Time series models: Continuous data

Atmospheric variables tend to be persistent, they have a lag-autocorrelation $r_1 > 0$.
Simplest time series model: $x_{t+1} - \mu = \phi_1(x_t - \mu) + \varepsilon_{t+1}$, or in terms of the anomalies, $x'_{t+1} = \phi_1 x'_t + \varepsilon_{t+1}$

This is an autoregressive model of order 1 (AR(1)). Such model can be used to:

a) Fit the time series and derive some of its properties. Similar to fitting a theoretical probability distribution to a sample.
b) To make a forecast: $\hat{x}_{t+1} - \mu = \phi_1(x_t - \mu)$

We need to determine $\phi_1 = \phi$ and the variance of the error $\mathrm{var}(\varepsilon)$ from the data that we want to fit with the AR(1) model.

Since $r_1 = corr(x'_t, x'_{t+1}) = \dfrac{\left(\overline{x'_t x'_{t+1}}\right)}{\sqrt{\sigma^2_{x'_t}\sigma^2_{x'_{t+1}}}} = \dfrac{\left(\overline{x'_t x'_{t+1}}\right)}{\sigma^2_{x'_t}}$, a linear regression forecast is simply $\hat{x}_{t+1} - \mu = r_1(x_t - \mu)$ or $\phi_1 = r_1$

Note that autocorrelations at longer lags are not zero for AR(1), even though we only need the last observation to make an AR(1) forecast:

$$P\left\{X_{t+1} \le x_{t+1} \mid X_t \le x_t, X_{t-1} \le x_{t-1}, \ldots, X_1 \le x_1\right\} = P\left\{X_{t+1} \le x_{t+1} \mid X_t \le x_t\right\}$$

i.e., just the last observation is enough to make a forecast, but

$$r_1 = \phi_1, \quad r_2 = \phi_1^2, \quad r_3 = \phi_1^3, \ldots$$

In the AR(1)

$$x'_{t+1} = \phi x'_t + \varepsilon_{t+1}.$$

Multiply this equation by x'_t and average over a long time series, divide by $\sigma_x^2 \approx s_x^2$, and use $\overline{x'_t x'_t} = \sigma_x^2 = \overline{x'_{t+1} x'_{t+1}}$, $\overline{x'_t x'_{t+1}} = r_1\sigma_x^2 = \phi\sigma_x^2$ and obtain

$\overline{x'_t \varepsilon_{t+1}} = 0$, i.e., the error is uncorrelated with the predictor.

We have determined the coefficient $\phi = r_1$, but we still need to determine the value of σ_ε^2:

Multiply $x'_{t+1} = \phi x'_t + \varepsilon_{t+1}$ now by x'_{t+1} and average over a long time series and obtain

$\overline{x'_{t+1} x'_{t+1}} = \phi \overline{x'_{t+1} x'_t} + \overline{\varepsilon_{t+1}^2}$ so that the unexplained variance of the prediction is

$$\sigma_\varepsilon^2 = (1-\phi^2)\sigma_x^2 .$$

The estimate of $\hat{\phi} = r_1$ obtained from a sample can be tested for significance (whether it is REALLY different from zero) as in linear regression. Recall that in linear regression $y_t - \overline{y} = b_1(x_t - \overline{x}) + \varepsilon_t$, and the variance of b_1 is

estimated as $\sigma_{b_1}^2 \approx \dfrac{\overline{\varepsilon_t^2}}{\sum_{t=1}^{n}(x_t - \overline{x})^2} \approx \dfrac{\overline{\varepsilon_t^2}}{n\sigma_x^2}$ where $\overline{\varepsilon_t^2} \approx \dfrac{1}{n}\sum_{t=1}^{n}\left(y_t - \hat{y}_t\right)^2$ is the forecast error squared (unexplained variance). Then we use a t-test

$$T = \frac{b_1 - 0}{\sigma_{b_1}} .$$

(<u>Note</u>: this is the same type of statistics that we would use to estimate the significance of a climate trend b_1 if we assume that the trend is linear with time, $Temp = Temp_{mean} + b_1\ time$.)

As shown above, for AR(1) $x'_{t+1} = \phi x'_t + \varepsilon_{t+1}$, the variance "unexplained" by regression is $\sigma_\varepsilon^2 = (1-\phi^2)\sigma_x^2$.

Using the same formula for the estimated variance of the linear coefficient ϕ as for linear regression coefficient b_1, $\sigma_{b_1}^2 \approx \dfrac{\overline{\varepsilon_t^2}}{n\sigma_x^2}$, the error variance of ϕ is estimated as

$$\sigma_\phi^2 = \frac{\sigma_\varepsilon^2}{n\sigma_x^2} = \frac{1-\phi^2}{n} .$$

One can then test whether the persistence is significantly different from zero using $z = \dfrac{\hat{\phi} - 0}{\sqrt{(1-\hat{\phi}^2)/n}}$ (we can use a Gaussian distribution because $\hat{\phi}$ determines both the mean and its standard deviation).

