Forecasting electricity consumption in a functional way

- **Statistical aim**

We focus on the electricity consumption time series. The data are recorded as a sequence of real numbers. Here, the dataset is componed of $N = 336$ real values $\{z_i, i = 1, \ldots, 336\}$ (not to mask the main of purpose, we will directly work with the differenciated log data) 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 28	z_{325}	$\cdots$	z_{324+j}	$\cdots$	z_{336}

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 electrical consumption by using the consumption 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 28^{th} year and we will predict it by mean of the data corresponding of the 27 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 $\{\chi_i, i = 1, \ldots, 26\}$ which will be loaded in the following 26×12 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_{301}	$\cdots$	z_{300+j}	$\cdots$	z_{312}

and a response real sample $\{y_i, i = 1, \ldots, 26\}$, which will be loaded in the following 26-dimensional vector:

z_{12+s}	$\cdots$	z_{12i+s}	$\cdots$	z_{312+s}

For fixed horizon s, we can predict the value $\widehat{z}_{324+s}$ 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 economic 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 $\widehat{z}_{324+s}$ 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 square errors ($q = 5$ for `funopare.knn.lcv`, $q = 2$ for `funopare.mode.lcv` and `funopare.quantile.lcv`).

- **Organizing electrical data**

```
CONSELDAT <- as.matrix(read.table("npfda-electricity.dat"))
attributes(CONSELDAT)$dimnames[[1]] <- character(0)
learning <- 1:26
testing <- 27
elec.past.learn<-CONSELDAT[learning,]  # sample of explanatory curves
elec.past.testing<-CONSELDAT[testing,] # The 27th year
s <- 1                                 # forecasting horizon 1
elec.futur.s <- CONSELDAT[2:27,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(elec.futur.s,
  elec.past.learn,elec.past.testing,5,
  kind.of.kernel="quadratic",semimetric="pca")
              # Kernel functional regression forecasting

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

result.pred.quantiles.s<-funopare.quantile.lcv(
  elec.futur.s,elec.past.learn,elec.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 28th 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){
        elec.futur.s <- CONSELDAT[2:27,s]
        result.pred.step.s <- funopare.knn.lcv(elec.futur.s,
          elec.past.learn,elec.past.testing,5,
          kind.of.kernel="quadratic",semimetric="pca")
        pred.reg[s] <- result.pred.step.s$Predicted.values
        result.pred.median.step.s <- funopare.quantile.lcv(elec.futur.s,
          elec.past.learn,elec.past.testing,2,alpha=0.5,
          Knearest=NULL, kind.of.kernel="quadratic", semimetric="pca")
        pred.median[s] <- result.pred.median.step.s$Predicted.values
        result.pred.mode.step.s <- funopare.mode.lcv(elec.futur.s,
          elec.past.learn,elec.past.testing,2,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 28th year obtained (Figure 1) by the various functional prediction methods and we compare them with the observed values (28th year):

```
year28 <- CONSELDAT[28,]
mse.reg <- round(sum((pred.reg-year28)^2)/12,4)
mse.median <- round(sum((pred.median-year28)^2)/12,4)
mse.mode <- round(sum((pred.mode-year28)^2)/12,4)
```

