FRAM-Clean Pasvik case

Rasmus E. Benestad, The Norwegian Meteorolgical Institute

2023-04-04

Introduction

This analysis aims at assessing links between climate and fresh water fish species in the Pasvik river in the far northeast of Norway. The analysis is carried out based on information from Amundsen et al. (2021) on expected links between fish, their habitat and climate: "*utvasking og avrenning fra det store nedslagsfeltet ved at de pågående klimaendringene også omfatter økte nedbørsmengder*", "*klar korrelasjon mellom den økte abbortetthet og økende vanntemperaturer i vassdraget*" (in Norwegian). Early growth in perch ("abbor") seems to be correlated with the preceding summer temperature, hence the summer temperature may be most critical. Such links can be utilised in climate change projections by applying empirical-statistical downscaling methods, described in Benestad (2021), directly to the population estimates directly using the large-scale summer mean temperature in the region as predictor. We use the inter-annual variations for calibrating the models.

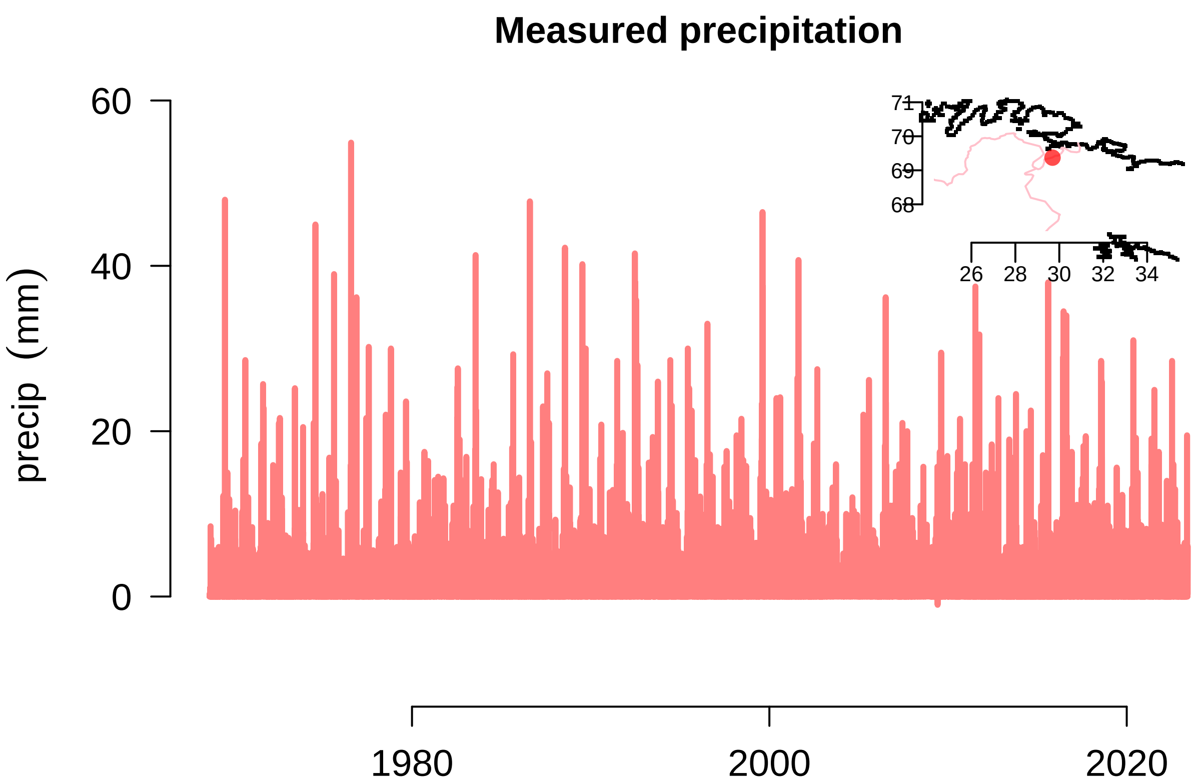
climate data

Fetch locally measured climate data near Pasvik and present it to get a visual impression about its quality.

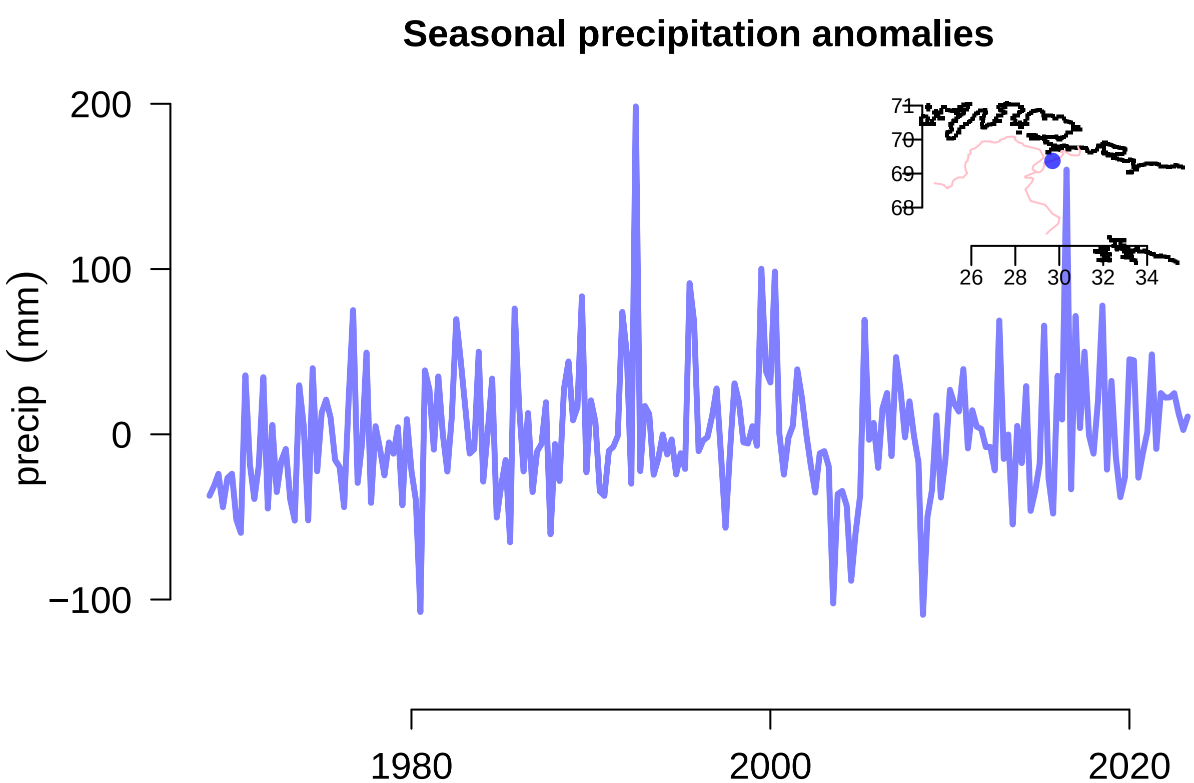
```
library(esd)
```

```
## Loading required package: ncdf4

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## Registered S3 methods overwritten by 'esd':
##   method        from
##   subset.matrix base
##   subset.zoo    zoo
```

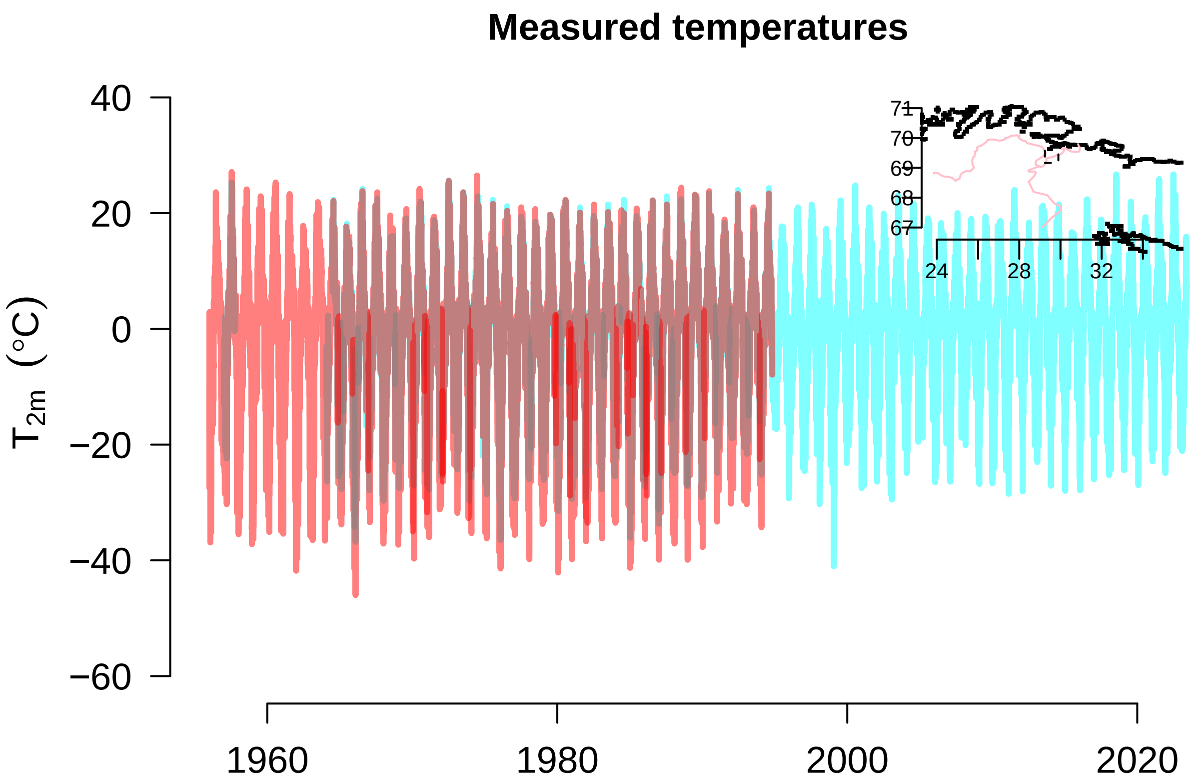
```
## Data from rain gauge at Skogfoss
precip <- station(param='precip',src='metnod.thredds',stid=99500)
## Plot daily precipitation
plot(precip,new=FALSE,main='Measured precipitation')
```



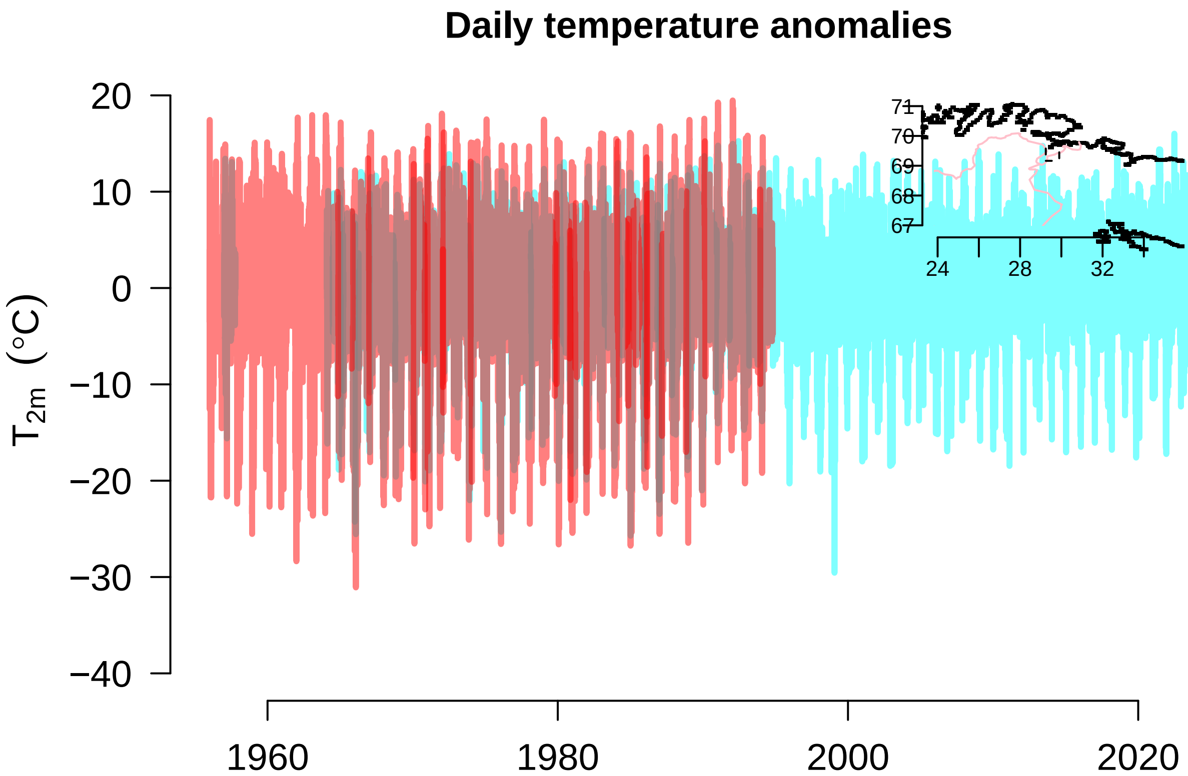

```r
## Plot seasonal precipitation anomalies
plot(anomaly(as.4seasons(precip,FUN='sum')),new=FALSE,main='Seasonal precipitation anomalies')
```



