Time series modeling, with application to SDEs and the biological sciences

Peter F. Craigmile

Workshop: Spatially-varying SDEs, with application to the biological sciences
Mathematical Biosciences Institute (MBI), The Ohio State University
7-10 Jul 2015

Thanks to MBI and STATMOS; Supported by the US National Science Foundation (DMS-1407604 and SES-1424481)
Analyzing time series

• A time series is a set of observations made sequentially in “time”.

  – R. A. Fisher: “One damned thing after another”.

• Time series analysis is the area of statistics which deals with analyzing dependencies between different observations in time.

• If we ignore the dependencies that we observe in time series data, then we can be led to incorrect statistical inferences.

• References:

  Good intro text: Brockwell and Davis [2002]

  The theory version: Brockwell and Davis [1991]

  Other popular textbooks: Cryer and Chan [2008], Shumway and Stoffer [2013]

  Older “wordier” texts: Kendall [1973], Diggle [1990], Chatfield [2013]
Time series processes and data

- A **time series process (model)** is a stochastic process \(\{X_t : t \in T\}\) where the index set \(T\) is usually either
  - a subset of the integers, \(\mathbb{Z}\) (**discrete-time time series**), or
  - a subset of the reals, \(\mathbb{R}\) (**continuous-time time series**).

- For these lectures we will focus on discrete-time time series.

- For a given \(\omega\), \(\{x_t\} = \{X_t(\omega)\}\), a realization or sample path of the process, yields our **time series data**.

- Sometimes a time series “model” will only be **summaries** of the process (for example, the means and covariances of the RVs).
Exploratory data analysis (EDA) for time series

- Start with time series plot of \( \{x_t : t = 1, \ldots, n\} \) versus \( t \).

  - Carefully think about the time scale and the axes.

- Look for the following:

  1. Are there changes in behavior over time?

     (e.g., are there shifts in mean and/or variance?)

     This is an indicator of nonstationarity – often we try to model the nonstationarity, remove it, and try to model the remainder.

  2. Are there outliers?

     (unusual values with respect to the rest of the data).
Exploratory data analysis (EDA) for time series, cont.

3. Are some data missing in time?

(complicates EDA and analysis).

4. Is the data non-Gaussian?

Sometimes we can just transform the data.

Other times we need to consider non-Gaussian time series models (e.g., Poisson or binomial time series models)
Classifying features of times series

1. Trend

2. Seasonal or periodic components

3. “The noise” (also called the residual, irregular, error, or random term).

Detecting and modeling changepoints are currently in vogue.
Combining trend, seasonality, and noise

• Most common: The **additive model**, 
  \[ x_t = \mu_t + s_t + \eta_t. \]

• Also common: the **multiplicative** model 
  \[ x_t = \mu_t s_t \eta_t. \]

If all the variables are **positive** then we obtain the additive model by **taking logarithms**: 
  \[ \log x_t = \log \mu_t + \log s_t + \log \eta_t. \]

Often want to **transform back** so that we can interpret on the original measurement scale.
Example: Monthly reported mumps cases (counts) of mumps in New York City

Source: Time Series Data Library
Example: Heart rates (every 15 seconds)

Source: Dave Scarpetti
Objectives of time series analysis

1. Analysis and modeling

2. Forecasting and prediction

3. Adjustment/transformation

4. Simulation/emulation

5. Control
Means and autocovariances

- The **mean function** $\mu_X(\cdot)$ of $\{X_t\}$ is

\[
\mu_X(t) = E(X_t).
\]

Think of $\mu_X(t)$ are being the theoretical mean at time $t$, taken over the possible values that could have generated $X_t$.

- The **autocovariance function** $\gamma_X(\cdot, \cdot)$ of $\{X_t\}$ is

\[
\gamma_X(s, t) = \text{cov}(X_s, X_t) = E[(X_s - \mu_X(s))(X_t - \mu_X(t))].
\]

Measures the strength of linear dependence between the two random variables (RVs) $X_s$ and $X_t$. 
Strictly stationary processes

- In **strict (narrow-sense) stationarity** the joint distribution of a set of RVs are unaffected by time shifts.

