Forecasting El Niño time series in a functional way

- **Statistical aim**

We focus on the El Niño time series which gives monthly Sea Surface temperatures. This dataset is not presented in the NPFDA book and for more details, see Ferraty, Rabhi and Vieu (2005, Sankhya, 67, 378-398). The data are recorded as a sequence of real numbers. Here, the dataset is componed of $N = 648$ real values $\{z_i, i = 1, \ldots, 648\}$ and organized as follows (see description of the datasets):

	Col 1	$\cdots$	Col j	$\cdots$	Col 12
Row 1	z_1	$\cdots$	z_j	$\cdots$	z_{12}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Row i	$z_{1+12(i-1)}$	$\cdots$	$z_{j+12(i-1)}$	$\cdots$	z_{12i}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Row 54	z_{633}	$\cdots$	z_{632+j}	$\cdots$	z_{648}

Firstly, one has to decide which past values have to be taken into account for prediction. In order to apply the functional methodology, one has to cut the original time series in a set of functional data. Here we have decided to predict future sea surface temperatures by using the El Niño data for the whole last year. That means that, with notations of our book, we have choosen $\tau = 12$. In order to illustrate our purpose, we will not use the 54^{th} year and we will predict it by mean of the data corresponding of the 53 previous years. To use the nonparametric functional methods, one has first to decide what is the horizon s of prediction that is wished. Then, for fixed s, the data will be reorganized into a functional explanatory sample $\{\chi_i, i = 1, \ldots, 52\}$ which will be loaded in the following 52×12 matrix:

z_1	$\cdots$	z_j	$\cdots$	z_{12}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$z_{1+12(i-1)}$	$\cdots$	$z_{j+12(i-1)}$	$\cdots$	z_{12i}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
z_{613}	$\cdots$	z_{612+j}	$\cdots$	z_{624}

and a response real sample $\{y_i, i = 1, \ldots, 52\}$, which will be loaded in the following 52-dimensional vector:

z_{12+s}	$\cdots$	z_{12i+s}	$\cdots$	z_{612+s}

For fixed horizon s, we can predict the value $\widehat{z}_{624+s}$ by using any technique among the three ones which are described in the book. Our goal is not to make a full analysis of this dataset, and to make things clearer we will just present the results obtained with $R/S+$ routines involving automatic smoothing parameter choices. More precisely, each among the three $R/S+$ routines `funopare.knn.lcv`, `funopare.mode.lcv` and `funopare.quantile.lcv` have been used to compute the predicted value $\widehat{z}_{624+s}$ obtained respectively by the regression operator estimation technique, by the conditional mode estimation technique and by the conditional median estimation technique. Concerning the semi-metric involved in the nonparametric forecasting procedures, the small number of discretization points for each curve (exactly 12) suggested the use of a semi-metric based on functional principal components ideas. Precisely, we used the PCA semi-metric $\boldsymbol{d}_q^{PCA}$ defined in the book and we took the parameter q which allows to get the best mean squares error ($q = 4$ for `funopare.knn.lcv`, $q = 3$ for `funopare.mode.lcv` and $q = 2$ for `funopare.quantile.lcv`).

- **Entering and organizing El Niño dataset**

```
ELNINODAT <- as.matrix(read.table("npfda-elnino.dat"))
attributes(ELNINODAT)$dimnames[[1]] <- character(0)
learning <- 1:52
testing <- 53
elnino.past.learn <- ELNINODAT[learning,]  # sample of explanatory curves
elnino.past.testing <- ELNINODAT[testing,] # The 53th year
s <- 1                                     # forecasting horizon 1
elnino.futur.s <- ELNINODAT[2:53,s]        # sample of real responses
```

Now, the R/S+ routines for prediction of a scalar response from a functional sample can be easily used in the following way.