```
par(mfrow=c(2,2))
plot(1:12,pred.reg,xlab='28th year', ylab='',
  main=paste('Regression: MSE=',mse.reg,sep=''),
  type='l',lty=2,ylim=range(c(pred.reg,year28)))
par(new=T)
plot(1:12,year28,type='l',lty=1,axes=F,xlab='', ylab='')
plot(1:12,pred.median,xlab='28th year', ylab='',
  main=paste('Median: MSE=',mse.median,sep=''),
  type='l',lty=2,ylim=range(c(pred.median,year28)))
par(new=T)
plot(1:12,year28,type='l',lty=1,axes=F,xlab='', ylab='',
  ylim=range(c(pred.median,year28)))
plot(1:12,pred.mode,xlab='28th year', ylab='',
  main=paste('Mode: MSE=',mse.mode,sep=''),
  type='l',lty=2,ylim=range(c(pred.mode,year28)))
par(new=T)
plot(1:12,year28,type='l',lty=1,axes=F,xlab='', ylab='',
  ylim=range(c(pred.mode,year28)))
```

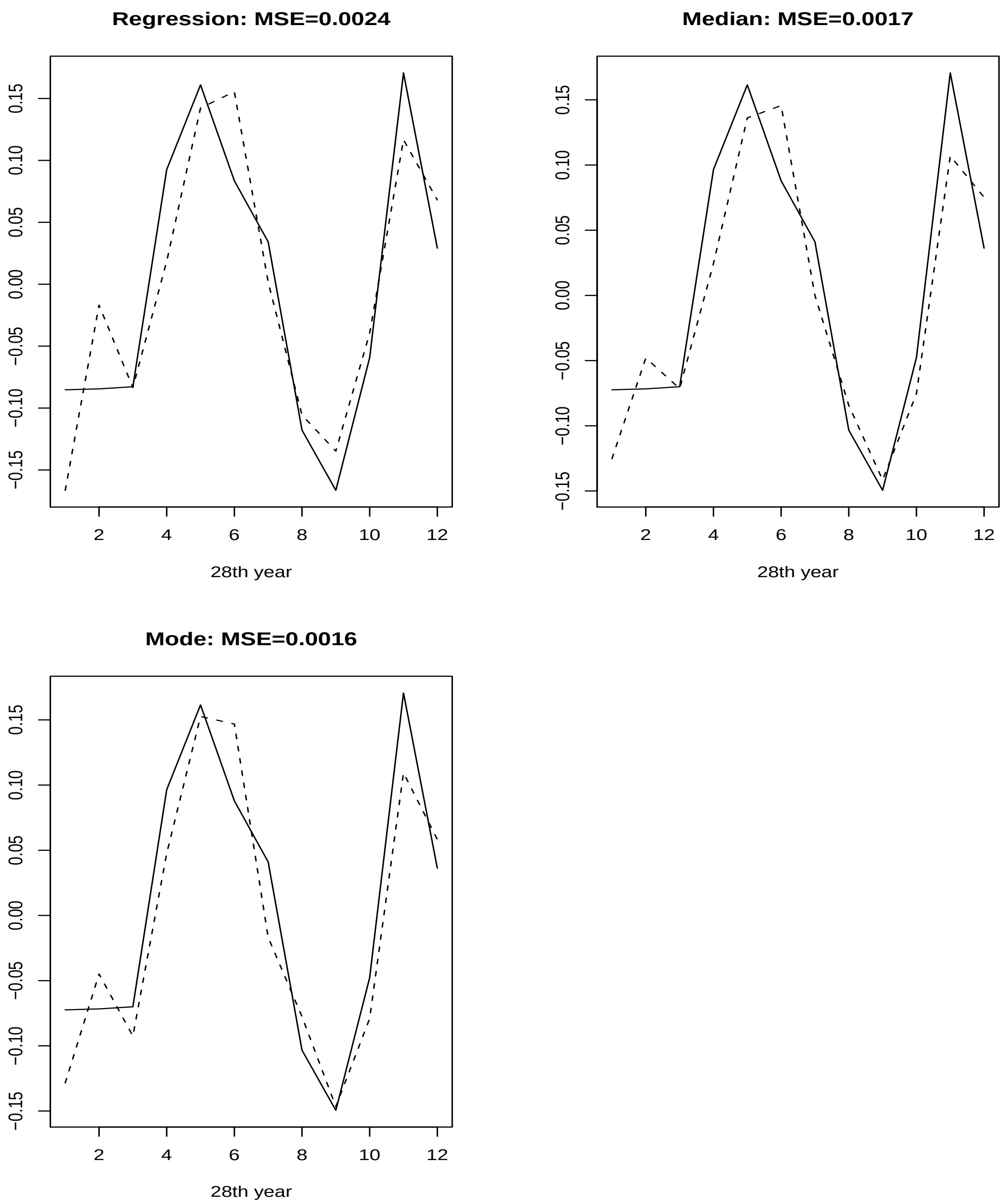

Figure 1: Forecasted 28th year for the three prediction methods