Models of Seasonality

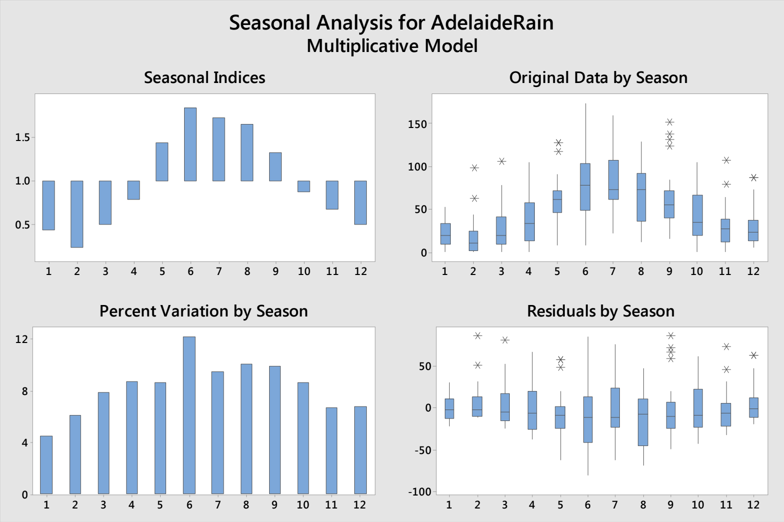
John Boland
University of South Australia

2022

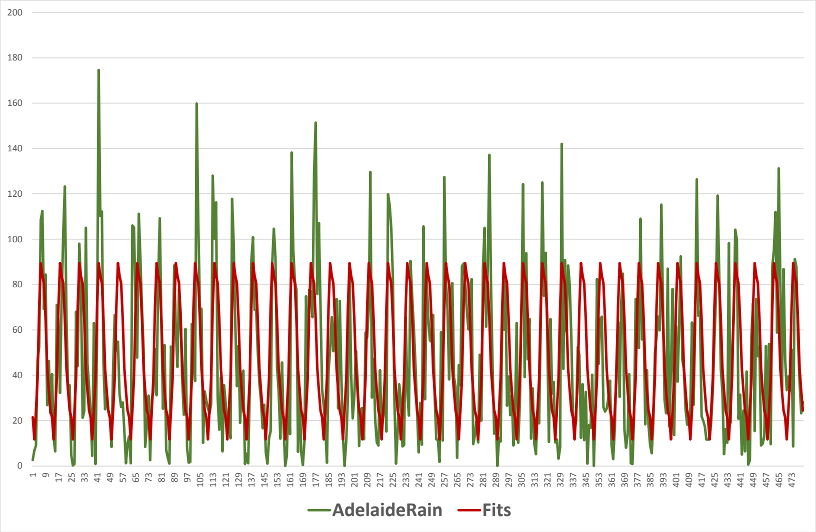
Dealing with Seasonality in Time Series

- One of the initial phases of time series analysis is to identify and somehow model the seasonal component of the series, if one exists.
- Then, one removes that seasonality, in order to transform the original series into a stationary one, in preparation for further analysis.
- There are two forms of ridding the original of seasonality, by either dividing by appropriate seasonal indices or by subtracting them.
- In other words, there are multiplicative and additive processes, and either one may be more appropriate for a particular situation.

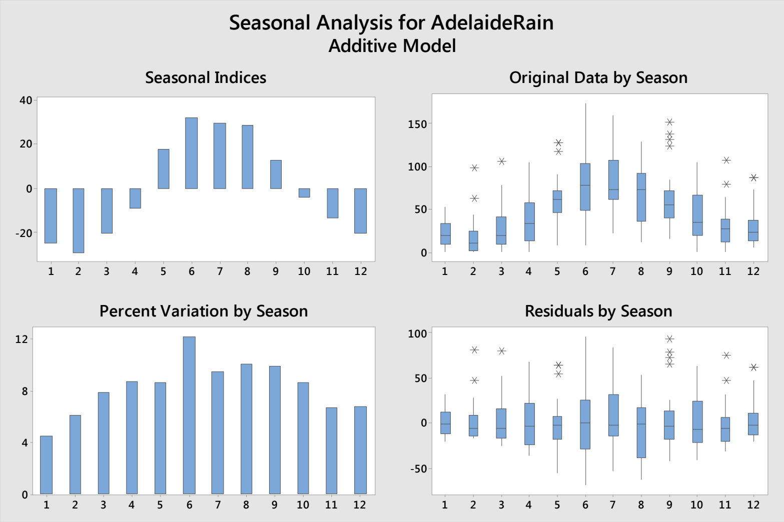
Multiplicative Deseasoning - Monthly Rainfall