- **The functional nonparametric forecasting**

```
result.pred.reg.step.s <- funopare.knn.lcv(elnino.futur.s,
  elnino.past.learn,elnino.past.testing,4,
  kind.of.kernel="quadratic",semimetric="pca")
              # Kernel functional regression forecasting

result.pred.median.step.s <- funopare.quantile.lcv(
  elnino.futur.s,elnino.past.learn,elnino.past.testing,2,
  alpha=0.5, Knearest=NULL, kind.of.kernel="quadratic",
  semimetric="pca")
              # Kernel functional median forecasting

result.pred.mode.step.s <- funopare.mode.lcv(
  elnino.futur.s,elnino.past.learn,elnino.past.testing,3,
  Knearest=NULL, kind.of.kernel="quadratic",
  semimetric="pca")
               # Kernel functional mode forecasting

result.pred.quantiles.s <- funopare.quantile.lcv(
  elnino.futur.s,elnino.past.learn,elnino.past.testing,
  2,alpha=c(0.05,0.5,0.95), Knearest=NULL,
  kind.of.kernel="quadratic",semimetric="pca")
               # Median estimation and 90% prediction band
```

- **Collecting the forecasting results**

  These R/S+ routines are recording several different results. The most important ones are the predicted responses which can be obtained in the following manner.

```
result.pred.reg.step.s$Predicted.values
              # Forecasted value with regression method

result.pred.median.step.s$Predicted.values
              # Forecasted value with median method

result.pred.mode.step.s$Predicted.values
              # Forecasted value with mode method
```

```
result.pred.quantiles.s$Predicted.values
              #.05, .5 and .95 estimated quantiles
```

- **Forecasting the 54th year**

  To do that, it suffices to repeat the previous stages for $s = 1, \ldots, 12$ (horizons):

```
pred.reg <- 0
pred.median <- 0
pred.mode <- 0
for(s in 1:12){
        elnino.futur.s <- ELNINODAT[2:53,s]
        result.pred.step.s <- funopare.knn.lcv(elnino.futur.s,
          elnino.past.learn,elnino.past.testing,4,
          kind.of.kernel="quadratic",semimetric="pca")
        pred.reg[s] <- result.pred.step.s$Predicted.values
        result.pred.median.step.s <- funopare.quantile.lcv(elnino.futur.s,
          elnino.past.learn,elnino.past.testing,2,alpha=0.5,
          Knearest=NULL, kind.of.kernel="quadratic", semimetric="pca")
        pred.median[s] <- result.pred.median.step.s$Predicted.values
        result.pred.mode.step.s <- funopare.mode.lcv(elnino.futur.s,
          elnino.past.learn,elnino.past.testing,3,Knearest=NULL,
          kind.of.kernel="quadratic",semimetric="pca")
        pred.mode[s] <- result.pred.mode.step.s$Predicted.values
}
```

- **Plotting the forecasted values**

  The following commandlines allow to display the forecasted 54th year obtained (Figure 1) by the various functional prediction methods and we compare them with the observed values (54th year):

```
year54 <- ELNINODAT[54,]
mse.reg <- round(sum((pred.reg-year54)^2)/12,4)
mse.median <- round(sum((pred.median-year54)^2)/12,4)
mse.mode <- round(sum((pred.mode-year54)^2)/12,4)
```

```
par(mfrow=c(2,2))
plot(1:12,pred.reg,xlab='54th year', ylab='',
  main=paste('Regression: MSE=',mse.reg,sep=''),
  type='l',lty=2,ylim=range(c(pred.reg,year54)))
par(new=T)
plot(1:12,year54,type='l',lty=1,axes=F,xlab='', ylab='')
plot(1:12,pred.median,xlab='54th year', ylab='',
  main=paste('Median: MSE=',mse.median,sep=''),
  type='l',lty=2,ylim=range(c(pred.median,year54)))
par(new=T)
plot(1:12,year54,type='l',lty=1,axes=F,xlab='', ylab='',
  ylim=range(c(pred.median,year54)))
plot(1:12,pred.mode,xlab='54th year', ylab='',
  main=paste('Mode: MSE=',mse.mode,sep=''),
  type='l',lty=2,ylim=range(c(pred.mode,year54)))
par(new=T)
plot(1:12,year54,type='l',lty=1,axes=F,xlab='', ylab='',
  ylim=range(c(pred.mode,year54)))
```

Remark: it is possible to get pointwise confidence prediction band with the routine `funopare.quantile.lcv`.