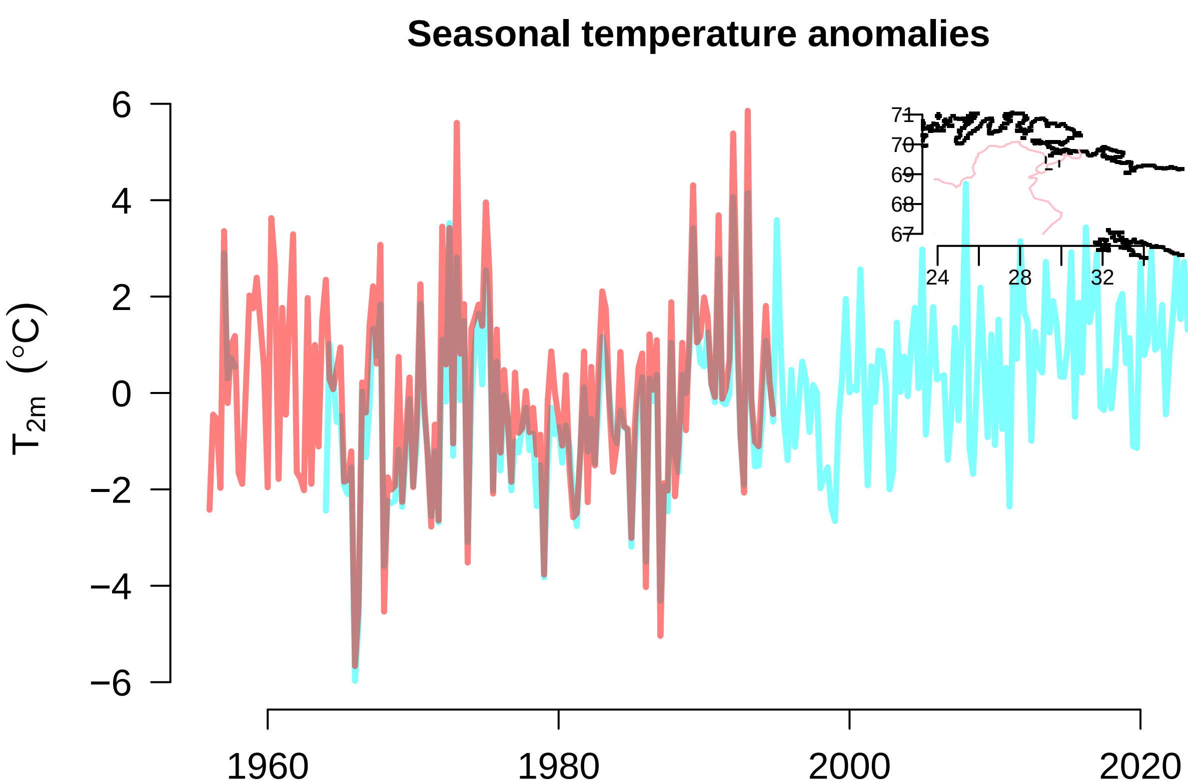
```r
## Data from thermometer at Pasvik
t2m <- station(param='t2m',src='metnod.thredds',stid=c(99530,99370))
## Plot the original data:
plot(t2m,new=FALSE,main='Measured temperatures')
```




```r
## plot daily anomalies:
plot(anomaly(t2m),new=FALSE,main='Daily temperature anomalies')
```




```
## plot seasonally aggregated anomalies:
plot(as.4seasons(anomaly(t2m)),new=FALSE,main='Seasonal temperature anomalies')
```



```r
## Estimate annual totals and means
PRECIP <- 365.25*annual(precip,nmin=300)
T2M <- subset(as.4seasons(t2m,nmin=70),it='jja'); index(T2M) <- year(T2M)
```

When we want to use results from global climate models (GCMs) to provide future climatic outlooks, we need to acknowledge the fact that they only are designed to reproduce large-scale meteorological phanomena, and that there is a need to downscale relevant information from such model simulations. Downscaling in general is discussed in https://shorturl.at/fpFGZ. The large-scale information used as inputs in downscaling are referred to as *predictors*, and to calibrate methods used in empirical-statistical downscaling (ESD), we often use global gridded data. The best choice is often datap product known as *reanalyses*, which consist of a synthesis between observations and weather forecasts.

The ERA5 reanalysis

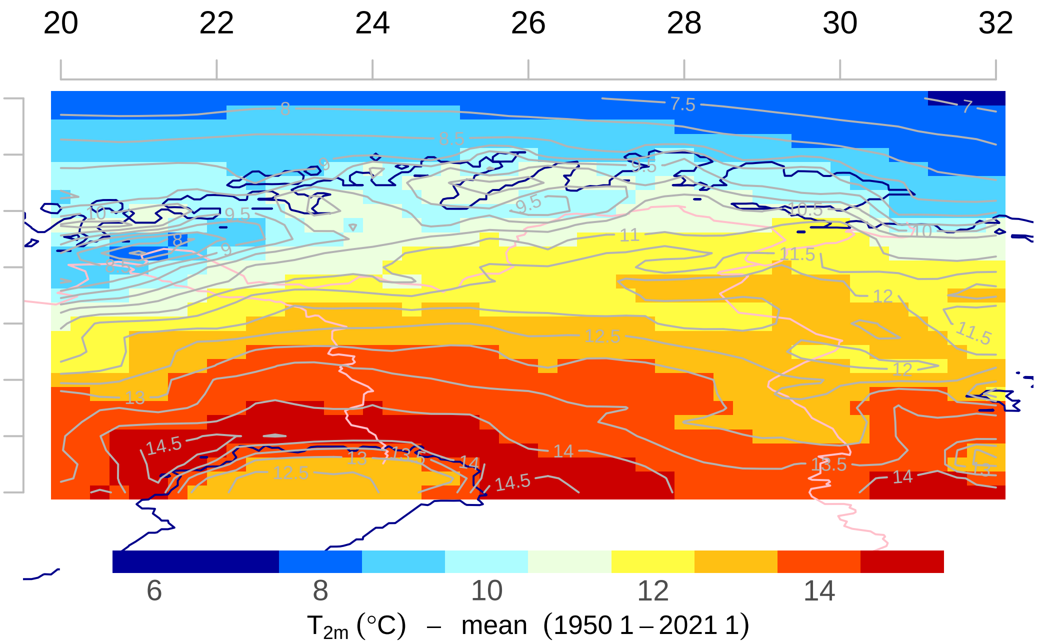
For predictor, we use summer temperature taken from the ERA5 reanalysis which is available from Copernicus Climate Climate Services ("C3S") Data Store ("CDS"): https://cds.climate.copernicus.eu/#!/home (https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels-monthly-means?tab=overview).

```r
## Read data that have been downloaded to local computer from the URL sated above.
file.remove('ERA5.fram-clean.pasvik.rda')
```

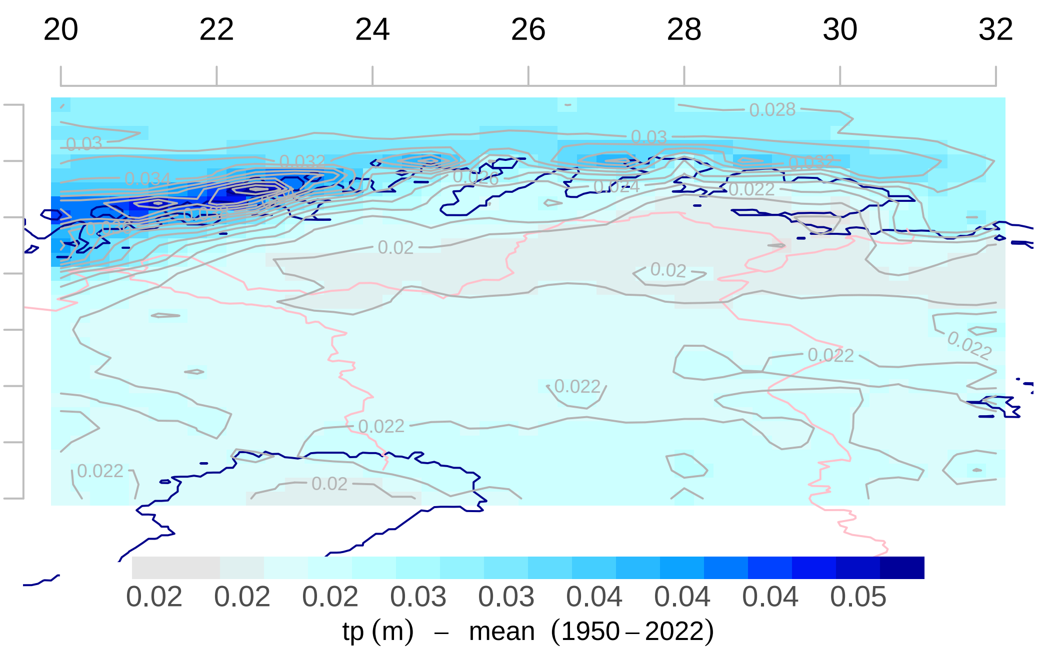
```
## Warning in file.remove("ERA5.fram-clean.pasvik.rda"): cannot remove file
## 'ERA5.fram-clean.pasvik.rda', reason 'No such file or directory'
```

```
## [1] FALSE
```

```r
if (!file.exists('ERA5.fram-clean.pasvik.rda')) {
  ## 2m surface air temperature
  era5.t2m <- retrieve('~/data/ERA5/ERA5_t2m_mon.nc',lon=c(20,32),lat=c(65,72))
  ## Total precipitation
  era5.tp <- retrieve('~/data/ERA5/ERA5_tp_mon.nc',lon=c(20,32),lat=c(65,72))
  ## Save a temporary buffer file for efficient re-run
  save(era5.t2m,era5.tp,file='ERA5.fram-clean.pasvik.rda')
} else load('ERA5.fram-clean.pasvik.rda')
## Extract the summer mean temperatures
era5.t2m <- subset(as.4seasons(era5.t2m),it='jja')
index(era5.t2m) <- year(era5.t2m)
map(era5.t2m,new=FALSE)
```



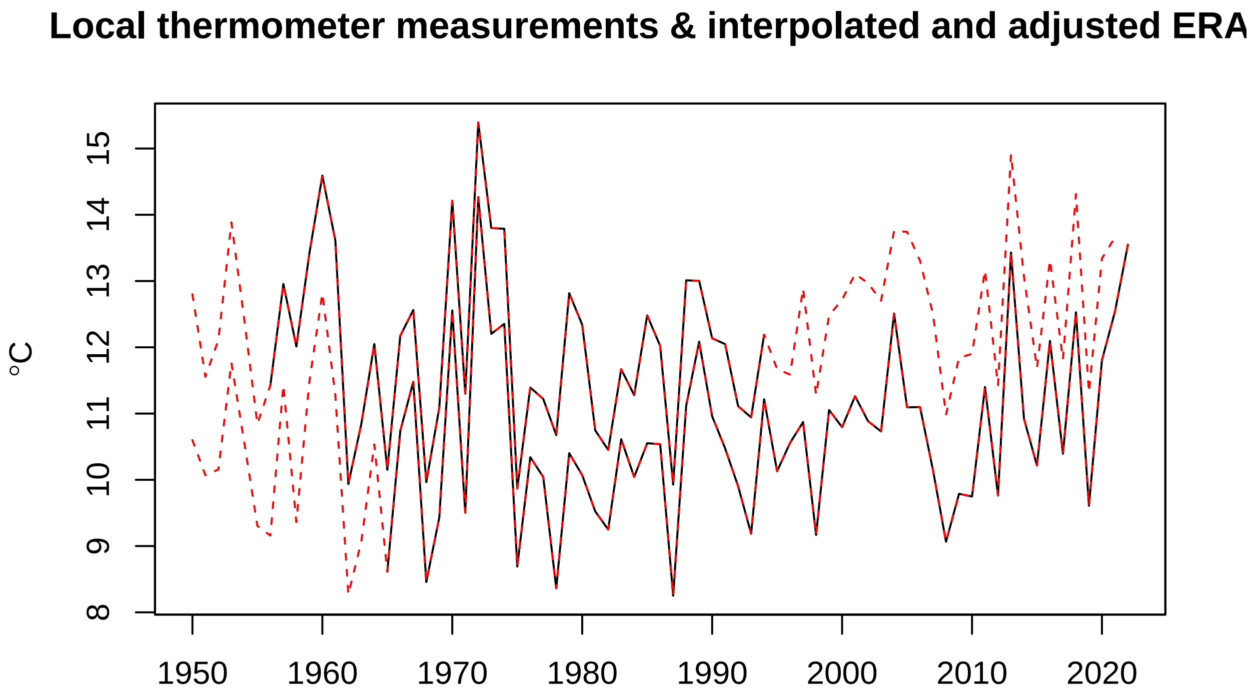

```r
era5.tp <- annual(era5.tp,FUN='sum')
map(era5.tp,new=FALSE)
```