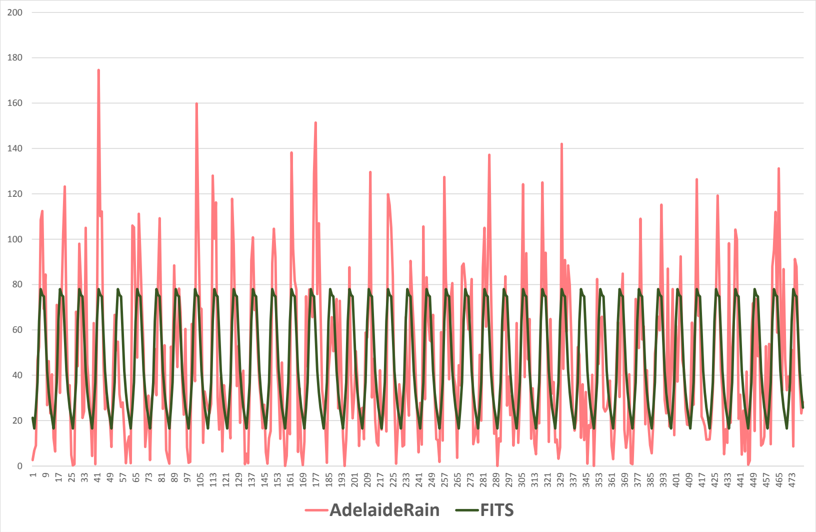
Multiplicative Deseasoning - Monthly Rainfall - Fits



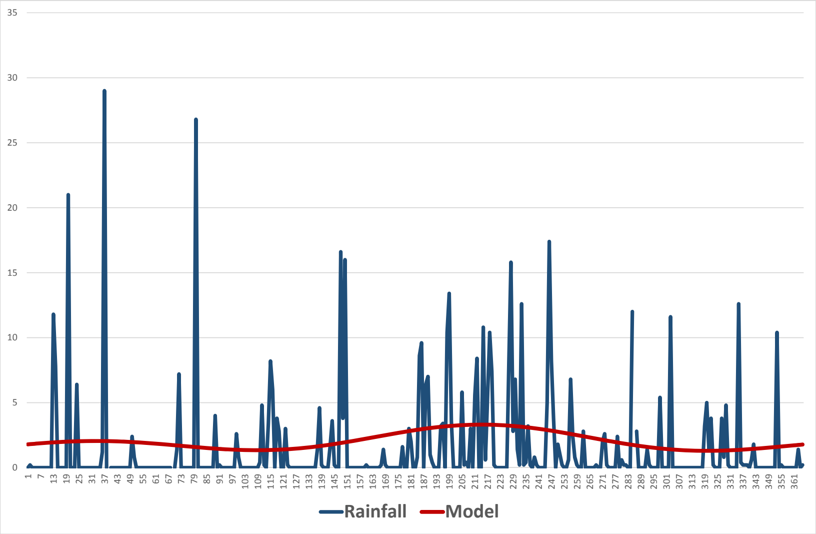
Additive Deseasoning - Monthly Rainfall



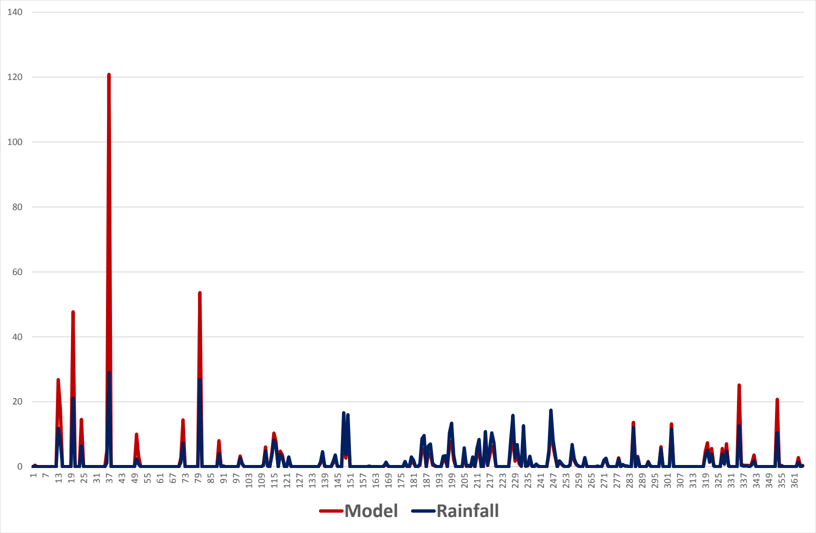
Additive Deseasoning - Monthly Rainfall - Fits



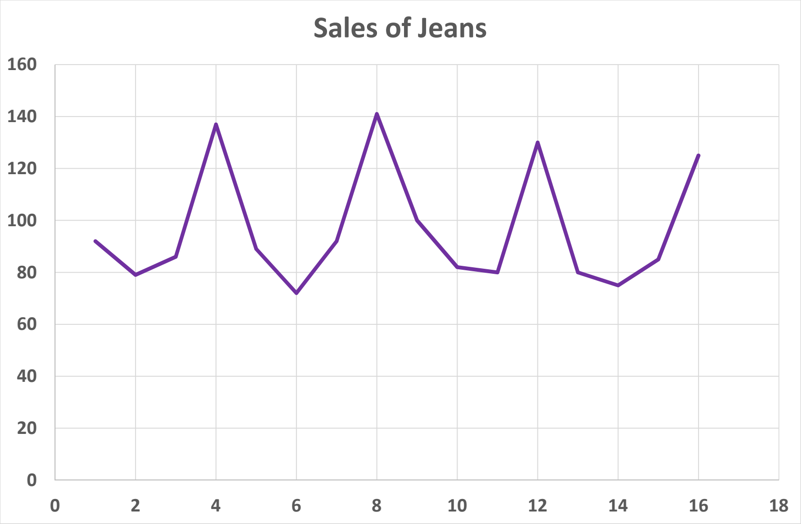
Additive Deseasoning - Daily Rainfall - Fits



