cts: An R Package for Continuous Time Autoregressive Models via Kalman Filter

Zhu Wang
UT Health San Antonio

Abstract

We describe an R package **cts** for fitting a modified form of continuous time autoregressive model, which can be particularly useful with unequally sampled time series. The estimation is based on the application of the Kalman filter. The paper provides the methods and algorithms implemented in the package, including parameter estimation, spectral analysis, forecasting, model checking and Kalman smoothing. The package contains R functions which interface underlying Fortran routines. The package is applied to geophysical and medical data for illustration.

Keywords: continuous time autoregressive model, state space model, Kalman filter, Kalman smoothing, R.

1. Introduction

The discrete time autoregressive model of order p, the $AR(p)$, is a widely used tool for modeling equally spaced time series data. It can be fitted in a straightforward and reliable manner and has the capacity to approximate a second order stationary process to any required degree of precision by choice of a suitably high order. Various information criteria such as the AIC (Akaike 1974) or the BIC (Schwarz 1978) can, in practice be used to determine a suitable order when fitting to a finite data set. This model is clearly not appropriate for irregularly sampled data, for which various authors have advocated the use of the continuous time autoregressive model of order p, the $CAR(p)$. For example Jones (1981) developed an estimation method for the $CAR(p)$ fitted to discretely sampled data through the application of the Kalman filter. However, the $CAR(p)$ model does not share with its discrete counterpart the same capacity to approximate any second order stationary continuous process. Belcher, Hampton, and Tunnicliffe Wilson (1994) modified the parameterization and structure of the $CAR(p)$ model in such a way that this general capacity for approximation was restored, requiring only a relatively minor modification of the estimation procedure of Jones. More recently, see Tunnicliffe Wilson and Morton (2004), this modified $CAR(p)$ model has been named the $CZAR(p)$ model and expressed in autoregressive form using the generalized continuous time shift operator with the alternative parameterization appearing in a natural manner. Wang, Woodward, and Gray (2009) utilized the methods in Belcher *et al.* (1994) for fitting time varying nonstationary models. Since the technical details for the Kalman filter on CAR models are scattered in the literature, we give a thorough presentation in this paper, which provides a foundation for the implementations in the R (R Development Core Team

2013) package **cts** (Tunnicliffe Wilson and Wang 2013) for fitting CAR models with discrete data. The **cts** package contains typical time series applications including spectral estimation, forecasting, diagnostics, smoothing and signal extraction. The application is focused on unequally spaced data although the techniques can be applied to equally spaced data as well. The paper is organized as follows. Section 2 summarizes the methods from which the **cts** package was developed. Section 3 outlines the implementations in the package. Section 4 illustrates the capabilities of **cts** with two data sets. Finally, Section 5 concludes the paper.

2. Methods

2.1. The CAR and CZAR model

Suppose we have data x_k observed on time t_k for $k = 1, 2, ..., n$. We assume noise-affected observations $x_k = y(t_k) + \eta_k$ and $y(t_k)$ follows a p-th order continuous time autoregressive process $Y(t)$ satisfying

$$Y^{(p)}(t) + \alpha_1 Y^{(p-1)}(t) + \ldots + \alpha_{p-1} Y^{(1)}(t) + \alpha_p Y(t) = \epsilon(t), \tag{1}$$

where $Y^{(i)}(t)$ is the ith derivative of $Y(t)$ and $\epsilon(t)$ is the formal derivative of a Brownian process $B(t)$ with variance parameter $\sigma^2 = \text{var}\{B(t+1) - B(t)\}$. In addition, it will be assumed that η_k is a normally distributed random variable representing observational error, uncorrelated with $\epsilon(t)$, and $E(\eta_k) = 0; E(\eta_j \eta_k) = 0, \text{for } j \neq k; E(\eta_k^2) = \gamma \sigma^2$.

The operator notation of model (1) is $\alpha(D)Y(t) = \epsilon(t)$ where

$$\alpha(D) = D^p + \alpha_1 D^{p-1} + ... + \alpha_{p-1} D + \alpha_p, \tag{2}$$

where D is the derivative operator. The corresponding characteristic equation is then given by

$$\alpha(s) = s^p + \alpha_1 s^{p-1} + ... + \alpha_{p-1} s + \alpha_s = 0. \tag{3}$$

To assure the stability of the model, a parameterization was constructed on the zeros $r_1, ..., r_p$ of $\alpha(s)$ (Jones 1981), i.e.,

$$\alpha(s) = \prod_{i=1}^{p} (s - r_i). \tag{4}$$

The model in the **cts** package follows the modified structure (Belcher *et al.* 1994):

$$\alpha(D)Y(t) = (1 + D/\kappa)^{p-1} \epsilon(t), \tag{5}$$

with scaling parameter $\kappa > 0$. This introduces a prescribed moving average operator of order $p - 1$ into the model, which makes the model selection convenient along with other theoretic benefits described in Belcher *et al.* (1994). In practice model (5) has been found to fit data quite well without the need for an observation error term (Belcher *et al.* 1994). Indeed, for large orders p, the observation error variance may become confounded with the parameters of model (5), thus leading to some difficulty in estimation (Belcher *et al.* 1994). As argued by these authors, model 5 can be sufficient unless there is a substantial variability for measurements at the same or very nearly coincident times.

The power spectrum of the pth order continuous process (5) is defined by

$$G_y(f) = 2\pi\sigma^2 \left| \frac{(1 + \mathrm{i}2\pi f/\kappa)^{p-1}}{\alpha(\mathrm{i}2\pi f)} \right|^2. \tag{6}$$

This may also be expressed in a form which defines the alternative set of parameters $\phi_1, ..., \phi_p$, by using a transformation $g = \arctan(2\pi f/\kappa)/\pi$ of the frequency f:

$$G_y(f) = \frac{\sigma^2/\kappa^{2p-2}}{\kappa^2 + (2\pi f)^2} \left| \frac{\phi(-1)}{\phi\{\exp(2\pi i g)\}} \right|^2 2\pi,$$

where

$$\phi(z) = 1 + \phi_1 z + ... + \phi_p z^p.$$

The link between the α and ϕ parameters is given in Belcher *et al.* (1994) (9) and the line which follows. The new parameter space is identical to that of the stationary discrete time autoregressive model. The CZAR model of Tunnicliffe Wilson and Morton is identical except that the right-hand side of (5) is modified to $(\kappa + D)^{p-1}\epsilon(t)$ so that σ^2/κ^{2p-2} is re-scaled to σ^2. The system frequencies are determined by the roots of (4). In fact, the representation of (4) breaks a pth order autoregressive operator into its irreducible linear and quadratic factors that have complex zeros. A quadratic factor $(s - r_{2k-1})(s - r_{2k})$ with complex zeros is associated with a component of the data having the nature of a stochastic cycle, with approximate frequency given by $f = \frac{|\Im(r_{2k})|}{2\pi}$, where $|\Im(r_{2k})|$ is the absolute value of the imaginary part of r_{2k}. This will be reflected as a component of the autocorrelations with the same frequency, decaying in amplitude with a rate equal to the absolute value of the real part of the same zero. The model can represent a strong cycle in the data if this decay rate is very low. A linear factor with zero at r_k is associated with a component of the data having the nature of a first order autoregression with autocorrelation function exponentially decaying at the rate r_k. If r_k is very large, this can give the appearance of a white noise component.