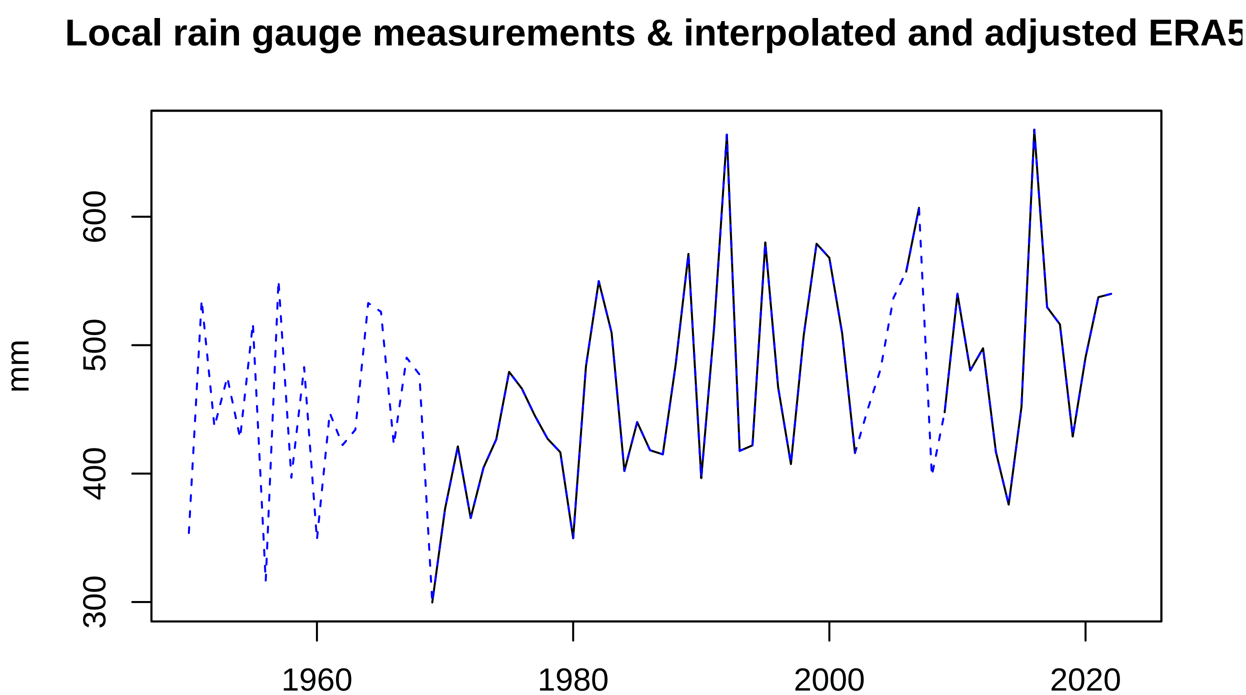
Evaluation of the ERA5 reanalysis

To evaluate the ERA5 reanalysis, we compare it with locally measured temperature and precipitation. Here we only look at seasonally or annually aggregated statistics.

```r
x.T2M <- reafill(T2M,era5.t2m)
plot(merge(zoo(T2M),zoo(x.T2M)),plot.type='single',lty=c(1,1,2,2),col=c('black','black','red','red'),
     main='Local thermometer measurements & interpolated and adjusted ERA5',ylab=expression(degree*C),xl
```




```
x.PRECIP <- reafill(PRECIP,era5.tp)
plot(merge(zoo(PRECIP),zoo(x.PRECIP)),plot.type='single',lty=c(1,2),col=c('black','blue'),
     main='Local rain gauge measurements & interpolated and adjusted ERA5',ylab='mm',xlab='')
```



The comparison reveals a close match between the aggregated locally measured temperature and precipitation and those interpolated from the ERA5 reanalysis. Hence, it doesn't matter much whether we use local station data or ERA5. One advantage with the ERA5 reanalysis is that it doesn't have gaps of missing data.

Import fish statistics from Pasvik

The statistics on fish in Pasvik are described in Excel files provided by Per-Arne Amundsen:

```
library(readxl)
Vaggatem <- read_excel("~/Downloads/Pasvik CPUE Vaggatem&Skrukkebukta AbborLagesildPlanktonsikBunnsik.xl
                       sheet = "CPUE Pasvik 1991-2020")
```

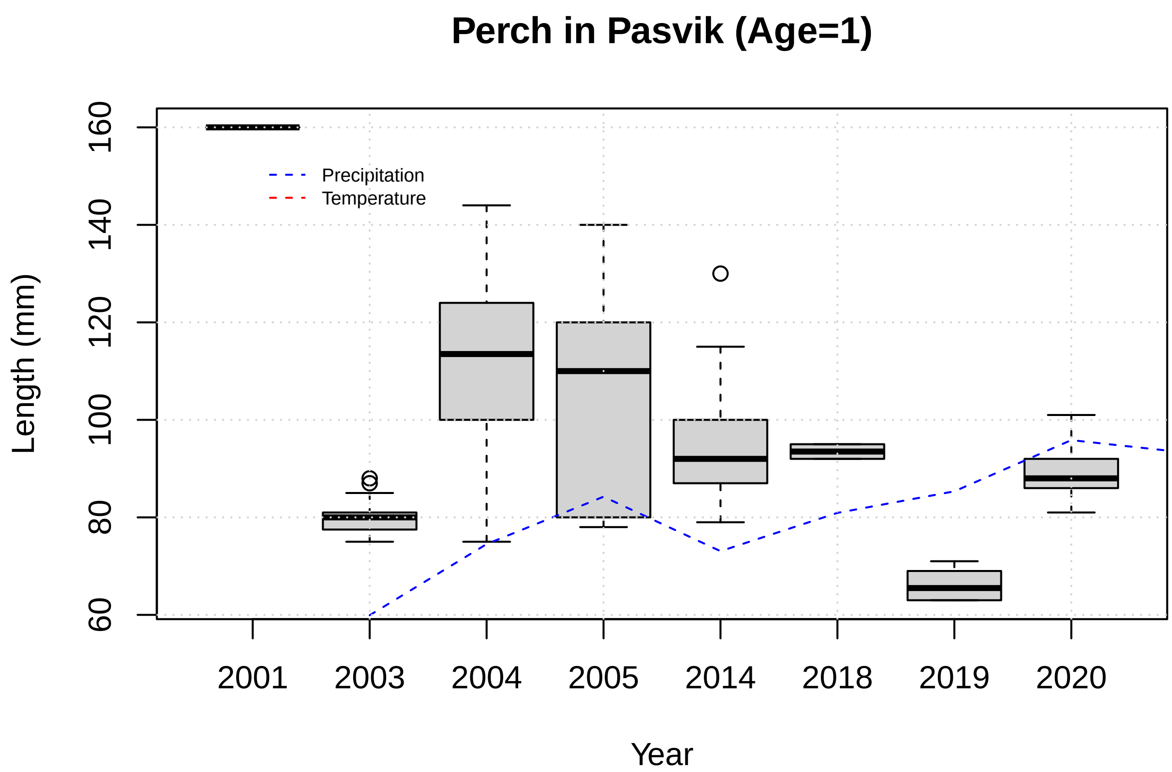
```
## New names:
## * `` -> `...1`
## * `` -> `...2`
## * `` -> `...4`
## * `` -> `...5`
## * `` -> `...6`
## * `` -> `...7`
## * `` -> `...8`
## * `` -> `...9`
## * `` -> `...10`
## * `` -> `...11`
## * `` -> `...12`
## * `` -> `...13`
## * `` -> `...14`
## * `` -> `...15`
## * `` -> `...17`
## * `` -> `...18`
## * `` -> `...19`
## * `` -> `...20`
## * `` -> `...21`
## * `` -> `...22`
## * `` -> `...23`
## * `` -> `...24`
## * `` -> `...25`
## * `` -> `...26`
## * `` -> `...27`
```

```
Pasvik <- read_excel("~/Downloads/Pasvik1991-2020 AbborLagesild&DR&LSRSik med aldersdata.xlsx",
                     sheet = "Pasvik1991-2020 Data")
```

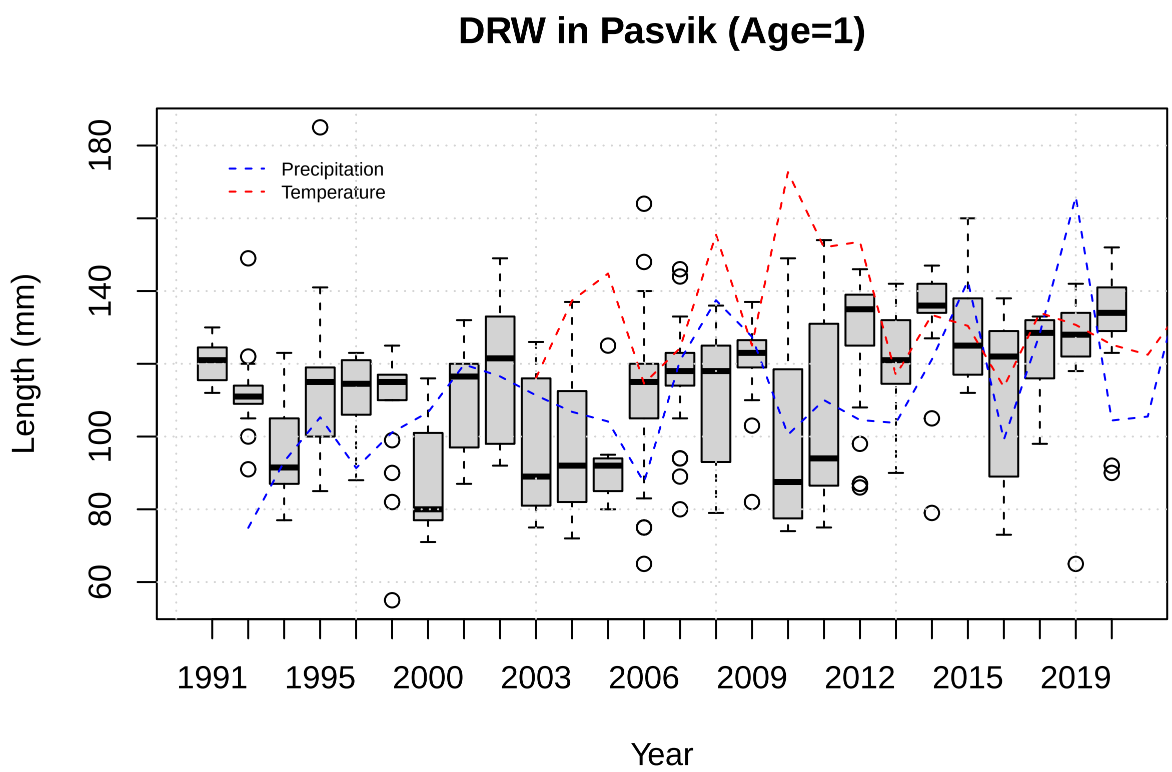
Time series check

The following plots give an impression of the nature of selected local fish data with which to work.