Multiplicative Deseasoning - Daily Rainfall - Fits



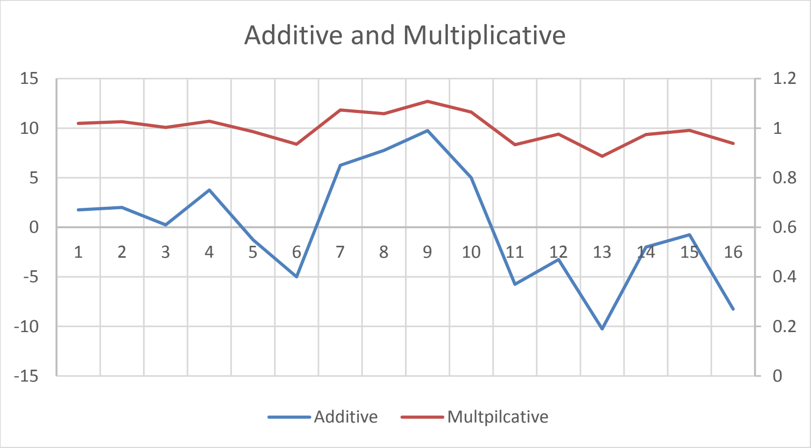
Seasonally Adjusted Data - Simple Example



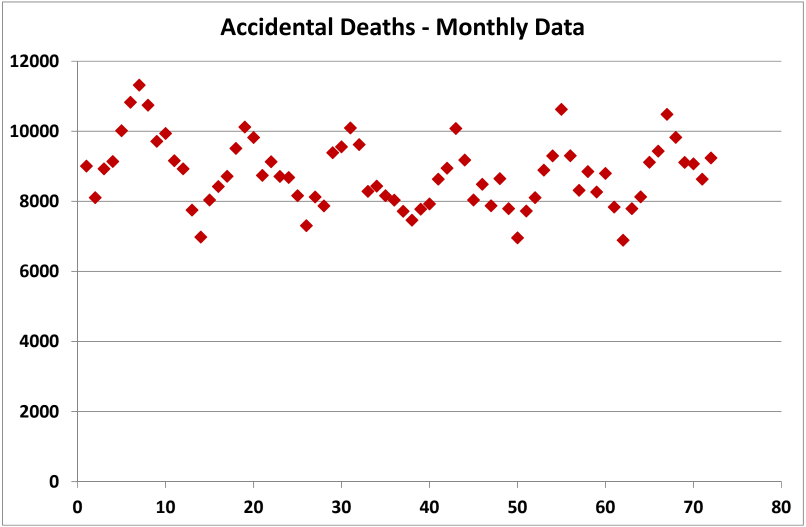
Method

- Form the mean for each season.
- $\bar{x}_i = \frac{1}{4}\sum_{j=1}^{4} x_{i+4j}$.
- $\hat{x}_i = x_i - \bar{x}_i$ for additive.
- $\hat{x}_i = \frac{x_i}{\bar{x}_i}$ for multiplicative.

Deseasoned Data



Seasonality for Periodic Data



Fourier Series Model for Seasonality

Any periodic function satisfies the relation:$f(t) = f(t+T)$ where T is the period of the function.
The simplest and most common periodic functions are the trigonometric functions. The functions $\sin nt$ and $\cos nt, \quad n \in 0, 1, 2, 3...$ being harmonic are by definition periodic with periods $\frac{2\pi}{n}$. Because trigonometric functions are relatively easy to work with and because they possess the important property of orthogonality they are useful to represent the seasonality of data of this type.

It is easy to verify that the set of functions

$$\frac{1}{\sqrt{2\pi}}, \frac{\sin t}{\sqrt{\pi}}, \frac{\cos t}{\sqrt{\pi}}, \frac{\sin 2t}{\sqrt{\pi}}, \frac{\cos 2t}{\sqrt{\pi}}, \frac{\sin 3t}{\sqrt{\pi}} \dots \tag{1}$$

constitutes an orthonormal set.

Because this set of functions is complete in the interval $0 \leq t \leq 2\pi$ every function $f(t)$ which is continuous in that interval can be represented by the *Fourier Series*

$$f(t) = \frac{1}{2}a_0 + \sum_{r=1}^{\infty}(a_r \cos(rt) + b_r \sin(rt))$$

where the constants a_r and b_r are known as the Fourier coefficients.

Now, if we multiply both sides by $\cos st$ and integrate we get

$$\begin{aligned}
\int_0^{2\pi} f(t)\cos st\,dt &= \frac{1}{2}a_0\int_0^{2\pi}\cos st \\
&+ \sum_{r=1}^{\infty} a_r\int_0^{2\pi}\cos(rt)\cos st\,dt \\
&= \sum_{r=1}^{\infty} b_r\int_0^{2\pi}\sin(rt)\cos st\,dt
\end{aligned}$$

For $s = 0$, this yields $a_o = 1/\pi \int_0^{2\pi} f(t)dt$ so a_o may be referred to as the average value of $f(t)$. If $s \neq 0$ then it reduces to $a_r = 1/\pi \int_0^{2\pi} f(t) \cos rt dt$ and $b_r = 1/\pi \int_0^{2\pi} f(t) \sin rt dt$. When $f(t)$ is an even series, that is,$f(t)$ is reflected at the origin, the b coefficients vanish and the series is known as a Fourier cosine series. Similarly when the series is an odd series, $f(t) = -f(-t)$ the a coefficients vanish and the series is known as a Fourier sine series.