2.2. Kalman filtering

This section deals with the details related to applying the Kalman filter to estimate the parameters of model (5), following Jones (1981) and Belcher *et al.* (1994). To apply the Kalman filter, it is required to rewrite model (5) to a state space form, which may be found in Wiberg (1971). Let the unobservable state vector $\theta(t) = (z(t), z'(t), z''(t)..., z^{(p-1)}(t))^\top$ and θ' the first derivative of $\theta(t)$. The state equation is then given by

$$\theta' = A\theta + R\epsilon, \tag{7}$$

where

$$A = \begin{bmatrix} 0 & 1 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & & & \\ 0 & 0 & \ldots & 1 \\ -\alpha_p & -\alpha_{p-1} & \ldots & -\alpha_1 \end{bmatrix} \tag{8}$$

and

$$R^\top = \begin{bmatrix} 0 & 0 & \ldots & 1 \end{bmatrix}. \tag{9}$$

The observation equation is given by

$$x_k = H\theta(t_k) + \eta_k, \tag{10}$$

where the elements of the $1 \times p$ vector H are given by

$$H_i = \binom{p-1}{i-1}/\kappa^{i-1} \qquad i = 1, ..., p. \tag{11}$$

Suppose that A can be diagonalized by $A = U\mathbf{D}U^{-1}$, where

$$U = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ r_1 & r_2 & \ldots & r_p \\ r_1^2 & r_2^2 & \ldots & r_p^2 \\ \vdots & & & \\ r_1^{p-1} & r_2^{p-1} & \ldots & r_p^{p-1} \end{bmatrix}, \tag{12}$$

$r_1, r_2, ..., r_p$ are the roots of $\alpha(s)$, and $\mathbf{D}$ is a diagonal matrix with these roots as its diagonal elements. In this case, we let $\theta = U\psi$, and the state equation becomes

$$\psi' = \mathbf{D}\psi + J\epsilon, \tag{13}$$

where $J = U^{-1}R$. Consequently, the observation equation becomes

$$x_k = G\psi(t_k) + \eta_k \tag{14}$$

where $G = HU$. The necessary and sufficient condition for the diagonalization of A is that A has distinct eigenvalues. The diagonal form not only provides computational efficiency, but also provides an interpretation of unobserved components. The evaluation of $T_{\theta,k} = e^{A\delta_k}$ (standard form) is required where $\delta_k = t_k - t_{k-1}$. For a review of computations related to the exponential of a matrix, see Moler and Loan (2003). For the diagonal form, $T_{\psi,k} = e^{\mathbf{D}\delta_k}$ is diagonal with elements $e^{r_i\delta_k}$. When a diagonal form is not available, a numerical matrix exponential evaluation is needed.

To start the Kalman filter recursions, initial conditions are in demand. For a stationary model, the unconditional covariance matrix of state vector $\theta(t)$ is known (Doob 1953) and used in Jones (1981) and Harvey (1990, §9.1). The initial state for both standard and diagonalized version can be set as $\theta_0 = 0$ and $\psi_0 = 0$, respectively. The stationary covariance matrix Q satisfies

$$Q = \sigma^2 \int_0^\infty e^{As} RR^\top e^{A^\top s} ds. \tag{15}$$

When A can be diagonalized, it is straightforward to show that

$$Q_{\psi_{i,j}} = -\sigma^2 J_i \bar{J}_j / (r_i + \bar{r}_j), \tag{16}$$

where $\bar{J}_j$ and $\bar{r}_j$ are complex conjugates of J_j and r_j, respectively.

The scale parameter κ can be chosen approximately as the reciprocal of the mean time between observations. The algorithm of Kalman filter for the diagonal form is presented below. Starting with an initial stationary state vector of $\psi_0 = \psi(0|0) = 0$ and the initial stationary state covariance matrix Q_ψ (16), the recursion proceeds as follows:

1. Predict the state. Let

$$T_{\psi,k} = e^{\mathbf{D}\delta_k} \tag{17}$$

   a diagonal matrix, then

$$\psi(t_k|t_{k-1}) = T_{\psi,k}\psi(t_{k-1}|t_{k-1}). \tag{18}$$

2. Calculate the covariance matrix of this prediction:

$$P_\psi(t_k|t_{k-1}) = T_{\psi,k}(P_\psi(t_{k-1}|t_{k-1}) - Q_\psi)\bar{T}_{\psi,k} + Q_\psi. \tag{19}$$

3. Predict the observation at time t_k:

$$x_\psi(t_k|t_{k-1}) = G\psi(t_k|t_{k-1}) \tag{20}$$

4. Calculate the innovation:

$$v_\psi(t_k) = x_\psi(t_k) - x_\psi(t_k|t_{k-1}) \tag{21}$$

   and variance

$$F_\psi(t_k) = GP_\psi(t_k|t_{k-1})\bar{G}^\top + V \tag{22}$$

5. Update the estimate of the state vector:

$$\psi(t_k|t_k) = \psi(t_k|t_{k-1}) + P_\psi(t_k|t_{k-1})\bar{G}^\top F_\psi^{-1}(t_k)v_\psi(t_k) \tag{23}$$

6. Update the covariance matrix:

$$P_\psi(t_k|t_k) = P_\psi(t_k|t_{k-1}) - P_\psi(t_k|t_{k-1})\bar{G}^\top F_\psi^{-1}(t_k)G\bar{P}_\psi^\top(t_k|t_{k-1}) \tag{24}$$

7. The unknown scale factor σ^2 can be concentrated out by letting $\sigma^2 = 1$ temporally. -2 log-likelihood is calculated by

$$\log L_{\psi,c} = \sum_{t=1}^n \log F_\psi(t_k) + n \log \sum_{t=1}^n v_\psi^2(t_k)/F_\psi(t_k) \tag{25}$$

   The log-likelihood function (25) thus can be evaluated by a recursive application of the Kalman filter, and a nonlinear numerical optimization routine is then used to determine the parameter estimation. The unknown scale factor can then be estimated by

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^n v_\psi^2(t_k)/F_\psi(t_k). \tag{26}$$

When a diagonal form is not stable, a standard form Kalman filter recursion may be found in Belcher *et al.* (1994) or Wang (2004). However the computational load is reduced dramatically with the diagonal form since matrix $\mathbf{D}$ is diagonal.