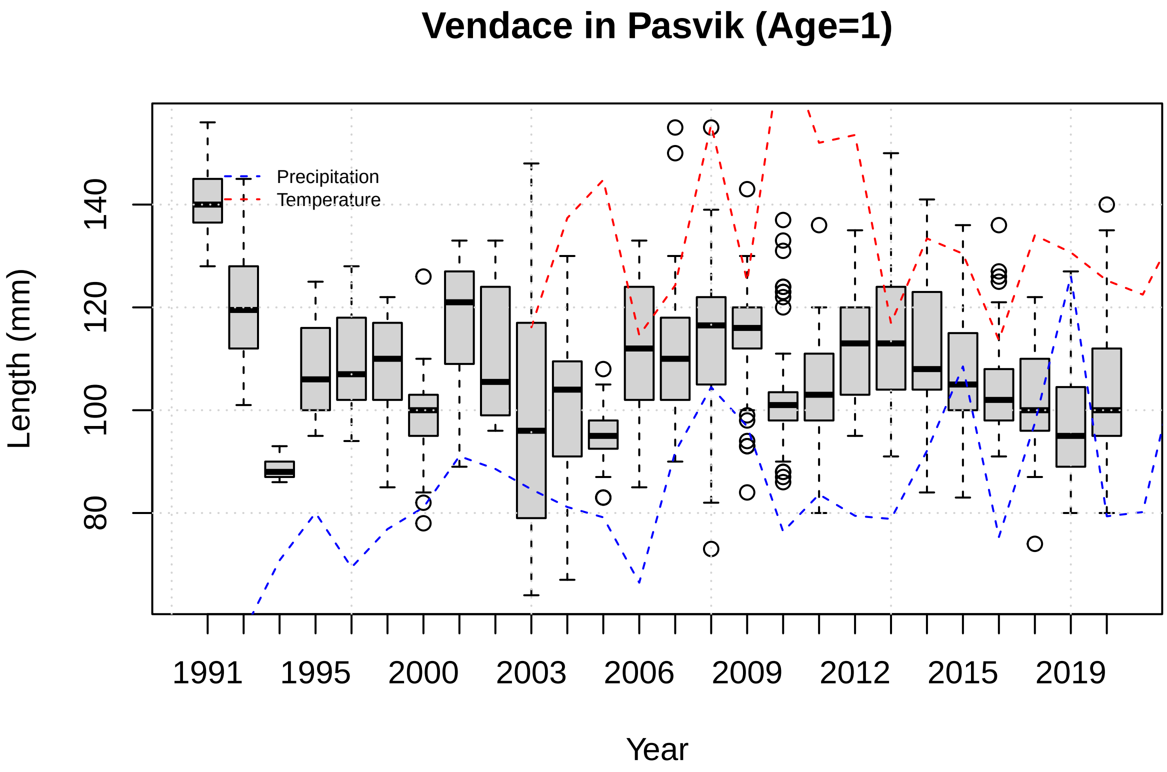
```
ip <- 1:length(index(PRECIP))
it <- 1:length(index(T2M))
Perch <- subset(Pasvik, Species == 'Perch' & Age == 1)
boxplot(`Length (mm)` ~ Year, data=Perch,main='Perch in Pasvik (Age=1)')
lines(ip,0.2*coredata(PRECIP),col='blue',lty=2)
lines(it,20+10*coredata(T2M[,1]),col='red',lty=2)
legend(1,155,c('Precipitation','Temperature'),col=c('blue','red'),lty=2,bty='n',cex=0.6)
grid()
```




```
DRW <- subset(Pasvik,Species == 'DRW' & Age == 1)
boxplot(`Length (mm)` ~ Year, data=DRW,main='DRW in Pasvik (Age=1)')
lines(ip,0.25*coredata(PRECIP),col='blue',lty=2)
lines(it,30+10*coredata(T2M[,1]),col='red',lty=2)
legend(1,180,c('Precipitation','Temperature'),col=c('blue','red'),lty=2,bty='n',cex=0.6)
grid()
```




```
Vendace <- subset(Pasvik,Species == 'Vendace' & Age == 1)
boxplot(`Length (mm)` ~ Year, data=Vendace,main='Vendace in Pasvik (Age=1)')
lines(ip,0.19*coredata(PRECIP),col='blue',lty=2)
lines(it,30+10*coredata(T2M[,1]),col='red',lty=2)
legend(1,150,c('Precipitation','Temperature'),col=c('blue','red'),lty=2,bty='n',cex=0.6)
grid()
```



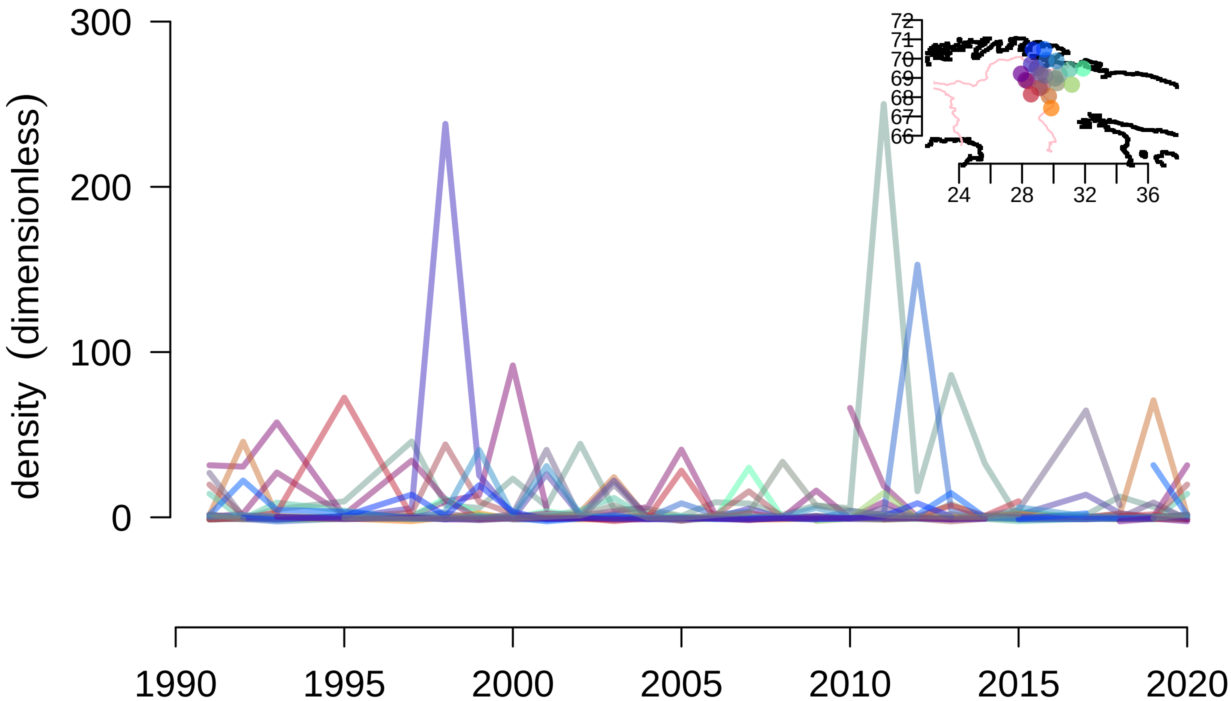
Fish statistics

The following lines extract statistics of the fish samples.

```
x <- matrix(round(as.numeric(as.matrix(Vaggatem[3:29,3:27])),2),26,24)
```

```
## Warning in matrix(round(as.numeric(as.matrix(Vaggatem[3:29, 3:27])), 2), : NAs
## introduced by coercion

## Warning in matrix(round(as.numeric(as.matrix(Vaggatem[3:29, 3:27])), 2), : data
## length [675] is not a sub-multiple or multiple of the number of rows [26]
```

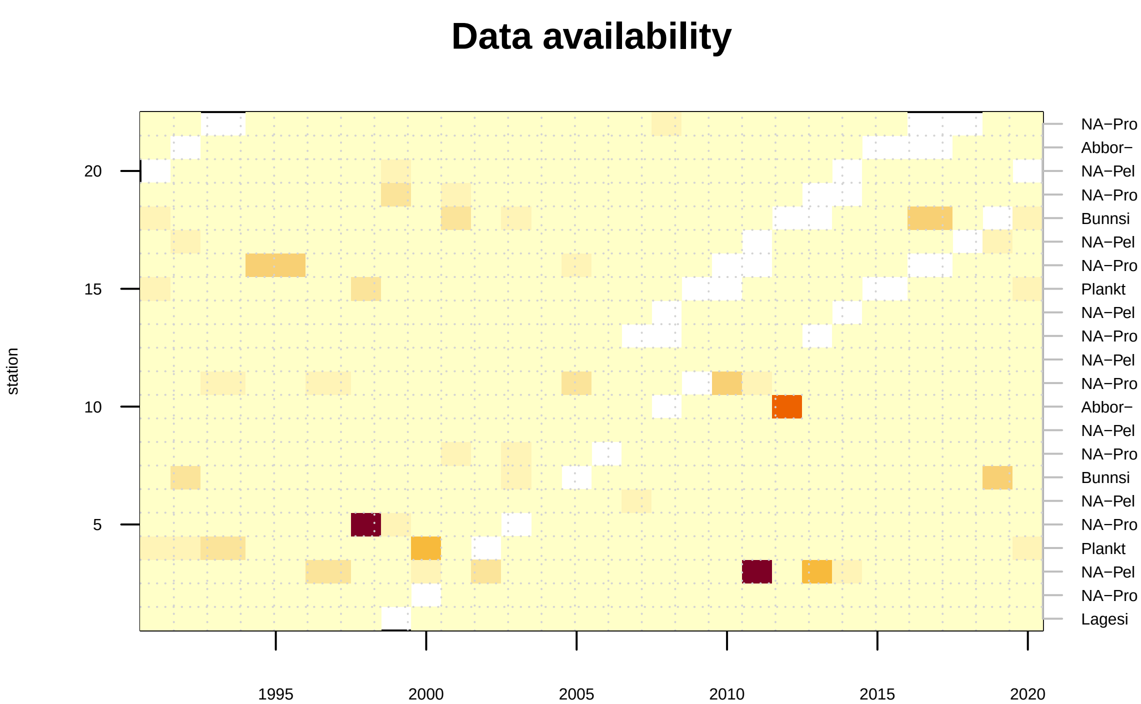
```
cm <- apply(x,2,'mean',na.rm=TRUE)
cs <- apply(x,2,'sd',na.rm=TRUE)
## Standardise the series so that one specie doesn't dominate the analysis:
x <- (x - cm)/cs
t <- as.numeric(as.character(Vaggatem$...1[3:29]))
X <- zoo(x,order.by=t)
X <- as.station(X,loc=paste(Vaggatem[1,3:27],Vaggatem[2,3:27],sep='-'),lon=rep(lon(precip),24),lat=rep(l
## For visualisation purposes
attr(X,'longitude') <- lon(X) + rnorm(length(lon(X)))
attr(X,'latitude') <- lat(X) + rnorm(length(lat(X)))
attr(X,'mean') <- cm
attr(X,'sigma') <- cs
nv <- apply(X,2,'nv')
X <- subset(X,is=nv > 20)
plot(X,new=FALSE)
```




```r
## Remove outliers
#coredata(X)[coredata(X) > 150] <- NA
```

The curves indicate that there are some spikes in some of the species in some years. The dots shown geographically are spread for the purpose of visualisation and their exact coordinate is the same for all of them. We use standardised series of fish data, as some have higher magnitudes and level than others.

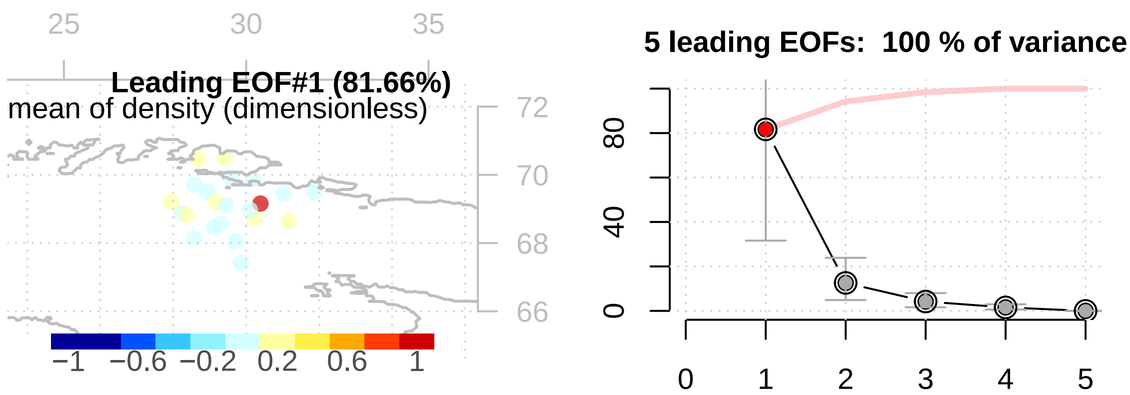
```r
diagnose(X)
```

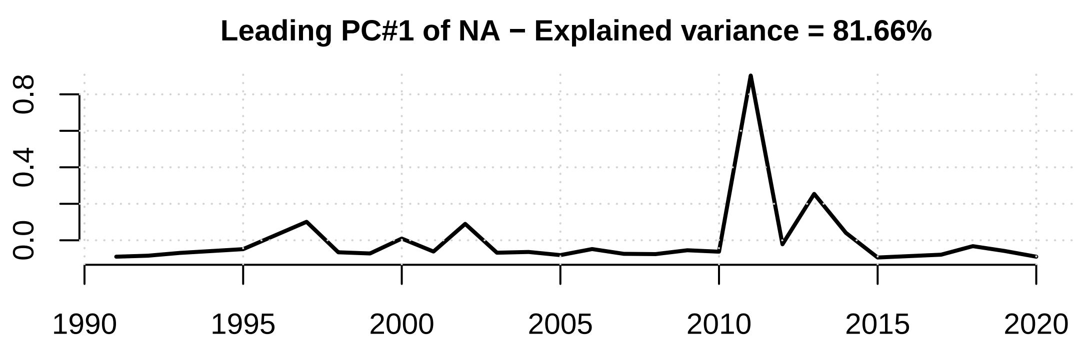


```
## Fill in gaps of missing data by assuming that the covariance structure in the data is stable
X <- pcafill(X)
```

There are some gaps with missing data. We fill in these gaps by assuming a constant co-variance tructure in the data.

