Spatial and Temporal Analysis in Ecology: A Primer

IBIO*6000 Advances in Ecology and Behaviour: A Class Project
Winter 2009 Semester

Josef Daniel Ackerman (Editor)

Contributions from J.D. Ackerman, T.J.Bartley, G.Bravener, M.J. D'Aguiar, M. Janssen, and J. Sheehy.

Department of Integrative Biology
University of Guelph

Table of Contents

Chapter 1: Preliminaries

Course Objective:

The objective of the course is to develop the ability to understand and communicate temporal and spatial concepts used in contemporary ecology. We will do so by reading and interpreting the peer-reviewed literature and through consultation with references on statistical techniques. This seminar course is not, however, designed as a statistics courses, rather it will endeavour to engage and develop numeracy in this important area. Each weekly session will focus on one of the selected topics[1] listed below by providing (1) a "primer" on the method and (2) a critical evaluation of selected readings[2]. All participants are expected to contribute to the discussion. Each student will provide one session on spatial and a second on temporal analysis.

[2]We will focus on peer-reviewed readings from the following journals: *Journal of Geophysical Research*, *Limnology & Oceanography*, *Nature*, *Proceedings of the National Academy of Science*, *Proceedings of the Royal Society*, and *Science*.

Course Readings:

Chapter 2 – Primer on Temporal Analysis

Lacasa L., B. Luque, F. Ballesteros, J. Luque, and J.C. Nuno 2008. From time series to complex networks: The visibility graph. *PNAS* 105: 4972-4975.

Raupach, M.R., G. Marland, P. Ciais, C. Le Quere, J.G. Canadell, G. Klepper, and C.B. Field 2007. Global and regional divers of accelerating C02 emissions. *PNAS* 104: 10288-10293.

Chapter 3 - Primer on Temporal Smoothing

Carrington E. 2002.Seasonal variation in the attachment strength of blue mussels: Causes and consequences. *Limnol. Oceanogr.* 47: 1723-1733.

Knudsen, E., A. Linden, T. Ergon, N. Jonzen, J.O. Vik, J. Knape, J.E. Roer, and N.C. Stenseth, 2007. Characterizing bird migration phenology using data from standardized monitoring at bird observatories. *Climate Research* 35:59-77.

Chapter 4 - Primer on Temporal Autocorrelation

Korpimaki, E., K. Norrdahl, O. Huitu, and T. Klemola 2005. Predator-induced synchrony in population oscillations of coexisting small mammal species. *Proc Royal Society* B 272: 193-202.

Thirgood, S.J., S.M. Redpath, D.T. Haydon, P. Rothery, I. Newton, and P.J. Hudson 2000. Habitat loss and raptor predation: disentangling long- and short-term causes of res grouse declines. *Proc Royal Society* B 267:651-656.

Chapter 5 - Primer on Spectral Analysis

Emerson, C.W., and J. Grant 1991. The control of soft-shell clam *(Mya arenaria)* recruitment on intertidal sandflats by bedload sediment transport. *Limnol. Oceanogr.* 36: 1288-1300.

Haydon, D.T., D.J. Shaw, I.M. Cattadori, P.J. Hudson, and S.J. Thirgood 2002. Analysing noisy time-series: describing regional variation in the cyclic dynamics of red grouse. *Proc Royal Society* B 269:1609-1617.

Chapter 6 - Primer on Wavelet Analysis

Cazelles, B., M. Chavez, D. Berteaux, F. Menard, J.O. Vik, S. Jenouvrier, and N.C. Stenseth 2008. Wavelet analysis of ecological time series. *Oecologia* 156: 287-304.

Rouyer, T., J.M. Fromentin, N.C. Stenseth, and B. Cazelles 2008. Analysing multiple time series and extending significance testing in wavelet analysis. *Mar. Ecol. Progr. Ser.* 359: 11-23.

Chapter 7 - Primer on Coherence

Aragao, L.E., Y. Malhi, N. Barbier, A. Lima, Y. Shimabukuro, L. Anderson, and S. Saatchi 2008. Interactions between rainfall, deforestation and fires during recent years in the Brazilian Amazonia. *Proc Royal Society* B 363: 1779-1785.

Platt, T., and K.L. Denman 1975. Spectral analysis in ecology. Annual Reviews 6: 189-210.

Rowe, P.M., and C.E. Epifanio 1994. Tidal stream transport of weakfish larvae in Delaware Bay, USA. *Mar. Ecol. Progr. Ser.* 110: 105-114.

Chapter 8 - Primer on Spatial Distribution

Arocena, J.M., and J.D. Ackerman 1998. Use of statistical tests to describe the basic distribution pattern of iron oxide nodules in soil thin sections. *Soil Sci. Soc. America J.* 62: 1346-1350

Frohlich, M., and H.D. Quednau 1995. Statistical analysis of the distribution pattern of natural regeneration in forests. *Forest Ecol. Manag.t* 73: 45-57.

Morales, J., J.J. Martinez, M. Rosetti, A. Fleury, V. Maza, M. Hernandez, N. Villalobos, G. Gragoso, A.S. de Aluja, C. Larralde, and E. Scuitto 2008. Spatial Distribution of *Taenia solium* Porcine Cysticercosis within a Rural Area of Mexico. *PLOS* 2: 284-290

Chapter 9 - Primer on the Use of Indices to Determine Spatial Patterns

Hurlbert, Stuart H., 1990. Spatial distribution of the montane unicorn. *OIKOS* 58: 257-271.

Stephanis, R., T. Cornulier, P. Verborgh, J.S. Sierra, N.P. Gimeno, and D. Guinet 2008. Summer spatial distribution of cetaceans in the Strait of Gibraltar in relation to the oceanographic context. *Mar. Ecol. Progr. Ser.* 353: 257-288.

Chapter 10 - Primer on Spatial Smoothing

Akhtari, R., S. Morid, M.S. Mahdian, and V. Smakhtin 2009. Assessment of areal interpolation methods for spatial analysis of SPI and EDI droughts indices. *Inter. J. Climatol.* 29: 135-145.

Conrad, K.F., I.P. Woiwod, J.N. Perry 2002. Long-term decline in abundance and distribution of the garden tiger moth (*Arctia caja*) in Great Britain. *Biol. Conserv.* 106: 329-337.

Chapter 11 - Primer on Spatial Autocorrelation

Koenig, W.D. 1997. Spatial autocorrelation in California land birds. *Conserv. Biol. 12*: 612-620.

Koenig, W.D., and J. Knops 1998. Testing for spatial autocorrelation in ecological studies. *Ecography* 21: 423-429

Chapter 12 - Primer on Spatial-Temporal Analysis

Aukema, B.H., A.L. Carroll, Y. Zheng, J. Zhu, K.F. Raffa, R.D. Moore, K. Stahl, and S.W. Taylor 2007. Movement of outbreak populations of mountain pine beetle: influences of spatiotemporal patterns and climate. *Ecography* 31: 348-358.

Moss R., D.A. Elston, and A. Watson 2000. Spatial asynchrony and demographic traveling waves during red grouse population cycles. *Ecology* 81: 981-989.

COMPONENTS OF A SCIENTIFIC PAPER
A GUIDE TO SCIENTIFIC COMMUNICATION

The Nature of Scientific Reports
This is the major form of scientific communications that exist for both students and professional scientists. It is the vehicle for reporting the results of scientific inquiry. These inquiries are based on the use of the scientific method, which aims to identify data through experimental methods (i.e., the hypothetico-inductive approach) or theoretical development (i.e., Modelling) that are objective, replicable (repeatable), and predictive of future inquiry. In this sense, the approach of science is incremental as it builds on existing data. It is essential that scientific reports identify the purpose or relevance of the work as well as provide an indication of the "take-home message" or conclusion of the study. This can only be achieved through a discussion of other works in the scientific literature.

The best way to improve your communications skills is to read as much as possible. You should refer to the journal articles to determine why the scientist used a particular form of communication (e.g., a scatter plot versus a bar chart). This will help you in developing your style.

Cover Page
A cover page with the appropriate information pertaining to the assignment and authorship (e.g., Title, Date, Name, Student Number) is required.

Presentation
The presentation of scientific reports is of great importance, and care must be taken to ensure that it is legible and consistent in style. The presentation should be double spaced throughout including the abstract. The pages should be numbered consecutively beginning on the first text page (i.e., the page after the cover page). The page number should appear centered at the bottom of the page (i.e., as a footer).

Abstract or Executive Summary
An abstract is a short/concise paragraph(s) that describes the motivation/ relevance, hypothesis examined, techniques, findings and conclusion/significance of the study. It is meant to provide the reader with a guide to what is reported in the main body of the paper. Figures, tables, and citations to the literature are not included in this section. With the large volume of literature, it may be the only part of a report that is read (e.g., on-line search) and is, therefore, probably the most important part of the report. Be sure to end the abstract with a concluding statement outlining the relevance and implication of the major finding.

Introduction
The Introduction provides the background for the study as well as identifying the relevant hypothesis that will be examined. It usually begins with a brief review of the scientific issue

related to the report and describes the findings of other researchers. The presentation is usually from the general to the more specific and this provides a perspective to the reader. The review also provides an opportunity to identify the theoretical foundations of the subject, gaps in our understanding, and areas of inquiry that require additional examination. In essence, it provides the motivation or purpose for undertaking the study and, presents the hypothesis to be examined.

Materials and Methods

The Materials and Methods section provides a relatively complete description of the methodology that was used in the study. It should be detailed in such a way that any other scientists can replicate the study, although it is not necessary to detail accepted/common techniques that exist in the literature (journal articles and/or texts). These techniques should be referred to as citations and any modifications to them should be stated clearly.

This section can begin with an explicit statement of the Null Hypothesis (i.e., the results of the different treatments are equal) to be tested or this can be incorporated within the text. In some cases, it is appropriate to provide section headings for particular portions of the study such as, “study site”, “survey design”, “laboratory analysis” and “statistical analysis”. The general approach is to report what was done chronologically. Generally, you should not report the motivation for your choice of techniques, nor should you report any results in this section.

Results

The Results section provides an opportunity to present the findings or the data revealed from your scientific inquiry. This is achieved through describing in prose what was observed as well as providing tabular and/or graphical results to illustrate the description. Both of these elements are necessary for the report and it is generally useful to include tables and/or figures in the text rather than at the end of the report. As was stated above, it can be useful to include subheadings if these are meaningful and help to clarify the report. The results should be presented but not discussed. Remember that your data are never wrong but your interpretation may be. Therefore, there should be no interpretation of the relevance/significance of the findings with respect to the Null Hypothesis or the literature, as these are included in the Discussion section.

Data - Data are plural and should be referred to in this manner (i.e., “the data were…”). Consequently and as a result of uncertainty in our techniques, measurement error, etc., it is not generally possible to refer to a single result. Rather, we refer to the distribution of the data by including the mean (central tendency) and the standard error (dispersion) of our observations. The standard error (standard error = standard deviation/square root of the sample size) allows us to compare the results of different experiments and is the basis of many statistical tests. It is also important to account for uncertainty and track the propagation of errors within reports (see section on Statistical Analysis).

Figures - Figures provide an illustration of what was undertaken or what was found in a scientific inquiry. It is important to note that schematic representations of equipment configurations and site maps are extremely useful to include in reports. In terms of the reporting of results, it should be recognized that different types of figures are used for different types of data. For example, bar charts are used for reporting categorical data, while scatter plots are used for data that vary with another factor such as time or concentration. You should note the mean and standard error in the graphical presentation. It is important to ensure that figures are as self-explanatory as possible so that a reader can understand them without referring to the text. This is achieved through careful consideration of (1) the labeling of constituent elements (e.g., axes), (2) the inclusion of units and equations (if necessary), and (3) the figure title and legend. Be sure to use scale bars for all drawings, and some form of direction indication (e.g., arrow) for all maps. It is generally acceptable to present figures at the end of the report to avoid formatting errors when there are a large number of figures.

***Tables* –** Tables provide an opportunity to present results that are not amenable to graphical presentation. These may be in the form of lists of results or summaries. As above, it is important to ensure that tables are as self-explanatory as possible so that a reader can understand them without referring to the text. This is achieved through careful consideration of the labeling of rows and columns along with units and, and the use of table title and legend.