Discrete Fourier Transform DFT

- What we need for our purposes is the discrete version of the series representation.
- We also need to adjust the period, where in the case of the deaths data, the fundamental period is 12 months.
- We will also consider how to find how many terms we use in the expansion.

The discrete Fourier Transform of a sequence of N real or complex numbers, $x_0, x_1, \ldots, x_{N-1}$ is the sequence of complex numbers $c_0, c_1, \ldots, c_{N-1}$ defined by

$$c_j = \frac{1}{N} \sum_{n=0}^{N-1} x_n e^{-2\pi j n / N}, j = 0, 1, \ldots, N-1$$

The original x_j can be recovered using the Inverse DFT or IDFT, defined by

$$x_j = \frac{1}{N} \sum_{n=0}^{N-1} c_n e^{2\pi j n/N}, j = 0, 1, \ldots, N-1$$

IDFT for real numbers

Assume $x_0, x_1, \ldots, x_{N-1}$ are real numbers and thus $c_{N-j} = c_j^*$.

$$\begin{aligned} x_j &= a_0 \\ &+ \sum_{n=1}^{N/2-1} \left(a_n \cos(2\pi jn/N) + b_n \sin(2\pi jn/N)\right) \\ &+ a_{N/2}(-1)^j, \qquad \text{for N even} \end{aligned}$$

$$\begin{aligned} x_j &= a_0 \\ &+ \sum_{n=1}^{(N-1)/2} \left(a_n \cos(2\pi jn/N) + b_n \sin(2\pi jn/N)\right), \text{for N odd.} \end{aligned}$$

Power Spectrum

- Note that is we used the formulae above we would capture all the values.
- But, the Principle of Parsimony leads us to construct the model that fits the data best with the least number of coefficients.
- Note that this will also the one that is most identifiable in a physical sense - it has meaning.
- We do this by identifying the most important frequencies to include via using the so-called Power Spectrum.

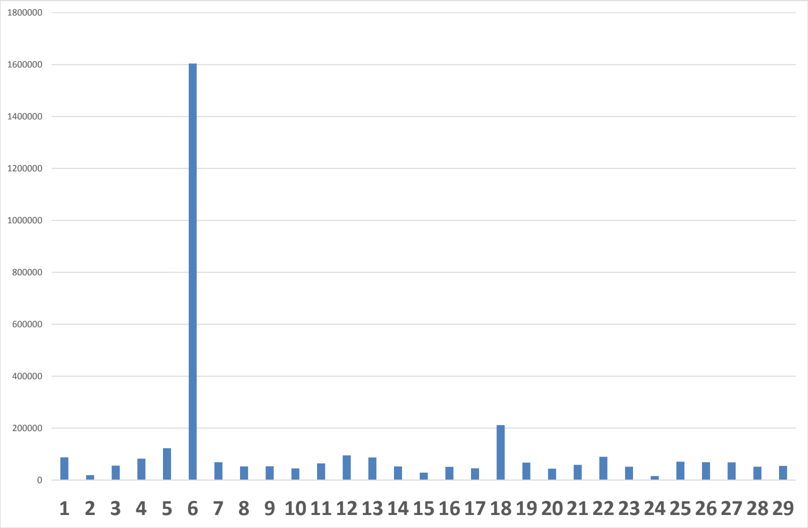
Power Spectrum or Periodogram

I will use a particular form that has useful corresponding meaning:

$$I_j = (a_j^2 + b_j^2)/2$$

In this form, it also gives the variance explained by the inclusion of that frequency in the Fourier series.

Power Spectrum of the Deaths Data



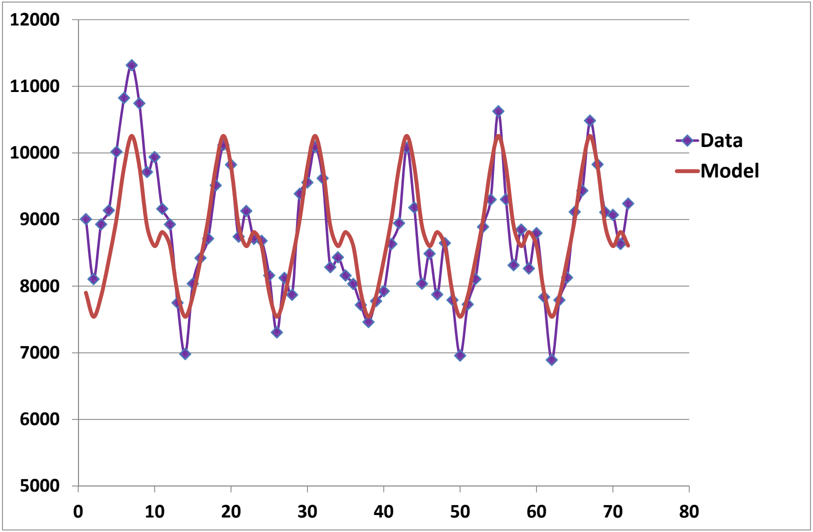
Fourier Series Representation

$$
\begin{aligned}
d_t &= a_0 \\
&+ a_1 \cos(2\pi t/12) + b_1 \sin(2\pi t/12) \\
&+ a_2 \cos(4\pi t/12) + b_2 \sin(4\pi t/12) \\
&+ a_3 \cos(6\pi t/12) + b_3 \sin(6\pi t/12)
\end{aligned}
$$

