```r
########## R-Script explaining manual forward selection ################################

# The manual forward selection has the objective to identify environmental parameters where each
# explains a separate dimension of variance in the taxa data (i.e. that show no collinearity between
# each other). By following this procedure, we will obtain a set of independent forcing variables that
# drive the taxa assemblage.

############################################################################################


# Ranking the 6 parameters by eigenvalue (the amount of taxa variance they explain):

rda(TAXA ~ PARAMETER.X, data=ENVI)

# Result: SSTsummer (12.7%), SeaIce (9.7%), SSSsummer (6.9%), PPsummer (4.6%), PPannual
# (4.2%), PPspring (2.5%)


# Evaluating the independence between all 6 parameters:
# To this end we calculated Variance Inflation Factors (VIF), expressing how much of the taxa
# variance explained by one environmental variable is already explained by another parameter. The
# VIF of a variable is calculated from the multiple correlations (r) among the environmental variables
# (ter Braak and Smilauer 2002), using the equation VIF=1/(1−r2). We chose a cut-off value of VIF ≤ 2
# for all parameters as also suggested in other studies (e.g., Lopes et al. 2010). Such a VIF value only
# allows collinearities of r2 ≤ 0.5 and thus not more than half of the variance in the taxa data explained
# by one variable to also be explained by another variable.


# Now we start with the actual forward selection:
# After ranking the 6 variables by their eigenvalue we start with the first one and add step-by-step
# another variable and check if the VIF's < 2.

n.rda <- rda(TAXA ~ SSTsummer+SeaIce, data=ENVI)
vif.cca(n.rda)

# VIF < 2? YES
# Result:        SSTsummer     SeaIce
#                1.78          1.78
# Taxa variance explained: 18.7%


# Proceede with the above model and add the next parameter:

n.rda <- rda(TAXA ~ SSTsummer+SeaIce+SSSsummer, data=ENVI)
vif.cca(n.rda)

# VIF < 2? NO
# Result:        SSTsummer     SeaIce        SSSsummer
#                2.13          2.60          1.46
# We find that two parameters show too much collinearity. To decide which parameters to keep and
# which to exclude from the set of independent forcing variables, we check which model explains more
# taxa variance and how large the VIFs are:

n.rda<-rda(TAXA ~ SSTsummer+SeaIce, data=ENVI)
# Taxa variance explained: 18.7%    VIFs: SSTsummer 1.78  SeaIce 1.78

n.rda<-rda(TAXA ~ SSSsummer+SeaIce, data=ENVI)
# Taxa variance explained: 17.5%    VIFs: SSSsummer 1.23  SeaIce 1.23

n.rda<-rda(TAXA ~ SSSsummer+SSTsummer, data=ENVI)
# Taxa variance explained: 19.6% VIFs: SSTsummer 1.00  SSSsummer 1.00

# The model including SSTsummer and SSSsummer (excluding SeaIce) is the better model explaining
# the largest amount of taxa variance and showing the highest amount of independence.
```

```
# Note:
# We exclude SeaIce from our dataset because it does not explain a separate dimension of variance
# in taxa assemblages within the regional Baffin Bay dataset. This does not mean, that it cannot be
# reconstructed or does not explain taxa variance! However, our analysis suggests that SSTsummer
# and SeaIce "play the same part" in driving assemblage compositions, meaning that by
# reconstructing one we actually reconstruct a "mixture" of SSTsummer and SeaIce. This is
# strengthened by the result of the MFA where we detected SST on the one end of the first dimension
# and SeaIce on the other end. However, to calibrate an independent dataset, we need to decide
# between SSTsummer and SeaIce. Statistics favour SSTsummer. But, we have to bear in mind when
# applying the independent calibration dataset that the SSTsummer parameter is a place holder for
# several temperature parameters like SSTsummer, SeaIce and presumably others.


# Proceede with the best performing model and add the next variable

n.rda <- rda(TAXA ~ SSTsummer+SSSsummer+Ppsummer, data=ENVI)
vif.cca(n.rda)

# VIF < 2? NO
# Result:          SSTsummer     SSSsummer     Ppsummer
#                  1.87          1.22          2.08
# We follow the same procedure as above: Which model explains more variance and how large are
# the VIFs?

n.rda <- rda(TAXA ~ SSSsummer+Ppsummer, data=ENVI)
# Taxa variance explained: 10.2%    VIFs: SSSsummer 1.11  Ppsummer 1.11

n.rda <- rda(TAXA ~ SSTsummer+SSSsummer, data=ENVI)
# Taxa variance explained: 19.6%   VIFs: SSTsummer 1.00 SSSsummer 1.00

n.rda <- rda(TAXA ~ SSTsummer+Ppsummer, data=ENVI)
# Taxa variance explained: 17.6%    VIFs: SSTsummer 1.71 Ppsummer 1.71


# Proceede with the best performing model and add the next variable

n.rda <- rda(TAXA ~ SSTsummer+SSSsummer+Ppannual, data=ENVI)
vif.cca(n.rda)

# VIF < 2? NO
# Result:          SSTsummer     SSSsummer     Ppannual
#                  2.11          1.09          2.19
# We follow the same procedure as above: Which model explains more variance and how large are
# the VIFs?

n.rda <- rda(TAXA ~ SSSsummer+Ppannual, data=ENVI)
# Taxa variance explained: 10.4%    VIFs: SSSsummer 1.04  Ppannual 1.04

n.rda <- rda(TAXA ~ SSTsummer+SSSsummer, data=ENVI)
# Taxa variance explained: 19.6%   VIFs: SSTsummer 1.00  SSSsummer 1.00

n.rda <- rda(TAXA ~ SSTsummer+Ppannual, data=ENVI)
# Taxa variance explained: 17.7%    VIFs: SSTsummer 2.01  Ppsummer 2.01


# Proceede with the best performing model and add the next (last) variable

n.rda <- rda(TAXA ~ SSTsummer+SSSsummer+Ppspring, data=ENVI)
vif.cca(n.rda)

# VIF < 2? YES
# Result:          SSTsummer     SSSsummer     Ppspring
#                  1.81          1.00          1.81
# Taxa variance explained: 23.6%
```

# Final results:

# -	The final model is SSTsummer+SSSsummer+PPspring

# -	Each of the three parameters explain an independent part of the variance in the assemblage
#	composition, while together explaining 23.6% of the total taxa variance

# -	SeaIce is excluded as it presumably explains the same part of the variance as SSTsummer
# -	SSTsummer reconstructions will reflect SeaIce reconstructions (if we had decided to keep
#	SeaIce, SeaIce reconstructions would reflect SSTsummer changes)

# -	PPannual and PPspring are excluded as they presumably explain the same part of the
#	variance as PPspring does

# -	In the local calibration dataset we find three "signals" in the taxa variance regarding the 6
#	parameters we took into account.

# References:

# ter Braak, C.J.F., Smilauer, P., 2002. CANOCO Reference Manual and User's Guide to Canoco for
# Windows: Software for Canonical Community Ordination (Version 4.5), Microcomputer Power.
# Microcomputer Power, Ithaca, New York, U.S.A.

# Lopes, C., Mix, A.C., Abrantes, F., 2010. Environmental controls of diatom species in northeast
# Pacific sediments. Palaeogeogr. Palaeoclimatol. Palaeoecol. 297, 188–200.
# https://doi.org/10.1016/j.palaeo.2010.07.029

##############################################################################