Discussion

The Discussion section is where you interpret the relevance/significance of your findings with respect to the Null Hypothesis and the scientific literature. In other words, this is where you explain the meaning of what you found. You should work from the specific to the general to show both what your results mean in the context of the study and within the context of the scientific literature. Generally you would begin with a statistical examination of the data and you would either accept or reject the Null Hypothesis. This would be followed by a treatment of the implications of acceptance/rejection of the Null Hypothesis and the significance of this evaluation.

This section is where you would address the questions posed in your Introduction. You should endeavor to incorporate your findings into the larger context of scientific understanding and literature. The limitations of your approach, ways to improve it, and potential future inquiries, can also be presented.

Conclusions

The Conclusion section provides you an opportunity to express the principal findings from your scientific inquiry. It generally is a brief paragraph(s) in which you report your findings and the relevance/significance of these from the perspective of your hypothesis and the literature. It concludes the report by addressing the issues, questions, and hypotheses posed in your Introduction section. In many cases, the conclusions may include new hypotheses, issues and hypotheses.

Literature Cited

This is where you provide the bibliographical information for literature that you referred to in the report. You should only include literature that was cited since this is not a bibliography. Please note that Internet sources and class notes are **not** permitted (i.e., do not cite any information that was not found in a credible scientific source). There are several accepted ways in which to cite literature (e.g., CBE) and these can vary according to the specific journal. Please use the following as a guide for your reports:

Journal Article.
Ackerman, J.D., Loewen, M.R., and P.F. Hamblin. 2001. Benthic-pelagic coupling over a zebra mussel bed in the western basin of Lake Erie. *Limnology and Oceanography* 46(4):892-904.

Book.
Lobban, C.S. and P.J. Harrison. 1994. *Seaweed Ecology and Physiology*. Cambridge University Press, New York. 366 pp.

Book Chapter.
Givnish, T.J. 1989. Ecology and evolution of carnivorous plants. pp. 243-290 *In*: W.G. Abrahamson (ed.) *Plant-Animal Interactions.* McGraw-Hill, New York. 480 pp.

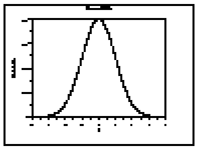
Technical Report.
French, T.D. and P.A. Chambers. 1995. *Environmental factors regulating the biomass and diversity of macrophyte communities in rivers, with emphasis on the Nechako River, British Columbia, Canada*. Report # 000 prepared for the BC Ministry of Environment, Prince George and Victoria. 114 pp.

Appendix
This is where you would include supplementary information that does not belong within the text of the report, but is relevant for the study. Copies of your original data forms, chart recordings, etc., would be included in an Appendix.

Note on non-experimental papers
Under certain circumstances (e.g., review articles, position papers), the traditional reporting methods described above may not necessarily be followed. (This is especially true for the Materials and Methods section, although in many situations, data selection and analysis are described in the M&M). In these cases, it is customary to deviate from these approaches through the use of headings to direct the reader. For example, headings may include: Abstract, Introduction, Literature Background, Position Statement, Alternative Views, Supporting Evidence, Discussion, Conclusions, Literature Cited.

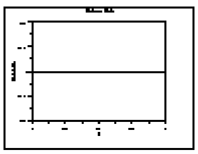
Statistical Distributions

(1) Continuous distributions



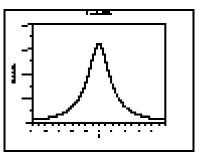
Normal Distribution $$f(x) = \frac{e^{-(x-\mu)^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}},$$

standard normal $$f(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$$ when □ = 0 and □= 1



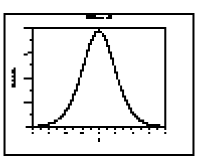
Uniform Distribution $$f(x) = \frac{1}{B-A} \qquad \text{for } A \le x \le B$$

standard uniform $$f(x) = 1 \qquad \text{for } 0 \le x \le 1$$ when A = 0, B = 1



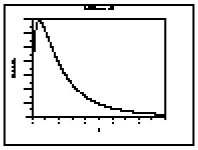
Cauchy Distribution $$f(x) = \frac{1}{s\pi(1 + ((x-t)/s)^2)},$$

standard Cauchy $$f(x) = \frac{1}{\pi(1+x^2)}$$ when ***t*** = 0 and ***s*** = 1

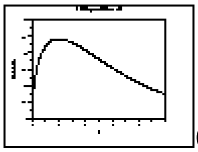


t Distribution $$f(x) = \frac{(1+\frac{x^2}{\nu})^{\frac{-(\nu+1)}{2}}}{B(0.5, 0.5\nu)\sqrt{\nu}}$$

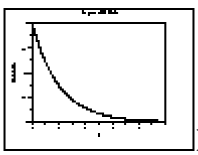
where $$B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt$$


F Distribution $$f(x)=\frac{\Gamma(\frac{\nu_1+\nu_2}{2})(\frac{\nu_1}{\nu_2})^{\frac{\nu_1}{2}}x^{\frac{\nu_1}{2}-1}}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})(1+\frac{\nu_1 x}{\nu_2})^{\frac{\nu_1+\nu_2}{2}}}$$

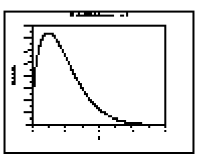
Where $\Gamma(a)=\int_0^\infty t^{a-1}e^{-t}dt$


Chi-Square Distribution $$f(x)=\frac{e^{\frac{-x}{2}}x^{\frac{\nu}{2}-1}}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} \qquad \text{for } x\geq 0$$

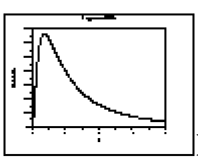
where $\Gamma(a)=\int_0^\infty t^{a-1}e^{-t}dt$


Exponential Distribution $$f(x)=\frac{1}{\beta}e^{-(x-\mu)/\beta} \qquad x\geq\mu;\beta>0$$

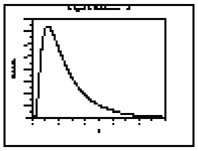
Standard exponential $f(x)=e^{-x}$ for $x\geq 0$, μ= 0 and β= 1


Weibull Distribution
$$f(x)=\frac{\gamma}{\alpha}(\frac{x-\mu}{\alpha})^{(\gamma-1)}\exp\left(-((x-\mu)/\alpha)^{\gamma}\right) \qquad x\geq\mu;\gamma,\alpha>0$$

Standard wiebull $f(x)=\gamma x^{(\gamma-1)}\exp(-(x^{\gamma}))$ $x\geq 0;\gamma>0$, μ= 0 and α= 1

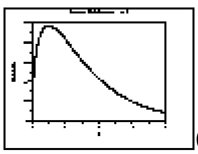

Lognormal Distribution $$f(x)=\frac{e^{-((\ln((x-\theta)/m))^2/(2\sigma^2))}}{(x-\theta)\sigma\sqrt{2\pi}} \qquad x\geq\theta;m,\sigma>0$$

Standard lognormal $f(x) = \frac{e^{-((\ln x)^2/2\sigma^2)}}{x\sigma\sqrt{2\pi}} \quad x \geq 0; \sigma > 0$, θ= 0 and m = 1


Fatigue Life Distribution

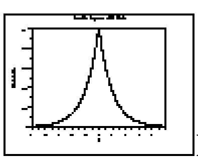
$$f(x) = \left(\frac{\sqrt{\frac{x-\mu}{\beta}} + \sqrt{\frac{\beta}{x-\mu}}}{2\gamma(x-\mu)}\right)\phi\left(\frac{\sqrt{\frac{x-\mu}{\beta}} - \sqrt{\frac{\beta}{x-\mu}}}{\gamma}\right) \quad x > \mu; \gamma, \beta > 0$$

Standard Fatigue life $f(x) = \left(\frac{\sqrt{x} + \sqrt{\frac{1}{x}}}{2\gamma x}\right)\phi\left(\frac{\sqrt{x} - \sqrt{\frac{1}{x}}}{\gamma}\right) \quad x > 0; \gamma > 0$ μ= 0, β= 1

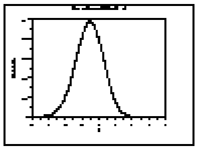

Gamma Distribution $f(x) = \frac{(\frac{x-\mu}{\beta})^{\gamma-1}\exp\left(-\frac{x-\mu}{\beta}\right)}{\beta\Gamma(\gamma)} \quad x \geq \mu; \gamma, \beta > 0$

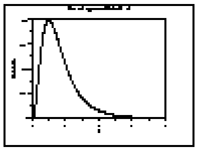
Where $\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt$

Standard Gamma $f(x) = \frac{x^{\gamma-1}e^{-x}}{\Gamma(\gamma)} \quad x \geq 0; \gamma > 0$, μ= 0 and β= 1

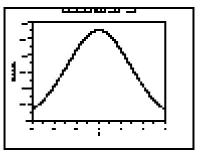

Double Exponential Distribution $f(x) = \frac{e^{-|\frac{x-\mu}{\beta}|}}{2\beta}$

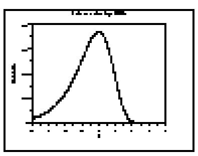
Standard double exponential $f(x) = \frac{e^{-|x|}}{2}$, μ= 0 and β= 1

 Power Normal Distribution $f(x,p) = p\phi(x)(\Phi(-x))^{p-1} \qquad x, p > 0$

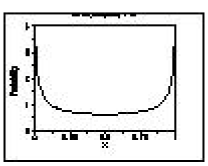
 Power Lognormal Distribution

$$f(x,p,\sigma) = (\frac{p}{x\sigma})\phi(\frac{\log x}{\sigma})(\Phi(\frac{-\log x}{\sigma}))^{p-1} \qquad x, p, \sigma > 0$$

 Tukey-Lambda Distribution No simple, closed form, so must be computed numerically

 Extreme Value Type I Distribution Gumbel $f(x) = \frac{1}{\beta}e^{-\frac{x-\mu}{\beta}}e^{-e^{-\frac{x-\mu}{\beta}}}$

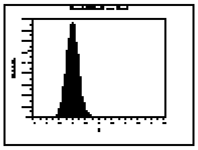
Standard Gumbel Min = $f(x) = \frac{1}{\beta}e^{\frac{x-\mu}{\beta}}e^{-e^{\frac{x-\mu}{\beta}}}$

 Beta Distribution $f(x) = \frac{(x-a)^{p-1}(b-x)^{q-1}}{B(p,q)(b-a)^{p+q-1}} \qquad a \le x \le b; p, q > 0$

Where $B(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt$

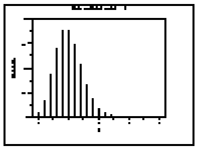
Standard beta $$f(x) = \frac{x^{p-1}(1-x)^{q-1}}{B(p,q)} \qquad 0 \le x \le 1; p, q > 0$$ $a = 0$, $b = 1$

(2) Discrete distributions


Binomial Distribution

$$P(x,p,n) = \binom{n}{x} (p)^x (1-p)^{(n-x)} \qquad \text{for } x = 0, 1, 2, \cdots, n$$

Where $$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$


Poisson Distribution $$p(x,\lambda) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad \text{for } x = 0, 1, 2, \cdots$$

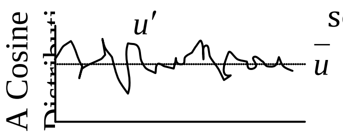
See:
NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/, date.

http://www.itl.nist.gov/div898/handbook/eda/section3/eda366.htm

CHAPTER 2: PRIMER ON TEMPORAL ANALYSIS

Josef Daniel Ackerman

- **Time Series**: Many variables fluctuate in time (e.g., water speed; bird sightings), when measurements are made through time



so we have $\bar{u}$ = mean speed

u' = fluctuation from $\bar{u}$

u = instantaneous speed

u = $\bar{u}$ + u'

  - **Fundamental Question**: Are observations close in time more related (dependent)?
  - **Discrete Time Series** vs. **Continuous Time Series** – type of sampling
  - **Eulerian Observation** vs. **Lagragian Observation** – frames of reference
  - **Underlying Process** vs. **Modeling/Forecasting** – motivation

- **Patterns in time series**

  I – Systematic Patterns

  - **(1) Seasonal effects: -** annual variation: summer vs. winter; diurnal effects, lunar
  - **(2) Cycles and Quasi-cycles:** - pattern that is not seasonal

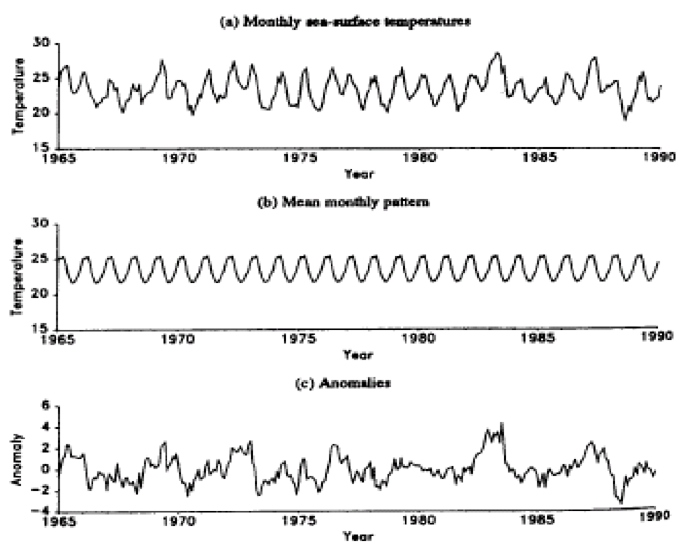


Figure 13.2 Time series of sea surface temperatures observed on the shore at Academy Bay, Santa Cruz, Galapagos, 1965–1990. (a) Original series; (b) mean monthly temperatures; (c) anomalies. Reprinted by permission of the Charles Darwin Research Station.