When the nonlinear optimization is successfully completed, in addition to the maximum likelihood estimation of the parameters and error variances, the Kalman filter returns the optimal estimate of the state and the state covariance matrix at the last time point. The

forecasting of the state, state covariance matrix and observation can be continued into future desired time points using equations from (17) to (20).

2.3. Model selection

To identify a model order, Belcher *et al.* (1994) proposed a strategy corresponding to the reparameterization. Start with a large order model, and obtain the parameter vector ϕ and its covariance matrix V_ϕ, we then make a Cholesky decomposition such that $V_\phi^{-1} = L_\phi L_\phi^\top$ where L_ϕ is a lower triangular matrix, and define the vector $t_\phi = L_\phi^\top \phi$ and construct the sequence $AIC_d = -\sum_{i=1}^{d} t_{\phi,i}^2 + 2d$ for $d = 1, ..., p$. The index of the minimum value of AIC_d suggests a preferred model order. In addition, if the true model order p is less than the large value used for model estimation, then for $i > p$ the t-statistics may be treated as normal-distributed variables, so that the deviation from their true values of 0 will be small. However, Belcher et al used a 33MHz maths co-processor, and with the speed of present day computers the best practice is to compute the classical AIC or BIC (Akaike 1974; Schwarz 1978) by fitting the models of increasing order p to the series. The AIC is defined as $n \log SS(p) + 2p$ and BIC is defined as $n \log SS(p) + p \log(p)$ where SS is the sum of squares function given by Belcher *et al.* (1994) equation 15. The AIC and BIC can be easily modified if an additional mean value of the series is estimated. The package **cts** has implemented the relevant functions for model selection and a data example will be illustrated.

2.4. Diagnostics

The assumptions underlying the model (7) and (10) are that the disturbances $\epsilon(t)$ and η_k are normally distributed and serially independent with constant variances. Based on these assumptions, the standardized one-step forecast errors

$$e(t_k) = v(t_k)/\sqrt{F(t_k)} \qquad k = 1, ..., n \tag{27}$$

are also normally distributed and serially independent with unit variance. Hence, in addition to inspection of time plot, the QQ-normal plot can be used to visualize the ordered residuals against their theoretical quantiles. For a white noise sequence, the sample autocorrelations are approximately independently and normally distributed with zero means and variances $1/n$. Note that for a purely random series, the cumulative periodogram should follow along a line $y = 2x$ where x is frequency. A standard portmanteau test statistic for serial correlation, such as the Ljung-Box statistic, can be used as well. The proposed calculation of the scaled innovation is frequently done in classical discrete-time models. This way a sequence of n numbers is calculated, and the auto-correlation and discrete-time spectrum of these numbers are calculated, i.e. the time between observations does not enter these calculations.

2.5. Kalman smoothing

For a structural time series model, it is often of interest to estimate the unobserved components at all points in the sample. Estimation of smoothed trend and cyclical components provides an example. The purpose of smoothing at time t is to find the expected value of the state vector, conditional on the information made available after time t. In this section, a fixed-interval smoothing algorithm (Harvey 1990, §3.6.2) is implemented with modifications for the model considered, though a more efficient approach is possible, see the discussion in Durbin and Koopman (2001, §4.3). Estimating unobserved components relies on the diagonal form which provides the associated structure with the corresponding roots $r_1, ... r_p$. The smoothing state and covariance matrix are given by

$$\psi_s(t_k|t_n) = \psi(t_k|t_k) + P^*(t_k)(\psi_s(t_{k+1}|t_n) - \psi(t_{k+1}|t_k)) \tag{28}$$
$$P_s(t_k|t_n) = P(t_k|t_k) + P^*(t_k)(P_s(t_{k+1}|t_n) - P(t_{k+1}|t_k))\bar{P}^*(t_k) \tag{29}$$

where

$$P^*(t_k) = P(t_k|t_k)\bar{T}_{\psi,k+1}P^{-1}(t_{k+1}|t_k) \tag{30}$$

and $T_{\psi,k+1} = e^{\mathbf{D}(t_{k+2}-t_{k+1})}$, and $\bar{T}_{\psi,k+1}$ and $\bar{P}(t_k|t_k)$ are complex conjugates. To start the recursion, the initial values are given by $\psi_s(t_n|t_n) = \psi(t_n|t_n)$ and $P_s(t_n|t_n) = P(t_n|t_n)$. The observed value x_k, in the absence of measurement error, is the sum of contributions from the diagonalized state variables ψ, i.e., $x_k = \sum_j G_j\psi_j(t_k)$. Therefore, the original data may be partitioned, as in Jenkins and Watts (1968, §7.3.5). Any pair of two complex conjugate zeros of (4) is associated with two corresponding state variables whose combined contribution to x_k represents a source of diurnal variation. The real zeros correspond to exponential decay. If a real zero is very large, this can provide an appearance of white noise component. Hence, the contributions $G_j\psi_j$ at every time point can be estimated from all the data using the Kalman smoother as described above.

3. Implementation

The `cts` package utilizes the Fortran program developed by the authors of Belcher *et al.* (1994), with substantial additional Fortran and R routines. In this process, two Fortran subroutines in Belcher *et al.* (1994) have to be substituted since they belong to commercial NAG Fortran Library, developed by the Numerical Algorithms Group. One subroutine was to compute the approximate solution of a set of complex linear equations with multiple right-hand sides, using an Lower-Upper LU factorization with partial pivoting. Another subroutine was to find all roots of a real polynomial equation, using a variant of Laguerre's Method. In the `cts` package, these subroutines have been replaced by their public available counterparts in the LAPACK & BLAS Fortran Library. All the Fortran programs were written in double precision.

