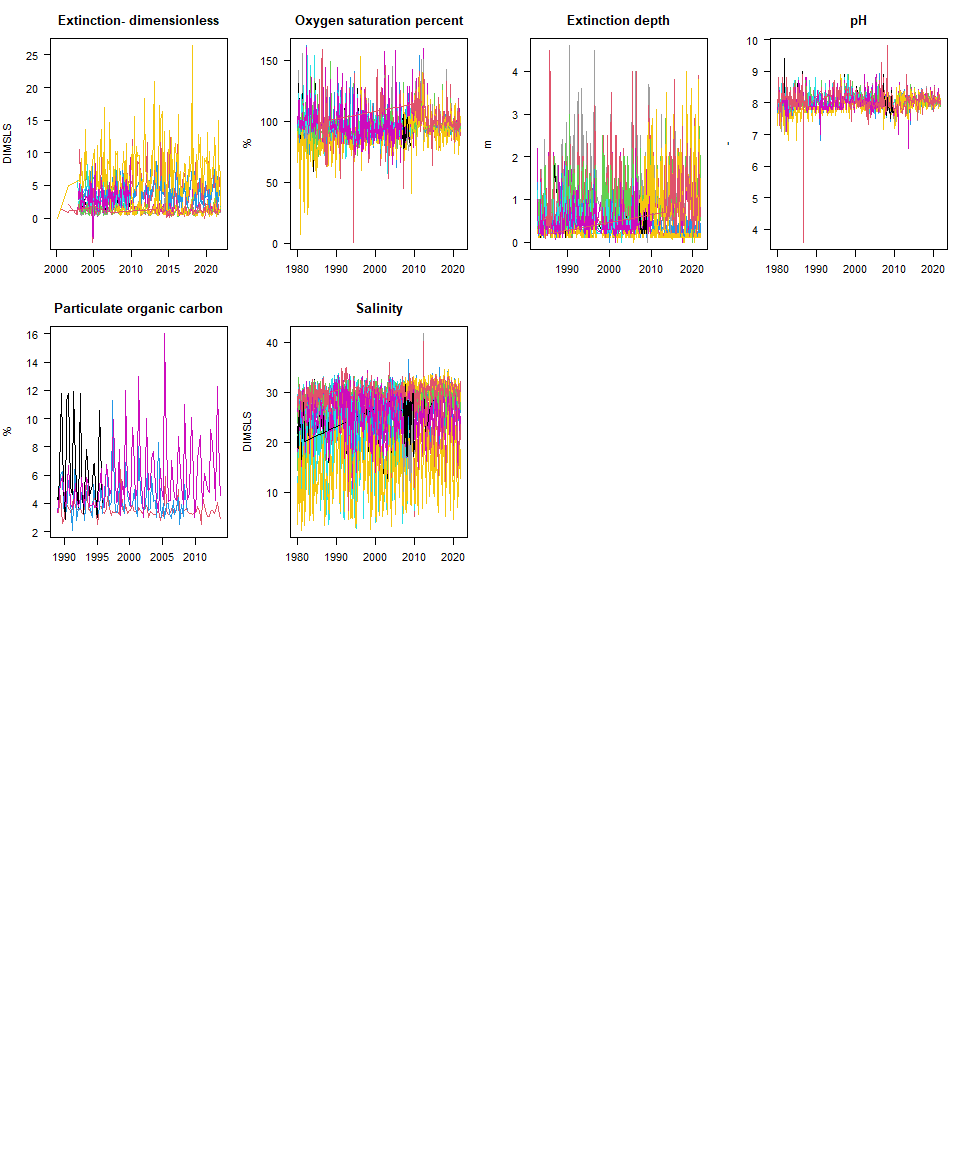
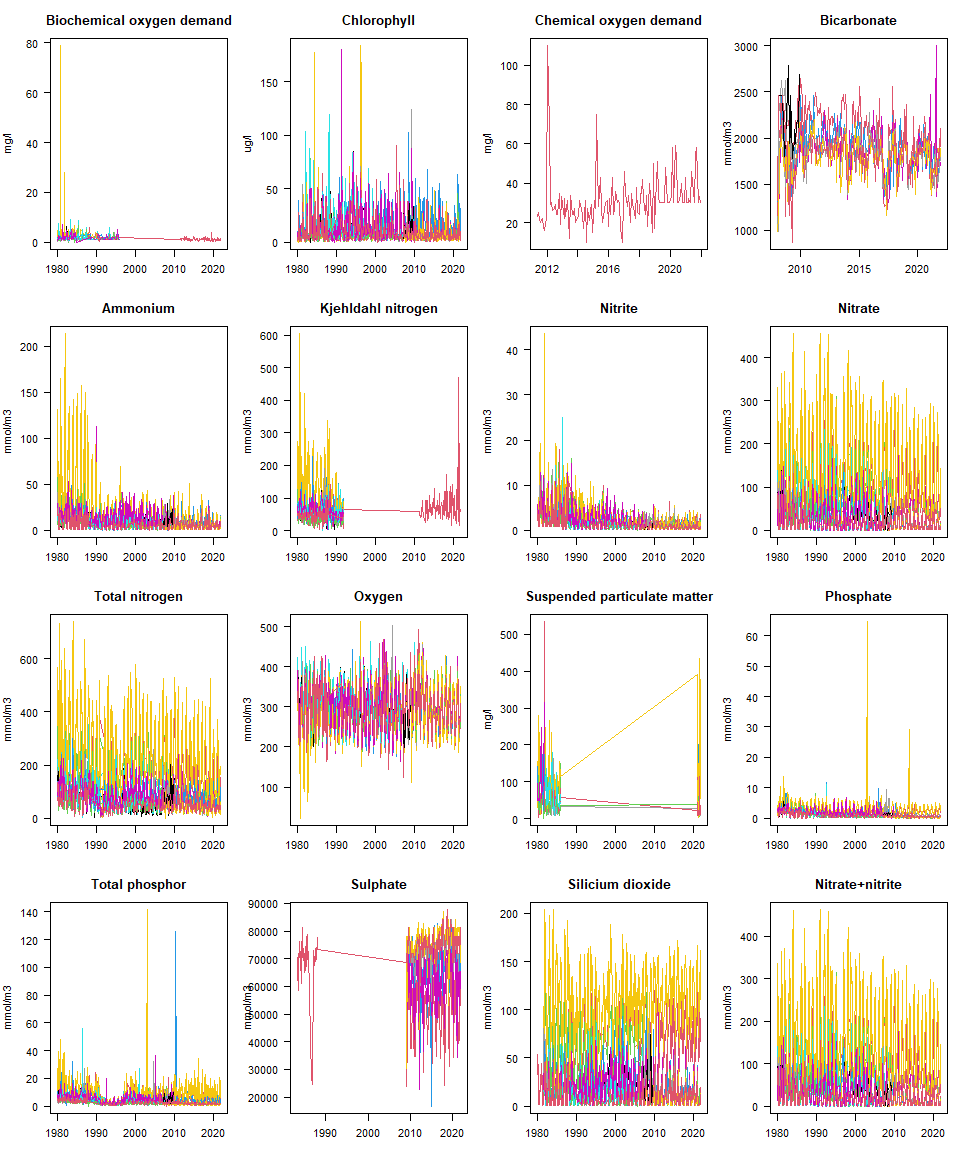
## Warning in reshapeWide(data, idvar = idvar, timevar = timevar, varying =  
## varying, : multiple rows match for variable=Cl: first taken