(Brown and Rothery 1993)

- **(3) Trends:** a smooth underlying, long-term change not seasonal
- **(4) Residual or Random Variation: -** irregular fluctuations

II. Types of Random Variation

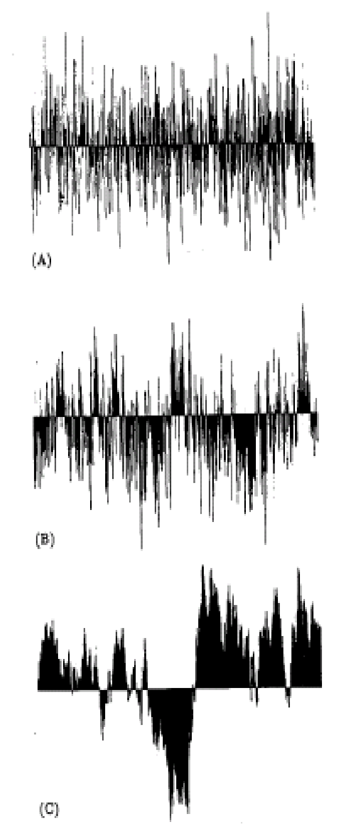
- **Recall:**
    - **Random Numbers** from a (1) **random number table (**tabulation) or via a (2) **pseudo-random number generator** (e.g. seed # to start the process).
    - **Random variable** is subject to "**chance fluctuations**" that are confined within the bounds as $N(0, \sigma^2)$ – i.e., "**uniformly distribution random**" variable, **stochastic**.

- **(1) Stationary Time Series:** - random variation with or without serial dependence.
    - Can be inherently stationary vs. where systematic trends have been removed.
    - (1) mean, (2) mean square, and (3) ACF (autocorrelation function) do not vary with changes in t.

- **(2) White Noise:** - this is an example of a stationary T.S., $1/t = \mu + z_t$ where $z_t \sim N(0,$ □2). Observations are serially independent; has a spectrum of $\sim f^{o}$

- **(3) Pink Noise –** intermediate between white and pink, has a spectrum of ~1/f

- **(4) Brown Noise** – values not independent but the increments are, has a spectrum of $\sim 1/f^{2}$ – may be good predictor of turbulence



(Schroeder 1992)

- **Analysis of Time Series**
    - **Time domain –** statistical approaches, autocorrelation etc
    - **Frequency (f = 1/t) domain–** spectral analysis, wavelet analysis
    - **Statistical Representation**
        - Graph the information using appropriate temporal scale
        - Measure central tendency, and variation (more important?)

- **Time average**
  - Mean value of the time series
  - **Ensemble Average**
  - Average the results of all occurrences that occur at time t = t as replicates, i.e., consider the separate time series as "replicates"
- **Ergotic**
  - time series is ergotic if it is stationary (mean is the same for any t_1) and the time average is the same as the individual ensemble averages

- **Transients**
  - sometime you get a burst (or gust) that looks like a non-stationary process
  - indicate that the series is not ergotic

- **Decomposition**

  - Decompose the time series into its components. (e.g., trend; cycle, season, “noise” or residual) e.g., Observation **= trend + cycle + seasonal variation + random variation**

  - **Same dataset as above:**

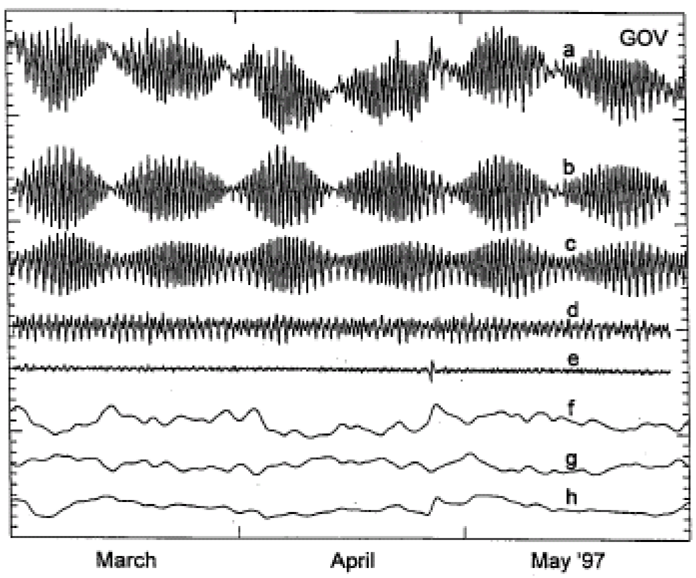
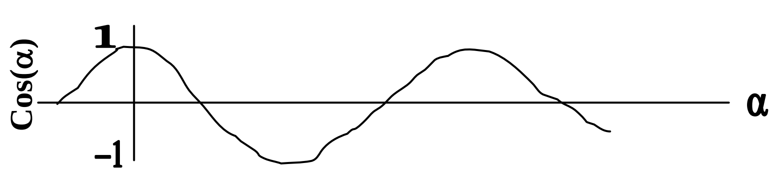


Fig. 2. Well level signal from site GOV (for example, see Figure 1) and extraction of residual pore pressure signal. a) Raw signal; b) raw signal, hf-part; c) tidal model fit; d) air pressure influence, hf-part; e) residual hf-signal; f) raw signal, lf-part; g) air pressure influence, lf-part; h) residual lf-signal. Tick marks on vertical axis denote centimeters.

Gupta, H.K., I. Radhakrishna, R.K. Chandra, H.J. Kumpel, and G. Greckshe. 2000. Pore pressure studies initiated in area of reservoir-induced earthquakes in India. *Eos* 81(14):145-151.

- **Identify patterns**

  (1) Cyclical Patterns ~ Review of Periodic Function



- Consider the cosine curve
- **Four steps to describe the function** ***y = cos α***

  - *t* = independent variable; α - phase angle (depends on t), *l* = unknown period

(1) $so\ \alpha = \frac{360°}{l} t \quad or \quad \alpha = \frac{2\pi}{l} t$

$if\ \omega = \frac{360°}{l}$ or $\frac{2\pi}{l}$, ω = **angular frequency**, which indicates how often *l* occurs in 1 rotation (360º or 2π rad.)

α = ωt **or** y = cos ωt

If there is no peak at *t* = ▯, then it occurs at to $0 < t_o < l$
So if there is a peak at $t' = t + t_o$ or at $t = (t' - t_o)$, but we drop t′ (i.e., a delay)

(2) Now $y = \cos \omega (t - t_o)$, where t_o = arc phase position when first peak occurs

(3) What about the amplitude (we not dealing with a perfect curve: $-1 \le \cos \omega \le 1$)
Add factor “c”, $y = c \cos\omega (t - t_o)$

(4) What happens if y oscillates around *c* by c_o?
$y = c_o + c \cos\omega (t - t_o)$

also recall that: $\cos \omega (t - t_o) = \cos (\omega t - \omega t_o)$

$= \cos \omega t \cos \omega t_o + \sin \omega t \sin \omega t$

but: $\cos \omega t_o + \sin \omega t_o$ → *constants:* $a = \cos \omega t_o$, $b = \sin \omega t_o$

so: $\boldsymbol{y = c_o + a \cos \omega t + b \sin \omega t}$ also $\mathbf{y = c_0 + a \cos 2\pi ft + b \sin 2\pi ft}$

- This is the general form that we use to describe any periodic function

References:

Bendat, J. S. and A. G. Piersol. 1986 Random data: Analysis and measurement procedure, 2nd edition, John Wiley.

Brown, D., Rothery, P. 1993. Models in Biology: mathematics, statistics and computing. J. Wiley.

Schroeder M. 1992 Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise. W.H. Freeman.

Chapter 3: TEMPORAL SMOOTHING

Justin Sheehy

What is smoothing?

Smoothing is a statistical technique in which an approximating function is created in an attempt to capture important patterns in a set of data, while leaving out noise or other fine-scale structures/rapid phenomena. It is a process by which data points are averaged with their neighbours in a series, such as a time series. This process tends to blur the sharp edges in the data, giving it a "smoother" appearance.

Benefits of Smoothing Data

- Great for seperating general trends and broad patterns from noisy data
- Gives a better visualisation of variation across space and time
- Maximises access to data that would otherwise be hidden

Problems with using Smoothing Data

- Smoothing data transforms the data
- Smoothed data is correlated based on kernel

Types of Smoothing

Note: There are several types of smoothing, but this primer will only focus on the statistical or mathematical forms of smoothing which are typically used in temporal (or spatial) studies.

Moving average

- Also called a rolling average or running average
- Can be applied to any data set, but commonly used with time series data to smooth out short-term fluctuations and highlight longer-term trends or cycles.
- Often used to try to capture important trends in repeated statistical trials
- Creates an average of one subset of the full data set at a time with each number in the subset given an equal statistical weight.
- Not a single number, but it is a set of numbers, each of which is the average of the corresponding subset of a larger set of data points.
- For example, if there is a data set (N = 50), the first value could be the moving average or mean of data points 1 through 10. The next value would be the mean of data points 2 through 11, and so forth, until the final value, which would be the mean of data points 40 to 50.
- The subset size being averaged is often constant, but does not need to be.

Simple Moving Average

A simple moving average (SMA) is the unweighted mean of the previous n data points. For example, a 5-day simple moving average of population abundance is the mean of the previous 5 days abundances. If those abundances are p_M, p_{M-1}… p_{M-4} then the formula is:

$$SMA = \frac{p_M + p_{M-1} + \ldots + p_{M-4}}{5}$$

Central Moving Average

A problem with simple moving averages is that they create a shift in the data, induced by only using "past" data. For many temporal and spatial studies, it is optimal to avoid this shifting, thus a central moving average can be determined using both "past" and "future" data. The "future" data in this calculation are not predictions, but simply data that was obtained after the time at which the average is to be calculated. If one was to calculate the CMA for a data set using the previous 5 days of information and the 5 days after the data point, the formula would be:

$$CMA = \frac{p_{m-5} + ... + p_{m-1} + p_{m+1} + ... + p_{m+5}}{10}$$

Example

For example, a fictitious researcher measured the number of termite larvae hatched daily for 80 days, and determined the proportion of larvae which survived each day. The original information was recorded and put into a histrogram (Figure 1), which can be seen on the following page. At first, no trend could be seen over the 80 days could be seen. Therefore, the researcher decided to perform a central moving average smoothing technique in an attempt to find any patterns.

The researcher found that performing a CMA with only 1 value on each side did not present any patterns but by using 5 (Figure 5) or 10 values (Figure 6) on each side presented a relatively clear trend. The proportion of surviving hatchlings appears to go through cycles of higher and lower survival rates.