If a constant term is estimated by the default setting `ccv="CTES"` in the `car` function, it represents the mean μ, and the model is formulated in terms of $(x - \mu)$. In the setting (`ccv="MNCT"`), the series is not actually mean corrected, but the sample mean is just used to estimate μ in the above model formulation. Several supporting R functions are available in the `cts` package that extract or calculate useful statistics based on the fitted CAR model, such as model summary, predicted values and model spectrum. In particular, the function `car` returns

objects of class `car`, for which the following methods are available: `print, summary, plot, predict, AIC, tsdiag, spectrum, kalsmo`. A detailed description of these functions is available in the online help files. Here a brief introduction will be given and the usage will be illustrated in the next section. Specifying `trace=TRUE` in `car_control` could trigger annotated printout of information during the fitting process and major results for the fitted model. The model fitting results can be graphical displayed with `plot` function. With argument `type` equal to `"spec"`, `"pred"` and `"diag"`, respectively, a figure can be plotted for spectrum, predicted values and model diagnostic checking, respectively. Three types of prediction exist: forecast past the end, forecast last L-step, forecast last L-step update. This can be achieved by invoking argument `fty=1, 2, 3`, respectively. For instance, `ctrl=car_control(fty=1, n.ahead=10)` can predict 10 steps past the end. Function `AIC` can generate both t-statistic and AIC values following section 2.3. Function `tsdiag` follows section 2.4 to provide model diagnostic checking. Indeed, this function provides the backbone for function `plot` with argument `type="diag"`. Function `kalsmo` implements the Kalman smoothing described in section 2.5.

The source version of the **cts** package is freely available from the Comprehensive R Archive Network (http://CRAN.R-project.org). The reader can install the package directly from the R prompt via

```
R> install.packages("cts")
```

All analyses presented below are contained in a package vignette. The rendered output of the analyses is available by the R-command

```
R> library("cts")
R> vignette("kf", package = "cts")
```

To reproduce the analyses, one can invoke the R code

```
R> edit(vignette("kf", package = "cts"))
```

4. Data examples

Two data examples in Belcher *et al.* (1994) are used to illustrate the capabilities of **cts**. A detailed description of the data can be found in the original paper. Since some analysis here reproduces the results in Belcher *et al.* (1994), we also ignore a lengthy discussion for brevity. These analyses were done using R version 3.0.0.

4.1. Geophysical application

Belcher *et al.* (1994) analyzed 164 measurements of relative abundance of an oxygen isotope in an ocean core. These are unequally spaced time points with an average of separation of 2000 years. Unequally spaced tick marks indicate the corresponding irregularly sampled times in Figure 1.

```
R> library("cts")
R> data("V22174")
```

```
R> plot(V22174, type = "l", xlab = "Time in kiloyears",
+       ylab = "")
R> rug(V22174[, 1], col = "red")
```
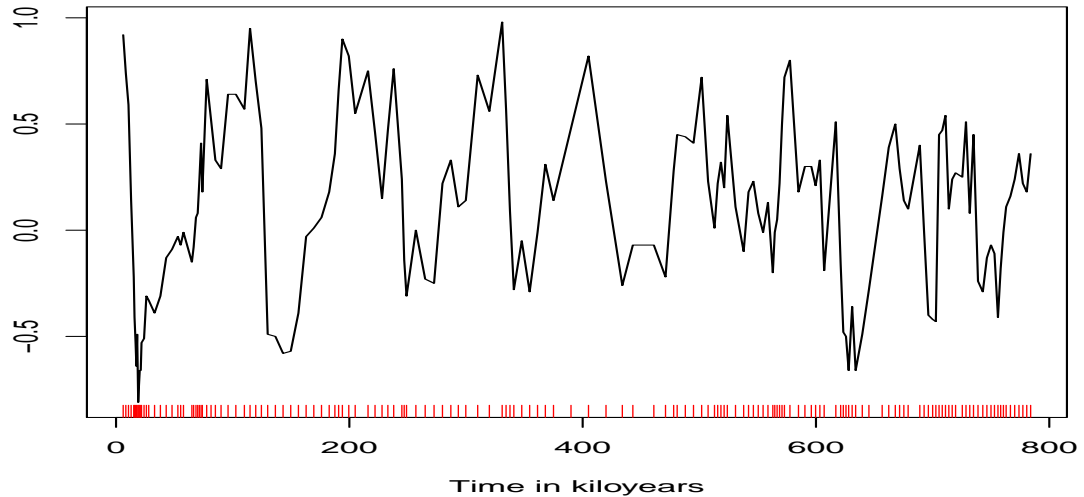

Figure 1: Oxygen isotope series.

We first fit a model of order 14 to the data, following Belcher *et al.* (1994). The scale parameter is chosen to be 0.2 as well. The estimation algorithm converges rather quickly as demonstrated in the following printout, which shows the sum of squares and the value of ϕ_{14} at each iteration. The results are similar to Table 1 of Belcher *et al.* (1994), which took 30 minutes on a PC386/387 machine to carry out the computing. These authors expected that simple improvements to the program's code could substantially speed up the procedure. Despite that the current **cts** package has no intent to accomplish such a task, running the above `car` function took only 0.3 second, on an ordinary desktop PC (Intel Core 2 CPU, 1.86 GHz). Such a dramatic efficiency improvement is unlikely driven by software change, but by hardware advancement in the last 20 years.

```
R> V22174.car14 <- car(V22174, scale = 0.2, order = 14)
R> tab1 <- cbind(V22174.car14$tnit, V22174.car14$ss, V22174.car14$bit[,
+       14])
R> colnames(tab1) <- c("Iteration", "Sum of Squares", "phi_14")

R> print(as.data.frame(round(tab1, 5)), row.names = FALSE,
+       print.gap = 8)
```

Iteration	Sum of Squares	phi_14
0	12.92737	0.00000
1	8.32272	-0.16453
2	8.24798	-0.23762
3	8.24156	-0.21668

4	8.24013	-0.23189
5	8.23935	-0.22256
6	8.23899	-0.23043
7	8.23877	-0.22493
8	8.23866	-0.22931
9	8.23859	-0.22613

Following section 2.3, a model selection was conducted with `AIC` which generates exactly the same results as Table 2 of Belcher *et al.* (1994). Accordingly, the first-order value for the AIC shows the most rapid drop from the base-line of 0. Consequently a large *t*-value of 3.20 suggests order 7 while the minimum AIC implies order 9. For illustration, a model order 7 was selected as in Belcher *et al.* (1994).

```
R> AIC(V22174.car14)


Call:
car(x = V22174, scale = 0.2, order = 14)

Model selection statistics

 order t.statistic      AIC
     1       -8.66   -72.93
     2        1.72   -73.89
     3       -1.35   -73.72
     4        3.56   -84.41
     5        3.61   -95.47
     6        0.89   -94.27
     7        3.20  -102.50
     8        2.15  -105.14
     9       -2.00  -107.16
    10       -0.82  -105.83
    11        0.71  -104.34
    12        0.04  -102.34
    13       -1.91  -103.99
    14       -1.92  -105.66


R> V22174.car7 <- car(V22174, scale = 0.2, order = 7)
R> summary(V22174.car7)


Call:
car(x = V22174, scale = 0.2, order = 7)

Order of model = 7, sigma^2 = 1.37e-09

Estimated coefficients (standard errors):
      phi_1 phi_2 phi_3  phi_4 phi_5  phi_6 phi_7
coef -0.501 0.355 0.085 -0.022 0.605 -0.371 0.483
```

```
S.E.   0.108 0.111 0.060  0.071 0.084  0.124 0.112

Estimated mean (standard error):
[1] 0.173
[1] 0.022
```