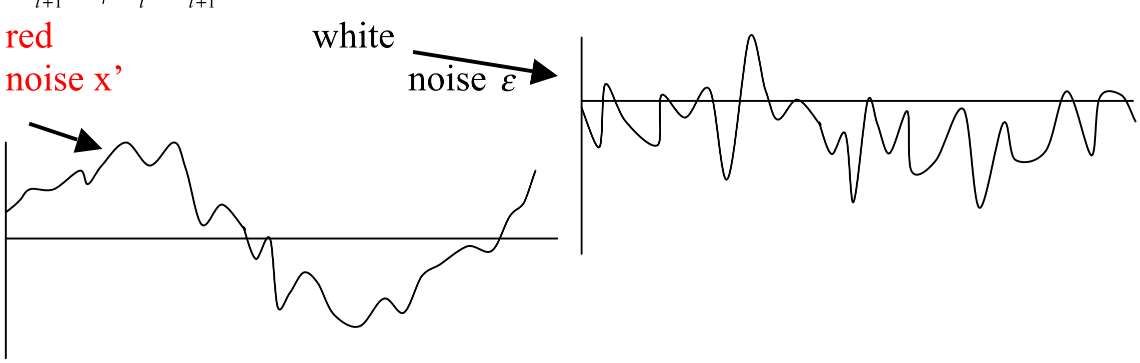
The variance of the noise (unexplained variance), when we use AR(1): $\sigma_\varepsilon^2 = (1-\phi^2)\sigma_x^2$ for the population. For a sample of size *n*, since we used a second degree of freedom for the coefficient ϕ ,

$$s_\varepsilon^2 = \frac{(1-\hat{\phi}^2)}{n-2}\sum_{t=1}^{n}(x_t - \overline{x})^2 = \frac{n-1}{n-2}(1-\hat{\phi}^2)s_x^2$$, but usually n is large so $\dfrac{n-1}{n-2} \approx 1$

Applications:

a) Create a persistent time series (red noise) that looks like nature:

$$x'_{t+1} = \hat{\phi}x'_t + \varepsilon_{t+1}$$



(How would blue noise look like? Hint: for white noise, persistence, $r_1=0$; for red noise, $r_1>0$; for blue noise, $r_1<0$, the anomaly changes sign very frequently)

b) Make a forecast: $x'_{t+1} = \hat{\phi} x'_t$

This is the basis for many forecasts derived from time series.

Higher order autoregressive model AR(K)

$$x_{t+1} - \mu = \sum_{k=1}^{K} \phi_k (x_{t-k+1} - \mu) + \varepsilon_{t+1}$$

For example, a 2nd order autoregressive model, AR(2):

$$x_{t+1} - \mu = \phi_1 (x_t - \mu) + \phi_2 (x_{t-1} - \mu) + \varepsilon_{t+1}$$

or $x'_{t+1} = \phi_1 x'_t + \phi_2 x'_{t-1} + \varepsilon_{t+1}$ (1)

Multiply both sides of (1) by x'_t, take an average over a long time series, divide by s_x^2, and obtain the following relationship:

$$r_1 = \hat{\phi}_1 + r_1 \hat{\phi}_2 \qquad (2)$$

Similarly, multiply both sides of (1) by x'_{t-1}, take an average over a long series, divide by s_x^2, and obtain the following:

$$r_2 = r_1 \hat{\phi}_1 + \hat{\phi}_2 \qquad (3)$$

From (2) and (3) can solve for $\hat{\phi}_1, \hat{\phi}_2$:

$$\hat{\phi}_1 = \frac{r_1(1-r_2)}{1-r_1^2}, \quad \hat{\phi}_2 = \frac{r_2 - r_1^2}{1-r_1^2}$$

Or , if we know $\hat{\phi}_1, \hat{\phi}_2$

$$r_1 = \frac{\hat{\phi}_1}{1-\hat{\phi}_2}, \quad r_2 = \hat{\phi}_2 + \frac{\hat{\phi}_2^2}{1-\hat{\phi}_2}$$

The expected variance of the error is

$$\sigma_\varepsilon^2 = (1 - \hat{\phi}_1 r_1 - \hat{\phi}_2 r_2)\sigma_x^2$$

For an AR(2) series to be stationary (so that it does not drift away), the following conditions have to be satisfied:

$$-1 \le \phi_2 \le 1, \quad \phi_1 + \phi_2 \le 1, \quad \phi_2 - \phi_1 \le 1$$

