Analysis of nonlinear time series

Douglas Nychka Statistics, NCSU
Stephen Ellner Biomath, NCSU
Barbara Bailey, Statistics project, NCAR

Main points:

- Nonlinear systems
- Nonparametric regression
- Data analysis tools and examples

`nychka@stat.ncsu.edu`
http://www.stat.ncsu.edu/

Nonlinear Autoregressive Processes

1-d Model:

$\{x_t\}$ is a univariate time series

$$x_t = f(x_{t-1}, x_{t-2}, \ldots, x_{t-d}) + e_t$$

$\{e_t\}$ are iid with $E(e_t) = 0$ and $Var(e_t) = \sigma^2$

State Space/Vector Model:

$$X_t = F(X_{t-1}) + E_t$$

$X_t \in \Re^d$ describes the state of system at time t and E_t is a random shock. $F : R^d \rightarrow R^d$

The 1-d model is a special case:

$X_t = (x_t, \ldots, x_{t-d+1})^T$ and $E_t = (e_t, 0, \ldots, 0)^T$

$$F(X_t) = \begin{pmatrix} f(x_{t-1}, x_{t-2}, \ldots, x_{t-d}) \\ x_{t-1} \\ \vdots \\ x_{t-d+1} \end{pmatrix}$$

Example: The logistic map

$$x_t = cx_{t-1}(1 - x_{t-1}) + e_t$$

A toy model for chaos that illustrates the difficulty of prediction in nonlinear systems (c $\approx$ 3.7)

What is the relationship of x_{t+m} to x_t ?

This behavior is in contrast to a linear autoregressive process

$$x_t = cx_{t-1} + e_t$$

For example, with $|c| < 1$ consider $x_{t+2} = e_t + ce_{t-1} + c^2 x_t$
The dependence on x_t dies away as a power of c!

Example: Rossler System

Coupled, nonlinear system of three first order differential equations:

$$\begin{aligned}\dot{x} &= -(y+z)\\ \dot{y} &= x + 0.15y\\ \dot{z} &= 0.2 + z(x-10)\end{aligned}$$

State-Space System:

$$\begin{aligned}x_t &= F_1(x_{t-1}, y_{t-1}, z_{t-1}) + e_{1,t}\\ y_t &= F_2(x_{t-1}, y_{t-1}, z_{t-1}) + e_{2,t}\\ z_t &= F_3(x_{t-1}, y_{t-1}, z_{t-1}) + e_{3,t},\end{aligned}$$

1-d system:

$$x_t = f(x_{t-1}, x_{t-2}, x_{t-3}) + e_t$$

f exists by Taken's theorem provided there are enough lags.

Lyapunov Exponents Measure Sensitivity to Initial Conditions

X_1 and Y_1 are two initial states close to each other. How different are X_2 and Y_2?

Use a Taylor series expansion to approximate the difference

$$\begin{aligned} X_2 - Y_2 &= F(X_1) - F(Y_1) \\ &\approx DF(X_1)(X_1 - Y_1) \end{aligned}$$

Let $J_t = DF(X_t)$, the Jacobian matrix of F .

By the chain rule for differentiation

$$X_n - Y_n \approx J_n \cdot J_{n-1} \cdots J_1(X_1 - Y_1)$$

for example

$$X_3 - Y_3 \approx J_2 J_1 (X_1 - Y_1)$$

where X and Y see the *same* random shocks.

Global Lyapunov Exponent

$$\lambda = \lim_{n\to\infty} \frac{1}{n} \ln \| J_n J_{n-1} \ldots J_1 \|$$

Interpretation: For small perturbations, the difference after n steps grows like

$$||Y_{t+n} - X_{t+n}|| \approx e^{\lambda n} ||Y_t - X_t||$$

For stationary, linear AR models the Lyapunov exponent is negative.

Local Lyapunov Exponent

$$\lambda_n(t) = \quad \frac{1}{n} \ln \| J_{n+t-1} J_{n+t-2} \cdots J_t \|$$

Interpretation: Large positive λ_n indicate times where the system is not predictable n steps ahead.

Chaos Facts

- If X_t is ergodic, stationary and bounded. Then λ exists and is independent of the trajectory (Multiplicative Ergodic Theorem of Oseledec)

- A system with $\lambda > 0$ has the property of

  "sensitive dependence on initial conditions"
  and is *chaotic*.

- By using the same sequences of random shocks this definition makes sense for systems with a random component.

- As m get large, the distribution of Local exponents will tend to a Normal (Gaussian) (with *scaling* law $\sqrt{n}$) centered at the global LE. Of course, $\lambda_n \to \lambda$ as $n \to \infty$.

Other measures of predictability (Tong 1995)

Let $\sigma^2 = VAR(e_t)$

$$\sigma_n^2(x) = VAR(X_n | X_1 = x)$$

and for small σ this can be tied to the local Lyapunov exponents over this period of time.

It is possible that $\sigma_n^2(x) > \sigma_{n+1}^2(x)$ in some cases!

Suppose that X_n is estimated by $\hat{f}(x)$

$$E\left((x_n - \hat{f}(x))^2 | X_1 = x + \delta\right) \approx \sigma_n^2(x+\delta) + \delta^T \nabla f_n(x)$$

Unlike the definition of the LE's, these quantities are averages over the future random shocks e_2,..., e_n. This feature may confuse stochastic and highly nonlinear systems.

The key with any of these measures when confronted with data is to *estimate* the unknown functions and parameters.

Nonparametric regression

$$Y = f(X) + \epsilon$$

Goal is to estimate f without making parametric assumptions on the functional form.

- Thin plate splines
- Neural Networks

These handle the dimension of the state vector in different ways and are complementary.

The curse of dimensionality or ...

Space is really big

If f is an M^{th} degree polynomial in d dimensions then there will be

$$\begin{pmatrix} M+d \\ d \end{pmatrix}$$

terms, growing like M^d. This exponential explosion of terms is one form of the curse of dimensionality.

Thin Plate Splines

A spline is the solution to a minimization problem.

Assume that e_i are i.i.d. $N(0, \sigma^2)$ and f is smooth (several times differentiable).

$J(f)$ is a roughness penalty.

A thin plate spline is the function that minimizes

$$\frac{1}{n} \sum_{i=1}^{n} (y_i - f(x_i))^2 + \rho J(f)$$

over all f.

For example if $x \in \Re^2$

$$J(f) = \int_{\Re^2} \left(\frac{\partial^2 f}{\partial x_1^2}\right)^2 + 2\left(\frac{\partial f}{\partial x_1}\frac{\partial f}{\partial x_2}\right)^2 + \left(\frac{\partial^2 f}{\partial x_2^2}\right)^2 dx$$

This is a rotationally invariant measure of the degree of curvature in the surface, f. (The bending energy if f was a thin plate.)

Remarks about spline estimates

- For fixed ρ, the spline depends in a linear way on the dependent variables.
- There is a matrix $A(\rho)$ so that $\hat{y} = A(\rho)y$
- The trace of $A(\rho)$ is a measure of the effective number of parameters used to fit the data.
- The smoothing can be found by a version of cross-validation. Find ρ that minimizes:

  $$V(\rho) = \frac{(1/n)RSS}{(1 - trA(\rho)/n)^2}$$

  This criterion is similar to Mallows C_p statistic and is an unbiased estimate of the average prediction variance for a given value of ρ.

Neural networks (single hidden layer)

The form of the model is

$$f(X) = \beta_0 + \sum_{i=1}^{k} \beta_i \varphi(X^T \gamma_i + \mu_i)$$

where $\varphi(u) = e^u/(1 + e^u)$, the logistic distribution function. Number of parameters is $p = 1 + k(d + 2)$.

The Demmler-Reinsch basis

Consider a linear combination of the radial basis functions and polynomials with some advantageous properties.

$\{g_\nu\}$ with eigenvalues $D_1, D_2, ..., D_n$

if

$$h(x) = \sum_\nu \alpha_\nu g_\nu(x)$$

then

$$\sum_{k=1}^{n} (Y_k - h(x_k))^2 = \sum_{\nu=1}^{n} (u_\nu - \alpha_\nu)^2$$

and

$$J_m(h) = \sum_{\nu=1}^{n} D_\nu \alpha_\nu^2$$

where $u_\nu = \Sigma_{k=1}^{n} g_\nu(x_k) Y_k$

$$\hat{f}(x) = \sum_{\nu=1}^{N} \frac{u_\nu}{1 + \lambda D_\nu} g_\nu(x)$$

$$tr(A(\rho)) = \sum_{k=1}^{n} \frac{1}{1 + \lambda D_k}$$

Finding a confidence band

For $C_1, C_2 > 0$ let

$$\mathcal{B} = \{h : \sum_{i=1}^{m} (\hat{f}(x_i) - h(x_i))^2 \leq C_1 \text{ and } J(h) \leq C_2\}$$

The upper and lower boundaries are derived by

$$U(x) = max_{h \in \mathcal{B}} h(x) \text{ and } L(x) = min_{h \in \mathcal{B}} h(x)$$

A computational simplification is to restrict B to functions spanned by $\{g_\nu\}$.

Neural networks (single hidden layer)

The form of the model is

$$f(X) = \beta_0 + \sum_{i=1}^{k} \beta_i \varphi(X^T \gamma_i + \mu_i)$$

where $\varphi(u) = e^u/(1+e^u)$, the logistic distribution function. Number of parameters is $p = 1 + k(d+2)$.