- A time series, \( \{ X_t \} \), is **strictly stationary** if

\[
(X_{t_1}, \ldots, X_{t_n}) =_d (X_{t_1+h}, \ldots, X_{t_n+h})
\]

for all integers \( h \) and \( n \geq 1 \), and time points \( \{ t_k \} \).

- Then:

1. \( \{ X_t \} \) is identically distributed
   - Not necessarily independent!

2. \( (X_t, X_{t+h}) =_d (X_1, X_{1+h}) \) for all \( t \) and \( h \);

3. When \( \mu_X \) is finite, \( \mu_X(t) = \mu_X \) is independent of time \( t \).

4. When the variance function exists,

\[
\gamma_X(s, t) = \gamma_X(s + h, t + h)
\]

for any \( s, t \) and \( h \).
(Weakly) stationary processes

• \( \{X_t\} \) is (weakly) stationary if

1. \( E(X_t) = \mu_X(t) = \mu_X \) for some constant \( \mu_X \) which does not depend on \( t \).

2. \( \text{cov}(X_t, X_{t+h}) = \gamma_X(t, t+h) = \gamma_X(h) \), a finite constant that can depend on the lag \( h \) but \textbf{not} on \( t \).

• Also called: \textit{second-order}, \textit{covariance}, \textit{wide-sense}.

• The \textbf{autocorrelation function (ACF)} of \( \{X_t\} \) is \( \rho_X(h) = \gamma_X(h)/\gamma_X(0) \).

• Relating weak and strict stationarity:

1. A strictly stationary process \( \{X_t\} \) is also weakly stationary as long as \( \mu_X(t) \) is finite for all \( t \).

2. \textbf{Weak} stationarity does \textbf{not} imply \textbf{strict} stationarity!
Gaussian processes

- \{X_t\} is a **Gaussian process**, if the joint distribution of any collection of the RVs has a multivariate normal (Gaussian) distribution.

- In this case, the distribution is completely characterized by \( \mu_X(\cdot) \) and \( \gamma_X(\cdot, \cdot) \). The joint probability density function of \( X = (X_1, \ldots, X_n)^T \) is

\[
f_X(x) = (2\pi)^{-n/2} \det(\Sigma)^{-1/2} \times \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)\right),
\]

where \( \mu = (\mu_1, \ldots, \mu_n)^T \) and the \((i, j)\) element of the covariance matrix \( \Sigma \) is \( \gamma_X(i, j) \).

- If these quantities exist and \( \{X_t\} \) is (weakly) stationary then the process is also strictly stationary with mean \( \mu_X \) and ACVF \( \gamma_X(\cdot) \).
Properties of the autocovariance function (ACVF)

• Suppose \( \{X_t\} \) is a stationary time series with ACVF \( \gamma_X(\cdot) \). Then

1. \( \gamma_X(0) \geq 0 \);

2. \( |\gamma_X(h)| \leq \gamma_X(0) \) for all \( h \);

3. \( \gamma_X(h) = \gamma_X(-h) \) for all \( h \) (\( \gamma_X(\cdot) \) is an even function);

4. \( \gamma_X(\cdot) \) is nonnegative definite.

• This is a strong limitation on the possible functions, \( \gamma_X(\cdot) \), allowed.

• How do we come up with stationary processes? Two possible ways:

1. Use Bochner’s theorem to create stationary processes from a certain class of functions.

2. Build a linear process.
What suffices to be an ACVF?

Bochner’s theorem:

A real-valued function $\gamma_X(\cdot)$ defined on the integers

is the ACVF of a stationary process

if and only if

it is even and nonnegative definite.
The white noise process is a useful building block

- Assume \( E(X_t) = \mu, \text{var}(X_t) = \sigma^2 < \infty \).

- Then \( \{X_t\} \) is a \textbf{white noise} or WN(\( \mu, \sigma^2 \)) process if

\[
\gamma_X(h) = 0,
\]

for \( h \neq 0 \).

