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

Zhu Wang
Connecticut Children's Medical Center
University of Connecticut School of Medicine

Abstract

We describe an R package **cts** for continuous time autoregressive model fitting, 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

A discrete time autoregressive model is a common tool for the analysis of time series data. An important alternative is the continuous time series (CTS) autoregressive model, which has unique advantages over a discrete time series model. For instance, a CTS model can provide interpolates between observations, generate signal derivatives and fit unequally sampled data. There is a rich literature on continuous autoregressive (CAR) model. Jones (1981) developed an estimation method through the application of the Kalman filter. Belcher, Hampton, and Tunnicliffe Wilson (1994) further extended to higher order model with rapid and reliable convergence of parameter estimates. Also see Tunnicliffe Wilson and Morton (2004) for additional description. Wang, Woodward, and Gray (2009) utilized the methods in Belcher *et al.* (1994) for fitting time varying nonstationary models. The CAR models have seen practical usages in many fields including astronomics, medicine and economics (Whitelock, Feast, Marang, and Breedt 2004; Belcher *et al.* 1994; Bergstrom 1990). 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 2011) package **cts** (Tunnicliffe Wilson and Wang 2011) 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. CAR model

Suppose we have data x_τ observed on time t_τ for $\tau = 1, 2, ..., n$. We assume noise-affected observations $x_\tau = y(t_\tau) + \eta_\tau$ and $y(t_\tau)$ 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_τ is a normally distributed random variable representing observational error, uncorrelated with $\epsilon(t)$, and $E(\eta_\tau) = 0; E(\eta_j \eta_k) = 0, \text{for } j \neq k; E(\eta_\tau^2) = \gamma\sigma^2$.

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 reparameterization (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.

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

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

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 first and quadratic factors that have complex roots. A quadratic factor $(s - r_{2k-1})(s - r_{2k})$ with complex poles is associated with "cyclic" behavior in data, given at $f = \frac{|\Im(r_{2k})|}{2\pi}$, where $|\Im(r_{2k})|$ is the absolute value of

the imaginary part of r_{2k}. Actually, the true cyclic behavior with these cycles is present in the autocorrelations and only approximately present in the realization itself. For a first order factor with the pole r_k, the system frequency is $f = 0$ if $r_k < 0$, and $f = 0.5$ if $r_k > 0$, which corresponds to a white noise process.

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 $\theta(t) = (z(t), z'(t), ..., z^{(p-1)}(t))^T$. The state equation is then given by

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

where

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

and

$$R' = \begin{bmatrix} 0 & 0 & \dots & 1 \end{bmatrix}. \tag{9}$$

The observation equation is given by

$$x_\tau = H\theta(t_\tau) + \eta_\tau, \tag{10}$$

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

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

$r_1, r_2, ..., r_p$ 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

$$x_\tau = G\psi(t_\tau) + \eta_\tau \tag{14}$$

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 $T_{\theta,\tau} = e^{A\delta_\tau}$ (standard form) is required where $\delta_\tau = t_\tau - t_{\tau-1}$. For a review of computations related to the exponential of a matrix, see Moler and Loan (2003). For the diagonal form, $T_{\psi,\tau} = e^{\mathbf{D}\delta_\tau}$ is diagonal with elements $e^{r_i\delta_\tau}$. 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

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

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
$$T_{\psi,\tau} = e^{\mathbf{D}\tau} \tag{17}$$
a diagonal matrix, then
$$\psi(t_k|t_{k-1}) = T_{\psi,\tau}\psi(t_{k-1}|t_{k-1}). \tag{18}$$

2. Calculate the covariance matrix of this prediction:
$$P_\psi(t_k|t_{k-1}) = T_{\psi,\tau}(P_\psi(t_{k-1}|t_{k-1}) - Q_\psi)\bar{T}_{\psi,\tau} + Q_\psi. \tag{19}$$

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

5. Update the estimate of the state vector:
$$\psi(t_k|t_k) = \psi(t_k|t_{k-1}) + P_\psi(t_k|t_{k-1})\bar{G}' F_\psi^{-1}(t_k) v_\psi(t_k) \tag{23}$$

6. Update the covariance matrix:

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

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 **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 $V_\phi^{-1} = L_\phi L'_\phi$ where L_ϕ is a lower triangular matrix, and define the vector $t_\phi = L'_\phi \phi$ and construct the sequence $AIC_k = -\sum_{i=1}^{k} t_{\phi,i}^2 + 2k$. The index of the minimum value of AIC_k 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.