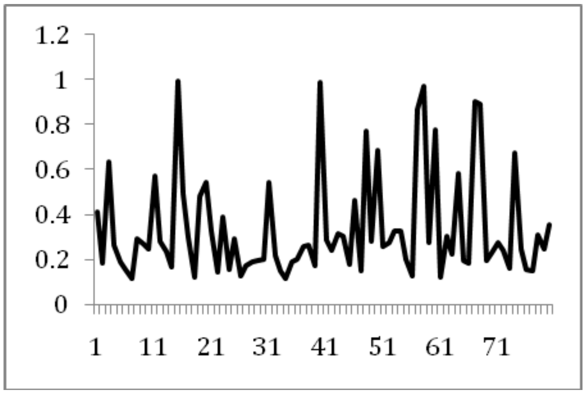


Figure 1: The raw data from the study. There does not appear to be any trends in the information.

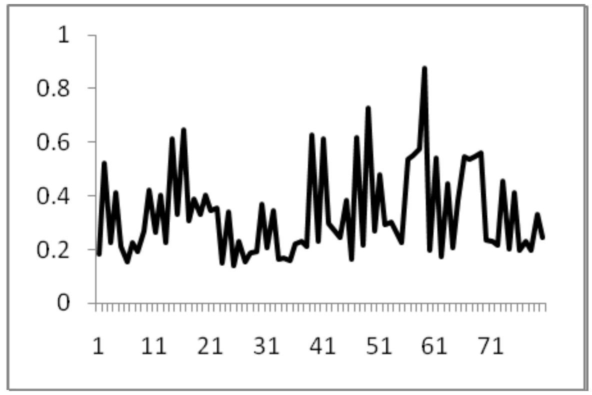


Figure 2: Data after CMA smoothing , using 1 value on each side of data point.

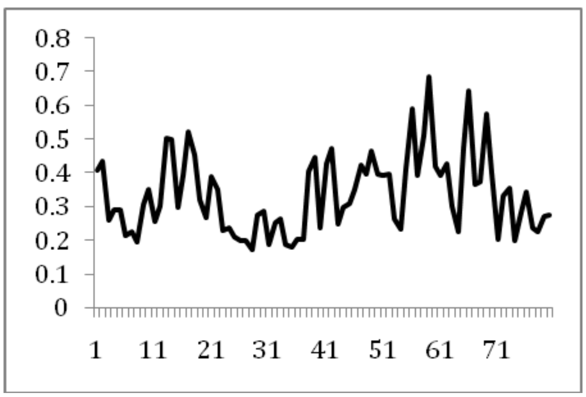


Figure 3: Data after CMA smoothing, using 2 values on each side of data point.

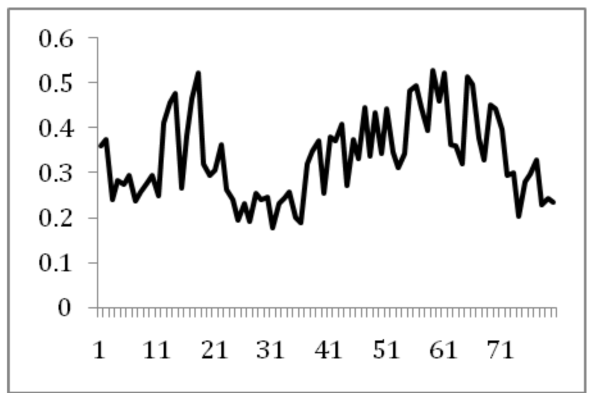


Figure 4: Data after CMA smoothing, using 3 values on each side of data point.

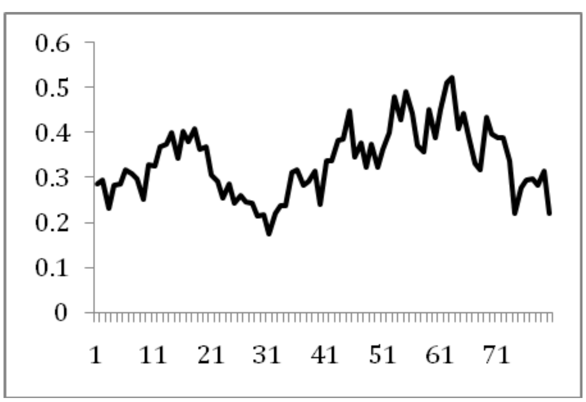


Figure 5: Data after CMA smoothing, using 5 values on each side of data point

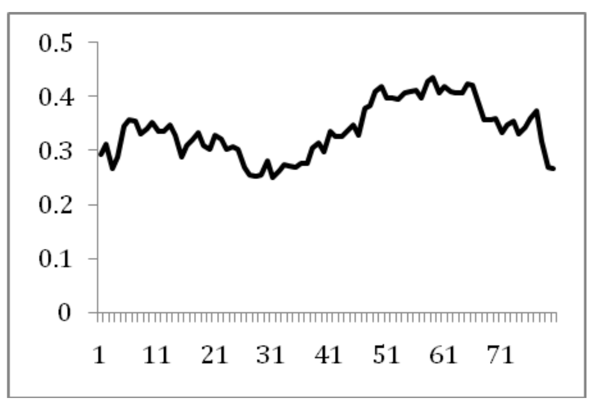


Figure 6: Data after CMA smoothing, using 10 values on each side of data point

Cumulative Moving Average

- Also called a long running average
- A type of moving average where each value is the average of all previous data points in the full data set.
- The size of the subset being averaged grows by one as each new value of the moving average is calculated.

The formula for the Cumulative Moving Average, which is usually an unweighted average, for *i* data points would be:

$$CMA_i = \frac{x_1 + x_2 + ... + x_i}{i}$$

The formula for the Cumulative Moving Average, for the following data point would equal:

$$CMA_{i+1} = CMA_i + \frac{x_{i+1} - CMA_i}{i+1}$$

Thus the current cumulative average for a new data point is equal to the previous cumulative average plus the difference between the latest data point and the previous average divided by the number of points received so far. When all of the data points arrive (i = N), the cumulative average will equal the final average.

Weighted Moving Average

- A moving average which has multiplying factors to give different weights to different data points.
- Has the specific meaning of weights which tend to decrease arithmetically.

An example of a weighted moving average would be;

$$WMA_M = \frac{np_M + (n-1)p_{M-1} + ...2np_{M-n+2} + p_{M-n+1}}{n + (n-1) + ... + 2 + 1}$$

where each day has a different weighting in the average.

Kernel Smoothing

- The statistical technique for estimating a real valued function with the use of a Kernel.
- A Kernel defines the shape of the function that is used to take the average of the neighbouring points. For example, a Gaussian kernel is a kernel with the shape of a Gaussian (normal distribution) curve.
- Basically, Kernal smoothing is fitting a shape into the random data, to remove some of the noise in the data

Steps to Kernel Smoothing

The following figure shows a data series, made of random numbers, taken over 40 days. At first, it appears that there are no trends in the information and therefore smoothing should be performed to see any trends which may be hiddin in the information.

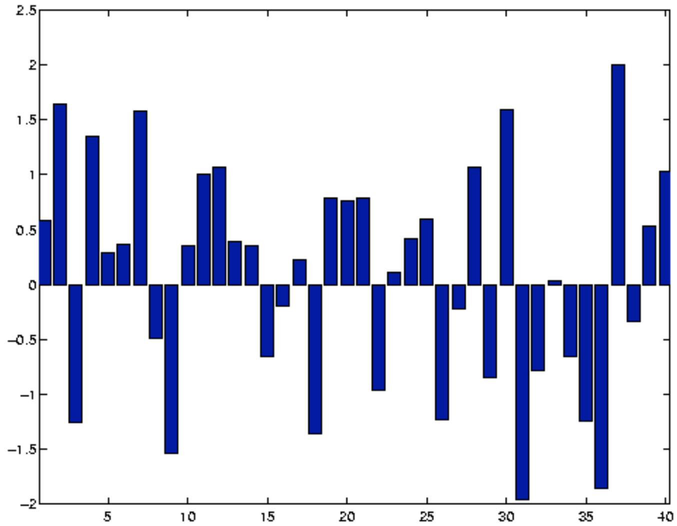
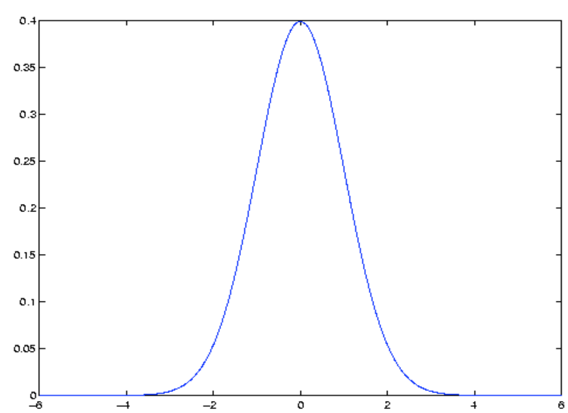


Figure 1: The raw data for the study. No trends can be seen and there appears to be Noise in the data. *
(From MRC, 1999)

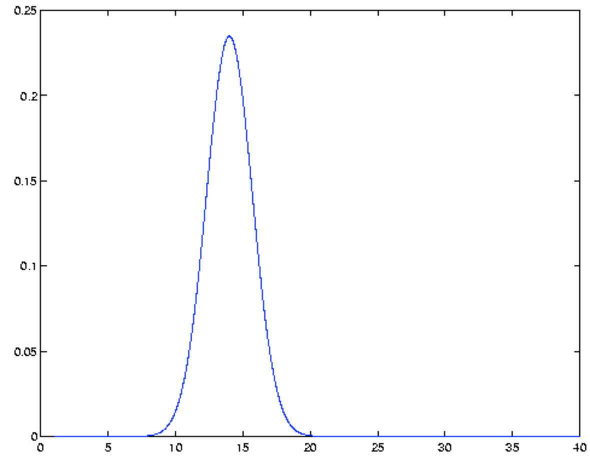
1) The kernel (or shape of the curve) in which the data will be fitted is chosen. In this example, the Gaussian (Normal) curve is used.



Note:
In most statistical analyses, the width of the Gaussian curve is in terms of sigma (δ) or standard deviations. However, when the Gaussian curve is used for smoothing, the width of the curve is defined using the Full

Width at Half Maximum (FWHM). The FWHM is the width of the kernel, at half of the maximum of the height of the Gaussian. Therefore, for this example, the maximum height is around 0.4. The width of the kernel at 0.2 is the FWHM. At x = -1.175 and 1.175 (when y = 0.2), the FWHM equals 2.35. Therefore, 2.35 is the FWHM for this example.

2) For each data point, a new, smoothed value (that is a function of the original value at that point and the surrounding data points) is calculated. For this example, a Gaussian smoothing, the function that is used is a Gaussian curve with a FWHM value of 4 x-axis units. To generate the Gaussian kernel average for a data point, the Gaussian shape is centred over that value on the x-axis. All of the values in the Gaussian curve are then divided by the total area under the curve, so that the values add up to 1.



The Gaussian Curve centred over the 14^{th} value in the data set. (From MRC, 1999)

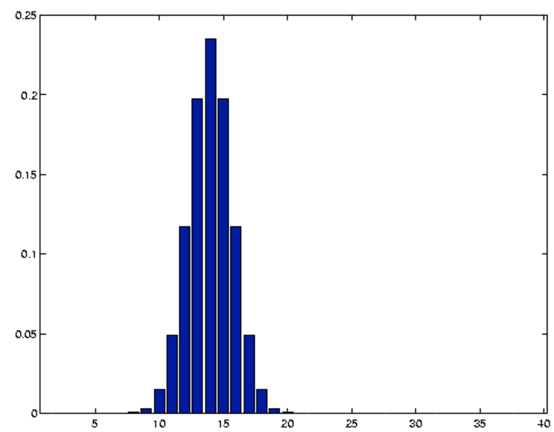
3) The values of the resulting function (in this case a Gaussian function) are generated for each of the points in the data.