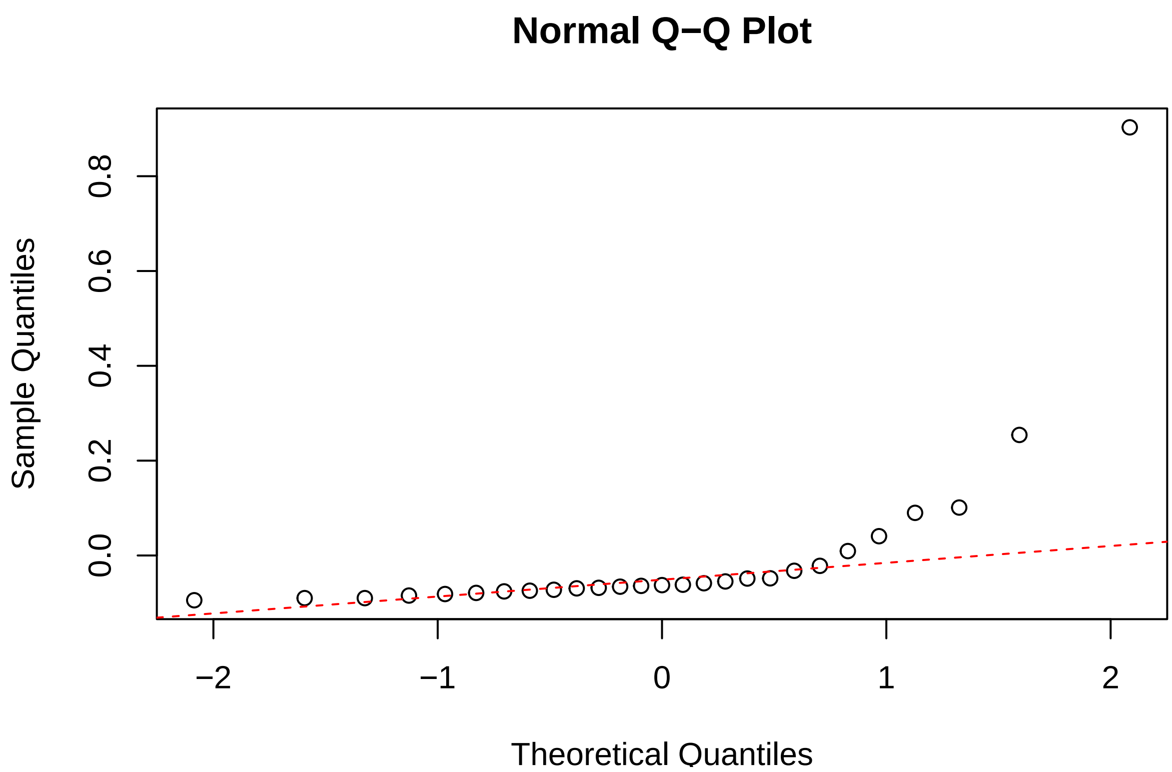
```
## Apply principal component analysis (PCA) to explore coherent structures in inter-annual variations i
## the PCA is also used to represent the predictands in the tests with empirical-statistical downscalin
pca <- PCA(X,n=5)
plot(pca,new=FALSE)
```





The PCA results suggest that there is a significant degree of coherent inter-annual variations in the different species, as the leding mode accounts for 82% of the variance. There is a spike in the data in 2011 - the data is not normally distributed.

```
qqnorm(coredata(pca[,1]))
qqline(coredata(pca[,1]),col='red',lty=2)
```



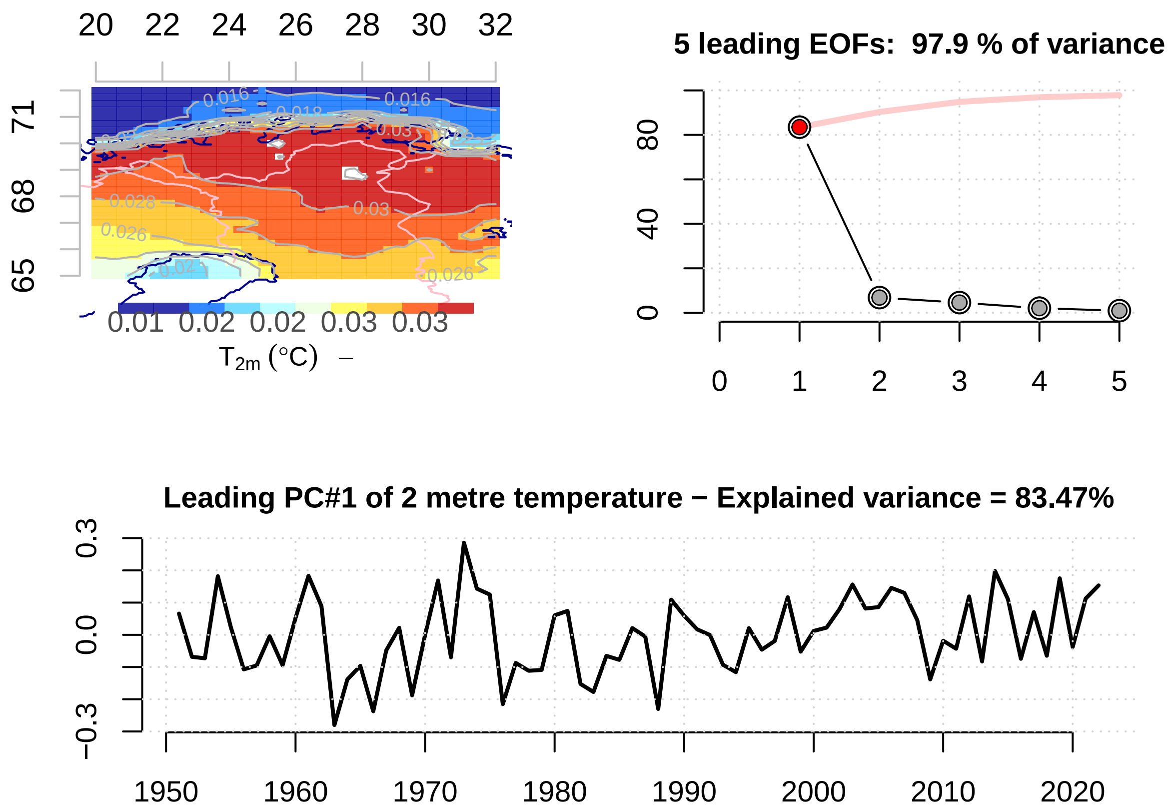
The data sample is relatively small which provides a constraint on the analysis.

Downscaling fish statistics

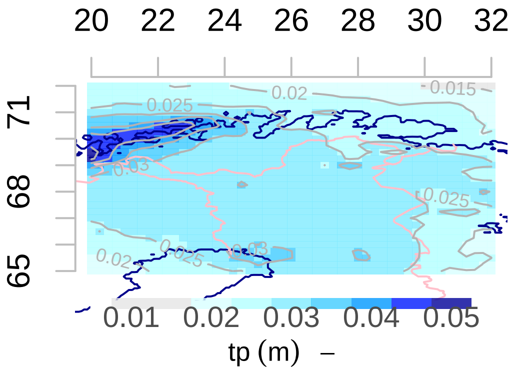
It is possible to downscale statistics of species directly in empirical-statistical downscaling by calibrating the methods on both large-scale meteorological data and local biological statistics. If there is a systematic effect of weather and climate on biological populations, then it should be found through such calibration exercises. One caveat is that the relationship between climate and biology may be non-linear, and here we have assumed a linear link between aggregated weather and fish statistics.

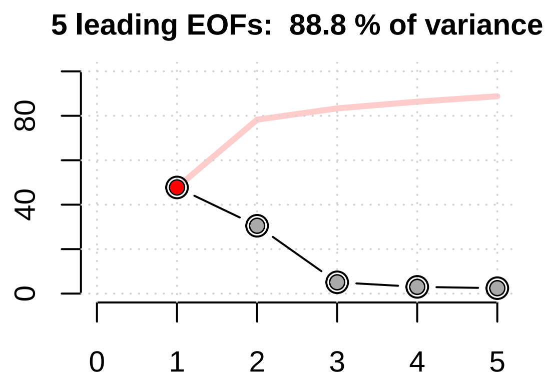
First we inspect the nature of potential predictands - we use empirical orthogonal functions (EOFs) to compress the data volume and extract salient information about inter-annual variations in summer mean temperature and annual precipitation totals. The maps presented below give an impression of how the strengths of these variations vary geographically.

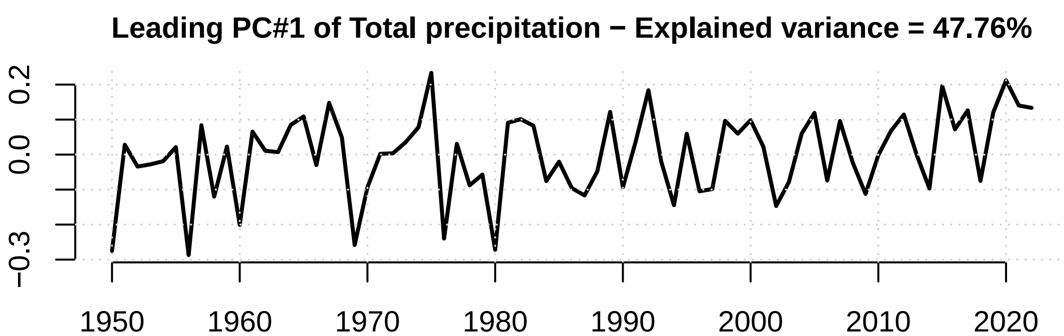
```
## Subset the region - the choice of region can sometimes matter.
# lons <- c(25,32); lats <- c(67,71)
# era5.t2m <- subset(era5.t2m,is=list(lon=lons,lat=lats))
# era5.tp <- subset(era5.tp,is=list(lon=lons,lat=lats))

## Estimate empirical orthogonal functions (EOFs) of the ERA5 temperature - this is a way to represent t
index(era5.t2m) <- year(era5.t2m) + 1 ## shift - so preceding summer temperature is compared with prese
eof.t2m <- EOF(era5.t2m,n=5)
plot(eof.t2m,new=FALSE)
```




```
## Repeat for precipitation
eof.tp <- EOF(era5.tp,n=5)
plot(eof.tp,new=FALSE)
```





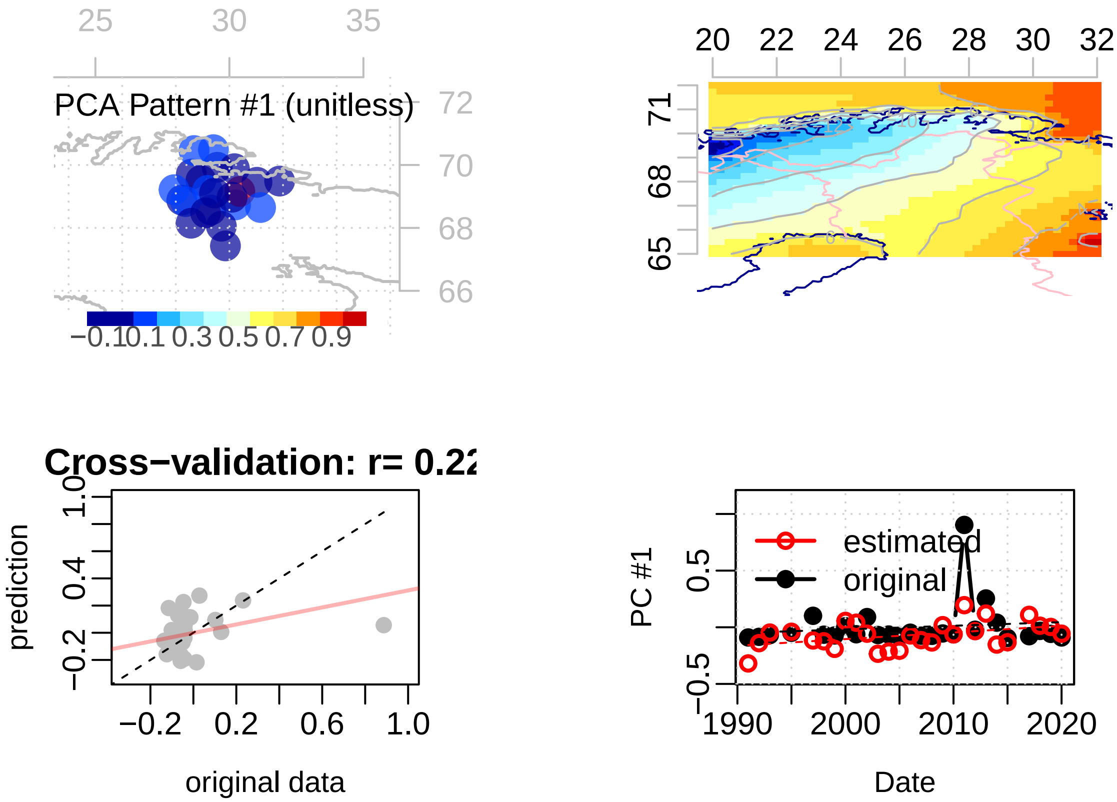


The EOF analyses indicate that the temperature varies most over land and the highest precipitation weights for the selected region is over northern Sweden. There is a strong spatial coherence is temperature (leading mode accounts for 92% of the variance), whereas there are more complicated spatio-temporal structure for precipitation. Nevertheless, 94% of the variance in annual precipitation sums can be captured by the 5 leading modes for the selected region.

```
## The PCA and EOF results are the inputs in the (stepwise multiple) regression-based downscaling (DS)
ds.t2m <- DS(pca,eof.t2m)
```