2.4. Diagnostics

The assumptions underlying the model (7) and (10) are that the disturbances $\epsilon(t)$ and η_τ 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.

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

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

and $T_{\psi,\tau+1} = e^{\mathbf{D}(t_{k+2}-t_{k+1})}$, and $\bar{T}_{\psi,\tau+1}$ 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_τ, in the absence of measurement error, is the sum of contributions from the diagonalized state variables ψ, i.e., $x_\tau = \sum_j G_j\psi_j(t_\tau)$. 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_τ represents a source of diurnal variation. One possible real zero contributes a low frequency component and the other possible real zero contributes a 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.

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. 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, `car(data, 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. In an earlier version of package, specifying `trace=TRUE` in `car_control` could trigger annotated printout of information during the fitting process and major results for the fitted model. However, since this call invoked print with Fortran, the printout is essentially suppressed in the current version.

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 2.13.1 (2011-07-08) and the operating system `i686-pc-cygwin`.

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")
```

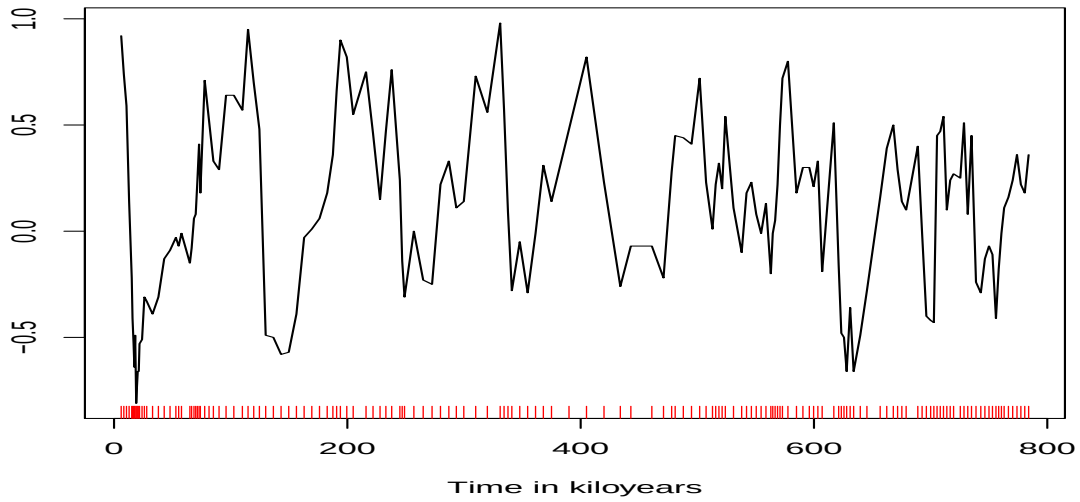

Figure 1: Oxygen isotope series.

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 0.9 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

Coefficients and mean (standard errors):
          1     2     3      4     5      6     7  Mean
     -0.501 0.355 0.085 -0.022 0.605 -0.371 0.483 0.173
S.E.  0.108 0.111 0.060  0.071 0.084  0.124 0.112 0.022
```

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.

