# Autoregressive Model to Parametrise Temperature Variability

Guy Munhoven

20th September 2021

**Abstract**

A series of autoregressive models is used to analyse sea-surface temperature time series in order to derive a simple parametrisation of temperature variability in a climate model.

# Introduction

# 1 Data

- Origin: `https://iridl.ldeo.columbia.edu/SOURCES/.NOAA/.NCDC/.OISST/.version2/.AVHRR/.anom/lat/%2819S%29%2816S%29RANGEEDGES/T/%281%20Jan%201982%29%2831%20Dec%202015%29RANGEEDGES/lon/%28148E%29%28154E%29RANGEEDGES/data.nc`
- time: 1982-01-01 to 2015-12-31 (12418 days)
- lat: $-18.875$ to $-16.125$ (12 latitudes at 0.25° resolution)
- lon: 148.125 to 153.875 (24 longitudes at 0.25° resolution)

# 2 Procedure

We fit a sequence of AR($p$) models to the time series extracted from the dataset at each lon-lat grid point, with $p = 1, \ldots, 6$ in order to determine the optimum order:

$$y_t = c + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \epsilon_t$$

where $c$ is a constant, $\phi_1, \ldots, \phi_p$ are the $p$ parameters of the model and $\epsilon_t$ is white noise (with zero mean and standard deviation $\sigma_\epsilon$).

The time series are first filtered using a Fast Fourier Transform (FFT) method with a high-pass triangular frequency filter in order to get rid of the seasonal cycle and any possible long-term trends. The cut-off frequency is set to 1.25 cycles/year. Since the total number of data points in each time series involves a comparatively large prime factor ($12418 = 2 \cdot 7 \cdot 887$), 130 points were dropped to reduce the time series length to $12288 = 2^{12} \cdot 3$, which allows for more efficient usage of the FFT algorithm.

# 3 Results

Each data series has a zero mean and therefore $c = 0$ for each model. We therefore fit an AR($p$) model

$$y_t = \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \sigma_\epsilon \upsilon_t$$

where $\upsilon_t$ is white noise with a standard deviation of 1. We fit such a model at each lon-lat point in the domain. Below, we report the averages of the parameter values $\phi_1, \ldots, \phi_p$ and of the required $\sigma_\epsilon$, together with the standard deviations of their distributions.

## 3.1 AR(1)

$$y_t = \phi_1 y_{t-1} + \sigma_\epsilon \upsilon_t$$

$$\begin{aligned} \phi_1 &= 0.8964 \pm 0.0060 \\ \sigma_\epsilon &= 0.2758 \pm 0.0094 \end{aligned}$$

## 3.2 AR(2)

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \sigma_\epsilon \upsilon_t$$

$$\begin{aligned} \phi_1 &= 1.0873 \pm 0.0097 \\ \phi_2 &= -0.21230 \pm 0.0069 \\ \sigma_\epsilon &= 0.2694 \pm 0.0091 \end{aligned}$$

### 3.3 AR(3)

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \sigma_\epsilon \upsilon_t$$

$$\begin{aligned} \phi_1 &= 1.115 \pm 0.010 \\ \phi_2 &= -0.354 \pm 0.015 \\ \phi_3 &= 0.130 \pm 0.011 \\ \sigma_\epsilon &= 0.2671 \pm 0.0089 \end{aligned}$$

### 3.4 AR(4)

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \phi_4 y_{t-4} + \sigma_\epsilon \upsilon_t$$

$$\begin{aligned} \phi_1 &= 1.112 \pm 0.011 \\ \phi_2 &= -0.345 \pm 0.015 \\ \phi_3 &= 0.101 \pm 0.015 \\ \phi_4 &= 0.0251 \pm 0.0077 \\ \sigma_\epsilon &= 0.2670 \pm 0.0088 \end{aligned}$$

### 3.5 AR(5)

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \phi_4 y_{t-4} + \phi_5 y_{t-5} + \sigma_\epsilon \upsilon_t$$

$$\begin{aligned} \phi_1 &= 1.111 \pm 0.011 \\ \phi_2 &= -0.348 \pm 0.016 \\ \phi_3 &= 0.112 \pm 0.015 \\ \phi_4 &= -0.009 \pm 0.011 \\ \phi_5 &= 0.0308 \pm 0.0062 \\ \sigma_\epsilon &= 0.2669 \pm 0.0088 \end{aligned}$$

### 3.6 AR(6)

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \phi_4 y_{t-4} + \phi_5 y_{t-5} + \phi_6 y_{t-6} + \sigma_\epsilon \upsilon_t$$

$$\begin{aligned}
\phi_1 &= 1.110 \pm 0.011 \\
\phi_2 &= -0.348 \pm 0.015 \\
\phi_3 &= 0.109 \pm 0.015 \\
\phi_4 &= 0.001 \pm 0.012 \\
\phi_5 &= -0.003 \pm 0.012 \\
\phi_6 &= 0.0300 \pm 0.0098 \\
\sigma_\epsilon &= 0.2668 \pm 0.0088
\end{aligned}$$

# 4 Discussion

The differences between the performances of the models of subsequent orders are generally small: $\sigma_\epsilon$, which is also the root mean square error of the model reduces by 2.2% from AR(1) to AR(2), by another 0.88% from AR(2) to AR(3), and by only 0.03% from AR(3) to AR(4).

Accordingly, there is little to no justification in calling upon a more complex model than AR(3).

Tests using a normally distributed random series with zero mean and unit variance for $v_t$ indicate that the temperature distributions in the time series and in the generated AR($p$) model series are very similar for $p = 1, \ldots, 6$. As a consequence, even the AR(1) model might be sufficient.