For example, In this case:
The Gaussian values for 12,13,14,15 and 16 are:

0.1174, 0.1975, 0.2349, 0.1975, 0.1174

and the data values for the points are:

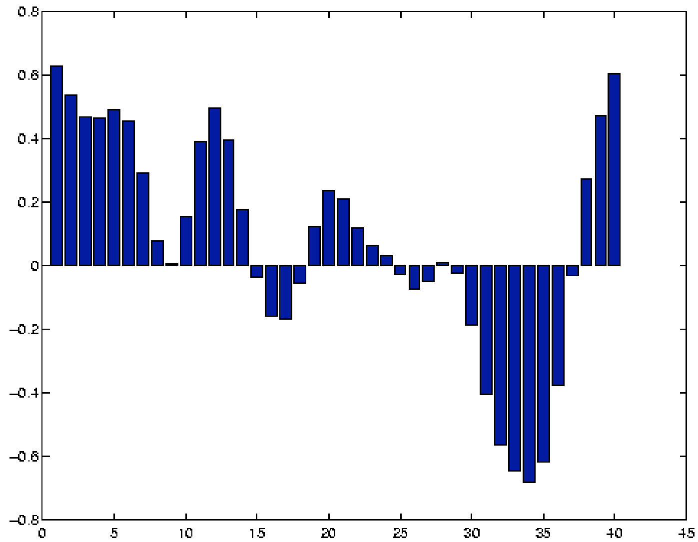
1.0645, 0.3893, 0.3490, -0.6566, -0.1946



4) The Gaussian values are multiplied by the data values, and the results are added up to get the new smoothed value for each point.

The values calculated using the Gaussian function centred over the 14^{th} value (From MRC, 1999)

5) The value of each new smooth point is kept and the smooth value is then calculated for the next data point. The result is a smooth version of the original data.



The new, smoothed data for the study. Most of the noise has been eliminated and patterns can be seen more clearly (From MRC, 1999)

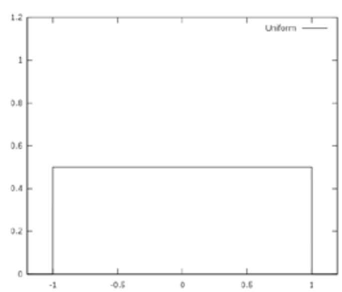
Kernel Functions

Several types of kernels functions are used in temporal and spatial studies. The following are the equations for each function and a graph of their shape/distribution.

Note: In the notation below, $1_{(p)}$ means that the function is multiplied by 1 when p is true, and 0 when p is false.

Uniform

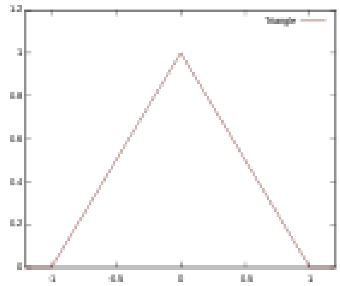
$$K(u) = \frac{1}{2} 1_{(|u| \leq 1)}$$



A uniform distribution (From Wikipedia, 2009)

Triangle

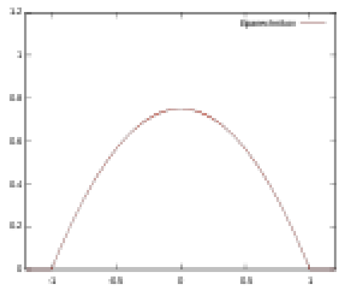
$$K(u) = (1 - |u|)1_{(|u| \le 1)}$$



A triangular Distribution (From Wikipedia, 2009)

Epanechnikov

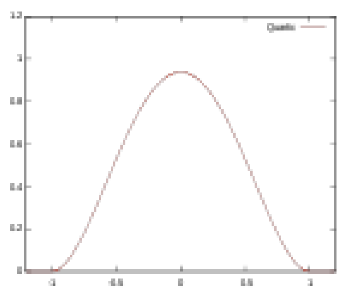
$$K(u) = \frac{3}{4}(1 - u^2)1_{(|u| \le 1)}$$



An Epanechnikov Distribution (From Wikipedia, 2009)

Quartic

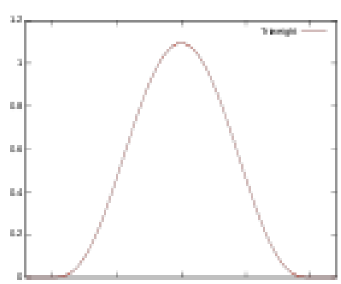
$$K(u) = \frac{15}{16}(1 - u^2)^2 1_{(|u| \le 1)}$$



A Quartic Distribution (From Wikipedia, 2009)

Triweight

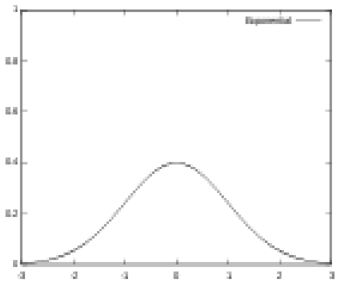
$$K(u) = \frac{35}{32}(1 - u^2)^3 1_{(|u| \le 1)}$$



A triweight Distribution (From Wikipedia, 2009)

Guassian

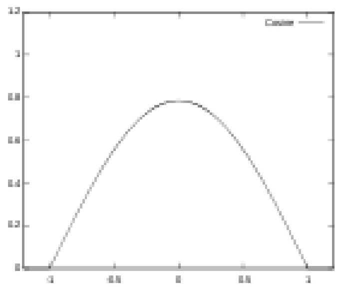
$$K(u)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}$$



A Gaussian Distribution (From Wikipedia, 2009)

Cosine

$$K(u)=\frac{\pi}{4}\cos(\frac{\pi}{2}u)1_{(|u|\leq 1)}$$



A Cosine Distribution (From Wikipedia, 2009)

Spline Smoothing

- The process of fitting a smooth curve to a set of noisy observations with a spline function

Unlike the previous smoothing types, spline smoothing has 2 goals:

1. To obtain a “smoother” set of observations
2. To maintain proximity to the actual sample data points

To accomplish goal number 2, a roughness penalty is defined. The smoothing spline estimate, $\hat{\mu}$, of the function is defined as a minimizer in the following formula:

$$\sum_{i=1}^{n}(Y_i-\hat{\mu}(x_i))^2+\lambda\int \hat{\mu}''(x)^2\,dx$$

The first part of the formula is the sum-of-squares and the second part is the roughness penalty, which includes the smoothing parameter, λ. The smoothing parameter controls this created trade-off between proximity to the original data and roughness of the function estimate. Therefore, if λ = 0, no smoothing will occur, and if λ is approaching an infinitely high value (∞) the roughness penalty would be infinitely large and the estimate converges to a linear least-squares estimate.

A typical sline used in statistics is the cubic spline, as this spline allows for easy formation of both first and second derivatives. An example of a cubic spline would be;

$$S_i(x)=a_i(x-x_i)^3+b_i(x-x_i)^2+c_i(x-x_i)+d_i$$

The first and second derivatives of this function would be;

$$S_i^{'}(x) = 3a_i(x - x_i)^2 + 2b_i(x - x_i) + c_i$$

$$S_i^{''}(x) = 6a_i(x - x_i) + 2b_i$$

Steps to fitting a Smoothing Spline Function

1. Using the spline function, derive the $\hat{\mu}(x_i)$ values for all of the values
2. From these calculated values, derive $\hat{\mu}(x)$ for all x

References

*Anonymous. An introduction to Smoothing. MRC. August 19 1999. Date Visited: Friday, January 9 2009.

http://imaging.mrc-cbu.cam.ac.uk/imaging/PrinciplesSmoothing

**Anonymous. Kernel (statistics). Wikipedia. Date Visited: Friday, January 9 2009.

http://en.wikipedia.org/wiki/Kernel_(statistics)

CHAPTER 4: PRIMER ON TEMPORAL AUTOCORRELATION

Mike Janssen

1. Recall Correlation?

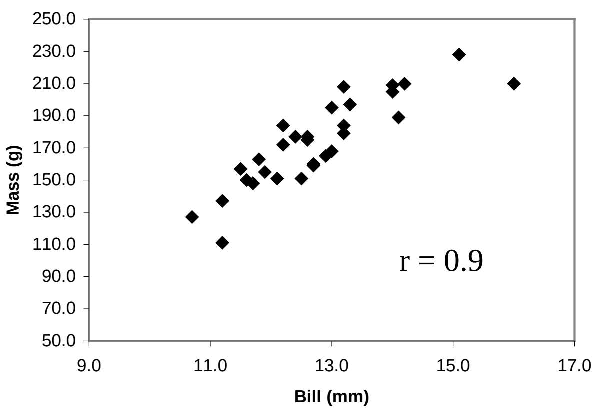
x	y
x_1	y_1
x_2	y_2
x_3	y_3
...	...

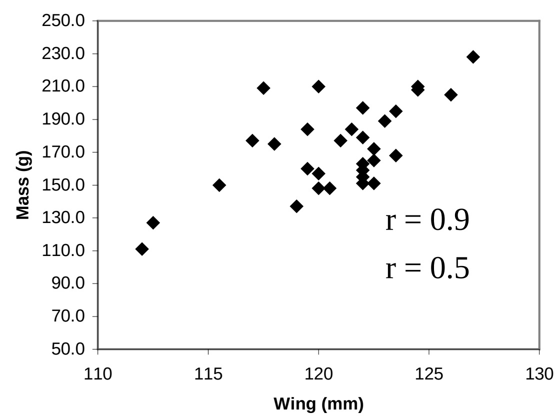
We want to know is x linearly associated with y, so a data point is created from each pair of (x_i, y_i) and we look for a trend.

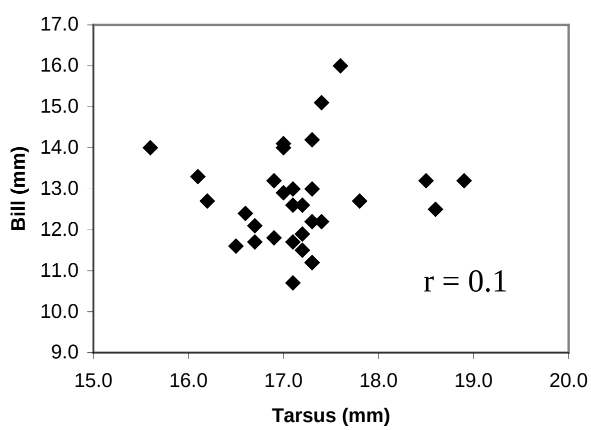
the sample correlation coefficient is given by:

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\left(\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2\right)^{0.5}}$$

Examples: Mike's Data on Marbled Murlets





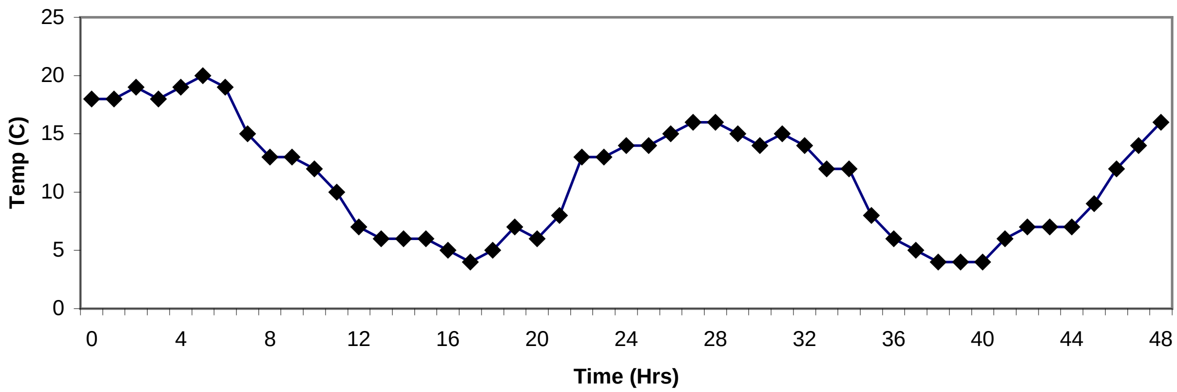


2. What is temporal autocorrelation?

Synonyms: serial dependence, serial correlation

Definition: Temporal autocorrelation occurs when the course of a time series is influenced by it's recent past, or put another way: when successive observations are correlated.

Example: Model Time Series



The temperature at any time (y_t) is the result of the temperature in the previous hour (y_{t-1}) and a host of other meteorological variables.

3. Why is it important?

Serial dependence violates the assumption of independence between observations, necessary to most of the statistics we are familiar with.

Example:
Comparing means:
- A sample average will tend to drift away from the long run mean
- values tend to be closer to each other than would be expected for independent observations.

If our goal is to decompose a time series, we need to be able to account for the influence of autocorrelation before we can properly understand the influence of our explanatory variables.