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

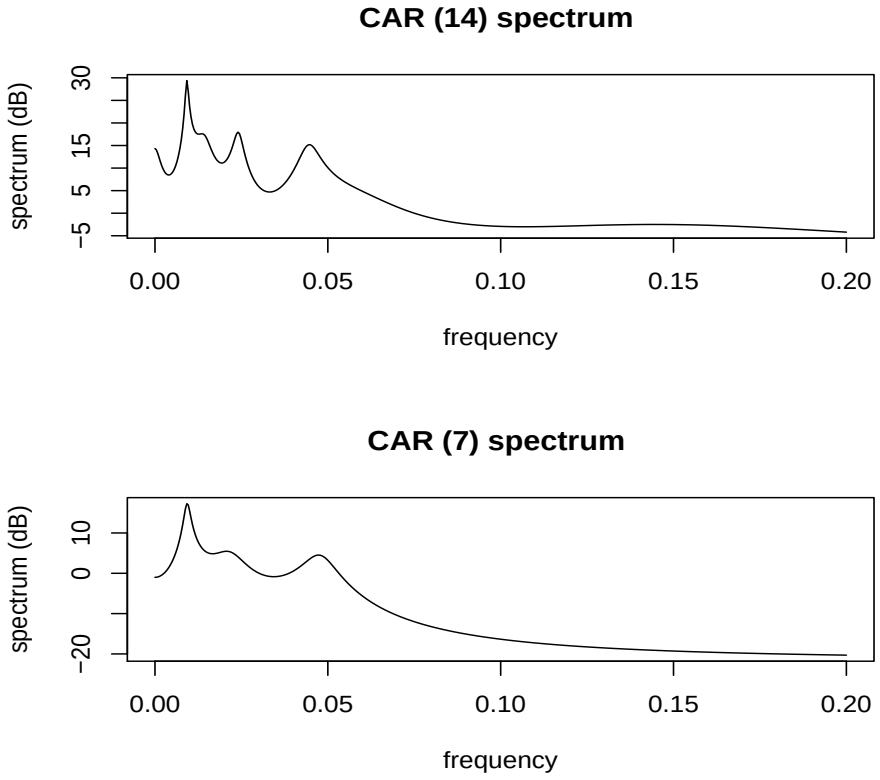

Figure 2: Spectra from fitted models for the oxygen isotope series.

