Decisions:

* Site must be treated as a fixed effect. Three factor levels is not enough for a random intercept (needs to be >6-7)
* My approach now will be to (1) fit GLM, (2) check for overdispersion and non-linear patterns in model residuals, (3) investigate all possible causes of overdispersion, (4) if all else fails then switch to a GAM framework. I can also compare GLM to GAM according to AIC provided the dataframes are exactly equal.
* I have eliminated most of the variables with lots of NAs. Therefore I can probably run models with and without those variables (and their NAs). The ‘without’ models will also be without CB? I think this only pertains to fluorescense.
* models will include correlation = corGaus(form = ~ lat+long| site) to account for spatial autocorrelation.
* Keep an eye on a potential Dbin~Temp interaction…data looks a bit sparse across those variables in two dimensions.
* Remember to omit the rows with large discrepancies between Seasoar versus CTD temp/sal (as per Amandine’s suggestion)

In summary, these are the variables I am/arn’t taking forward in the analyses:

Y variables

* Zooplankton biomass
* Zooplankton abundance
* NBSS slope
* NBSS intercept
* Mean ESD

X variables

* Temperature - no na
* Fluorescence
* Bathymetry - no na
* Sample depth bin (categorical), no na
* Transect site (categorical), no na

Omitted X variables

* Salinity
* Nitrate
* Phosphate
* Silicate
* Oxygen
* Distance from coast
* Seafloor slope (categorical)
* Vbin (categorical)

For details see html file