- Net parameters are estimated (with some hardship) by nonlinear least squares.
- Complexity of the model chosen by Cross Validation:

  $$V_{\mathcal{C}}(p) = \frac{\frac{1}{n} RSS}{\left(1 - p\frac{\mathcal{C}}{n}\right)^2}$$

  Our favorite cost per parameter is $\mathcal{C} = 2.0$.
- The method of projection pursuit replaces φ by arbitrary functions

Estimates of Lyapunov exponents

Estimate the map, (f or F)

Calculate partial derivatives at the observed points.

Form the Jacobian matrices and plug these in the product formula

Explicit form for the spline

Solution is a linear combination of radial basis functions and low order polynomials. In 2-dimensions

$$bold\beta_1 u_1 + \beta_2 u_3 + \sum_{i=1}^{n} \delta_i \psi_i(u)$$

and for this 2-d case

$$\psi_i(u) = E(\|u - x_i\|) = log(\|u - x_i\|)\|u - x_i\|^2$$

Confidence regions for weird stuff

Because the LLE's are highly nonlinear functions of the map, simple approximations to get confidence intervals are not available.

Confidence set approach: Suppose one finds a set (of maps or parameters), $\mathcal{A}$, so that

$$P(f \in \mathcal{A}) = .95$$

where f is the true function.
Then if $\psi(h)$ is the LE based on the map h,

$$\mathcal{I} = \{\psi(h) : h \in \mathcal{A}\}$$

is a confidence set (interval) for $\psi(f)$

Operationally:

For neural network estimates we find 500 representative models that are within the 95% joint confidence region based on the likelihood ratio statistic. Then the LE's are calculated for each of these models and the minimum and maximum values determine the interval.

Regional Oxidant Model (ROM)

ROM is a large photochemical model, forced by emissions and meteorology. It is used for determining the effect of emission reductions on a regional scale.

The data here is part of the ROM model (version 2.2) output for the summer 1987 run. Hourly ozone concentrations at 128 × 128 grid points (spacing of 18.5km).

Goal: Use Local LE's to characterize when ozone is more or less predictable.

Object-oriented ideas for data analysis in SPLUS

- Common operations can be applied to different types of data: *overloading*
- Using lists to bundle together and identify information as a single unit: *making an object*

Example of function overloading

```
> x<- c( 1,2,2,3)
> y<- c(-4, -5, 8,25)
> print(x)
 1 2 2 3
> print( y)
  -4  -5   8  25
> var( x)
 0.6666667
> var( x,y)
  9.666667
```

A FUNFITS Example in SPLUS

```
> plot( ozone$lon.lat)

> US( add=T)

> ozone.tps<- tps(ozone$x, ozone$y)

> summary( ozone.tps)

> plot(ozone.tps)

> surface(ozone.tps,'Fitted ozone surface')
```

```
Call:
tps(x = ozone$x, y = ozone$y)

 Number of Observations:                          20
 Degree of polynomial null space ( base model): 1
 Number of parameters in the null space         3
 Effective degrees of freedom:                    4.5
 Residual degrees of freedom:                     15.5
 Root Mean Square Error:                          4.047
 Pure Error:                                      NA
 RMSE (PRESS):                                    4.4594
 GCV (Prediction error)                           21.17
 Mutiple R-squared:                               0.3521
 Adjusted R-squared:                              0.07126
 Q-squared:                                       -0.1868
 PRESS:                                           397.7244
 Log10(lambda):                                   -0.034
 m, power, d, cost:                               2, 2, 4, 1
 Residuals:
    min   1st Q  median 3rd Q   max
 -6.712 -1.426 -0.4725 1.326 7.912
 Method for specifying the smoothing parameter is  GCV
```

Generic functions for surface fitting

- `predict, residuals, summary`
- `plot (diagnostic plots)`
- `surface`
- `predict.se`
- `lle ( Lyapunov exponents)`

Fitting the ROM series using `nlar`

```
> plot(ROM, xlab="hours", ylab="ozone in PPB")
> hours<- rep( 1:24,89)
> ROM.fit<- nlar( ROM, lags=c(1,2,24),
+       cov= hours, method='nnreg', k1=1,k2=4)
> summary( ROM.fit)
> plot( ROM.fit)
> ROM.lle<- lle( ROM.fit)
> summary( ROM.lle)
> plot( ROM.lle)
```