- \( \{X_t\} \) is clearly stationary.

- But, distributions of \( X_t \) and \( X_{t+1} \) can be different!

- All \textbf{independent and identically distributed (IID) noise} with finite variance is \textbf{white noise} (but, the converse need not be true).
Linear processes

• **Most common** classes of time series models.

• A set of real-valued constants \( \{\psi_j : j \in \mathbb{Z}\} \) is **absolutely summable** if \( \sum_{j=-\infty}^{\infty} |\psi_j| < \infty \).

• A time series \( \{X_t\} \) is a **linear process** with mean \( \mu \) if we can write it as

\[
X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j},
\]

for all \( t \), where \( \mu \) is a real-valued constant, \( \{Z_t\} \) is a WN(0, \( \sigma^2 \)) process and \( \{\psi_j\} \) is a set of absolutely summable constants;

i.e., we can write \( X_t \) as a linear combination of **all** the noise terms (the past, current and future), plus a mean.

• **Absolute summability** of the constants guarantees infinite sum **converges**.
Example: Moving average process of order $q$, MA($q$)

• Let $\{Z_t\}$ be a WN(0, $\sigma^2$) process. For an integer $q > 0$ and constants $\theta_1, \ldots, \theta_q$ with $\theta_q \neq 0$ define

$$X_t = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}$$

$$= \theta_0 Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}$$

$$= \sum_{j=0}^{q} \theta_j Z_{t-j},$$

where we let $\theta_0 = 1$.

• $\{X_t\}$ is known as the moving average process of order $q$, or the MA($q$) process.

• Clearly, by definition, the MA($q$) process is a linear process.
Defining linear processes with backward shifts

• The **backward shift** operator, $B$, is defined by $BX_t = X_{t-1}$.

• We can represent a linear process using the backward shift operator as

$$X_t = \mu + \psi(B)Z_t,$$

where we let $\psi(B) = \sum_{j=-\infty}^{\infty} \psi_j B^j$.

• $\psi(B)$ is a **polynomial** with the backward shift operator as the argument (not a number).

We will think of it as a type of **linear filter**.

• **Example**: a mean zero MA(1) process can be written as

$$X_t = \mu + \psi(B)Z_t,$$

where $\mu = 0$ and $\psi(B) = 1 + \theta B$. 
Linear filtering preserves stationarity

- Let \( \{Y_t\} \) be a time series and \( \{\psi_j\} \) be a set of absolutely summable constants that does not depend on time.

- **Definition**: A linear time invariant (LTI) filtering of \( \{Y_t\} \) with coefficients \( \{\psi_j\} \) that do not depend on time is defined by

\[
X_t = \psi(B)Y_t,
\]

where again \( \psi(B) = \sum_{j=-\infty}^{\infty} \psi_j B^j \).

- **Theorem**: Suppose \( \{Y_t\} \) is a zero mean stationary series with ACVF \( \gamma_Y(\cdot) \). Then \( \{X_t\} \) is a zero mean stationary process with ACVF

\[
\gamma_X(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_Y(k - j + h).
\]
Example: The MA($q$) process is stationary

- By the filtering preserves stationarity result, the MA($q$) process defined previously is a stationary process with mean zero and ACVF

\[ \gamma_X(h) = \sigma^2 \sum_{j=0}^{q} \theta_j \theta_{j+h} \]

(let $\theta_j = 0$ for $j < 0$ and $j > q$ for this to make sense).

- Let $I(A)$ denote the **indicator function** defined by

\[
I(A) = \begin{cases} 
1, & \text{if the event } A \text{ is true;} \\
0, & \text{if the event } A \text{ is false.}
\end{cases}
\]

Then

\[
\gamma_X(h) = \sum_{j=0}^{q} \sum_{k=0}^{q} \theta_j \theta_k \gamma_Z(j - k + h)
\]

\[
= \sigma^2 \sum_{j=0}^{q} \sum_{k=0}^{q} \theta_j \theta_k I(k = j + h) = \sigma^2 \sum_{j=0}^{q} \theta_j \theta_{j+h}.
\]
Another example: Differencing an MA(1) process

• Suppose $Y_t$ is an MA(1) process defined in the usual way. Let $X_t = Y_t - Y_{t-1}$ denote the differenced MA(1) process.