Fourier Series Representation

$$
\begin{aligned}
d_t &= 8787.7 \\
&+ -734.0\cos(2\pi t/12) - 711.6\sin(2\pi t/12) \\
&+ 409.3\cos(4\pi t/12) + 99.2\sin(4\pi t/12) \\
&+ 145.1\cos(6\pi t/12) - 185.6\sin(6\pi t/12)
\end{aligned}
$$

ModelFit



12000
11000
10000
9000
8000
7000
6000
5000
0
10
20
30
40
50
60
70
80
Data
Model

Methods of Estimating Parameters

- Direct calculation.
- Optimisation.
- Linear Regression.
- I will show how to do all of these, but in the next slides I will show a particular form of how to use regression for this.

Normal Equations for Regression

$$\begin{aligned}\hat{D}(t) &= a_0 + a_1 cos(2\pi t/12) + a_2 sin(2\pi t/12) \\ &+ a_3 cos(4\pi t/12) + a_4 sin(4\pi t/12) \\ &+ a_5 cos(6\pi t/12) + a_6 sin(6\pi t/12) \qquad (2)\end{aligned}$$

Let $x_1 = cos(2\pi t/12), x_2 = sin(2\pi t/12), x_3 = cos(4\pi t/12), x_4 = sin(4\pi t/12), x_5 = cos(6\pi t/12), x_6 = sin(6\pi t/12)$. Therefore we can rewrite [2] as

$$\hat{D}(t) = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + a_4 x_4 + a_5 x_5 + a_6 x_6 \qquad (3)$$

Derivation

The problem is to find $a_0, a_1, \ldots a_6$ to minimise the sum of squared deviations between the model and the data over the day [4].

$$E = \sum_{t=t_0}^{t_n} (\hat{S}(t) - S(t))^2 = \sum_{t=t_0}^{t_n} (a_0 + \sum_{j=1}^{6} a_j x_j(t) - S(t))^2 \quad (4)$$

To calculate $a_0, a_1, \ldots a_6$, we take the partial derivatives of E with respect to each and set these to zero to form a system of linear equations which we then solve.

$$
\begin{aligned}
\frac{\partial E}{\partial a_0} &= 2\sum_{t=t_0}^{t_n}(a_0 + \sum_{j=1}^{6} a_j x_j(t) - S(t)) = 0 \\
\frac{\partial E}{\partial a_1} &= 2\sum_{t=t_0}^{t_n} x_1(a_0 + \sum_{j=1}^{6} a_j x_j(t) - S(t)) = 0 \\
\ldots &= \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
\frac{\partial E}{\partial a_6} &= 2\sum_{t=t_0}^{t_n} x_4(a_0 + \sum_{j=1}^{6} a_j x_j(t) - S(t)) = 0
\end{aligned}
$$

This set of equations can be written in the form [5]

$$A\Phi = B \tag{5}$$

$A(7,7) =$

$$\begin{bmatrix} n & \sum_{t=t_0}^{t_n} x_1(t) & \dots & \sum_{t=t_0}^{t_n} x_6(t) \\ \sum_{t=t_0}^{t_n} x_1(t) & \sum_{t=t_0}^{t_n} x_1^2(t) & \dots & \sum_{t=t_0}^{t_n} x_1(t)x_6t) \\ \vdots & \ddots & \ddots & \vdots \\ \sum_{t=t_0}^{t_n} x_6(t) & \sum_{t=t_0}^{t_n} x_1(t)x_6(t) & \dots & \sum_{t=t_0}^{t_n} x_6^2(t) \end{bmatrix}$$

$$\Phi = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_6 \end{bmatrix}$$

$$B = \begin{bmatrix} \sum_{t=t_0}^{t_n} S(t) \\ \sum_{t=t_0}^{t_n} x_1(t)S(t) \\ \vdots \\ \sum_{t=t_0}^{t_n} x_6(t)S(t) \end{bmatrix}$$

Therefore, the solution is

$$\Phi = A^{-1}B \tag{6}$$

Intraday Solar Radiation

- One can deal with the seasonalities inherent in climate variable in a multiplicative or additive modelling framework.
- Interestingly, though I will argue that the multiplicative approach is problematic, it is the approach most often used for solar forecasting models in particular. There are two versions of the multiplicative approach with respect to solar radiation, calculating the clearness index, and estimating the clear sky index.
- To form the clearness index, one divides the global solar radiation by the extraterrestrial radiation, a quantity determined only via astronomical formulae. On the other hand the clear sky index involves dividing the global radiation by a clear sky model.

Continued

- Note that the wind resource is not as seasonally dependent as the solar radiation, and both multiplicative and additive versions of dealing with seasonality are used.
- Additive de-seasoning is enacted through subtracting a mean function from the solar radiation, that function formed usually through the addition of terms involving a basis of the function space.
- I will argue that an appropriate way to perform this operation is through the use of a Fourier set of basis functions.