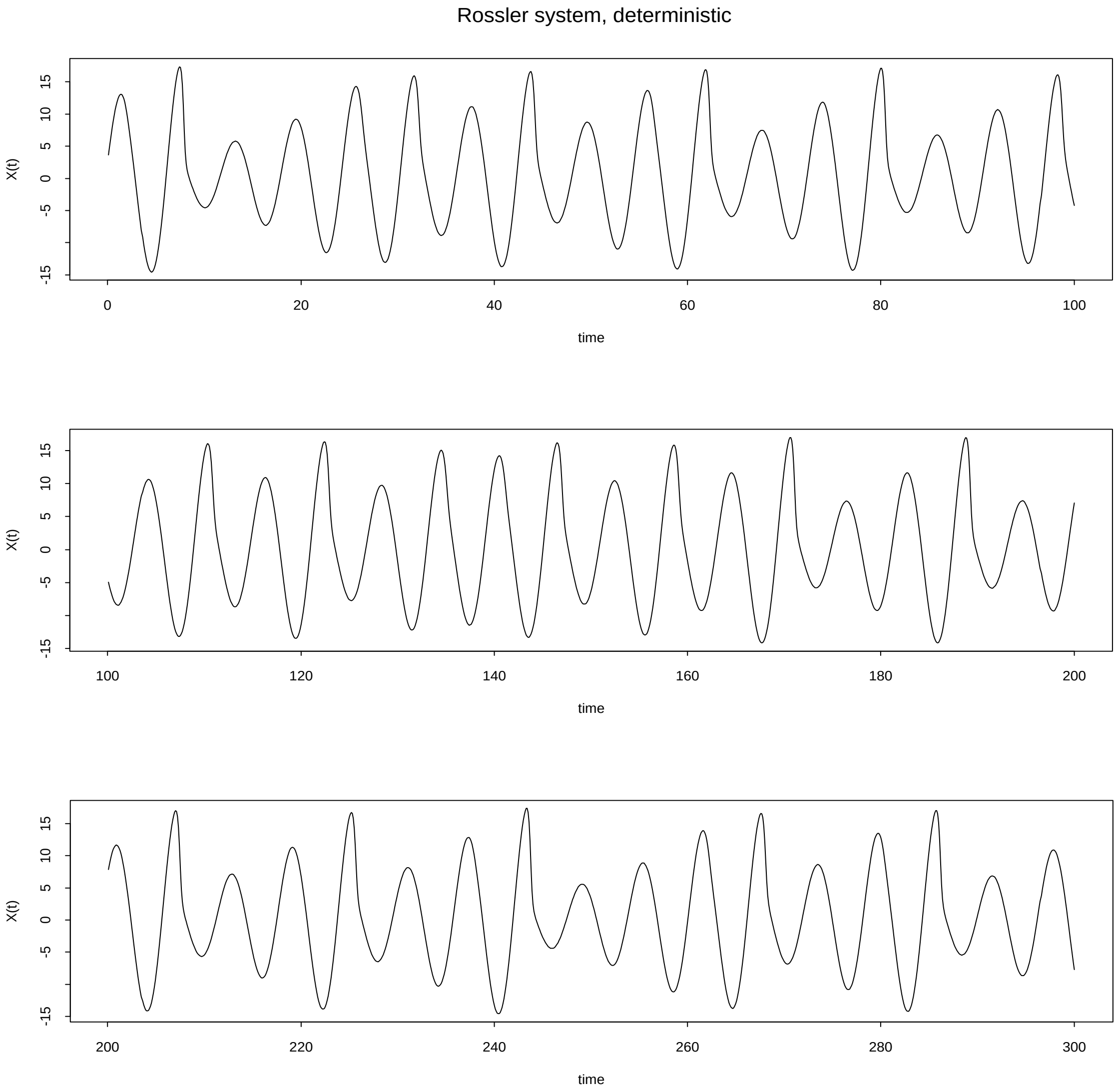


Figure 1: Time series plot of the Rossler X variable

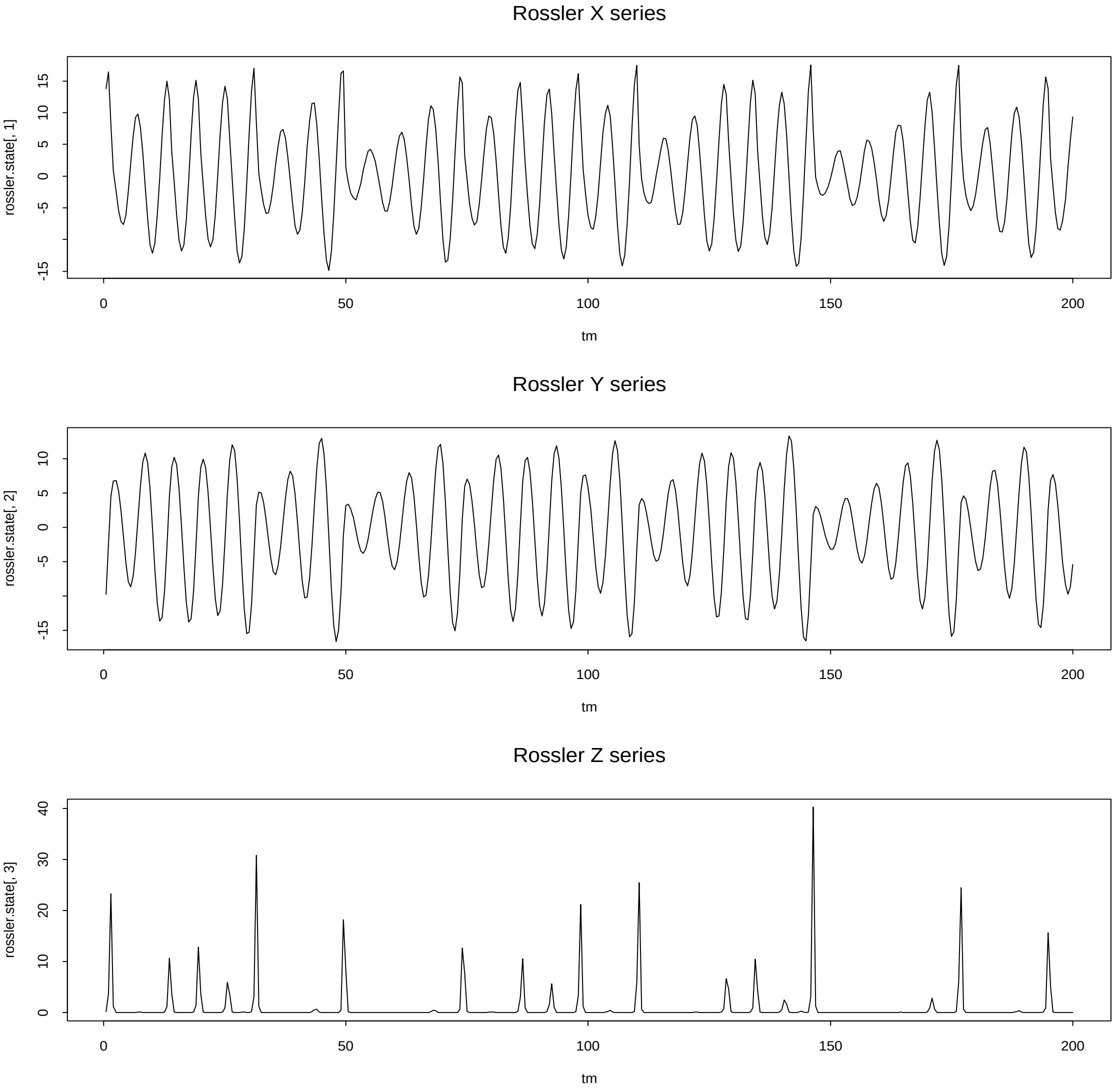
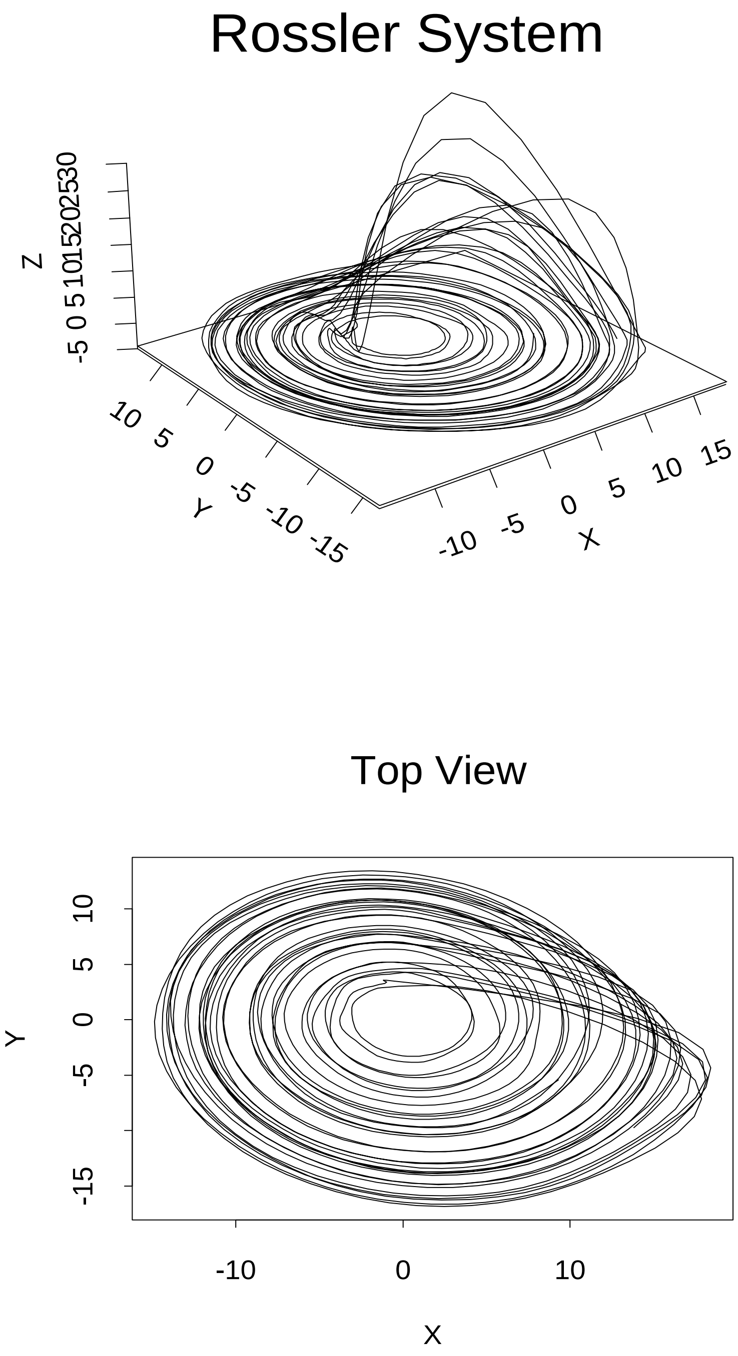


Figure 2: Time series plots of the three Rossler state variables. System simulated with 5% noise added



Rossler System

Top View

Figure 3: Perspective view of trajectory of the state vector

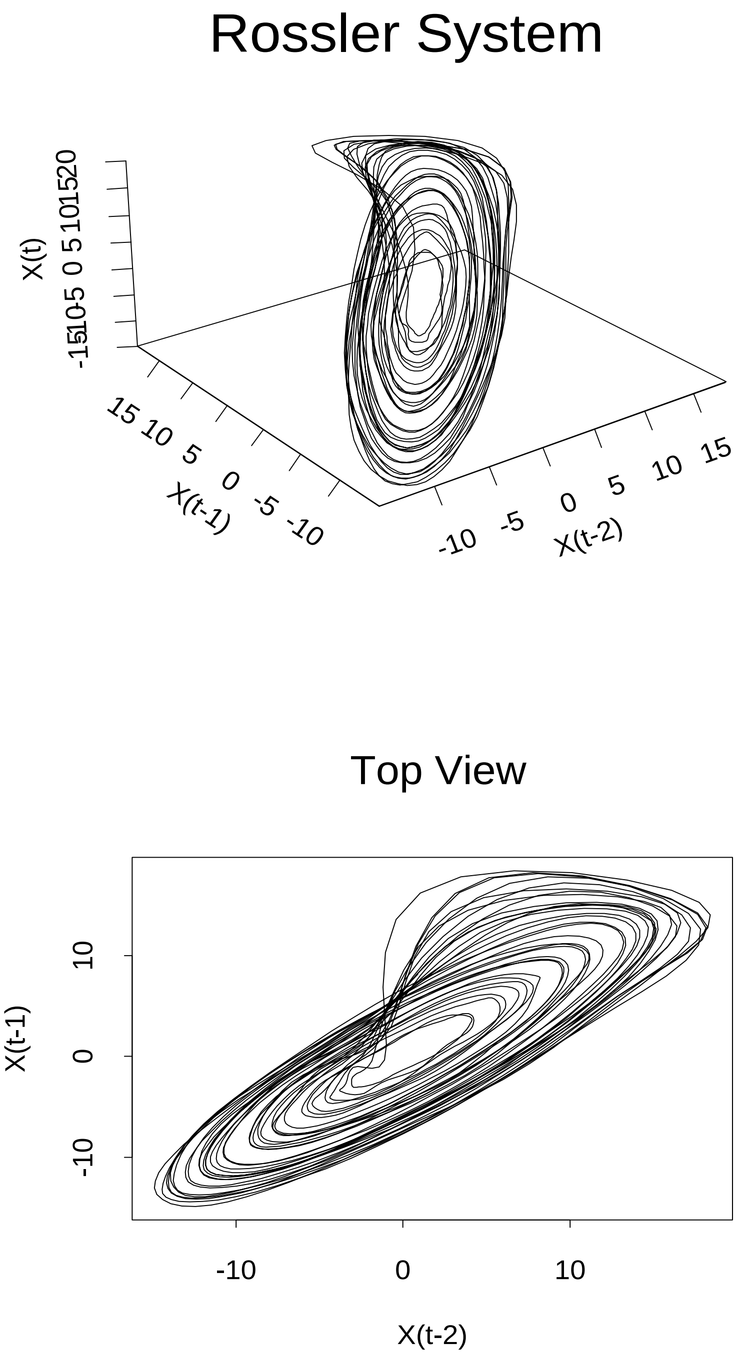


Figure 4: Perspective of the reconstructed state vector based on lags of just the X variable