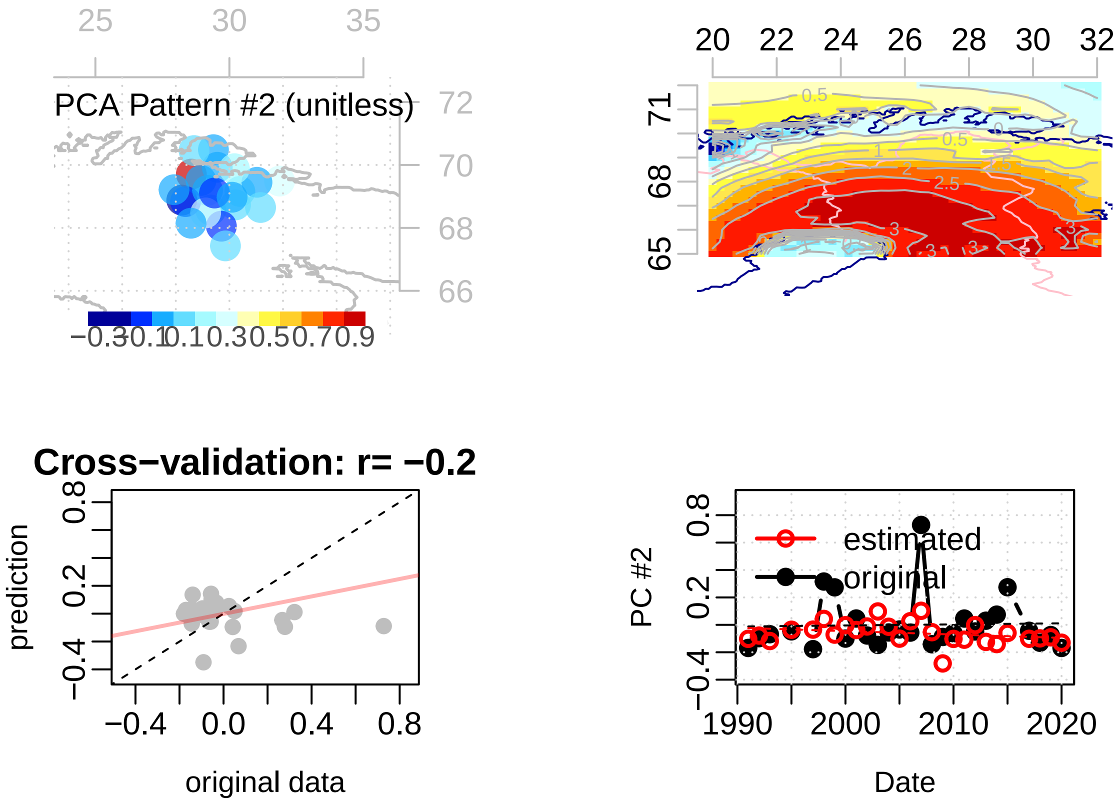
```
##   |                                                                              |
```

```
## The leading mode:
plot(ds.t2m,new=FALSE)
```



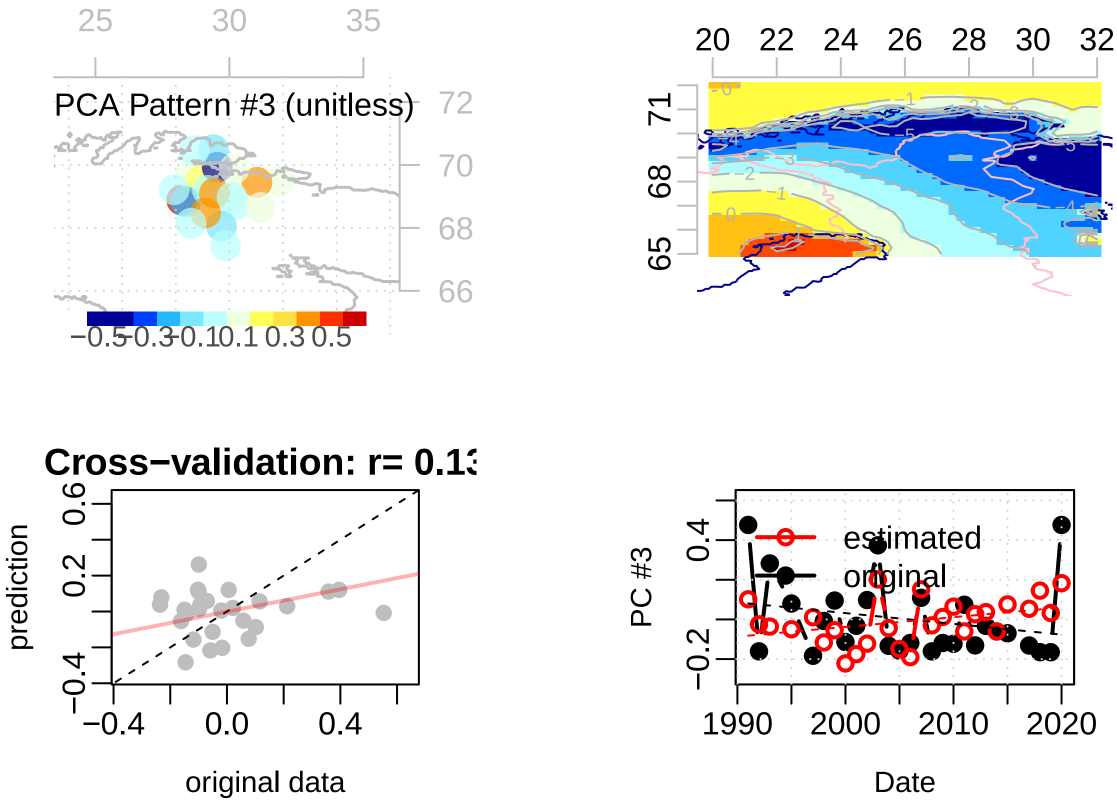

```
## NULL
```

```
## The secondary mode:
plot(ds.t2m,ip=2,new=FALSE)
```



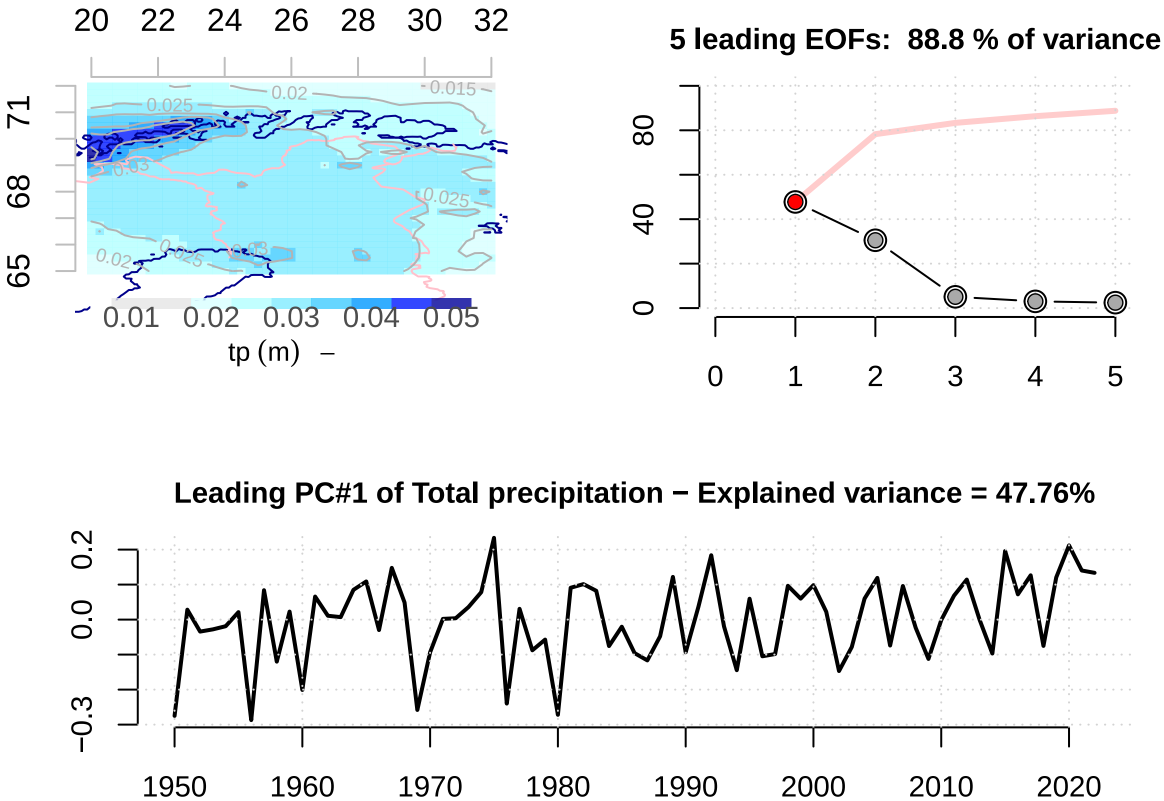

```
## NULL
```

```
## The third mode:
plot(ds.t2m,ip=3,new=FALSE)
```




```
## NULL
```

```
plot(eof.tp,new=FALSE)
```



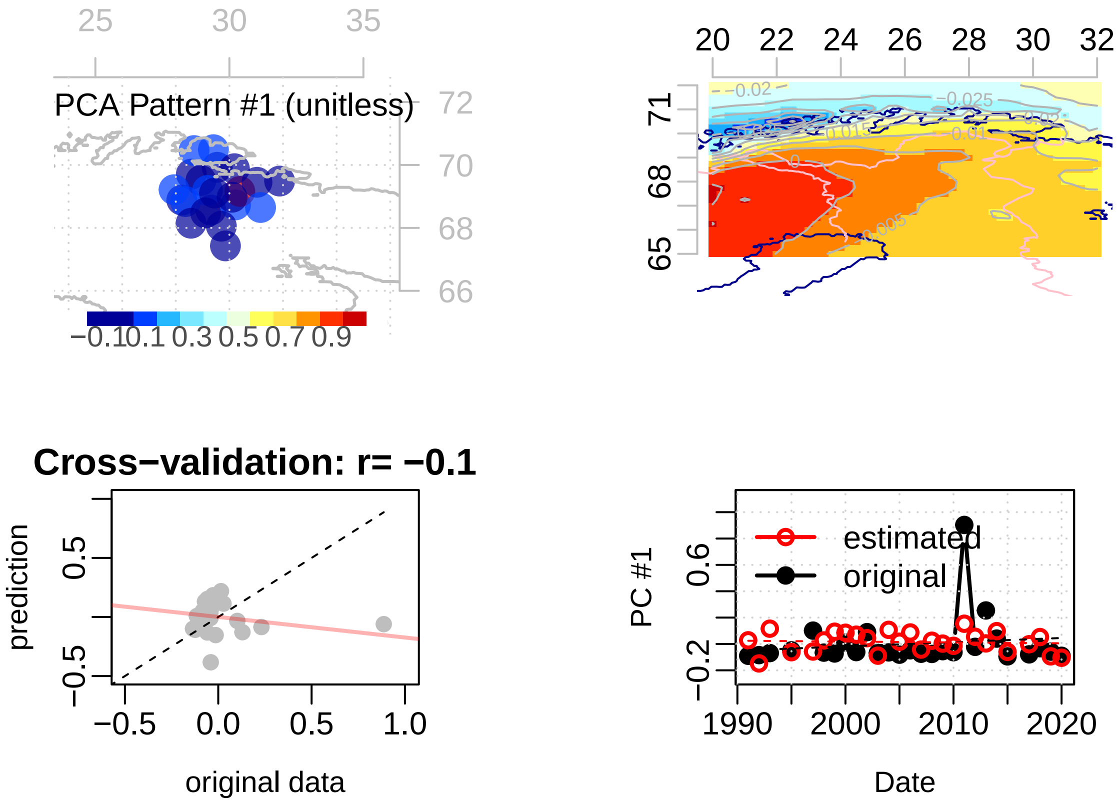
When using ERA5 summer mean temperature from the preceding year as predictor for annual statistics on the fish species in the Pasvik river, we find a weak connection, with a cross-variation correlation of 0.22 for the leading mode. One of the series dominate (red dot in the map), suggesting that there may be some issues with the data. The temperature pattern also reveals a west-east dipole structure, which may be difficult to associate with effects on the fish population.

