

Time Series Analysis

Sam Urmy (Fisheries)

urmy@uw.edu

4/21/11

Time Series

- Sequence of observations, one after another
- Ideally, equally-spaced w/ no missing values
- Found in just about every field
- Techniques often applied to other dimensions (e.g., distance along a transect)

Traditional stats

- Assume independent observations
- Repeated samples from *population* => infer characteristics of population

Time series

- Observations **not** independent
- Typically, can only observe a single time series
- Infer *process* from relationships *between* observations

Time domain & Frequency domain

- Two approaches to TSA, complimentary
- Time domain:
 - Autocorrelation, statistical models, forecasting
- Frequency domain:
 - Fourier transforms, power spectra, periodic signals, multi-resolution analysis

Time Domain

Process / Realization

- Think of observed series as *one possible realization* of an underlying process
- Process usually stochastic - i.e., has a random component
- Can also be deterministic, chaotic, etc.

Example: Stochastic Processes

Stationarity

- Assumption of most time series stats.
- Strong stationarity: mainly theoretical.
 - joint distribution of observations identical at all times
- Weak stationarity: what we usually use.
 - Expected value doesn't change w/ time
 - Autocovariance doesn't change w/ time

Stationarity

- If you know the mathematical process, can determine stationarity analytically
- Real data: often easy to tell by eye
- First step of analysis often removing trend and/or transforming series to be stationary

Example: Stationarity

Removing a trend

- Linear regression
- Local regression, smoothing
- Differencing

$$\nabla X_t = X_t - X_{t-1}$$

- (alternate notation: *backshift operator*)

$$\nabla X_t = X_t - BX_t = (1 - B)X_t$$

Example: Removing a trend

Autocovariance

- Fundamental descriptive statistic for time series
- Covariance of a series/process with a lagged version of itself
- Theoretical vs. sample autocovariance
- How similar are observations as a function of the distance between them?

Autocovariance Function (ACVF)

- Definition:

$$\gamma(h) = E[(X_t - \mu_t)(X_{t-h} - \mu_{t-h})]$$

- Sample ACVF (to estimate from real series)

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (X_t - \mu)(X_{t+h} - \mu)$$

Autocorrelation Function (ACF)

- Normalized version of the ACVF (more commonly used)

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}$$

- Available in R via function `acf()`

Partial Autocorrelation Function (PACF)

- It's kind of a long story...
- Think of it as the ACF, but with intermediate dependence factored out
- Useful in choosing time series model
- Available in R via `pacf()`

Example: `Autocorrelation`

Time Series Models

- Idea: find simple representation of the process that produced the series
- Explain all autocorrelation: residuals should be white noise
- Use current & past values to forecast future

Autoregressive (AR) models

- Each value of series: linear combination of past values, plus random error
- AR(1)

$$X_t = \phi X_{t-1} + W_t$$

- AR(p)

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + W_t$$

Moving average (MA) models

- Each value a linear combination of past values of white noise

- MA(1)

$$X_t = W_t + \theta_1 W_{t-1}$$

- MA(q)

$$X_t = W_t + \theta_1 W_{t-1} + \theta_2 W_{t-2} + \cdots + \theta_p W_{t-p}$$

Autoregressive Moving average (ARMA)

- Combination of AR and MA terms

$$X_t = W_t + \sum_i^p \phi_i X_{t-i} + \sum_j^q \theta_j W_{t-j}$$

- NB: can express any AR process as MA and vice versa, but involves an infinite sum.

Integrated ARMA (ARIMA)

- Difference the series d times before fitting model

$$\nabla^d X_t = W_t + \sum_i^p \phi_i X_{t-i} + \sum_j^q \theta_j W_{t-j}$$

- Most general of classic time series models

Fitting TS models

- Workhorse function:

```
arima(x, order = c(0, 0, 0),  
      seasonal = list(order = c(0, 0, 0), period = NA),  
      xreg = NULL, include.mean = TRUE,  
      transform.pars = TRUE,  
      fixed = NULL, init = NULL,  
      method = c("CSS-ML", "ML", "CSS"),  
      n.cond, optim.method = "BFGS",  
      optim.control = list(), kappa = 1e6)
```

- Needs: series, order of model to fit
- Provides: estimates of parameters, fitted values, residuals, AIC, etc.

Choosing model order

- Detrend series if necessary (regression, differencing, etc.)
- Look at ACF and PACF

	AR(p)	MA(q)	ARMA(p,q)
ACF	Tails off gradually	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off gradually	Tails off