1. Is $\{X_t\}$ stationary?

2. What is the ACVF of $\{X_t\}$?
Processes with a correlation/dependence that cuts off

- A time series \( \{X_t\} \) is \textit{q-correlated} if

\[
X_t \text{ and } X_s \text{ are uncorrelated for all } |t - s| > q,
\]

i.e., \( \text{cov}(X_t, X_s) = 0 \) for all \( |t - s| > q \).

- \textbf{Theorem:} if \( \{X_t\} \) is a \textit{stationary q-correlated} time series with a \textit{zero mean},

then it can always be represented as an \textbf{MA}(q) process.

- A time series \( \{X_t\} \) is \textit{q-dependent} if

\[
X_t \text{ and } X_s \text{ are independent for all } |t - s| > q,
\]
**Examples**

1. A WN process is 0-correlated.

2. An IID noise process, \( \{Z_t\} \), is 0-dependent.

3. An MA(1) process is 1-correlated.

4. Is an AR(1) process \( q \)-correlated?

5. A process, \( \{Z_t Z_{t-1}\} \), where \( \{Z_t\} \) is IID noise, is
The autoregressive process of order $p$, $\text{AR}(p)$

- This process is attributed to George Udny Yule (1871–1951) (See the biography by O’Connor and Robertson and look at Yule [1921] and Yule [1927]. The AR(1) process has also been called the Markov process).

- Let $\{Z_t\}$ be a WN($0, \sigma^2$) process and let $\{\phi_1, \ldots, \phi_p\}$ be a set of constants for some integer $p > 0$. Assume $\phi_p \neq 0$.

- The $\text{AR}(p)$ process is defined to be the solution to the equation

\[ X_t = \phi(B)X_t = Z_t, \]

\[ \text{i.e., } \sum_{j=1}^{p} \phi_j X_{t-j} + Z_t. \]

where we let $\phi(B) = 1 - \sum_{j=1}^{p} \phi_j B^j$. 

The autoregressive process of order $p$, $\text{AR}(p)$
A stationary solution

- In general we want the solution to the AR equation to yield a stationary process.

  - This will restrict the possible \{\phi_j\} values allowed.

- Consider the AR(1) process.

  We will demonstrate that a stationary solution exists for \(|\phi_1| < 1\). To prove this we need the following math result regarding geometric series: For \(|a| < 1\),

  \[
  \sum_{j=0}^{\infty} a^j = \frac{1}{1 - a}.
  \]
AR(1) example, continued

• We first write

\[ X_t = \phi_1 X_{t-1} + Z_t = \phi_1 (\phi_1 X_{t-2} + Z_{t-1}) + Z_t \]

\[ = \phi_1^2 X_{t-2} + \phi_1 Z_{t-1} + Z_t \]

\[ \vdots \]

\[ = \sum_{j=0}^{\infty} \phi_1^j Z_{t-j}, \]

suitably neglecting the influence of \( X_{t-h} \) as \( h \to \infty \).

• Now let \( \psi_j = \phi_1^j \) for each \( j \). We then have

\[ X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}. \]

• Using the math result, for \( |\phi_1| < 1 \), the sequence \( \{\psi_j\} \) is absolutely summable.
AR(1) example, continued

• Thus, since \{X_t\} is a \textbf{linear} process, it follows by the filtering preserves stationarity result that \{X_t\} is a zero mean stationary process with ACVF

\[
\gamma_X(h) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h} \\
= \sigma^2 \sum_{j=0}^{\infty} \phi_1^j \phi_1^{j+h} \\
= \sigma^2 \phi_1^h \sum_{j=0}^{\infty} (\phi_1^2)^j.
\]

• Now \(|\phi_1| < 1\) implies that \(|\phi_1^2| < 1\) and so using our math result again with \(a = \phi_1^2\),

\[
\gamma_X(h) = \frac{\sigma^2 \phi_1^h}{1 - \phi_1^2}.
\]

• Need a little more work to prove that this is the \textbf{unique} stationary solution.
Other values for the AR(1) parameter

- No stationary solutions exist for $|\phi_1|=1$ (forms of random walk).