Alternatively, the following code illustrates how to conduct model selection via the classical AIC or BIC by fitting the models of increasing order p to the series. Indeed, the model with order $p = 7$ is the second best model selected by the AIC and the best model by the BIC.

```
R> norder <- 14
R> V22174.aic <- V22174.bic <- rep(NA, norder)
R> for (i in 1:norder) {
+       fit <- car(V22174, scale = 0.2, order = i)
+       V22174.aic[i] <- fit$aic
+       V22174.bic[i] <- fit$bic
+  }
R> res <- data.frame(order = 1:norder, AIC = V22174.aic,
+       BIC = V22174.bic)
R> print(res, row.names = FALSE, print.gap = 8)
```

order	AIC	BIC
1	395.8212	402.0210
2	395.4665	404.7661
3	397.1475	409.5470
4	395.2427	410.7420
5	380.2526	398.8518
6	381.2466	402.9457
7	371.6226	396.4216
8	373.5190	401.4178
9	374.2532	405.2519
10	375.2333	409.3318
11	376.8200	414.0184
12	378.8153	419.1136
13	362.1151	405.5132
14	375.8480	422.3460

The estimated spectra for both models of order 14 and 7 are displayed on logarithmic (base 10) scale in Figure 2. Both two models indicate three peaks, while for the model of order 14 the resolution is much improved. In spectrum calculation, the default frequency range is set as from zero to `scale` parameter κ in `n.freq=500` intervals. It is convenient to plot the spectrum for a new range of frequencies with the argument `frmult` which can be used to multiply the frequency range.

```
R> par(mfrow = c(2, 1))
R> spectrum(V22174.car14)
R> spectrum(V22174.car7)
```

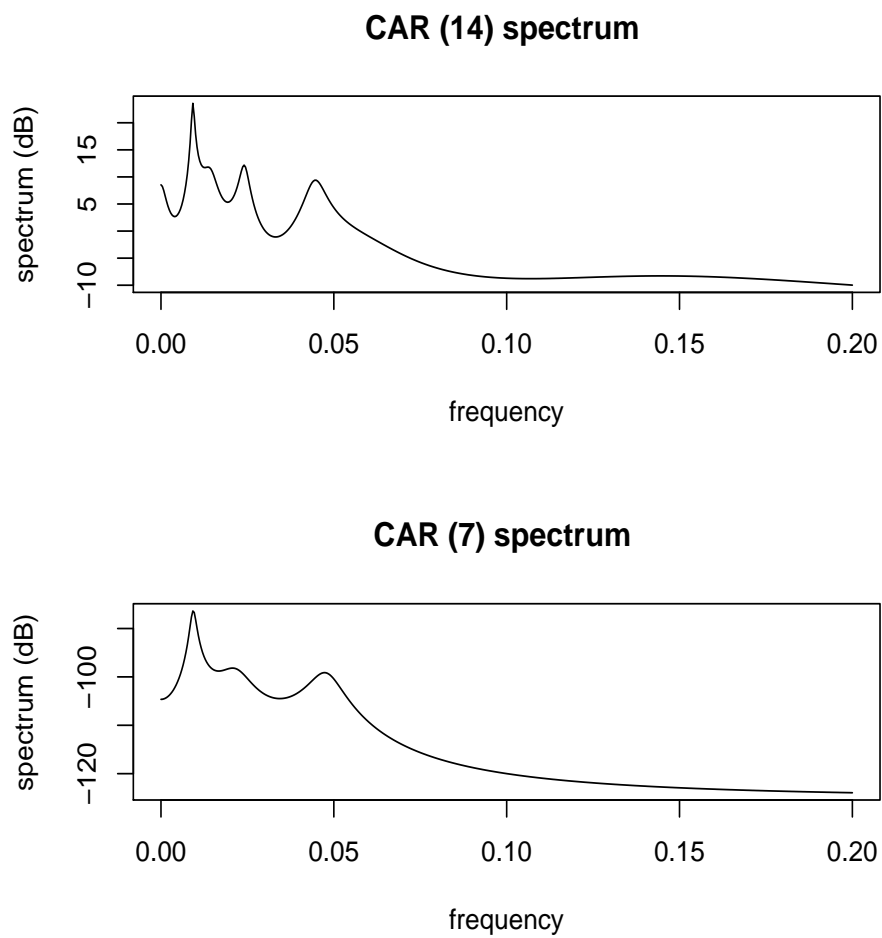**CAR (14) spectrum**

CAR (7) spectrum

Figure 2: Spectra from fitted models for the oxygen isotope series.