Seasonality modelling methods

- Multiplicative, dividing the data by clearness index.
- Multiplicative, dividing the data by clearness sky model.
- Additive, using Fourier series or wavelets.

Clearness Index - advantages and disadvantages

- It is a calculated, not modelled value as the divisor is the so-called extraterrestrial radiation H_t.
- H_t is the result of a calculation using spherical trigonometry applied to the 'solar constant'.
- From my experience it does not pick up all the seasonal effects.

Clear Sky Index - advantages and disadvantages

- It is based on a physical model of a so-called clear sky, and varies throughout the year.
- It uses some parameter values that vary at high frequency but are assumed to vary at low frequency.
- There are several clear sky models - Ineichen evaluates thirteen.

Maximum value of Unity?

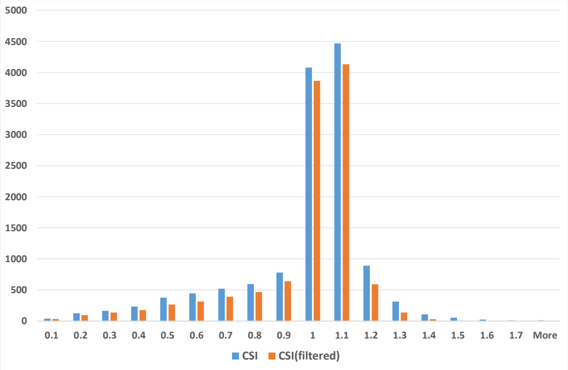


Figure: CSI values for Los Vegas

Horns

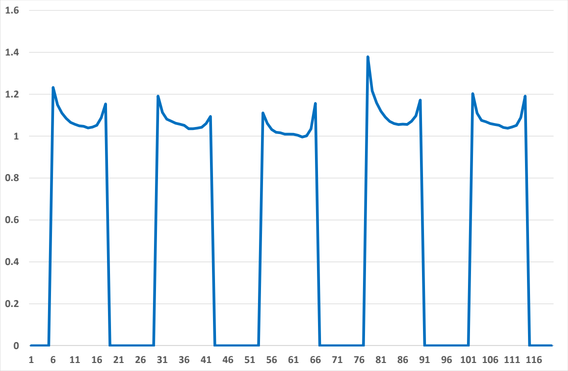
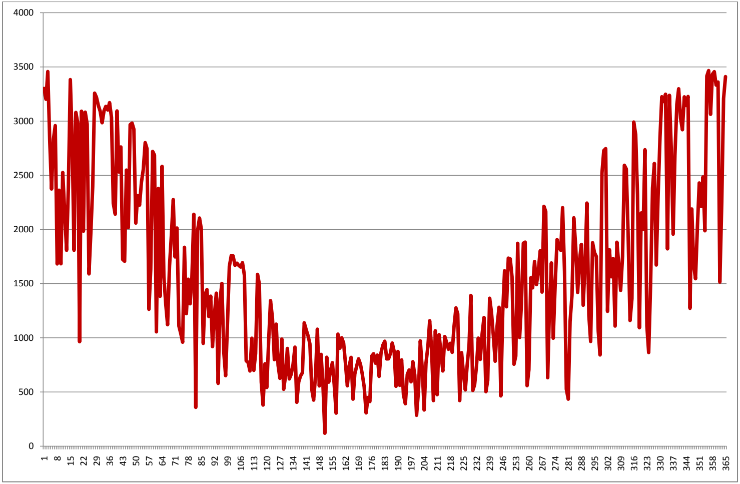