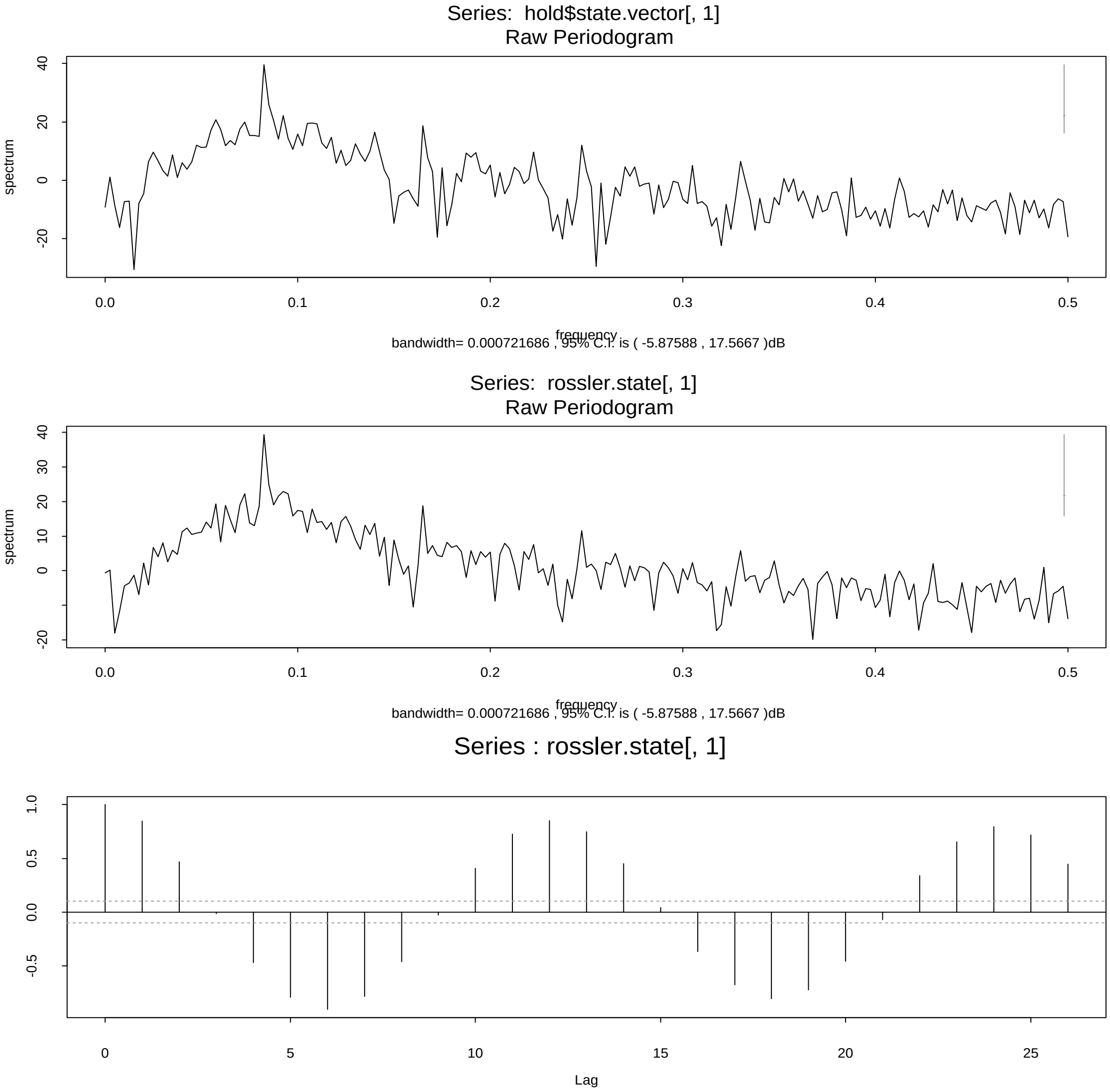


Figure 5: Spectral density of the Rossler X series top is exact, middle is with 5% noise. The last plot is the ACF for the noisy series.

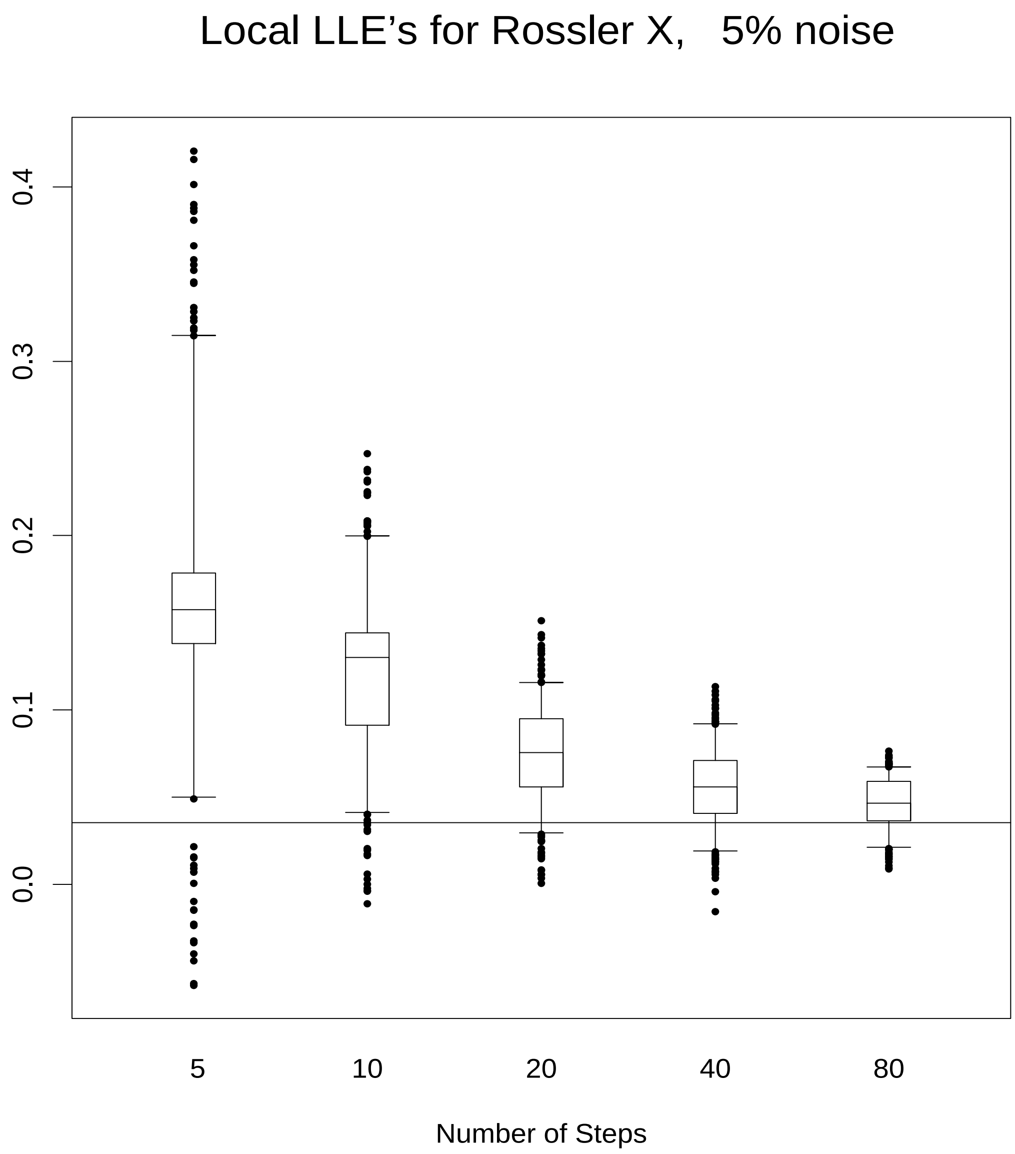


Figure 6: Boxplots for the local Lyapunov exponents estimated for the noisy Rossler X series. The global LE is estimated to be approximately .035 (horizontal line).