The higher ordre modes do not yield any sensible results with temperature.

```
## Repeat to explore a precipitation link
ds.tp <- DS(pca,eof.tp)
```

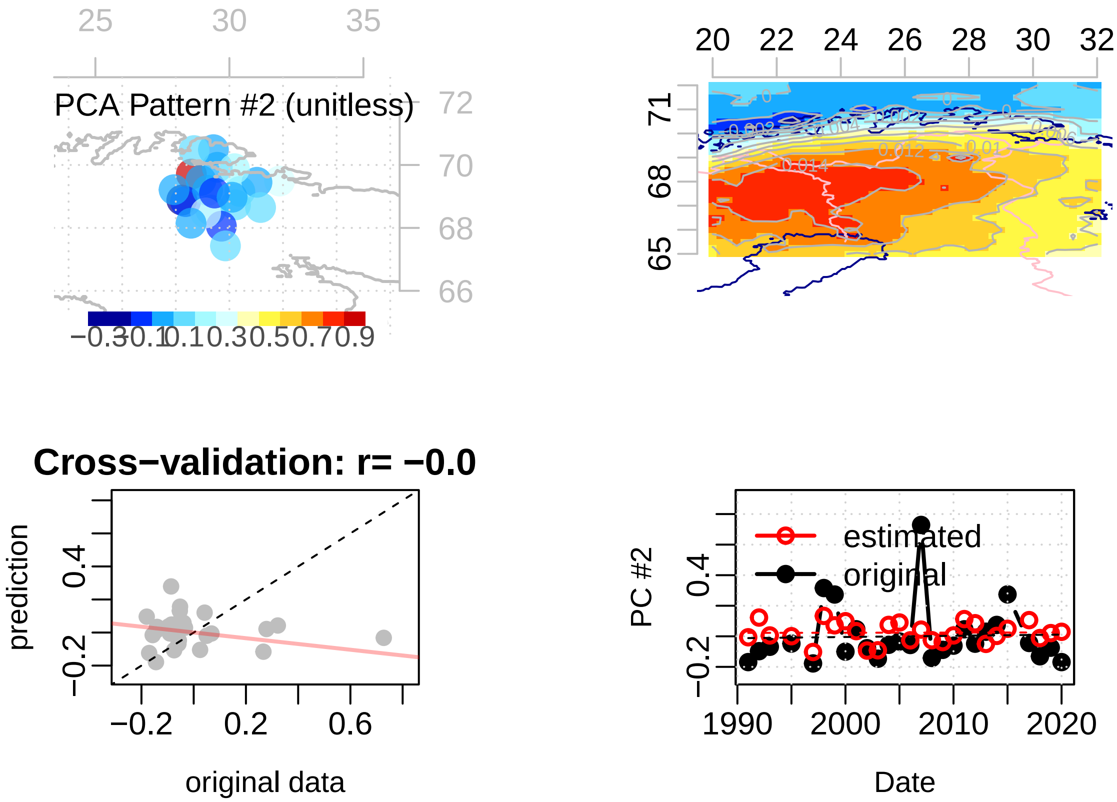
```
##   |                                                                              |
```

```
## The leading mode:
plot(ds.tp,new=FALSE)
```



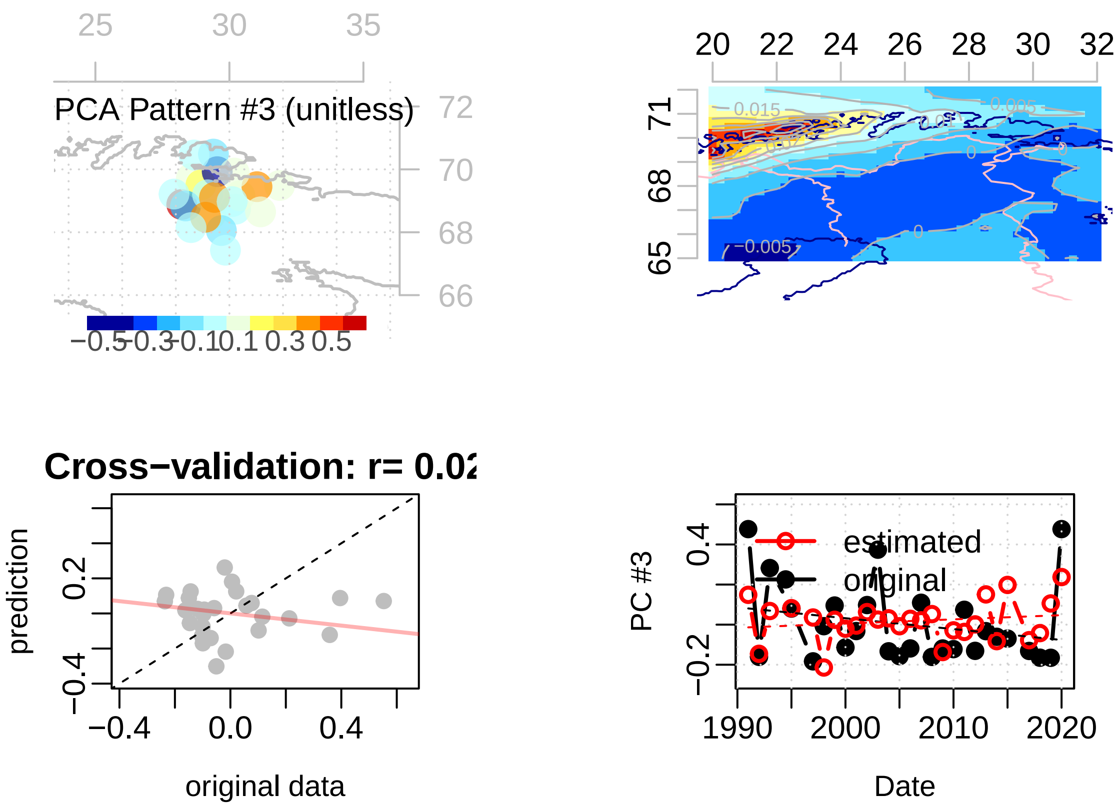

```
## NULL
```

```
## The secondary mode:
plot(ds.tp,ip=2,new=FALSE)
```




```
## NULL
```

```r
## The third mode:
plot(ds.tp,ip=3,new=FALSE)
```

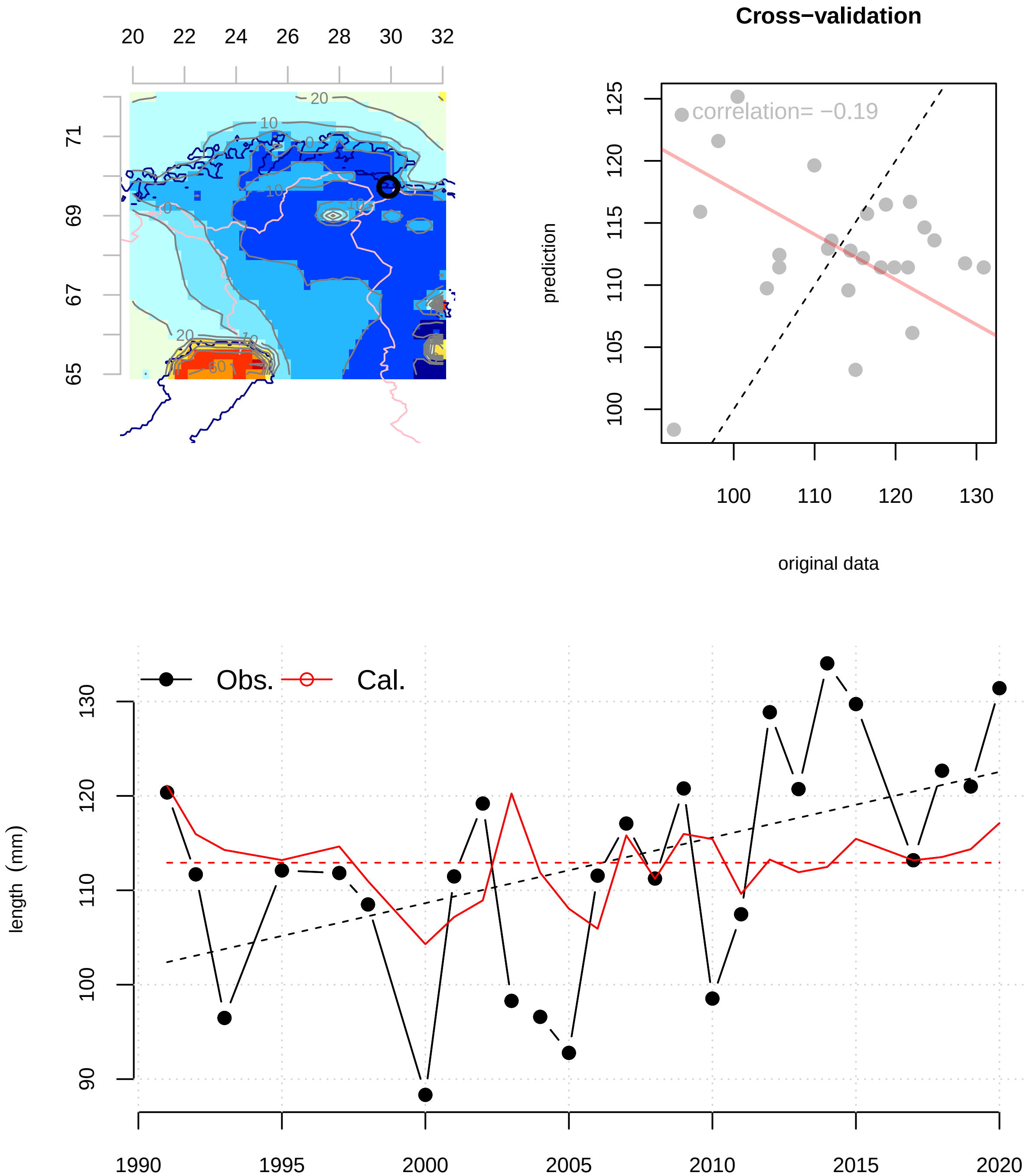



```
## NULL
```

The downscaling analysis with annual precipitation totals as predictor did not pick up any link that was deemed credible. The cross-validation correlation was less than zero. Possibly, there was a spike in the data that poluted the calibration.