To check model assumptions as described in section 2.4, Figure 3 displays a plot of the standardized residuals, the ACF of the residuals, cumulative periodogram of the standardized residuals, and the p-values associated with the Ljung-Box statistic. Visual inspection of the time plot of the standardized residuals in Figure 3 shows no obvious patterns, although one outlier extends 3 standard deviations. The ACF of the standardized residuals shows no apparent departure from the model assumptions, i.e., approximately independently and normally distributed with zero means and variances $1/n$ at lag > 0. The cumulative periodogram of standardized residuals follows the line $y = 2x$ reasonably well. The Ljung-Box statistic is not significant at the lags shown.
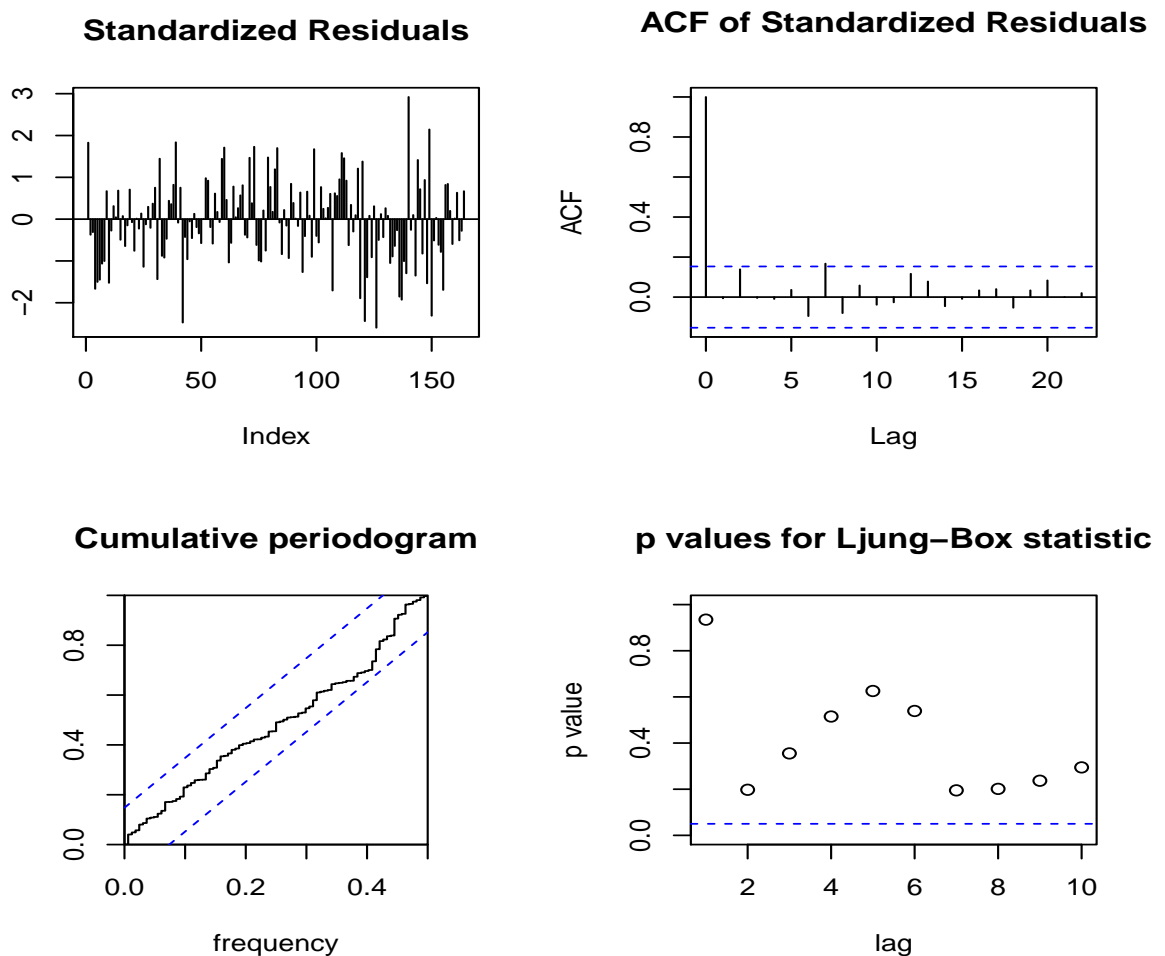
```
R> tsdiag(V22174.car7)
```


Figure 3: Model diagnostics for the oxygen isotope series.

4.2. Medical application

Belcher *et al.* (1994) analyzed 209 measurements of the lung function of an asthma patient. The time series is measured mostly at 2 hour time intervals but with irregular gaps, as demonstrated by the unequal space of tick marks in Figure 4.

```
R> data("asth")
```

```
R> plot(asth, type = "l", xlab = "Time in hours", ylab = "")
R> rug(asth[, 1], col = "red")
```
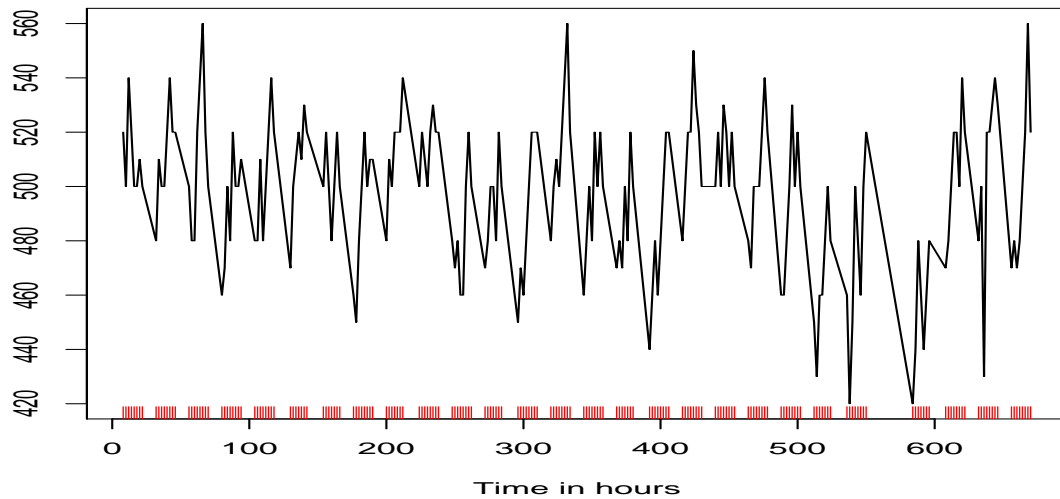

Figure 4: Measurements of the lung function.

To apply **cts**, a scale parameter 0.25 was chosen and a model of order 4 was fitted to the data (Belcher *et al.* 1994).

```
R> asth.car4 <- car(asth, scale = 0.25, order = 4, ctrl = car_control(n.ahead = 10))
R> summary(asth.car4)

Call:
car(x = asth, scale = 0.25, order = 4, ctrl = car_control(n.ahead = 10))

Order of model = 4, sigma^2 = 0.779

Estimated coefficients (standard errors):
     phi_1 phi_2 phi_3  phi_4
coef 0.093 0.037 0.015 -0.701
S.E. 0.075 0.071 0.077  0.096

Estimated mean (standard error):
```

```
[1] 495.544
[1] 4.524
```

The log-spectrum (base 10) of the fitted model is shown in Figure 5. The spectral peak indicates a strong diurnal cycle in the data.

It is possible to fit a model with an observation error term by setting `vri=TRUE` in the parameter control statement. The following code shows how to fit such a model. The estimated observation error variance can be found with the `summary` command and the corresponding spectrum in Figure 5 is compared with the model without an observation error.

```
R> asth.vri <- car(asth, scale = 0.25, order = 4, ctrl = car_control(vri = TRUE))
R> summary(asth.vri)

Call:
car(x = asth, scale = 0.25, order = 4, ctrl = car_control(vri = TRUE))

Order of model = 4, sigma^2 = 0.000158

Observation error variance: 243

Estimated coefficients (standard errors):
      phi_1 phi_2  phi_3 phi_4
coef -1.489 1.556 -1.462 0.680
S.E.  0.128 0.130  0.137 0.125

Estimated mean (standard error):
[1] 494.249
[1] 3.128
```