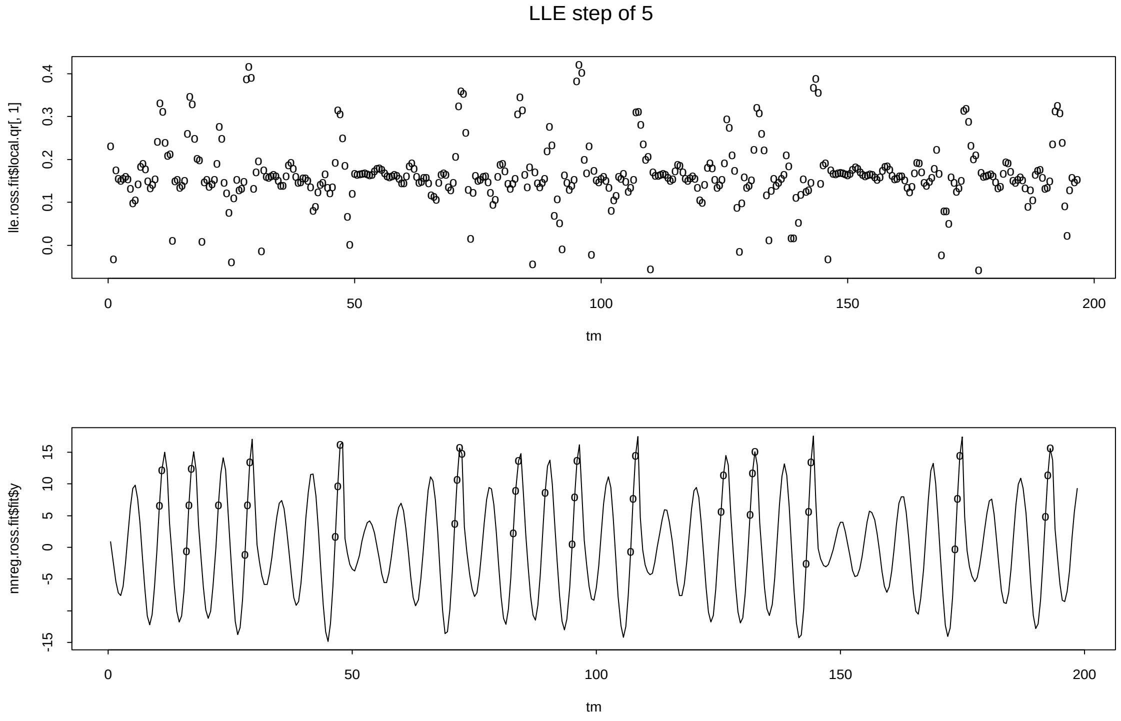


Figure 7: Time series plot of the LLE for the noisy Rossler system and the locations of the top 10% on the original X series.

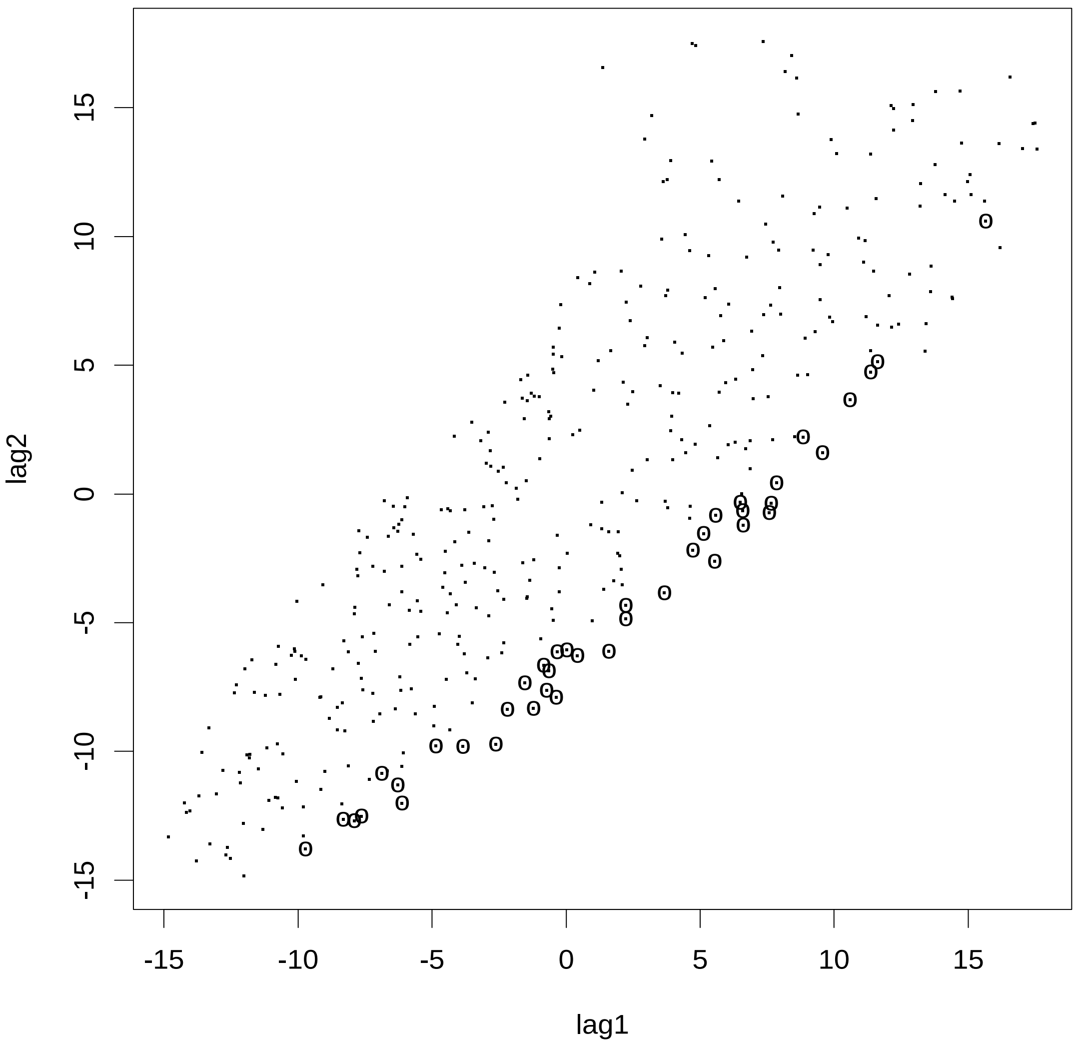


Figure 8: Locations of the top 10% LLE 5-step with respect to X and its lagged value.

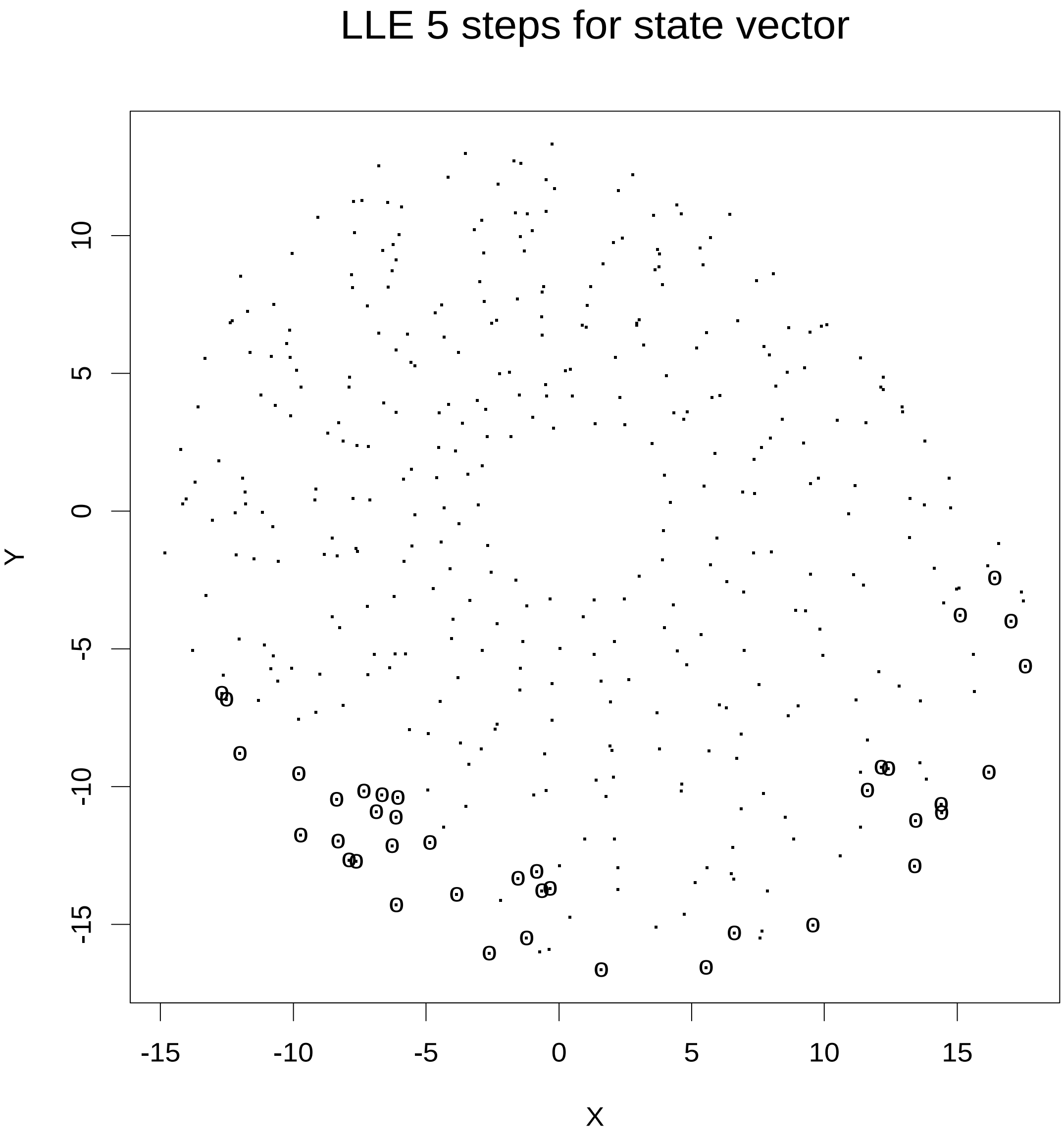


Figure 9: Locations of the top 10% LLE 5-step with respect to X and Y components of the state vector.