4. Measuring Autocorrelation

Consider a time series: $y_1, y_2, y_3, y_4 \ldots y_n$ or: 1,3,5,7….n

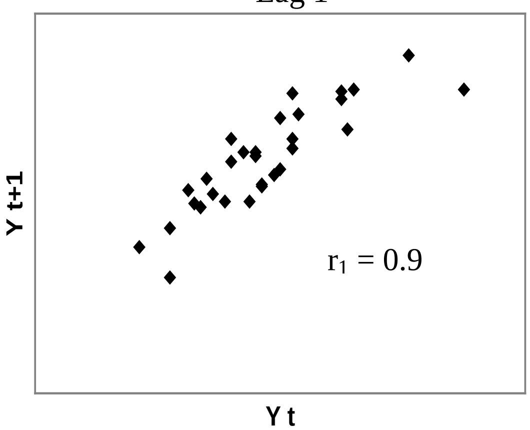
y	y_{t+1}
y_1	y_2
y_2	y_3
y_3	y_4
…	…

or:

y	y_{t+1}
1	3
3	5
5	7

We want to know whether x_t is linearly associated with x_{t+1}, so a data point is created from each pair of (x_t, x_{t+1}) and plotted.

Lag 1



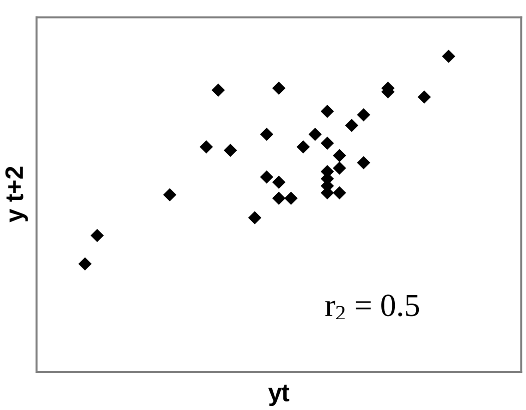
r_1 is the autocorrelation coefficient for a time series with lag of 1

$$r_1 = \frac{\sum_{t=1}^{n-1} (y_t - \bar{y})(y_{t+1} - \bar{y})}{\sum_{t=1}^{n} (y_t - \bar{y})^2}$$

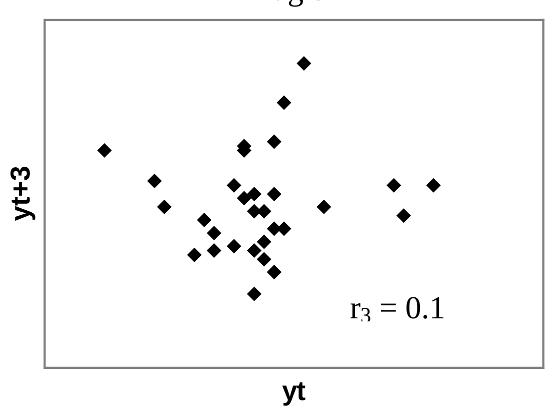
We can calculate the autocorrelation function for any lag ‘k’ using the equation:

$$r_k = \frac{\sum_{t=1}^{n-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^{n} (y_t - \bar{y})^2}$$

Lag 2

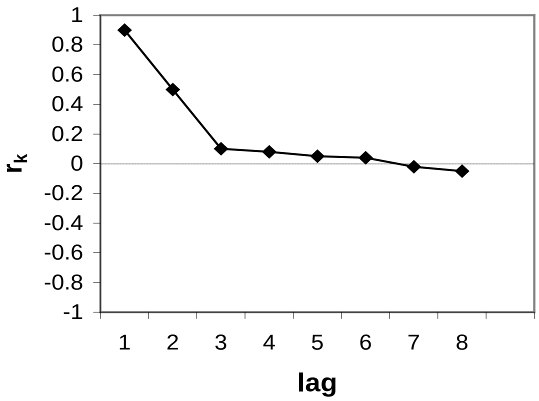


Lag 3



5. The Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF)

The autocorrelation coefficients can be plotted to create the autocorrelation function (or correlogram)



Visual inspection of a correlogram can provide useful information about the nature of our data series.

Alternating series: If successive observations lie on opposite sides of the mean then the ACF will alternate between negative and positive values.

Trends: If a series has a trend (therefore not stationary) then values of r_k will not come down to zero except for very large lag values

Cyclical or Seasonal series: If the data is cyclical, the ACF will also oscillate with similar frequency to the data series.

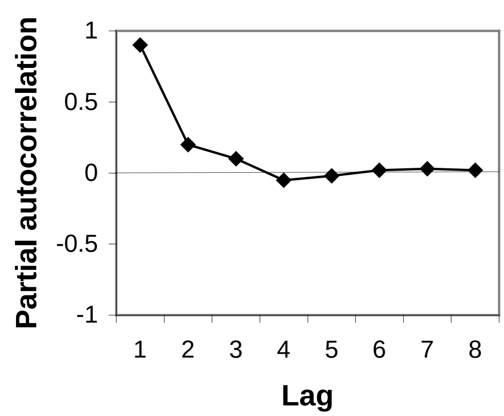
White Noise: If the observations are independent, then r_k values will all be near zero.

But is the effect of r_2 real, or is it the result of the influence of y_0 on y_1 and y_1 on y_2?

Partial correlation refers to the autocorrelation present at a given lag while controlling for the autocorrelation at intermediate lags.
Using the autocorrelation coefficients of lag 1 (r_1) and lag2 (r_2) we can calculate the partial autocorrelation coefficient of lag 2:

$$r_{2.1} = \frac{r_2 - r_1^2}{1 - r_1^2}$$

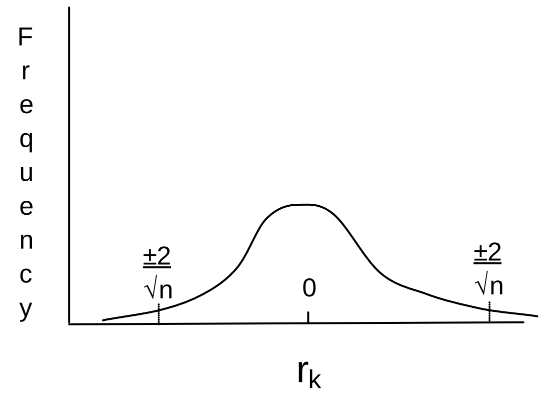
We can plot the result to give the Partial Autocorrelation Function (PACF)



6. Determining if autocorrelation is present (and significant)

For a long series of white noise with *n* observations:

r_k is normally distributed with mean 0 and variance of 1/n. So in 95% of cases r_k will lie between $\pm 2/\sqrt{n}$. Therefore, if r_k is outside these boundaries, you can be 95% sure that you do not have white noise, and your data are not independent of each other.



- Data that goes for long "runs" away from the long term mean is likely autocorrelated.

Another test of serial dependence is the runs test, for details see pp. 448 -450 in Ramsey and Shafer (2002).

7. Autoregression (AR) and the Moving Average (MA) model

- Often a scientist will seek to create a mathematical model that "fits" observed data, or produces an output that matches well with observed data.
- One can use an *autoregressive scheme* to incorporate serial dependence into a model.

An example of a linear 1st order autoregressive scheme AR(1) is:

$$\mathbf{Y_t - \mu = \alpha(Y_{t-1} - \mu) + Z_t}$$

Where:

μ = series mean, α = autoregressive parameter, Z_t = random error term

- Another way to incorporate serial dependence into a model is to use the moving average model.

An example of a 1st order moving average model MA(1) is:

$$\mathbf{Y_t - \mu = \Phi_1 Z_{t-1} + Z_t}$$

Where:

μ = series mean, Φ_1 = Moving average parameter, Z_t = random error term at time t, Z_{t-1} = random error term at time t-1.

Often an autoregressive scheme is used together with a moving average to create an **ARMA** model**.**

Example:
Combining an AR model of order p = 1 AR(1), and a MA model of order q = 1 MA(1), gives an ARMA(1,1) of:

$$\mathbf{Y_t - \mu = \alpha(Y_{t-1} - \mu) + \Phi_1 Z_{t-1} + Z_t}$$

8. Autoregressive Integrated Moving Average Models (ARIMA)

- ARMA models require a stationary time series
- Often a time series can be *differenced* until it appears to be stationary
- When this differencing gets incorporated into the ARMA, it becomes an ARIMA model

References/ Further readings:

Brown, D. & Rothery, P. Chapter 13: Statistical Analysis of Pattern and Sequence: Temporal and Spatial Series and DNA Sequences, *in* Models in Biology: Mathematics, Statistics and Computing. John Wiley and Sons. Ontario.

Chatfield, C. 2004. The analysis of time series, An Introduction, 6th Ed. Chapman and Hall, New York.

Diggle, P.J. 1990. Time Series: a Biostatistical Introduction. Clarendon Press, Oxford.

Ramsey, F.L. & Schafer, D.W. 2002. Chapter 15: Adjustment for Serial Correlation, *in* The Statistical Sleuth, A course in methods of data analysis. Duxbury, New York.

Scheiner, S.M. & Gurevitch, J. 2001. Design and Analysis of Ecological Experiments, 2nd Ed. Oxford University Press. New York. pp 158-177.

CHAPTER 5: PRIMER ON SPECTRAL ANALYSIS

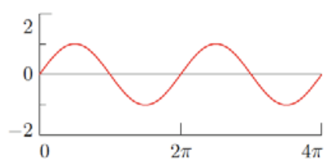
Timothy J. Bartley

I. What is Spectral Analysis?

- A method of decomposing a time series into functions representing the underlying cyclical components of variable frequency and determine which periodic components contribute to the variance of the variable of interest
- Tools like autocorrelation are used to test for serial dependence, but spectral analysis is used to quantify the underlying mathematical patterns
- Accomplished by partitioning the variance between different cycle lengths
- Also known as spectrum analysis

II. Some Reminders

§

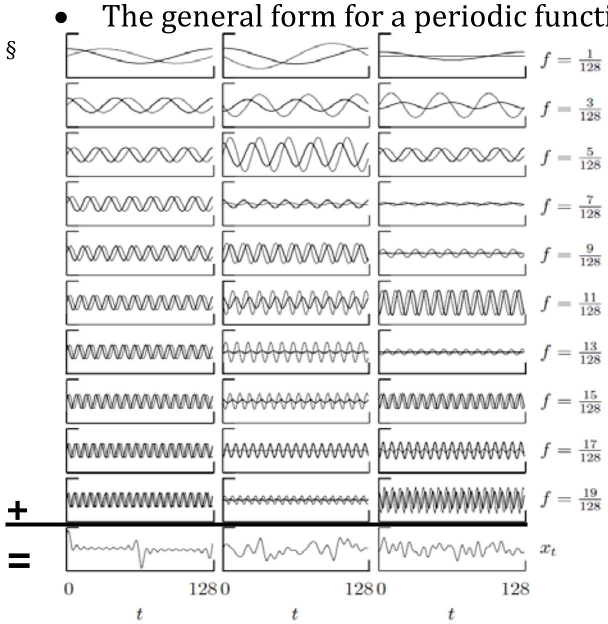


$$y = c + a\sin\omega(t \pm t_0)$$

Where: *c* is equivalent to the mean of the time series
a is equivalent to the amplitude
ω is equivalent to $2\pi \bullet$ frequency or $\frac{2\pi}{\text{period}}$ (because $f = \frac{1}{\tau}$)
t_0 is equivalent to the phase (or lag)
t is the counter of time through N observations

- The general form for a periodic function is:

§



$$y = c + a\cos(\omega t) + b\sin(\omega t)$$

A mathematically convenient and simple representation of cycles, but other waveform functions can be used for these analyses

The Fourier Theorem: *a periodic function can be expressed as the sum of a series of sine and cosine terms.*

III. Some Considerations

- Does the time series contain:
  - At least one full period(required for the following analyses)
  - Sufficient number of data points (a minimum of 50 is recommended)
  - A number of observations that is a multiple of the expected period (if this is known)
  - Variance in y (otherwise these analyses may not be useful)
  - A normal distribution in measurements (transforming the data may be required)
  - Outliers (these will interfere with the detection of periodicity)
  - Equally spaced time intervals (required for the following analyses)
  - Missing data points (these will need to be estimated)
  - Stationarity(required for the following analyses)
  - Trends (these must be removed before the following analyses)
  - Autocorrelation (to be sure a pattern exists in the data)
  - Frequencies expressed in radians (required for the following analyses)

Synopsis: Before doing spectral analysis, examine your data.