- Now consider for $|\phi_1| > 1$. Remember we have

\[ X_t = \phi_1 X_{t-1} + Z_t. \]

Dividing by $\phi_1$ we get

\[ \phi_1^{-1} X_t = X_{t-1} + \phi_1^{-1} Z_t, \]

i.e., \[ X_{t-1} = \phi_1^{-1} X_t - \phi_1^{-1} Z_t. \]

- Can write this as a linear combination of future $Z_t$’s. As this is also a filtering of a stationary process we have a stationary solution.

  - BUT, $X_t$ depends on future values of the $\{Z_t\}$ – not very practical!

- If we assume that $X_s$ and $Z_t$ are uncorrelated for each $t > s$, $|\phi_1| < 1$ is the only stationary solution to the AR equation.
Two interesting AR processes

- Two (commonly-used) stationary examples are:

  \[ X_t = 0.75X_{t-1} - 0.5X_{t-2} + Z_t. \]

  \[ X_t = 2.7607X_{t-1} - 3.8106X_{t-2} + 2.6535X_{t-3} - 0.9238X_{t-4} + Z_t. \]

- These realizations exhibit quasi-periodic behavior.
Example: Canadian lynx counts

Annual number of Canadian lynx trapped around the MacKenzie River between 1821 and 1934 [Elton and Nicholson, 1942]
How might you model the dynamics of Canadian lynx?
Approximating SDEs with an AR process

• Consider the informal form of a stochastic differential equation (SDE) from yesterday:

\[ dX_t = s(X_t)dt + \sigma(X_t)dW_t, \quad X_{t=0} = x_0. \]

• Remember \( s(\cdot) \) is the drift function, \( \sigma(\cdot) \) is the diffusion function and \( \{W_t\} \) is a standard Brownian motion (BM).

• Consider a fine time grid: \( 0 = t_0 < t_1 < \ldots t_N = T \) with \( t_{k+1} - t_k = \Delta. \)
The Euler approximation

• On this grid the Euler approximation is

\[ X_{t_{k+1}} \approx X_{t_k} + s(X_{t_k})\Delta + \sigma(X_{t_k}) U_{t_{k+1}} \]

where \( \{U_{t_k}\} \) is a set of IID N(0, \( \Delta \)) RVs.

• When does an AR(1) process provide a good approximation?

\[ X_t = \phi X_{t-1} + Z_t. \]
**ARMA**(\(p, q\)) processes

- Assume \(\{Z_t\}\) is a WN(0, \(\sigma^2\)) process. Then \(\{X_t\}\) is an **ARMA**(\(p, q\)) process if it satisfies

\[
X_t - \sum_{j=1}^{p} \phi_j X_{t-j} = Z_t + \sum_{j=1}^{q} \theta_j Z_{t-j},
\]

assuming \(\phi_p \neq 0\) and \(\theta_q \neq 0\).

- Let

\[
\phi(B) = 1 - \sum_{j=1}^{p} \phi_j B^j \quad \text{and} \quad \theta(B) = 1 + \sum_{j=1}^{q} \theta_j B^j.
\]

Then we can write

\[
\phi(B)X_t = \theta(B)Z_t.
\]

- To add a constant mean term \(\mu\), let \(Y_t = X_t + \mu\) for each \(t\). Equivalently \(\{Y_t\}\) is an ARMA(\(p,q\)) process with mean \(\mu\) if

\[
\phi(B)(Y_t - \mu) = \theta(B)Z_t.
\]
A stationary solution to the ARMA equation

- Why use (stationary) ARMA?
  - Stationary time series may be described by an stationary ARMA process using fewer parameters than if we used only a stationary AR or MA process.