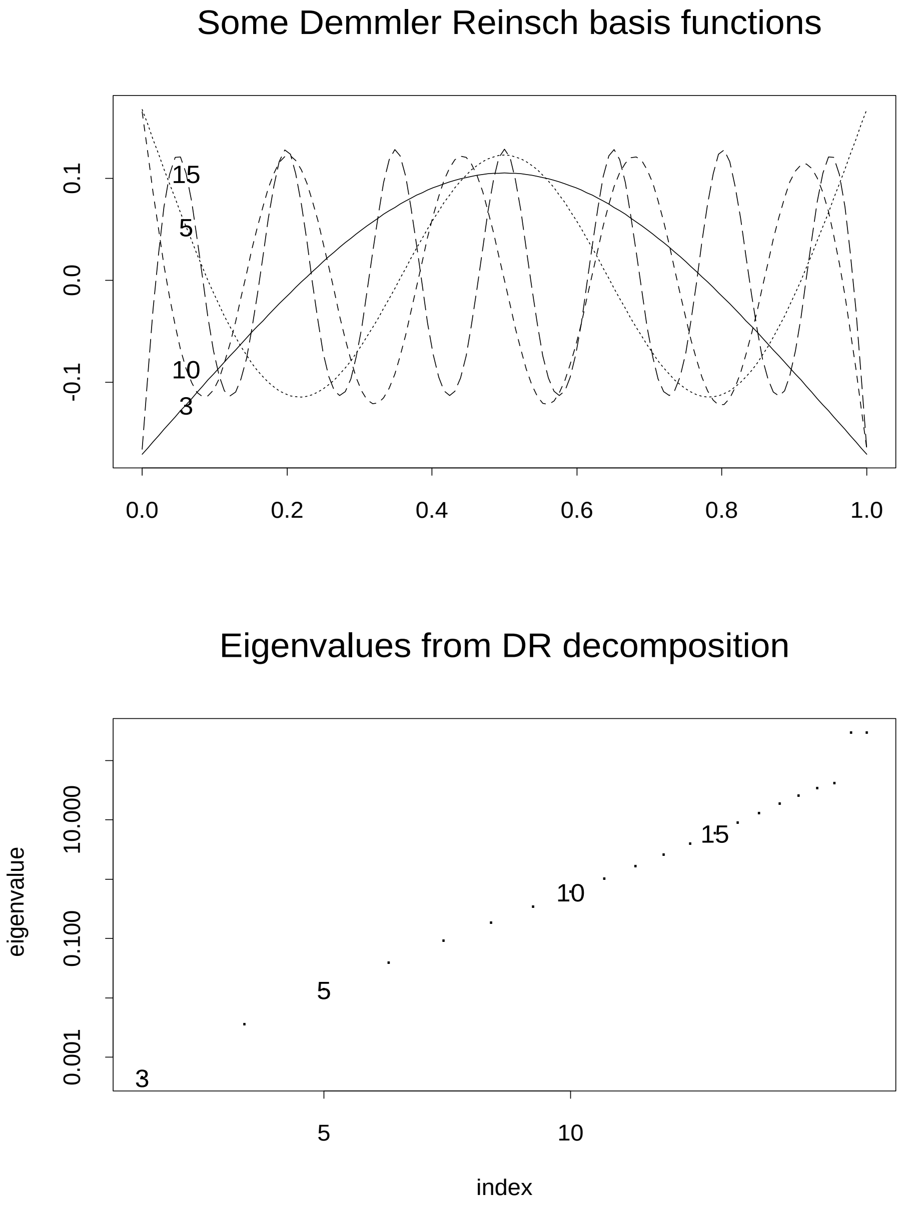


Figure 10: Illustration of several Demmler Reinsch basis functions and the associated eigenvalues for 20 equally spaced knots, 133 observations (as in the motorcycle data) and second divided differences ($k = 2$). The upper plot are $\{\psi_\nu\}$ for $\nu = (3, 5, 10, 15)$. The numerals identify the order of these basis functions and in the second plot identify the eigenvalues for these functions.

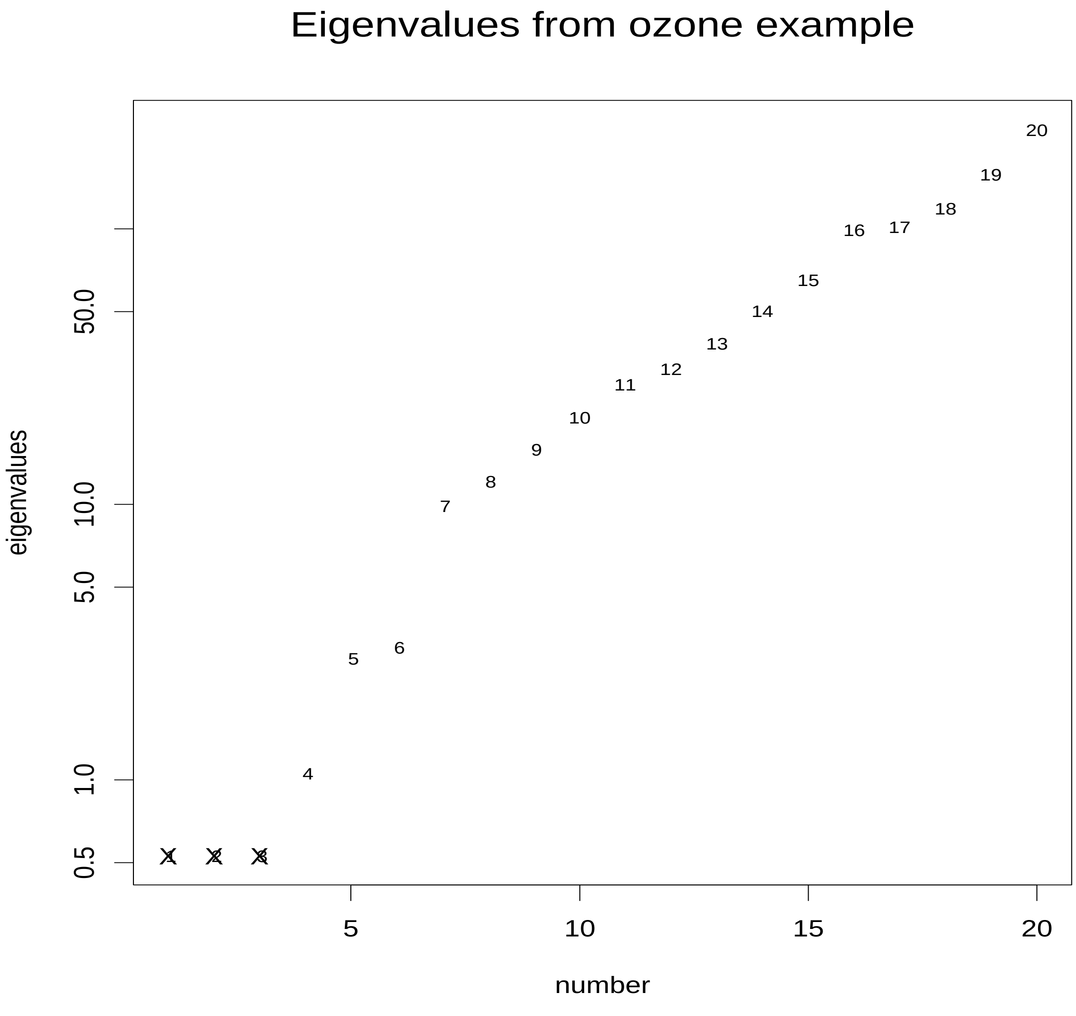


Figure 11: Eigenvalues for the DR decomposition for the a tps fit to 20 ozone locations