IV. Univariate (Single) Spectral Analyses

a. Harmonic Analysis

- Used if the period τ is known *a priori* or from the literature
- A type of regression analysis that estimates the mean, phase and amplitude

$$X_t = \mu + a\cos(\omega t) + b\sin(\omega t) + \varepsilon_t \text{ for } t = (0,\ 1,\ 2,\ \ldots,\ N)$$

Where: μ is the mean of the time series of X
ω is equivalent to $2\pi \cdot f$ or $\frac{2\pi}{\tau}$
ε_t are residuals uncorrelated with the periodic terms
t is the counter of time through N observations

- Varying *a* and *b* varies the relative weight of the sine and cosine functions
- The total amplitude (*r*) of the time series is:

$$r = \sqrt{a^2 + b^2}$$

- The mean and coefficients are estimated as follows:

$$\bar{\mu} = \frac{1}{N} \cdot \sum_{t=1}^{n} X_t \qquad \bar{a} = \frac{2}{N} \sum_{t=1}^{n} (X_t - \bar{X}) \cos \omega t \qquad \bar{b} = \frac{2}{N} \sum_{t=1}^{n} (X_t - \bar{X}) \sin \omega t$$

- To asses how well the period τ fits the observed data, calculate the expected values using that period and conduct an ordinary least squares regression, with R representing the goodness of fit of the model, and overall amplitude estimated with r
- To test for the significance of the model, a test for white noise in the residuals is required, then the significance of the R^2 can be tested

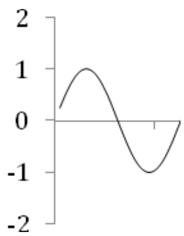
Synopsis: Harmonic analysis estimates the parameters of an underlying periodic function when the period is known.

b. Periodogram Analysis

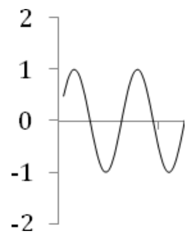
- An expansion of harmonic analysis for a series of periods
- Used when there is no known period
- Related to ANOVA
- Used to estimate which frequencies account for a large percentage of variance in the variable of interest

1. Divides the time series of length N into $\frac{N}{2}$ sinusoidal waveforms

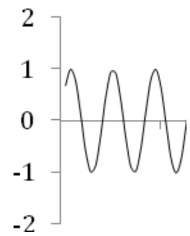
with cycle length $\frac{N}{1}, \frac{N}{2}, \frac{N}{3}, \ldots, \frac{N}{N/2}$ (or 2)



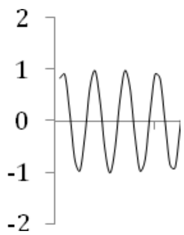
Period = α



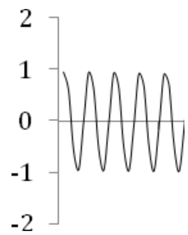
Period = α/2



Period = α/3



Period = α/4



Period=α/5

2. For each of the $\frac{N}{2}$ cyclic components, *a* and *b* coefficients are calculated as follows:

$$\text{For: } X_t = \mu + \sum_k (a_k \cos\omega_{k,t} + b_k \sin\omega_{k,t})$$

$$\bar{\mu} = \frac{1}{N} \cdot \sum_{t=1}^{n} X_t \qquad \bar{a} = \frac{2}{N}\sum_{t=1}^{n}(X_t - \bar{X})\cos\omega_k t \qquad \bar{b} = \frac{2}{N}\sum_{t=1}^{n}(X_t - \bar{X})\sin\omega_k t$$

Where: μ is the mean of the time series of X
ω is $\frac{2\pi k}{N}$ for $k = 1, 2, 3, \ldots, \frac{N}{2}$ (N must be even)
t is the counter of time through N observations

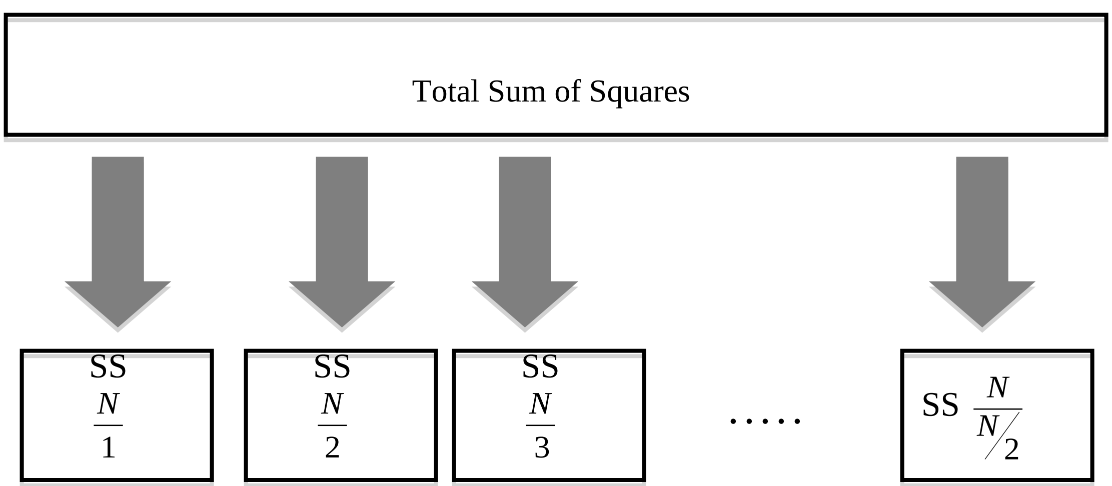
3. The periodogram ordinate, or S_k (the sum of squares accounted for by each of the $\frac{N}{2}$ periodic components) is calculated:

$$S_k = \frac{N}{2} \cdot (a_k{}^2 + b_k{}^2)$$

- The sum of all periodogram ordinates equals the total Sum of Squares:

$$\sum_k S_k = \frac{1}{N}\sum_{t=1}^{N}(X_t - \bar{X})^2 = \sigma_x^2$$

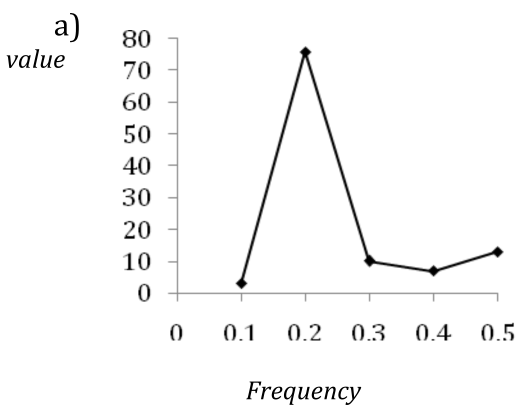
- Sometimes expressed as a percentage of the total Sum of Squares

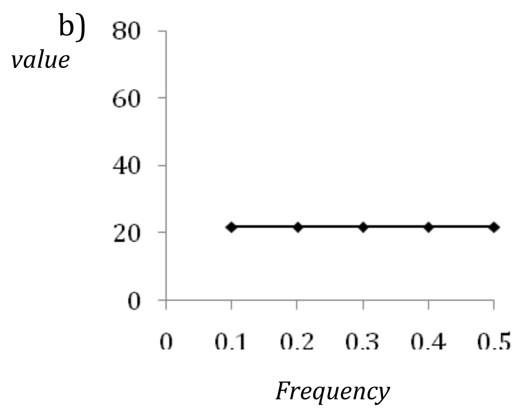


Time	Variable X	Period τ	Frequency f	PDG Value S_k	Percentage $S_k/\Sigma S$
1	x	-	0	0	0
2	x	10	0.1	3	0.027522936
3	x	5	0.2	76	0.697247706
4	x	3.33	0.3	10	0.091743119
5	x	2.5	0.4	7	0.064220183
6	x	2	0.5	13	0.119266055
7	x	-	-	-	-
7	x	-	-	-	-
9	x	-	-	-	-
10	x	-	-	-	-

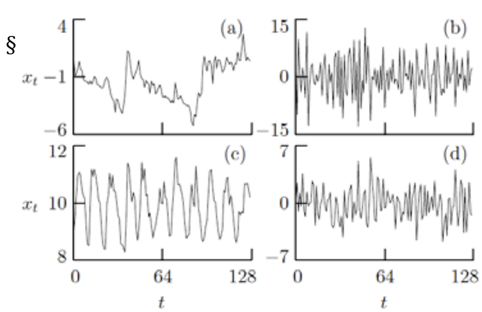
SUM= SS_{total} SUM=100%

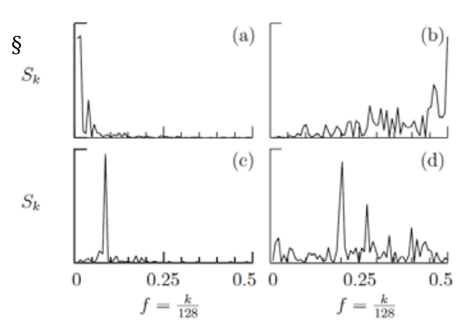
4. Graph the results





- Peaks at frequencies that explain large amounts of variation (see 'a')
- To test for the significance, conduct a Fisher test for each peak to get a CI
- If the variability is due to white noise, the line will be flat (see 'b')
- Watch out for leakage
- Compare to the original series to be sure that the result makes sense





example time series… ⟹ …and their spectra

Synopsis: A periodogram shows you how much variance in your variable of interest is accounted foreach of a number functions of varying frequencies.

c. **Power Spectra**

- Periodograms are susceptible to sampling errors, so smoothing is used on periodogram ordinate values
- However, this smoothing might make detecting the role of distinct periodic components more difficult
- A periodogram with smoothed values is known as a power spectrum
- Typically expressed as 'power density' or 'spectral density', which is calculated by dividing each spectral estimate by the overall power
- Power spectra are tested for significance using a confidence intervals and a χ^2 distribution

value *density*

$\Longrightarrow$

Periodogram Power Spectrum

Synopsis: A power spectrum is a periodogram with smoothed values.

V. Bivariate/Multivariate Spectral Analyses

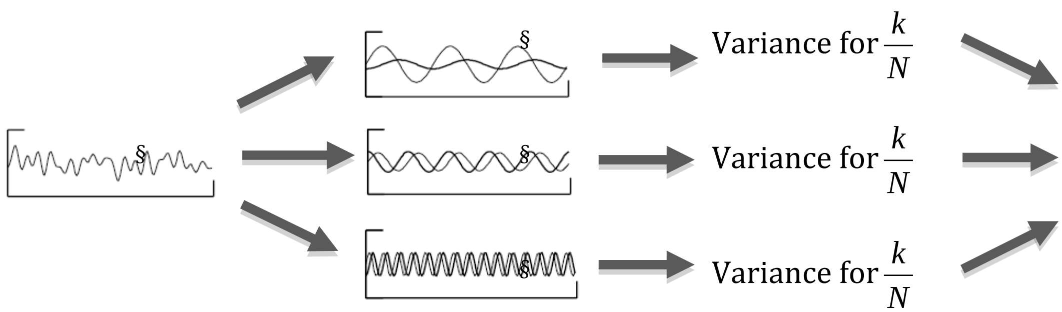
- These analyses are used to compare two variables measured concurrently
- Two time series may correlate in their trends, cycles, residuals or any combination of the three
- One could use a simple Pearson's correlation, but there are many statistical issues with this analysis including serial dependence, spurious correlations between time series and lagged correlations

a. Cross-Spectral Analysis

- Starts by conducting univariate spectral analysis for each time series
- Examines the correlation between time series for each frequency and the lag between time series

Synopsis: Multivariate spectral analyses allow the comparison of concurrently measured variables from time series for which univariate spectral analysis has already been conducted.

VI. Summary



time series → → → spectrum

VI. References and Further Readings:

Fuller, W. A. 1996. Introduction to Statistical Time Series. Second Edition. New York, NY: John Wiley and Sons, Inc. Chapter 4, Spectral Theory and Filtering; p 143-205.

Koopmans, L. H. 1974. The spectral Analysis of Time Series. First Edition. New York, NY: Acedemic Press. 366 p.

§Percival, D. Introduction to Spectral Analysis. Retreived January 26, 2009 from <faculty.washington.edu/dbp/PDFFILES/GHS-AP-Stat-talk.pdf>

Preistly, M. B. 1981. Spectral Analysis and Time Series. First Edition. New York, NY: Acedemic Press. 890 p.