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)
```

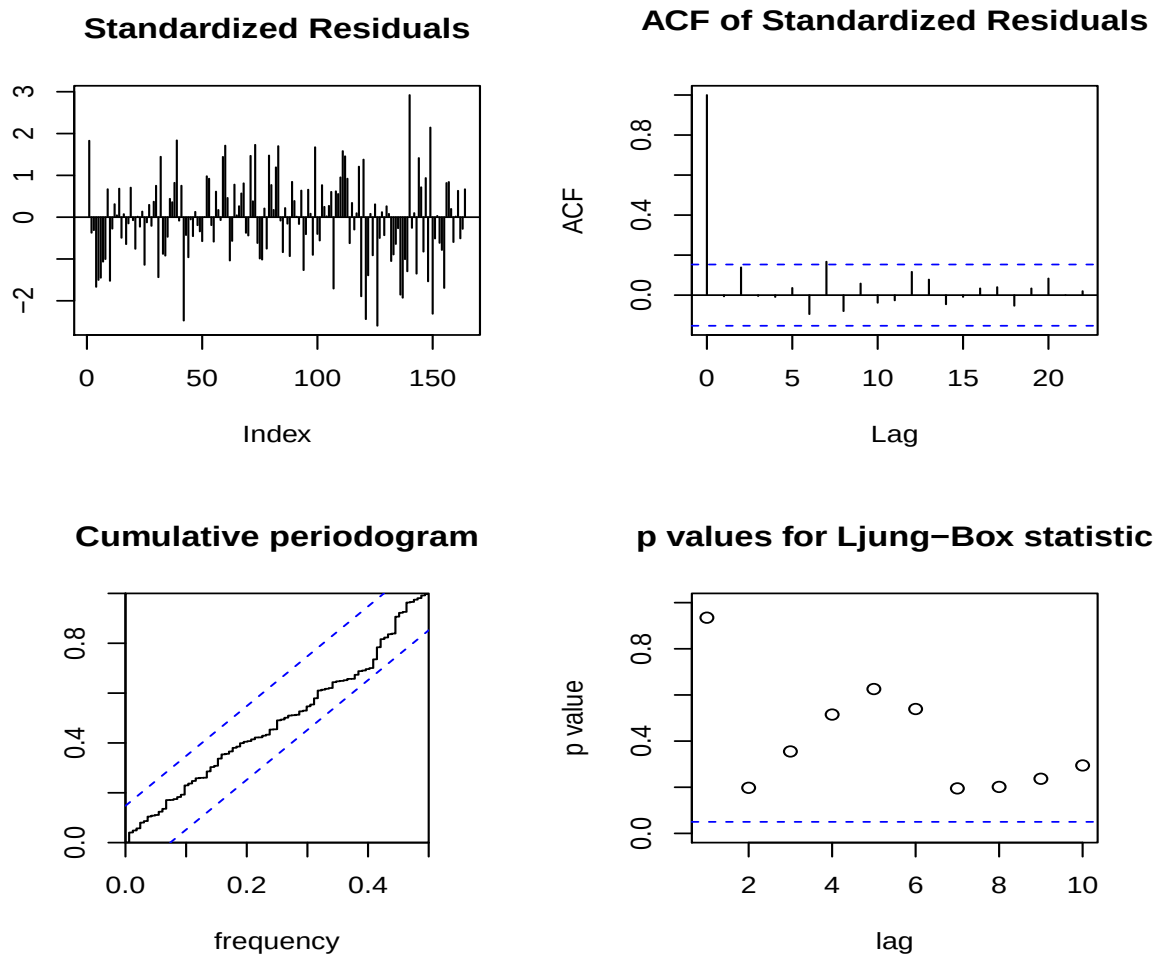

Figure 3: Model diagnostics for the oxygen isotope series.

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

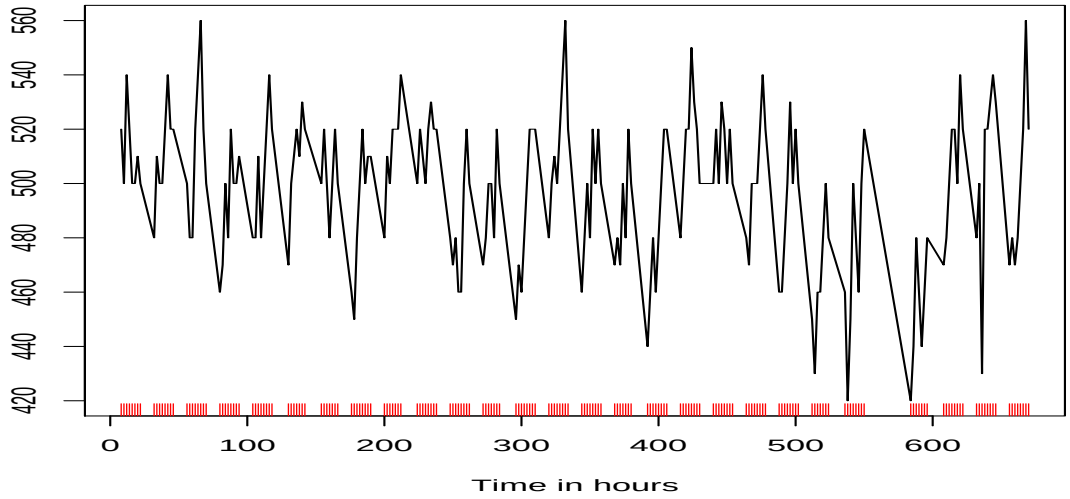

Figure 4: Measurements of the lung function.

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")
```