- The zero mean ARMA process is stationary if can write it as a linear process, \( X_t = \psi(B)Z_t \), where \( \psi(B) = \sum_{j=-\infty}^{\infty} \psi_j B^j \) for an absolutely summable sequence \( \{\psi_j\} \).

- This only happens if we can divide by \( \phi(B) \) in some sense; i.e., it is stationary only if this makes sense:
  \[
  (\phi(B))^{-1}\phi(B)X_t = (\phi(B))^{-1}\theta(B)Z_t.
  \]

- Forget about the backshift operator \( B \) now – we replace with a number \( z \) and consider whether we can divide \( \theta(z) \) by \( \phi(z) \).
Polynomial roots

- A polynomial of degree $p$ has the form,

$$f(z) = \sum_{j=0}^{p} a_j z^j.$$ 

- A root of the polynomial is a value $\zeta$, such that $f(\zeta) = 0$.
  
  - Problem: $\zeta$ can be real-valued or complex-valued!

- Aside: Let $i$ be the square root of negative one, $i = \sqrt{-1}$.

  Then any complex number $\zeta$ can be written as $\zeta = a + b \, i$, for real numbers $a$ and $b$.

  The modulus of a complex number $|\zeta|$ is defined by $|\zeta| = \sqrt{a^2 + b^2}$. 

Checking stationarity in general

• For any ARMA($p,q$) process, a **stationary** and **unique** solution exists if and only if

\[ \phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p \neq 0, \]

for all $|z| = 1$.

• Note: Stationarity of the ARMA process has **nothing** to do with the MA polynomial!

• General strategy:

1. Find the **roots** of $\phi(\cdot)$,

   i.e., the values of $\zeta$ such that $\phi(\zeta) = 0$.

2. Show that the roots **do not** have modulus 1;

   i.e., show $|\zeta| \neq 1$ for each root $\zeta$. 
AR(4) example

• Consider the following AR(4) example:

\[ X_t = 2.7607X_{t-1} - 3.8106X_{t-2} + 2.6535X_{t-3} - 0.9238X_{t-4} + Z_t, \]

where \( \{Z_t\} \) is a WN process, with AR polynomial

\[ \phi(z) = 1 - 2.7607z + 3.8106z^2 - 2.6536z^3 + 0.9238z^4. \]

• Hard to find the roots numerically – we use the `polyroot` function in R:

```r
# calculate the roots of the polynomial, and store in 'zetas'
zetas <- polyroot(c(1, -2.7607, 3.8106, -2.6535, 0.9238))
[1] 0.6499635+0.7859373i 0.7862240+0.6500408i 0.6499635-0.7859373i
[4] 0.7862240-0.6500408i

# calculate the modulus of the roots, zetas
Mod(zetas)
[1] 1.019877 1.020148 1.019877 1.020148
```
Causal and invertible ARMA processes

• An ARMA process is **causal** if there exists constants \( \{\psi_j\} \) with \( \sum_{j=0}^{\infty} |\psi_j| < \infty \) and

\[
X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j};
\]

that is, we can write \( \{X_t\} \) as an MA\((\infty)\) process depending **only** on the current and past values of \( \{Z_t\} \); equivalently the process is **causal** if and only if

\[
\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p \neq 0, \quad \text{for all } |z| \leq 1.
\]

• An ARMA process is **invertible** if there exists constants \( \{\pi_j\} \) with \( \sum_{j=0}^{\infty} |\pi_j| < \infty \) and

\[
Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j};
\]

i.e., we can write \( \{Z_t\} \) as an MA process depending **only** on the current and past values of \( \{X_t\} \); equivalently the process is **invertible** if and only if

\[
\theta(z) = 1 + \theta_1 z + \ldots + \theta_p z^q \neq 0, \quad \text{for all } |z| \leq 1.
\]
References