Nevertheless, we focus on the model fit `asth.car4` without the measurement error. Applying function `factab` to this model returns one complex zero and two real zeros.

```
R> factab(asth.car4)

Call:
factab(object = asth.car4)

Characteristic root of original parameterization in alpha

            1                2                3                4
-0.016+0.000i   -0.020+0.255i   -0.020-0.255i   -7.246+0.000i

Frequency

     1       2       3       4
0.000   0.041   0.041   0.000
```

```
R> par(mfrow = c(2, 1))
R> spectrum(asth.car4)
R> spectrum(asth.vri)
```

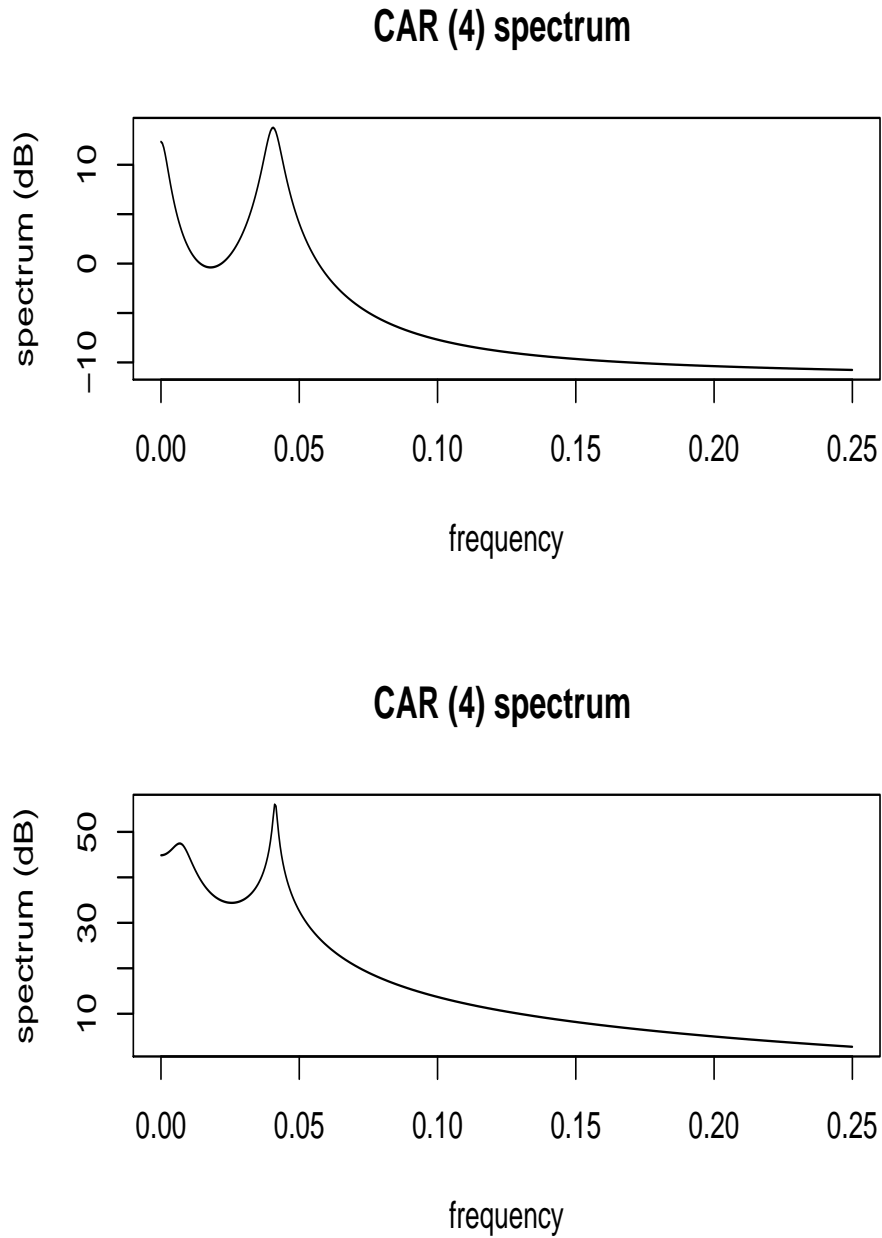**CAR (4) spectrum**

CAR (4) spectrum

Figure 5: Spectra from fitted models without and with an observation error term (top and bottom panel, respectively), for the lung function measurements.

```
R> asth.kalsmo <- kalsmo(asth.car4)
R> par(mfrow = c(3, 1))
R> kalsmoComp(asth.kalsmo, comp = 1, xlab = "Time in hours")
R> kalsmoComp(asth.kalsmo, comp = c(2, 3), xlab = "Time in hours")
R> kalsmoComp(asth.kalsmo, comp = 4, xlab = "Time in hours")
```
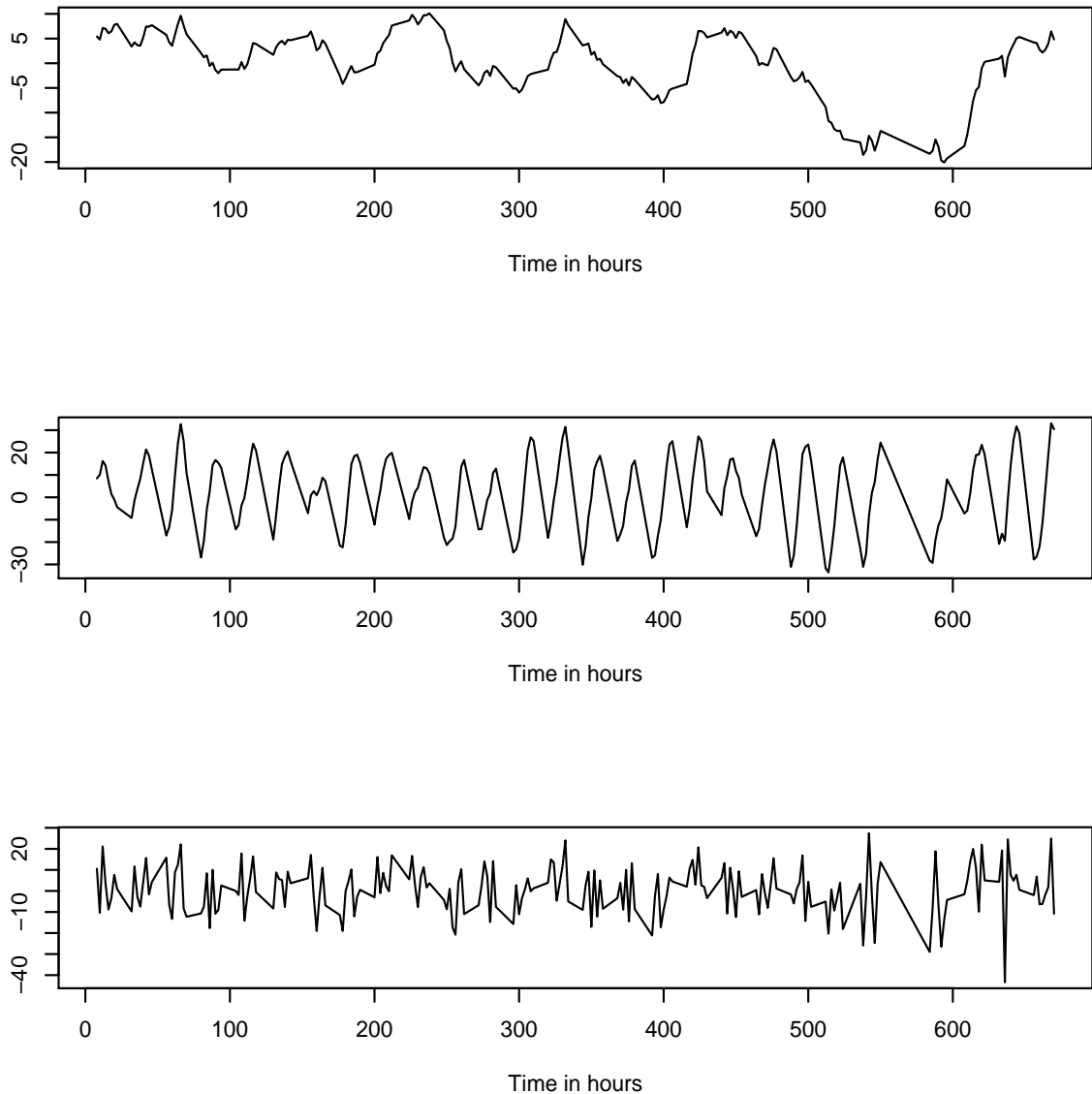