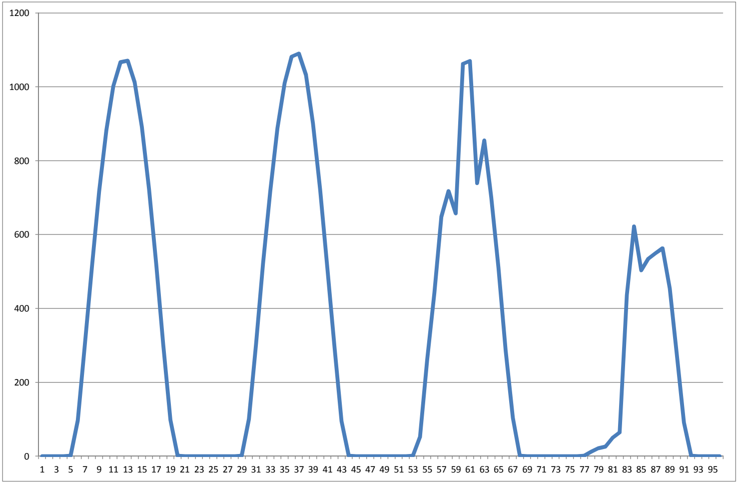
Figure: Five days of CSI for Los Vegas

Forecasting solar radiation on a short time scale - for example one hour

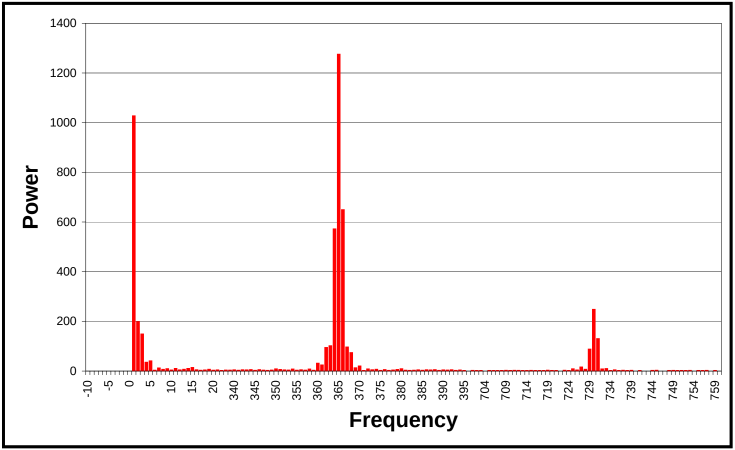
Daily Seasonality



Hourly Seasonality



Power Spectrum



Seasonality

- The first step is to identify and model the seasonality. We have identified several significant cycles using spectral analysis. Fourier series will be used in this step.

$$
\begin{aligned}
S_t \quad = \quad & \alpha_0 + \alpha_1 \times \cos\frac{2\pi t}{8760} + \beta_1 \times \sin\frac{2\pi t}{8760} + \\
& \alpha_2 \times \cos\frac{4\pi t}{8760} + \beta_2 \times \sin\frac{4\pi t}{8760} + \\
& \sum_{i=3}^{11}\sum_{n=1}^{3}\sum_{m=-1}^{1}[\alpha_i \times \cos\frac{2\pi(365n+m)t}{8760} + \\
& \beta_i \times \sin\frac{2\pi(365n+m)t}{8760}]
\end{aligned}
$$

  here S_t is seasonal component.

- Note that in examples we have tested, the amount of the variance explained by the Fourier Series is approximately 80-85%.

Amplitude Modulation

- With solar radiation series one significant aspect is the varying amplitude of the daily cycle as one progresses through the year.
- The amplitude is higher in summer than winter but the progression is systematic over the year.
- The inclusion of the beat frequencies in the Fourier series specifically captures this amplitude modulation, and this is a compelling reason for the representation of the series via this method.