# remove some columns  
RWSbiogeo$unknown <- RWSbiogeo$PIC <- RWSbiogeo$Ctot <- RWSbiogeo$Cl <- NULL  
  
# select stations that actually have data and were also sampled in 2000  
LS <- tapply(RWSbiogeo$datetime,   
 INDEX = RWSbiogeo$station,   
 FUN = max,   
 na.rm = TRUE)  
  
LastSample <- as.POSIXct(LS, origin="1970-01-01")  
  
NumNutrients <- tapply(RWSbiogeo$NO3,   
 INDEX = RWSbiogeo$station,   
 FUN = function(x) sum(! is.na(x))  
 )  
  
Tokeep <- LastSample[LastSample > "2000-01-01" & NumNutrients > 0]  
  
RWSbiogeo <- subset(RWSbiogeo, subset=station %in% names(Tokeep))  
  
RWSbiogeo$O2[RWSbiogeo$O2 > 1000 ] <- NA  
RWSbiogeo$pH[RWSbiogeo$pH > 100 ] <- NA  
  
save(file="../processed\_data/RWSbiogeo.rda", RWSbiogeo)

From the original data (comprising 64 stations), 18 stations were retained.

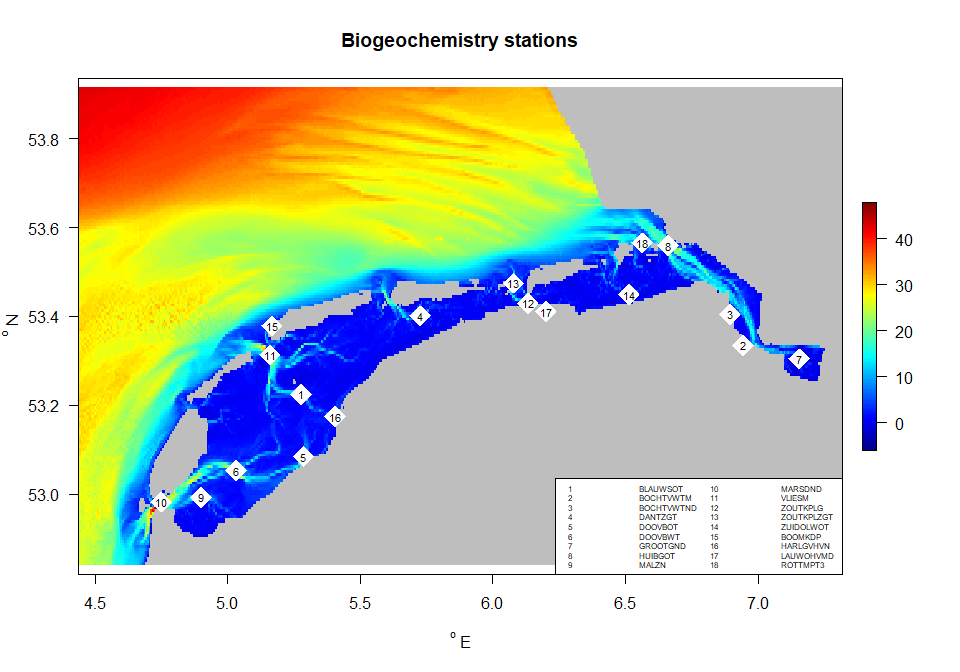
## Plotting all data

par(mar=c(3,4,3,1), las=1)  
plot(RWSbiogeo, type="l", cvar="station", mfrow=c(4,4))



The stations with the data are

stats <- attributes(RWSbiogeo)$stations  
plotBathymetry(WadDepth,   
 pts=stats[,c("longitude", "latitude")],   
 ptlist=list(cex=3, col= "white"), NAcol="grey", type="image",  
 main="Biogeochemistry stations")  
text(stats$longitude, stats$latitude, labels=1:nrow(stats), cex=0.7)  
#plot(1, type="n", axes=FALSE, xlab="", ylab="")  
nr <- nrow(stats)  
legend("bottomright", legend=c(1:(nr/2), stats$station[1:(nr/2)],   
 (nr/2+1):nr, stats$station[(nr/2+1):nr]),   
 ncol=4, cex=0.5)



## Subsetting the data

The data for 2021 are extracted and saved.

WadBiogeo <- subset(RWSbiogeo,   
 subset=datetime >= "2021-01-01" & datetime < "2022-01-01")  
  
# order according to station number  
WadBiogeo <- WadBiogeo[order(WadBiogeo$station),]  
  
# remove some columns  
WadBiogeo$SO4 <- WadBiogeo$POC <- WadBiogeo$KjN <- NULL  
WadBiogeo$BOD <- WadBiogeo$COD <- NULL  
  
save(file="../processed\_data/WadBiogeo.rda", WadBiogeo)

# References

The following R-sources were used for this work:

R-core:

* R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

R-package dtWad,

* Soetaert K (2024). dtWad: Waddensea Digital Twin: general utilities. R package version 0.0.1.