Figure 6: Components of the lung function measurements. From top to bottom: trend component, diurnal component, and approximate white noise component.

We thus decomposed the original time series into three corresponding components via the

Kalman smoother as shown in Figure 6.

Finally, we predicted the last 10 steps past the end of time series in Figure 7.

```
R> predict(asth.car4, xlab = "Time in hours")

Call:
car(x = asth, scale = 0.25, order = 4, ctrl = car_control(n.ahead = 10))

                1       2       3       4       5       6       7
Time      671.000 672.000 673.000 674.000 675.000 676.000 677.000
Predict   527.692 522.959 516.956 510.116 502.904 495.786 489.208
                8       9      10
Time      678.000 679.000 680.000
Predict   483.561 479.165 476.245
```
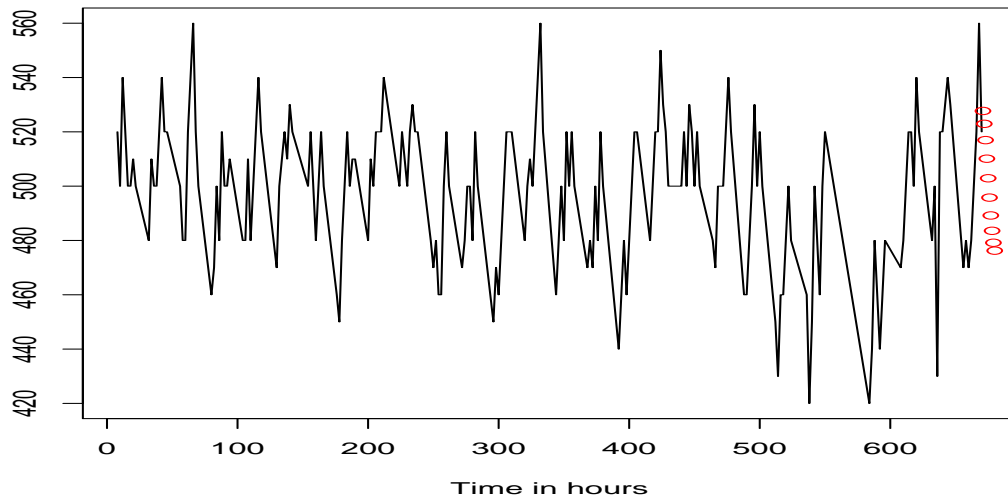

Figure 7: Forecasts (circles) for lung function measurements.

5. Conclusion

In this article we have outlined the methods and algorithms for fitting continuous time autoregressive models through the Kalman filter. The theoretical ingredients of Kalman filter have their counterparts in the R package **cts**, which can be particularly useful with unequally sampled time series data.

References

Akaike H (1974). "A new look at the statistical model identification." *IEEE Transactions on Automatic Control*, **19**(6), 716–723.

Belcher J, Hampton JS, Tunnicliffe Wilson G (1994). "Parameterization of Continuous Time Autoregressive Models for Irregularly Sampled Time Series Data." *Journal of the Royal Statistical Society B*, **56**, 141–155.

Doob JL (1953). *Stochastic Processes.* John Wiley & Sons, New York.

Durbin J, Koopman SJ (2001). *Time Series Analysis by State Space Methods.* Oxford University Press, Oxford.

Harvey AC (1990). *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press, Cambridge.

Jenkins GM, Watts DG (1968). *Spectral Analysis and its Applications.* Holden-Day, San Francisco, California.

Jones RH (1981). "Fitting a Continuous Time Autoregression to Discrete Data." In *Applied Time Series Analysis II*, pp. 651–682.

Moler C, Loan CV (2003). "Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later." *SIAM Review*, **45**(1), 3–49. Society for Industrial and Applied Mathematics.

R Development Core Team (2013). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org.

Schwarz G (1978). "Estimating the dimension of a model." *Annals of Statistics*, **6**(2), 461–464.

Tunnicliffe Wilson G, Morton AS (2004). "Modelling Multiple Time Series: Achieving the Aims." In J Antoch (ed.), *Proceedings in Computational Statistics, 2004*, pp. 527–538. Physica Verlag, Heidelberg.

Tunnicliffe Wilson G, Wang Z (2013). ***cts:*** *Continuous Time Autoregressive Models.* R package version 1.0-15, URL http://CRAN.R-project.org/package=cts.

Wang Z (2004). *The Application of the Kalman Filter to Nonstationary Time Series through Time Deformation.* Ph.D. thesis, Southern Methodist University.

Wang Z, Woodward WA, Gray HL (2009). "The Application of the Kalman Filter to Nonstationary Time Series through Time Deformation." *Journal of Time Series Analysis*, **30**(5), 559–574.

Wiberg DM (1971). *Schaum's Outline of Theory and Problems of State Space and Linear Systems.* McGRAW-HILL, New York.

Affiliation:

Zhu Wang
Department of Epidemiology & Biostatistics
UT Health San Antonio
E-mail: wangz1@uthscsa.edu