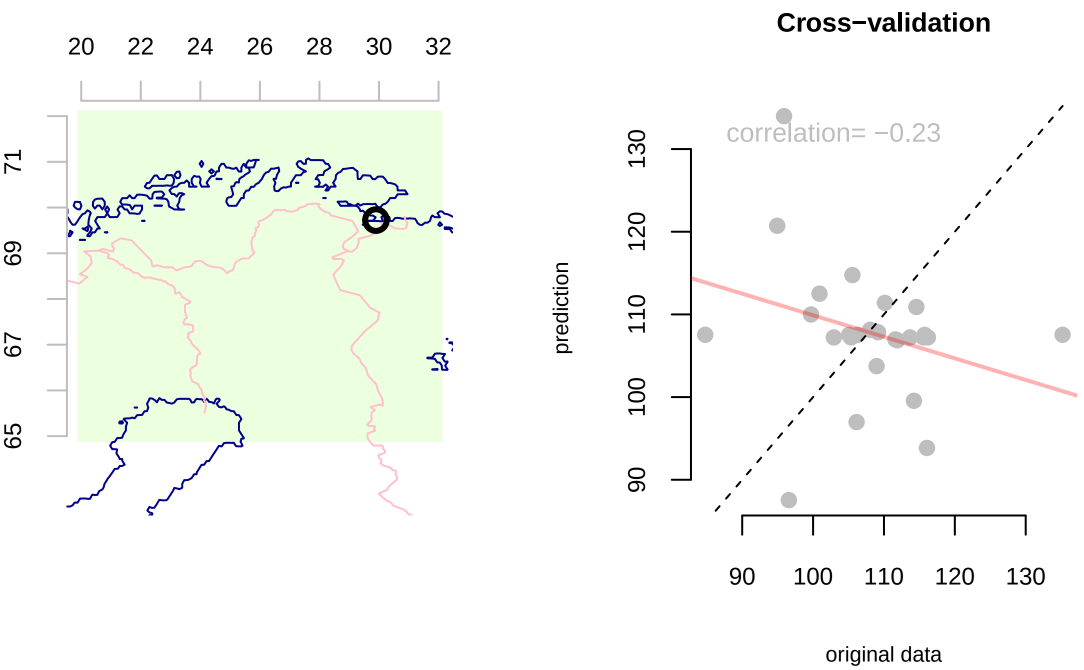
Fish length statistics from Pasvik

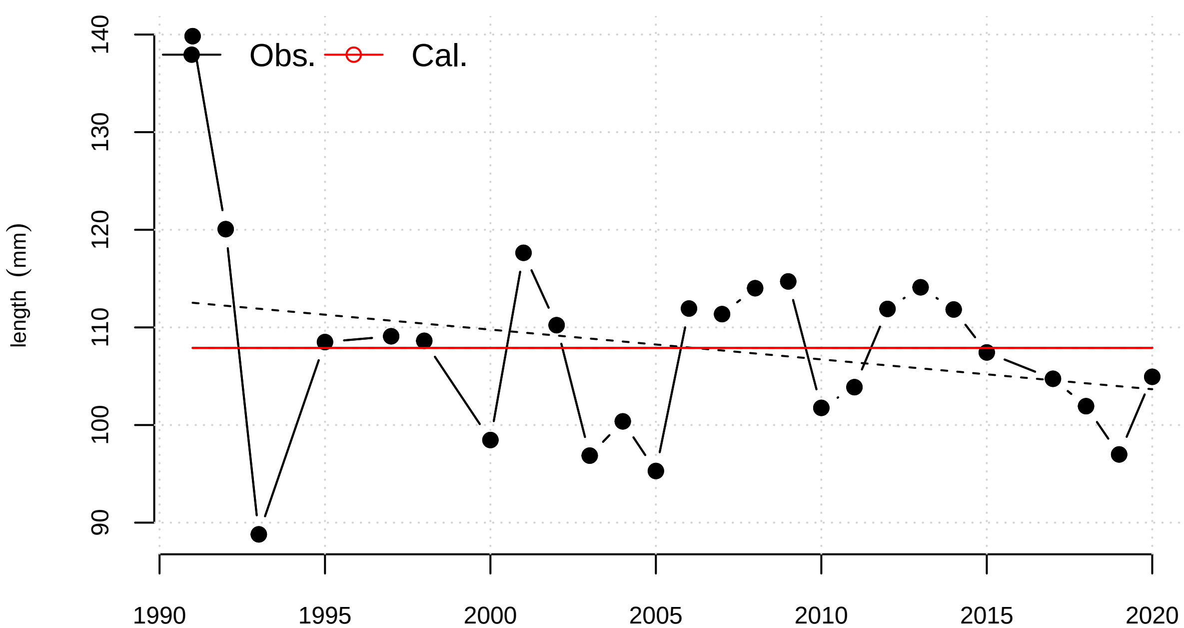
```
y1 <- aggregate(`Length (mm)` ~ Year, data=Perch,FUN='mean')
y1 <- zoo(y1$`Length (mm)`,order.by=y1$Year)
y2 <- aggregate(`Length (mm)` ~ Year, data=DRW,FUN='mean')
y2 <- zoo(y2$`Length (mm)`,order.by=y2$Year)
y3 <- aggregate(`Length (mm)` ~ Year, data=Vendace,FUN='mean')
y3 <- zoo(y3$`Length (mm)`,order.by=y3$Year)

## Make thetime series into esd-station objects that can be used as predictands in downscaling
y1 <- as.station(y1,loc='Perch',lon=lon(T2M),lat=lat(T2M),param='length',unit='mm')
## ds1 <- DS(y1,era5.t2m) # too short - only 8 years
y2 <- as.station(y2,loc='DRW',lon=lon(T2M),lat=lat(T2M),param='length',unit='mm')
ds2.t2m <- DS(y2,era5.t2m)
plot(ds2.t2m,new=FALSE)
```




```
y3 <- as.station(y3,loc='Vendace',lon=lon(T2M),lat=lat(T2M),param='length',unit='mm')
ds3.t2m <- DS(y3,era5.t2m)
plot(ds3.t2m,new=FALSE)
```

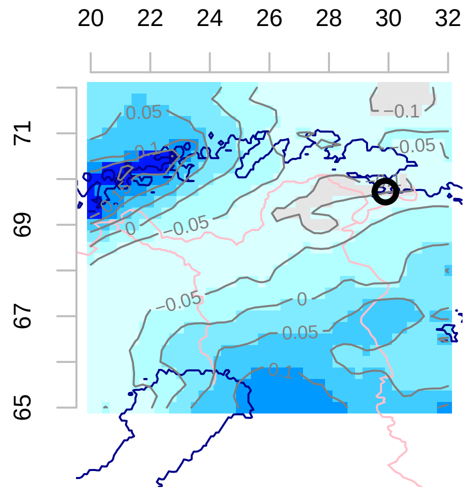
```
## Warning in summary.lm(trend1): essentially perfect fit: summary may be
## unreliable
```

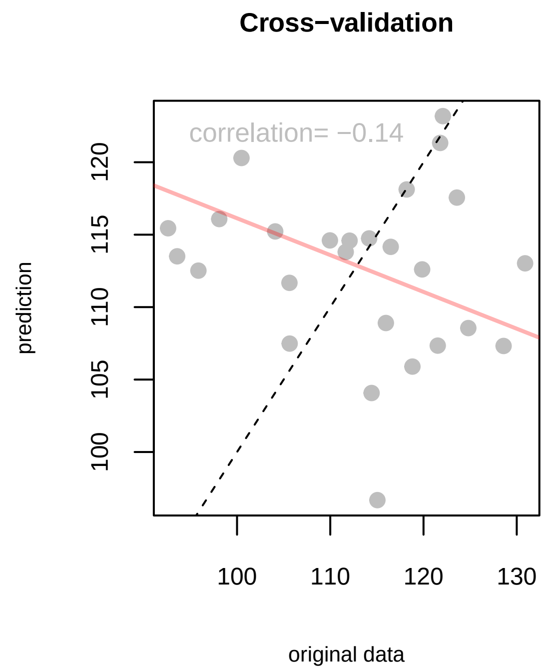


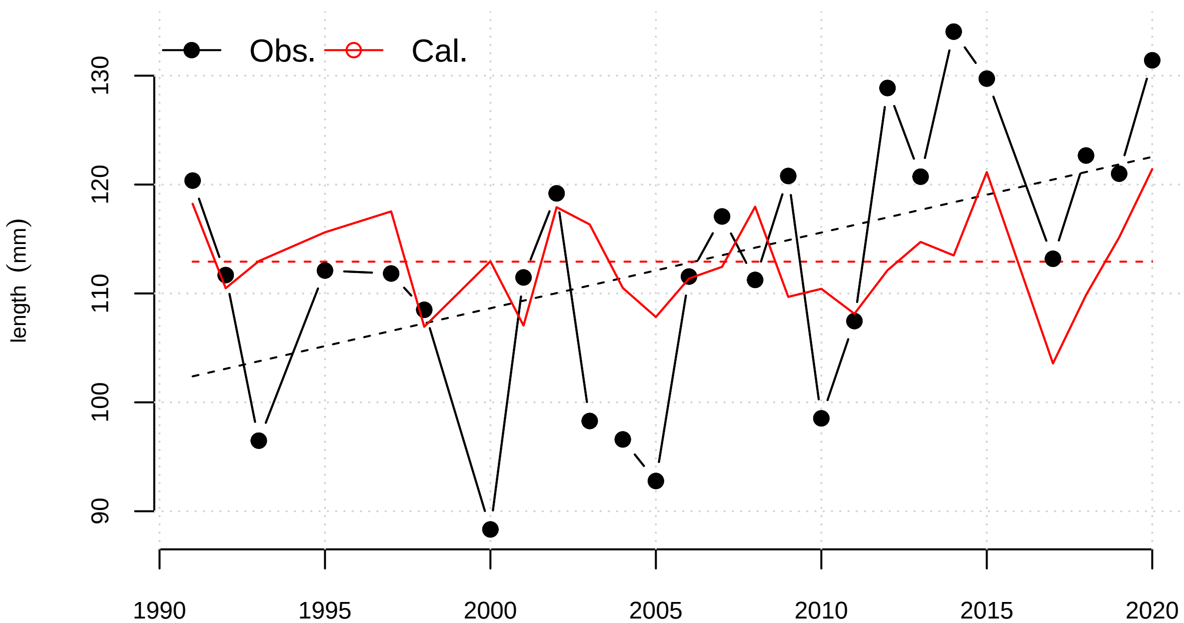


A downscaling analysis against DRW length for Age 1 gave a cross-valiation correlation of -0.19, and indicates no link. Similar score was obtained for Vendace. The data for Perch was insufficient for any analysis of this kind.

```
## Repeat the downscaling exercises with annual precipitation totals:
ds2.tp <- DS(y2,era5.tp)
plot(ds2.tp,new=FALSE)
```

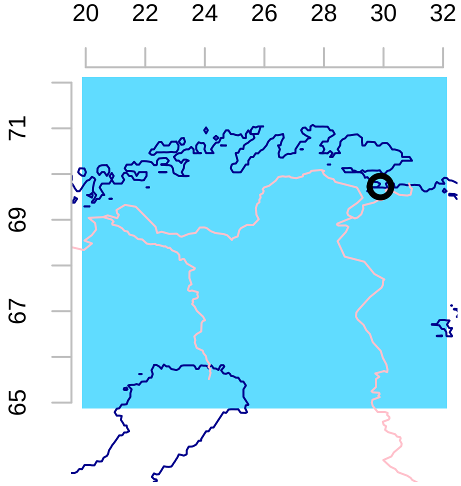


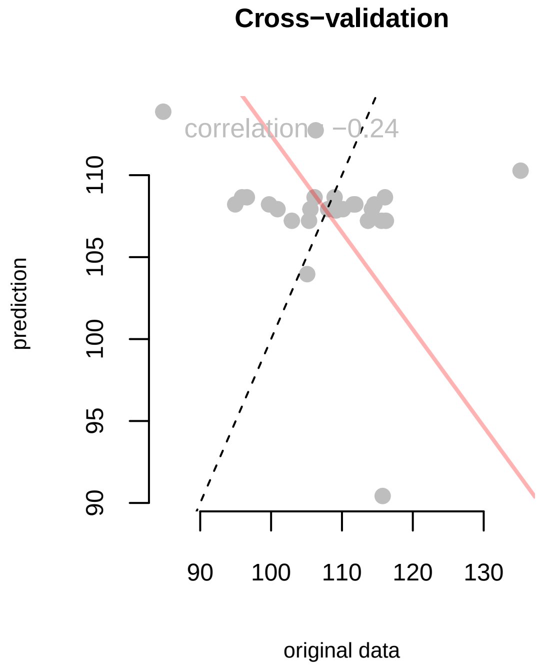


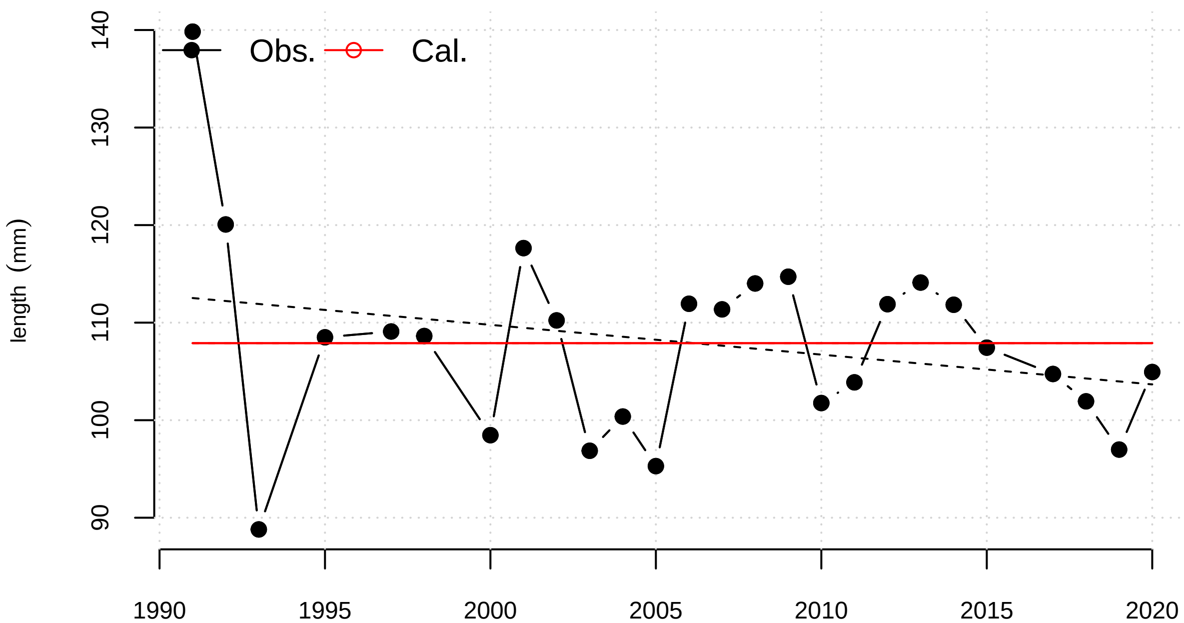



```
ds3.tp <- DS(y3,era5.tp)
plot(ds3.tp,new=FALSE)
```

```
## Warning in summary.lm(trend1): essentially perfect fit: summary may be
## unreliable
```







The results from the downscaling exercise using length of 1-year-old fish caught in Pasvik did not indicate any clear connection to the preceding summer mean temperature.

The analysis carried out here involved linear methods and did not evaluate any non-linear relationships between temperature/precipitation and the fish data.

A negative result is a valid scientific outcome, and here we used the ERA5 reanalysis to explore potential links between available (elected) fish statistics and seasonal temperature/precipitation.

References

Amundsen et al., NA. 2021. "Langtidsendringer i Økologi Og Miljøstatus for Fisk i Pasvikvassdraget 1991-2020." *UiT Norges Arktiske Universitet.* UiT. https://doi.org/NA.

Benestad, R.E. 2021. "A Norwegian Approach to Downscaling." *Geosci. Model Dev. Discuss.* Copernicus. https://doi.org/10.5194/gmd-2021-176.