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

Coefficients and mean (standard errors):
         1     2     3      4    Mean
     0.093 0.037 0.015 -0.701 495.544
S.E. 0.075 0.071 0.077  0.096   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, along with a low frequency. With function `factab`, the model has one diurnal frequency 0.041, a low frequency component and the other close to 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> spectrum(asth.car4)
```

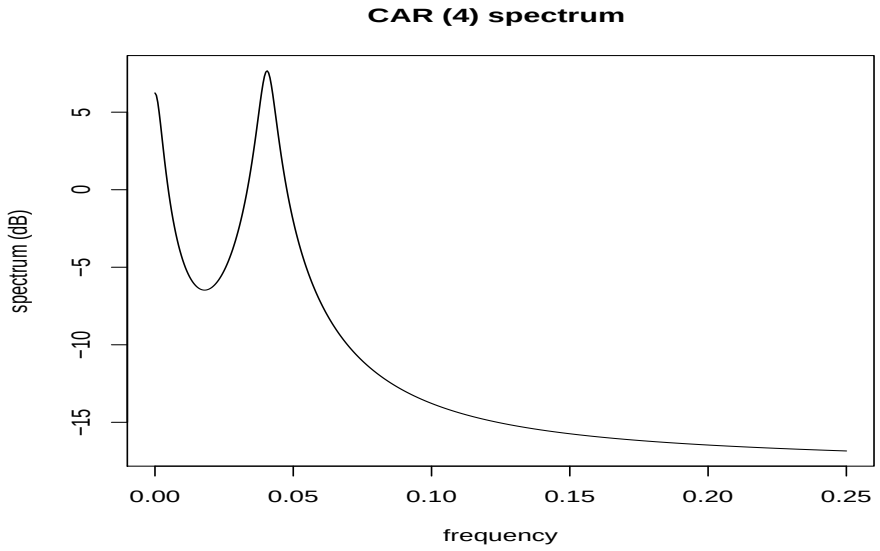

Figure 5: Spectrum from fitted model for the lung function measurements.

```
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> 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")
```

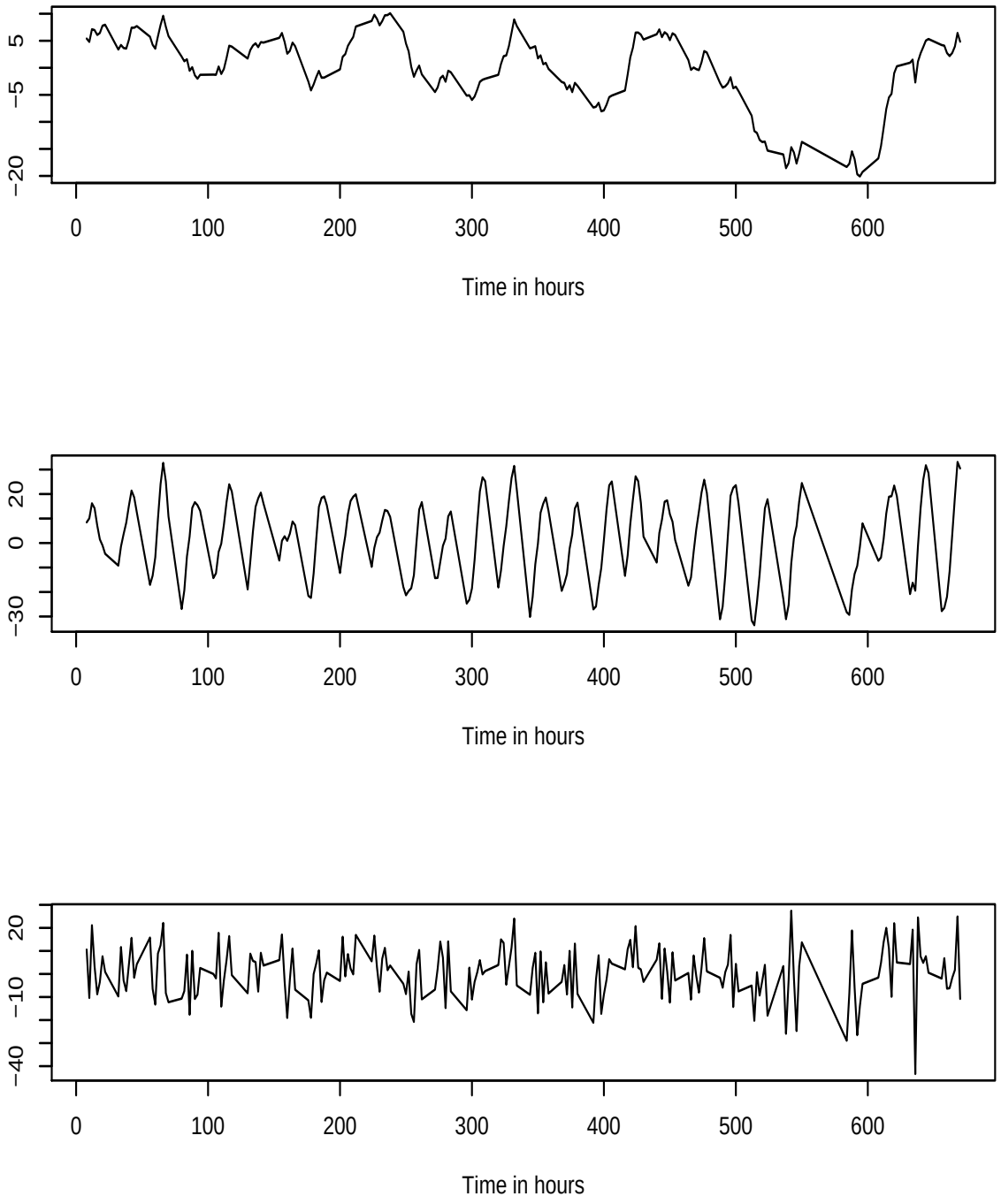

Figure 6: Components of the lung function measurements. From top to bottom: trend (low frequency) component, diurnal component, and white noise component.