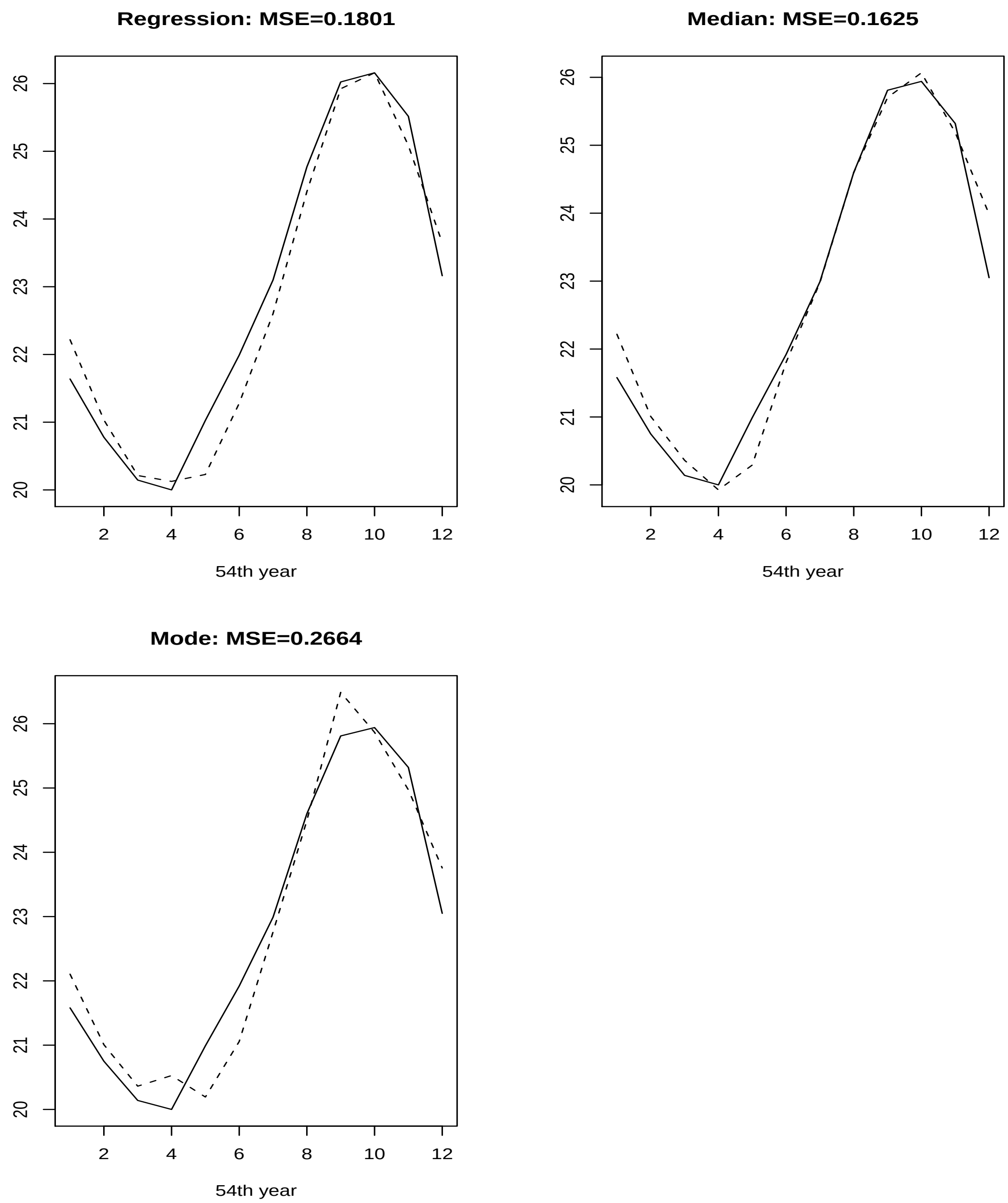

Figure 1: Forecasted 54th year for the three prediction methods