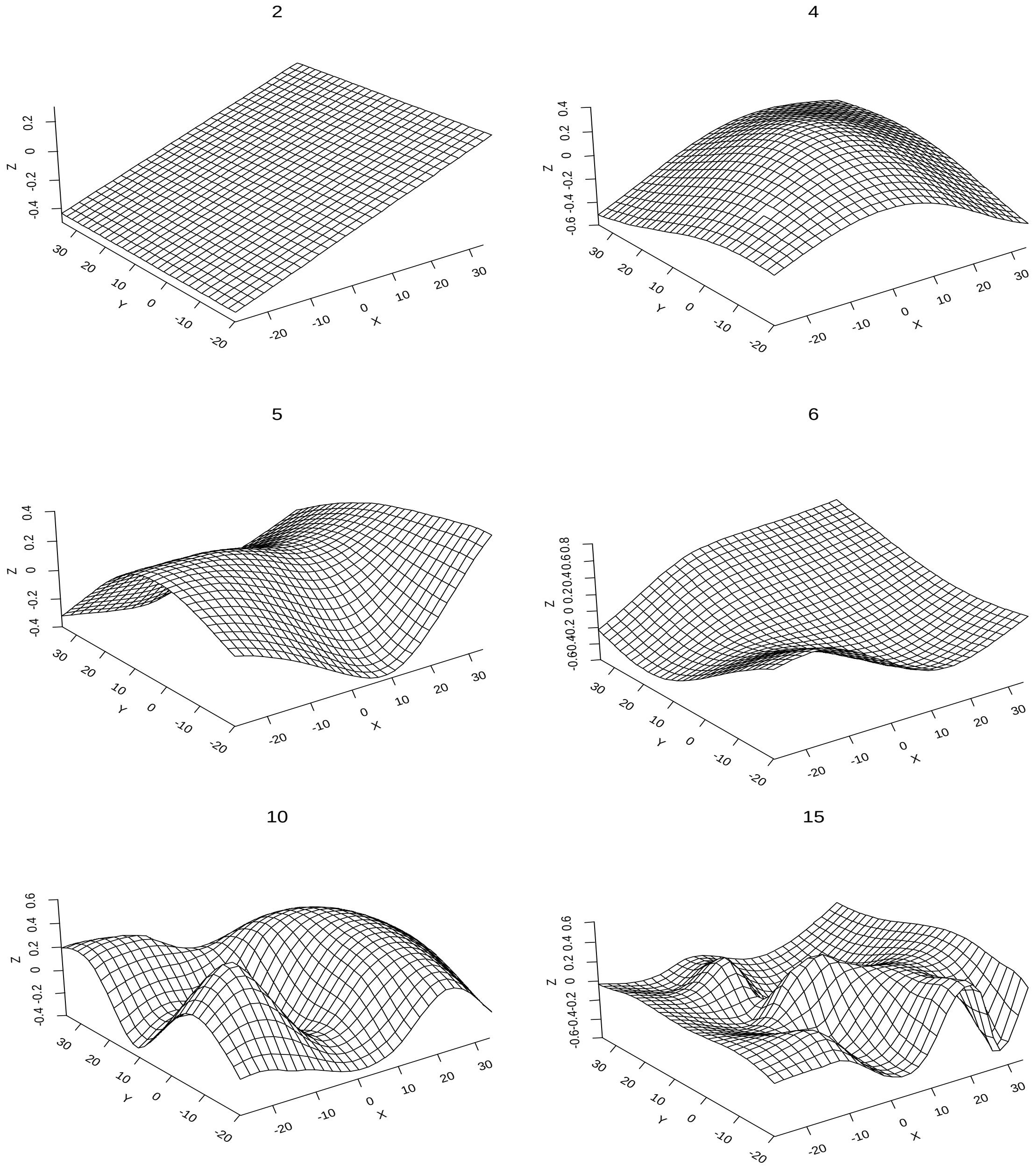


Figure 12: Some DR basis functions for the Chicago ozone example

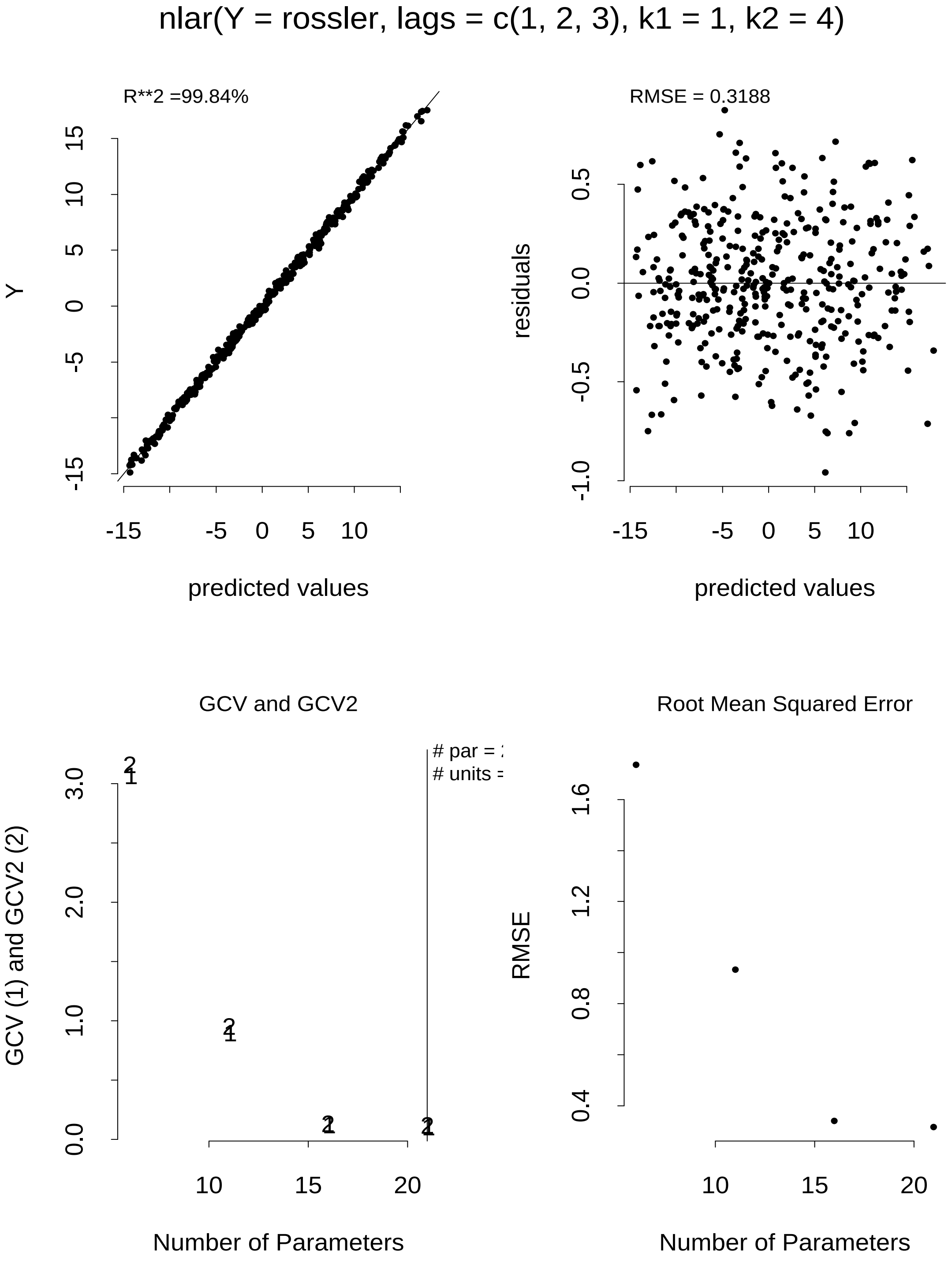
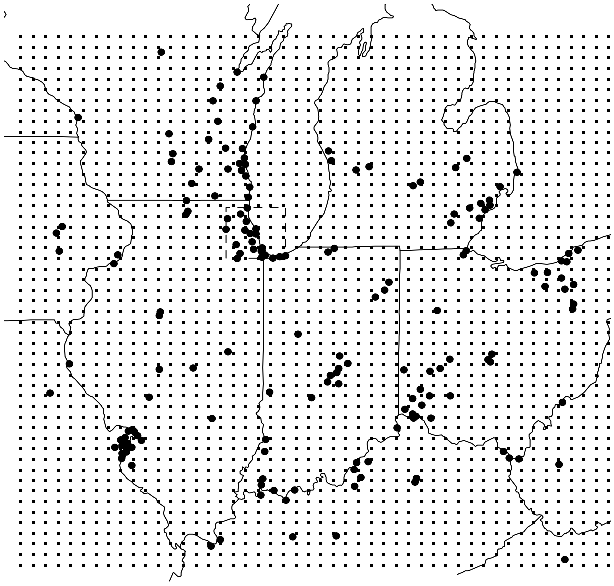
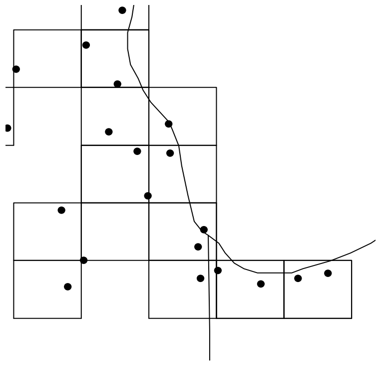


Figure 13: Summary of the fit to the Rossler data using a neural network model.



ROM study region



Chicago urban area

Figure 14: Subset of locations from output of the Regional Oxidant Model and an enlargement with the grid cells and monitoring locations for the Chicago urban area.