- Use AIC (or, preferably, AICc) to choose between models

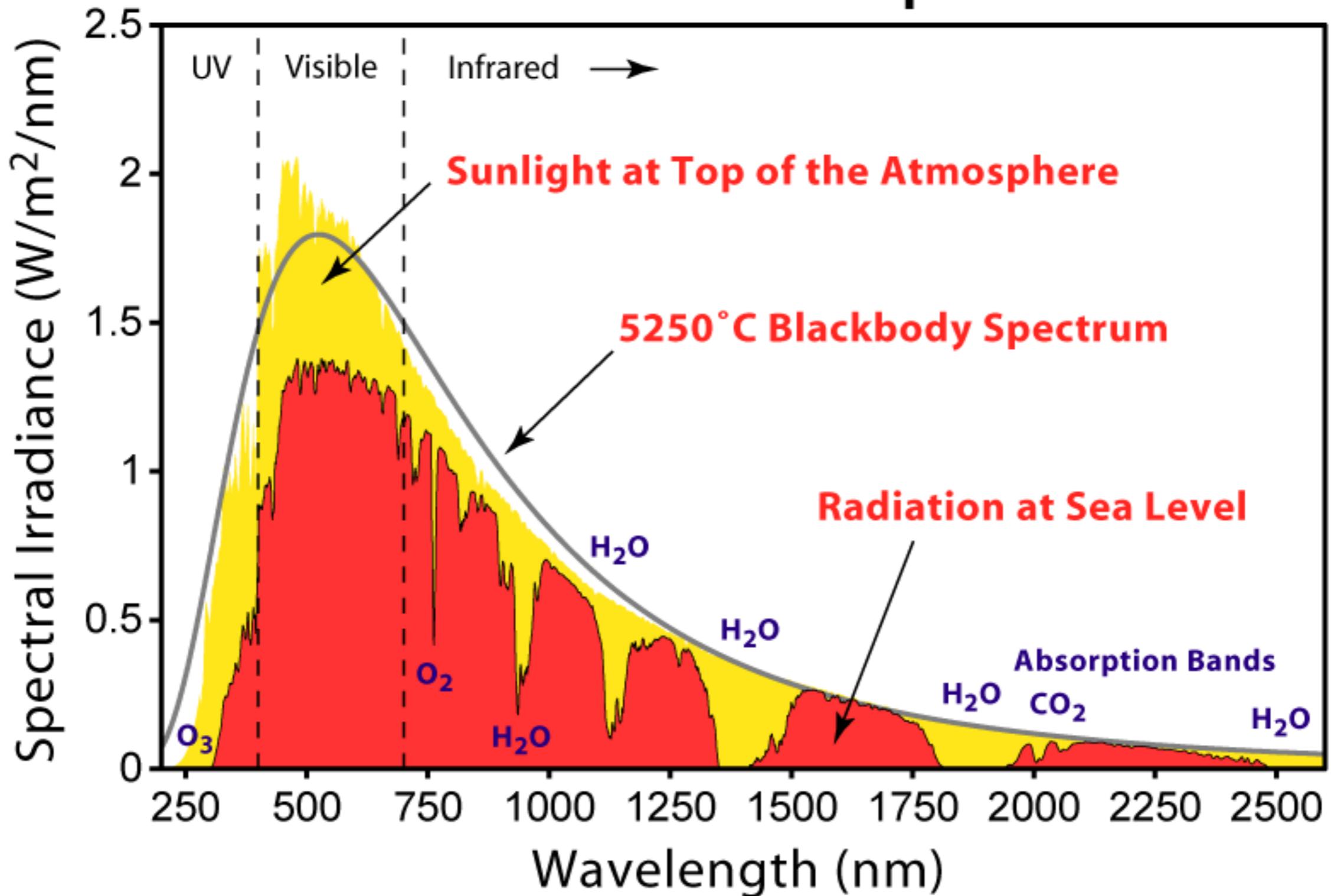
Example: Fitting models

Frequency Domain

Frequency Domain

- Different perspective: characteristics of time series as function of *frequency* instead of *time*
- Fourier/harmonic/spectral analysis
- Central idea: *spectrum* of a process
- Can be hard to wrap your head around, but you're already familiar with the idea

Solar Radiation Spectrum





The concept of a spectrum can be applied to most time series of sufficient length (more than about 50 observations).

Fourier Transform

- Any sequence can be expressed as a sum of sinusoids. In math:

$$X'_k = \sum_{t=0}^{N-1} X_t e^{-\frac{2\pi i}{N} tn}, \quad k = 0, \dots, N - 1$$

- In R: `fft()` (Fast Fourier Transform)

Fourier Transform

- Complex-valued: real and imaginary parts
- Symmetrical: one side mirrors the other
- Reversible: can transform FT back to original series.
- FT contains *exactly the same information* as the original series, but in different form

Example: `Fourier transforms`

Why do we care?

- The Fourier Transform tells us how much energy (variability) a series contains *as a function of frequency*.
- Can help us:
 - Discover hidden periodicities in the data
 - Infer processes going on at different scales
 - Find links between processes

Power Spectrum

- Square of the absolute value of the Fourier transform.
- A.K.A. the “periodogram”
- Proportional to energy or variance (different normalizations used)
- In R: `spectrum()`

Example: Spectral density and the periodogram

References

There are many more references out there. These (plus trial, error, R documentation, and relentless Googling) are the ones I've been using.

- Chatfield, *The Analysis of Time Series: An Introduction*.
- Less detail (not necessarily a bad thing...)
- Shumway and Stoffer, *Time Series Analysis and Its Applications: With R Examples*.
- R examples. May be available for free download if your library has a Springer subscription.
- Brockwell and Davis, *Introduction to Time Series and Forecasting*.
- Very detailed.