```
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
```

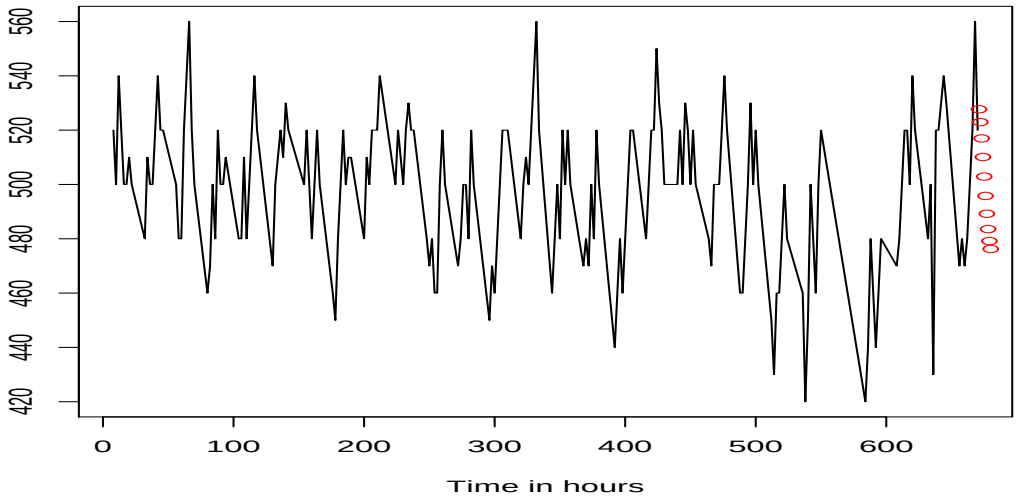

Figure 7: Forecasts (circles) for lung function measurements.

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

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.

Bergstrom AR (1990). *Continuous Time Econometric Modeling*. Oxford University Press, New York.

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 (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

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 (2011). ***cts**: Continuous Time Autoregressive Models*. R package version 1.0-9, 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.

Whitelock PA, Feast MW, Marang F, Breedt E (2004). "The 2003 Shell Event in η Carinae." *Monthly Notices of the Royal Astronomical Society*, **352**(2), 447–456.

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 Research
Connecticut Children's Medical Center
Department of Pediatrics
University of Connecticut School of Medicine
Connecticut 06106, United States of America
E-mail: zwang@ccmckids.org