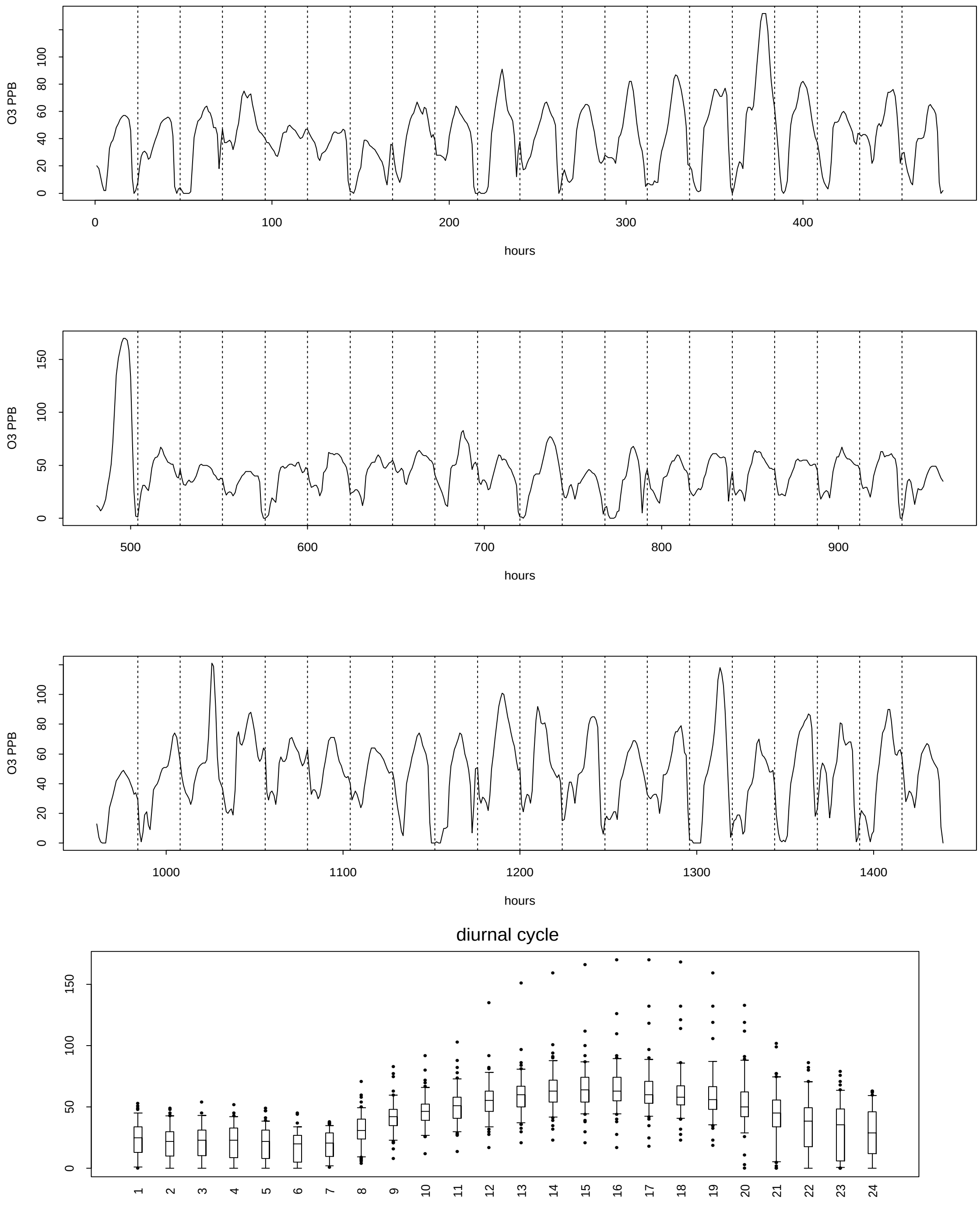


Figure 15: ROM output for a grid cell in the Chicago area.

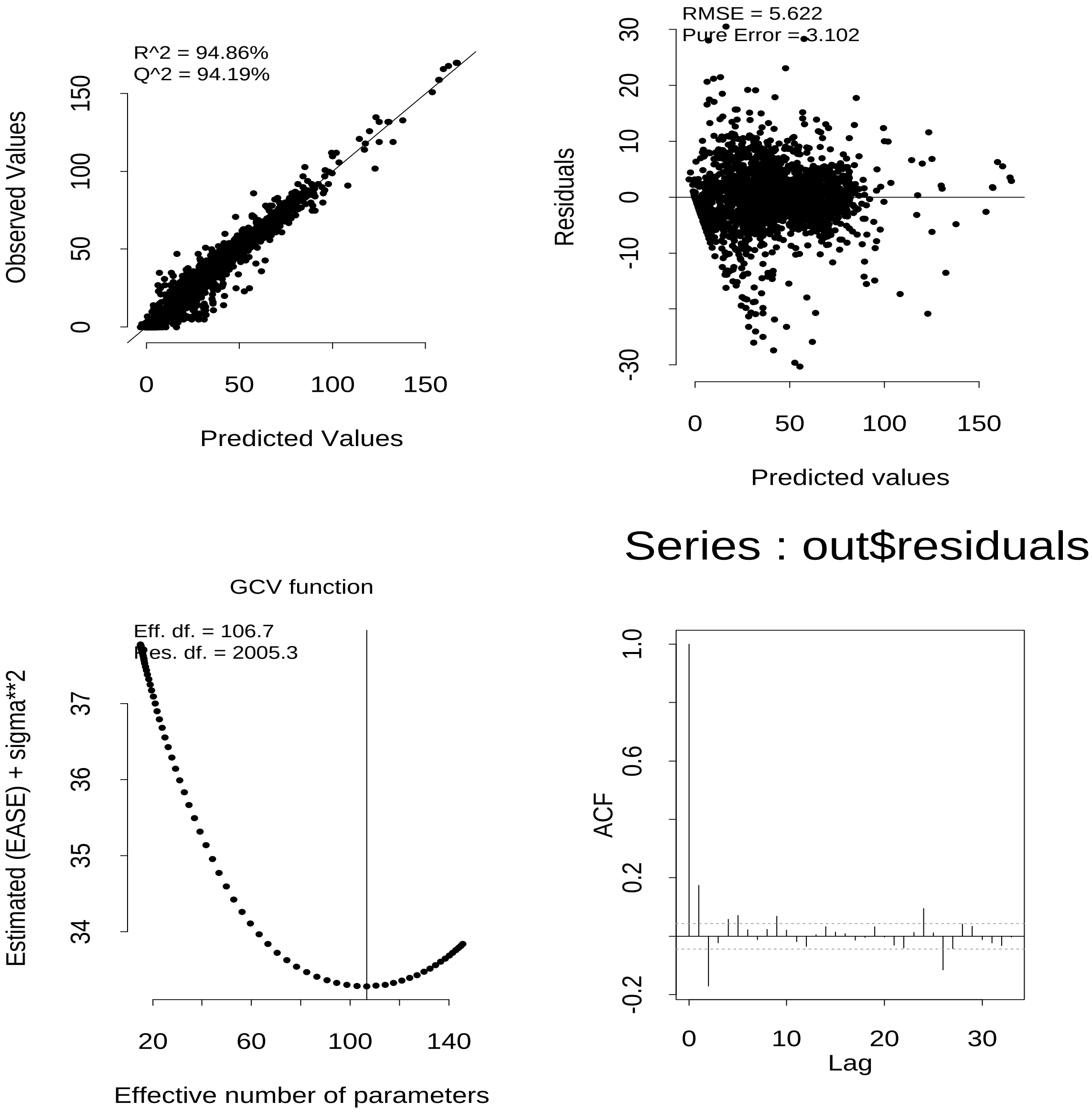


Figure 16: Plot indicating fit of a thin plate spline estimate to the ROM series. Model consists of 4 variables to explain current ozone: lagged values at 1, 2 and 24 hours plus a 24 hour clock.

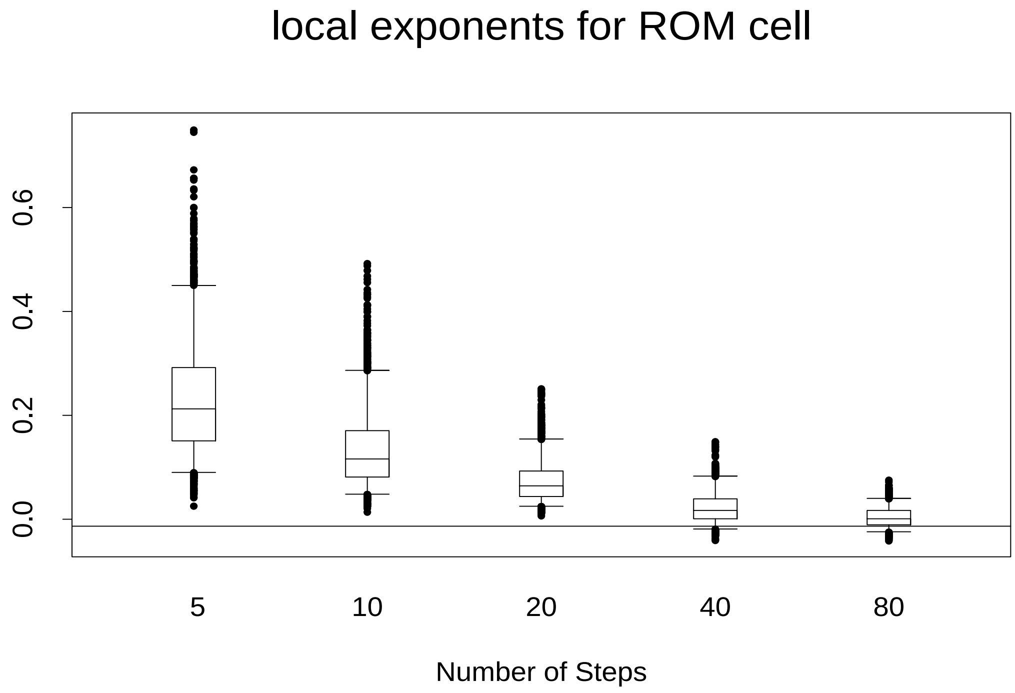


Figure 17: Distribution of LLE's for the ROM data