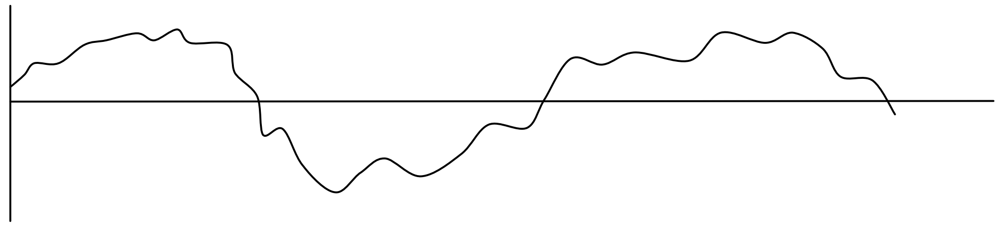
(see figures 8.7, 8.8 in Wilks).

Note: Variance of a time series

If a time series has zero autocorrelation, the variance of the mean is the familiar

$\sigma_{\bar{x}}^2 = \frac{s_x^2}{n}$ which shows that the mean of the time series measured by n time steps has a variance n times smaller than the individual measurements.

However, if $r_1 > 0$ then the time mean has a larger variance than indicated by this formula! This is because there are fewer independent measurements than n:



This effect can be estimated by a variance inflation factor V :

$s_{\bar{x}}^2 = V\frac{s_x^2}{n}$ where $V = \frac{1+\phi_1}{1-\phi_1}$ for AR(1). In other words, the effective number of independent observations in an AR(1) time series is

$$n' = \frac{1-\phi_1}{1+\phi_1}n$$

This is very important when estimating the number of degrees of freedom for, say, daily observations. If $\phi_1 = 0.5$, then $n' \approx \frac{0.5}{1.5}n = n/3$. These means that we should consider only observations every third day, or conversely, assume that the number of degrees of freedom is $n/3$.

Note on autoregression modeling (not sure who is the author but it is a nice summary)

Autoregression Modeling:

- mathematical model used to try and explain a time series of observations: $\{y_{i\Delta t}\}$, i=1,2,...,N.

AR(p) process:

$$y_t = a_1 y_{t-\Delta t} + a_2 y_{t-2\Delta t} + \ldots + a_p y_{t-p\Delta t} + a_0 \upsilon_t$$

Gaussian white noise

- an AR(p) model is a discretized p^{th}-order ordinary differential equation:

e.g. AR(1): $\frac{d}{dt}y(t) + \frac{y(t)}{\tau} = \upsilon(t)$ ← Gaussian white noise

$$y(t) = y_0 \exp\left(-\frac{t}{\tau}\right)$$

τ = characteristic timescale, or 'memory'
~ (heat capacity) / (cooling rate)

AR(1), aka 'red noise' = simplest self-consistent model for a geophysical system (e.g., Hasselman, 1976).

AR(2+): combination of oscillations/growing and decaying exponentials.

Autoregressive, moving-average models ARMA(K,M)

In these models we assume that the noise also has some persistence, and persist the last few observed "noises":

$$x_{t+1} - \mu = \sum_{k=1}^{K} \phi_k (x_{t-k+1} - \mu) + \varepsilon_{t+1} + \sum_{m=1}^{M} \theta_m \varepsilon_{t-m+1}$$

The simplest ARMA model is ARMA(1,1):

$$x_{t+1} - \mu = \phi_1 (x_t - \mu) + \varepsilon_{t+1} + \theta_1 \varepsilon_t$$

The lag-1 autocorrelation for this model is

$$r_1 = \frac{(1 - \phi_1\theta_1)(\phi_1 - \theta_1)}{1 + \theta_1^2 - 2\phi_1\theta_1},$$

and the expected error variance is given by

$$\sigma_\varepsilon^2 = \frac{1 - \phi_1^2}{1 + \theta_1^2 + 2\phi_1\theta_1} \sigma_x^2 .$$

So, in practice, to make an ARMA(1,1) forecast, we need to fit the data and compute the first two lag autocorrelations r_1, r_2. Then, obtain ϕ_1 from

$r_2 = \phi_1 r_1$, and finally θ_1 from $r_1 = \dfrac{(1 - \phi_1\theta_1)(\phi_1 - \theta_1)}{1 + \theta_1^2 - 2\phi_1\theta_1}$.

Then the forecast for future data becomes:

$$\hat{x}_{t+1} - \mu = \phi_1 (x_t - \mu) + \theta_1 \varepsilon_t$$

where ε_t is the last observed forecast error $\varepsilon_t = x_t - \hat{x}_t$.