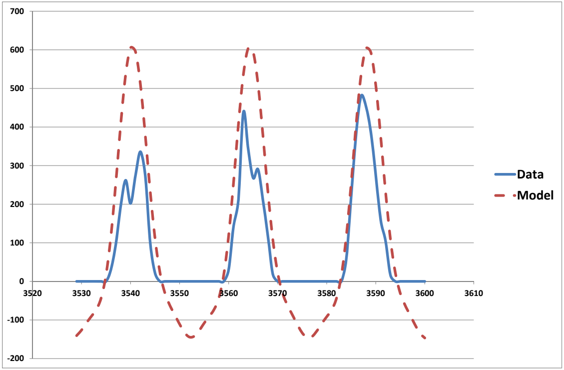


Figure: Effect of no sidebands on winter solar radiation

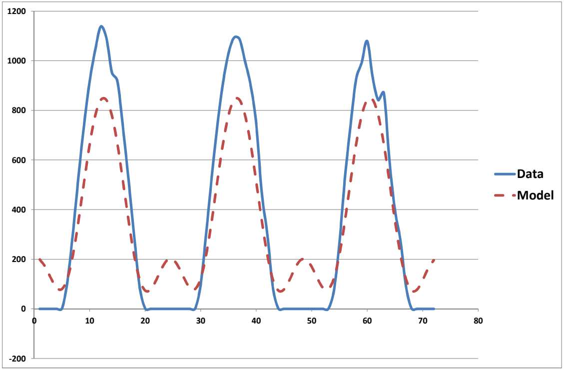


Figure: Effect of no sidebands on summer solar radiation

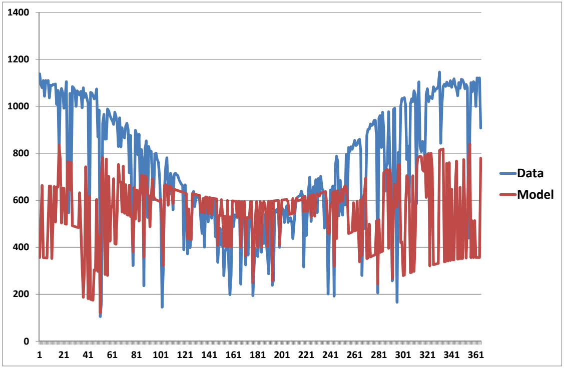


Figure: Amplitude with no sidebands

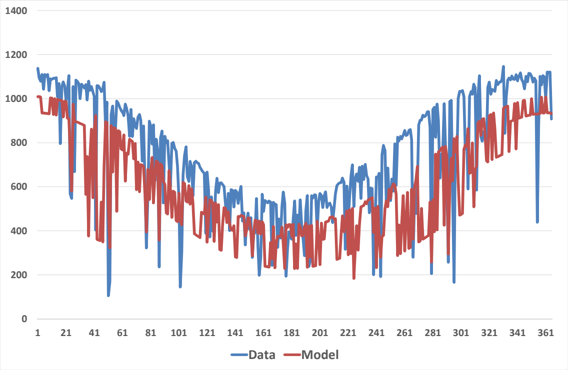


Figure: Amplitude with sidebands included

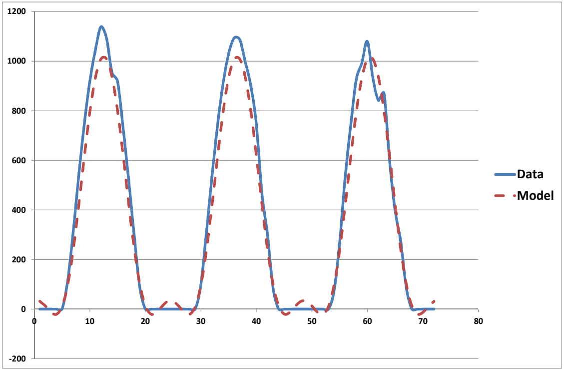


Figure: Effect of including sidebands on summer solar radiation

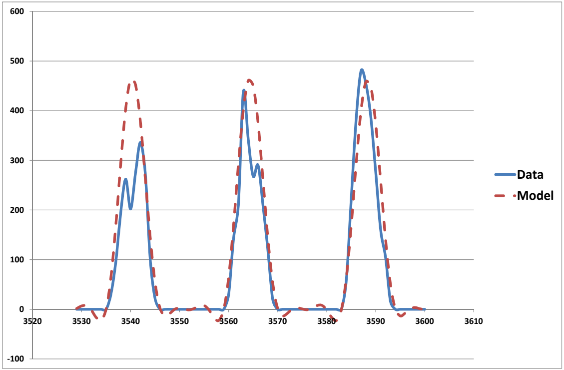


Figure: Effect of including sidebands on winter solar radiation

Latitude Effects

- There is an interesting contrast in the analysis for a tropical location, the island of Desirade, part of Guadeloupe in the French West Indies, latitude 16.32^o.
- Inspection of the power spectrum hints at the fact that there may not be a significant change in the daily amplitude over the year.
- There are no apparent sidebands present in this graph.

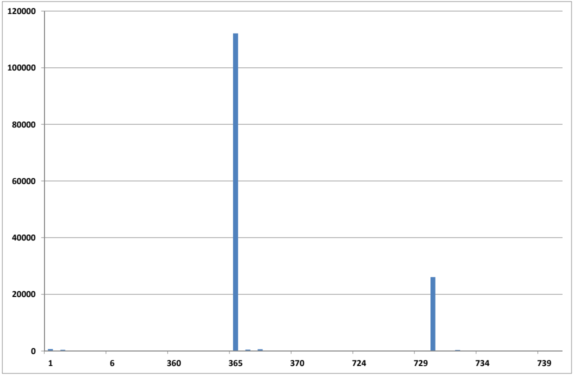


Figure: Power spectrum for Desirade

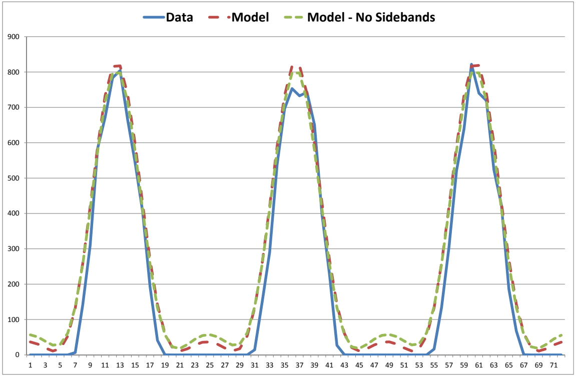


Figure: Comparison of Fourier series model with and without sidebands - summer in Desirade

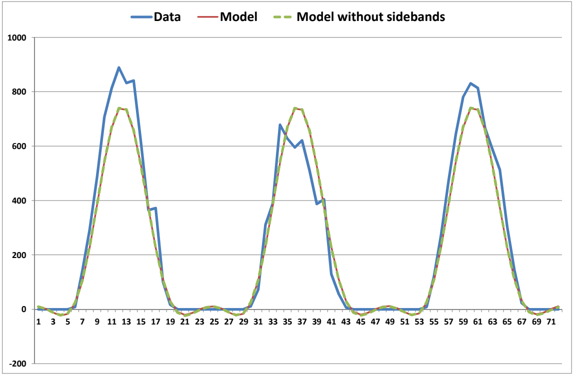


Figure: Comparison of Fourier series model with and without sidebands - winter in Desirade

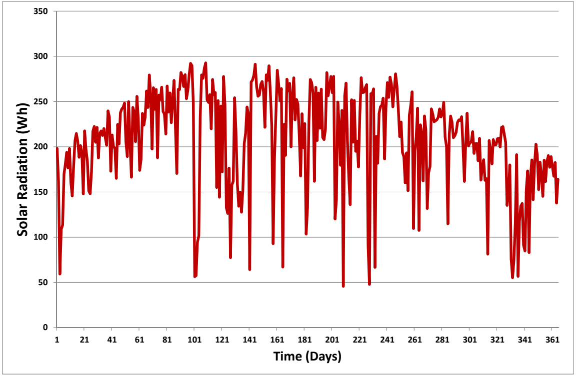


Figure: Daily mean solar radiation for Desirade

Conclusion

- Additive deaseasoning seems a more reasonable proposition than using CSI for solar radiation time series.
- Fourier Series seems to capture all the aspects of the seasonality.
- It is difficult to convince engineers and physicists who traditionally use CSI.
- We will see another advantage when we get to the next step - looking at the series with seasonality removed.
- Note that Fourier Series is not so good if a series has really sharp peaks.