Time Series Analysis. 2008. StarSoft Inc. Accessed January 26, 2009, URL <http://www.statsoft.com/textbook/sttimser.html>

Warner, R. M. 1998. Spectral Analysis of Time Series Data. First Edition. New York, NY: Guilford Press. 225 p.

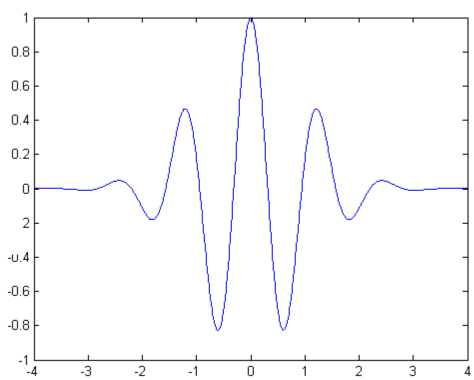
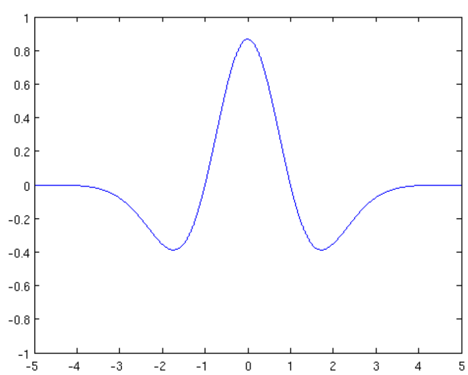
Chapter 6: Primer on WAVELET ANALYSIS
Mark D'Aguiar

Wavelet Analysis: decomposing a time series into time-frequency space.

A Wavelet: The term **wavelet** means a **small wave** . The smallness refers to the condition that this (window) function is of finite length (**compactly supported**). The wave refers to the condition that this function is oscillatory .

- The mother wavelet is a **prototype** for generating the other window functions.
- a mathematical function used to divide a given function or continuous-time signal into different scale components.
- The wavelets are scaled and translated copies (known as "daughter wavelets") of a finite-length or fast-decaying oscillating waveform (known as the "mother wavelet").
- Must have a mean of zero.

Wavelet Transform: is the representation of a function by wavelets.
-Wavelet transforms are classified into discrete wavelet transforms (DWTs) and continuous wavelet transforms (CWTs).
Examples of wavelets (mother wavelets):

Morlet wavelet: $\varphi(t) = \pi^{-1/4} \exp(-i2\pi f_0 t) \exp(-t^2/2)$,	**Mexican Hat: Continuous** $g_2(z) = (z^2 - 1)\exp(-z^2/2)$,
Scale  Time	 Time

NOTE: An admissibility condition must be satisfied; satisfied as long as

$$\int_{-\infty}^{\infty} g_2(t)\, dt = 0 \qquad (2)$$

<u>Why Transform?</u> Transformations are applied to time series (signals) to obtain further information from that signal that is not readily available in the raw signal.

- The parameters of' scale' and 'translation' make it possible to ZOOM IN on the transient behavior of a signal. I.e. parameters make it possible to analyze the behavior of a signal at a dense set of time locations and with respect to a large range of scales

Wavelet Analysis:

Recall:

Spectral analysis

- Decomposes a signal into its harmonic component based on the Fourier analysis. Which is regarded as the partition of the Variance of the series in its different oscillating components with different frequencies (periods).
- Peaks in the 'periodogram' indicated which frequencies are contributing the most to the variance of the series.
- Spectral and Fourier analysis can determine all spectral components in a signal, but does not provide any information to when they are present.

Assumption: statistical properties of the time series **do NOT vary** with time (AKA stationary).

Wavelets:

- Wavelet analysis is similar to Fourier analysis in the sense that it breaks a signal down into its constituent parts for analysis.
- The wavelet transform breaks the signal into its "wavelets", scaled and shifted versions of the "mother wavelet".
- It is these properties of being irregular in shape and compactly supported that make wavelets an ideal tool for analyzing signals of a non-stationary nature.
- Their irregular shape lends them to analyzing signals with discontinuity's or sharp changes, while their compactly supported nature enables temporal localization of a signals features.

Performs **local time-scale** decompositions of the signal = estimation of its **spectral characteristics in time.**

Advantages of Wavelet Analysis:

1. Wavelet analysis overcomes the problems of non-stationarity in time series by performing a local time-scale decomposition of the signal, i.e., the estimation of its spectral characteristics as a function of time. Through this approach one can track how the different scales related to the periodic components of the signal change over time.

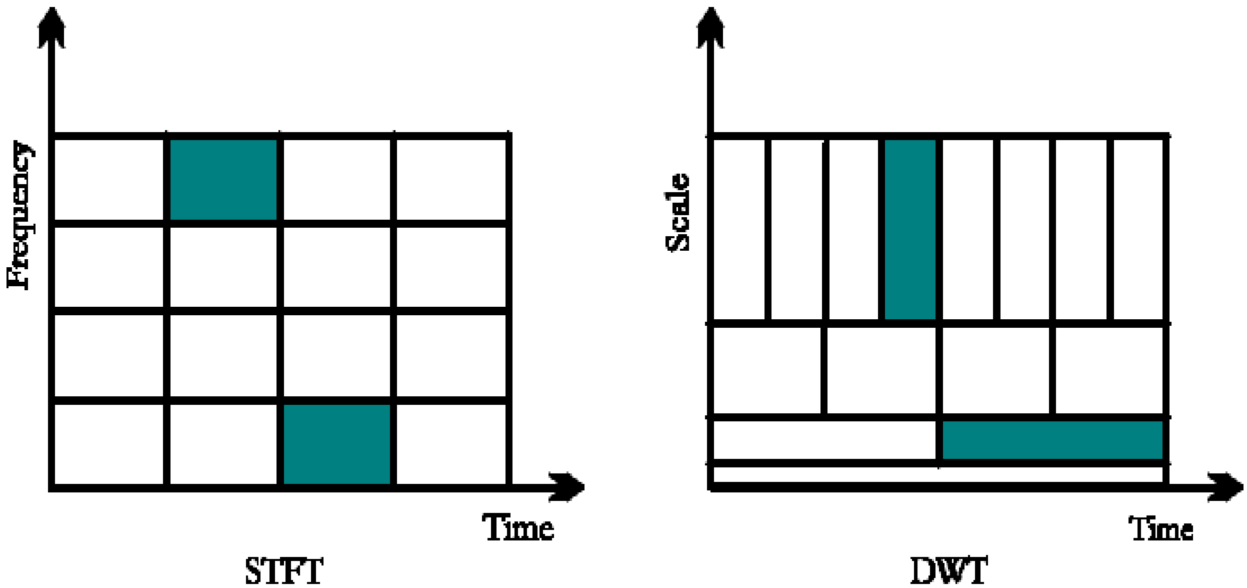
2. Wavelet analysis permits analysis of the relationships between two signals, and it is especially appropriate for following gradual change in forcing by exogenous variables.

General Summary of Fourier vs. wavelets

If you are not interested in **at what times these frequency components occur**, but only interested in what frequency components exist in a signal, then FT can be a suitable tool to use.

Fourier transform (FT) assumes stationary. i.e frequencies present at all time intervals

- STFT assumes small time intervals of stationary, and is basically the FT multiplied by a 'Window Function' 'w'.



Short term Fourier Transport **Discrete Wavelet Transform**

Figure: Comparison of STFT with Discrete Wavelet Transform-

- Windowed Fourier transform of **fixed time and frequency resolution.**
- The **wavelet transform offers superior temporal resolution of the high frequency components and scale (frequency) resolution of the low frequency components.**

This is often beneficial as it allows the low frequency components, which usually give a signal its main characteristics or identity, to be distinguished from one another in terms of their frequency content, while providing an excellent temporal resolution for the high frequency components which add to the signal.

Fourier series Gives frequency information. Basis functions last the entire interval.	**Wavelets** Wavelet basis functions give frequency info but are *local* in time.

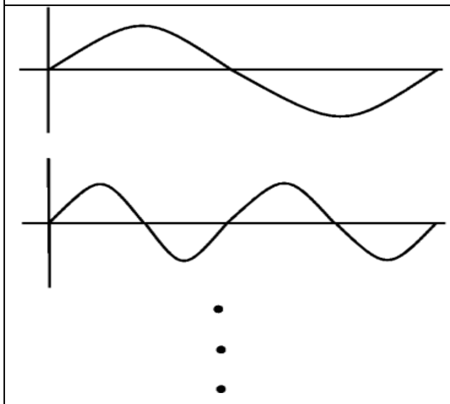
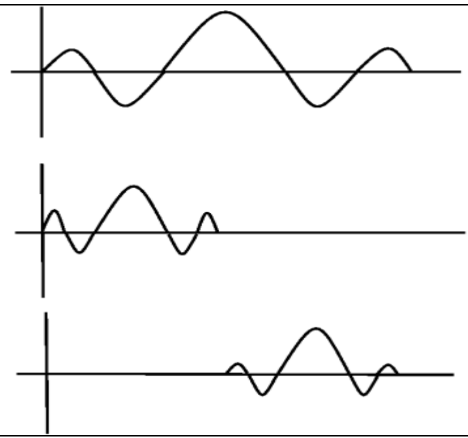

	

Figure: Fourier basis functions	**Figure :** Wavelet basis functions

In Fourier basis, the basis functions are *harmonic multiples*

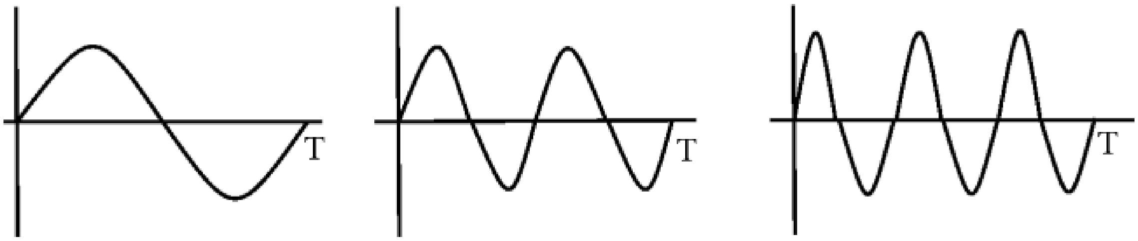


Figure: Fourier basis

In [Wavelets](), the basis functions are *scaled and translated* versions of a "mother wavelet" $\psi(t)$. Where ‘j’ is the coefficient, and ‘k’ is the ‘translation’ (or shift) coefficient.

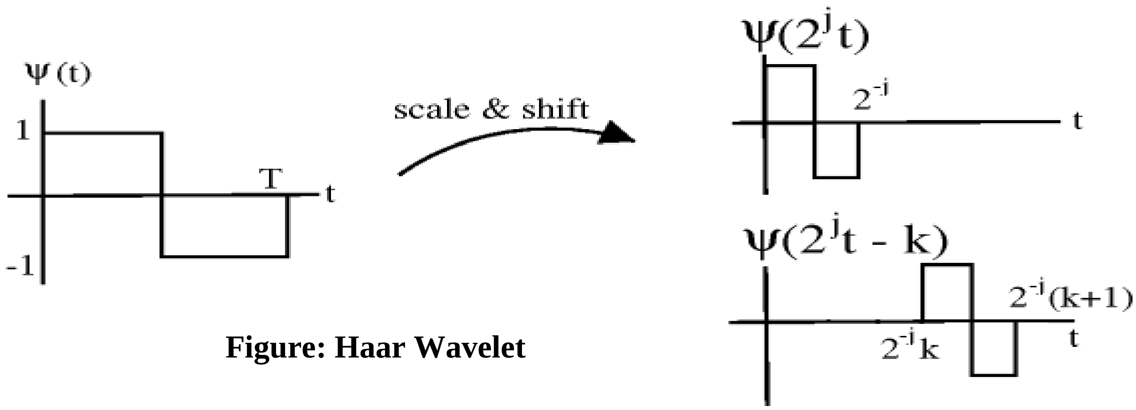


Figure: Haar Wavelet

Continuous wavelet approach

The wavelet transform decomposes signals over dilated and translated functions called "mother wavelets" $\varphi(t)$ that can be expressed as the function of two parameters, one for the time position (τ), and the other for the scale of the wavelets (a). More explicitly, wavelets are defined as

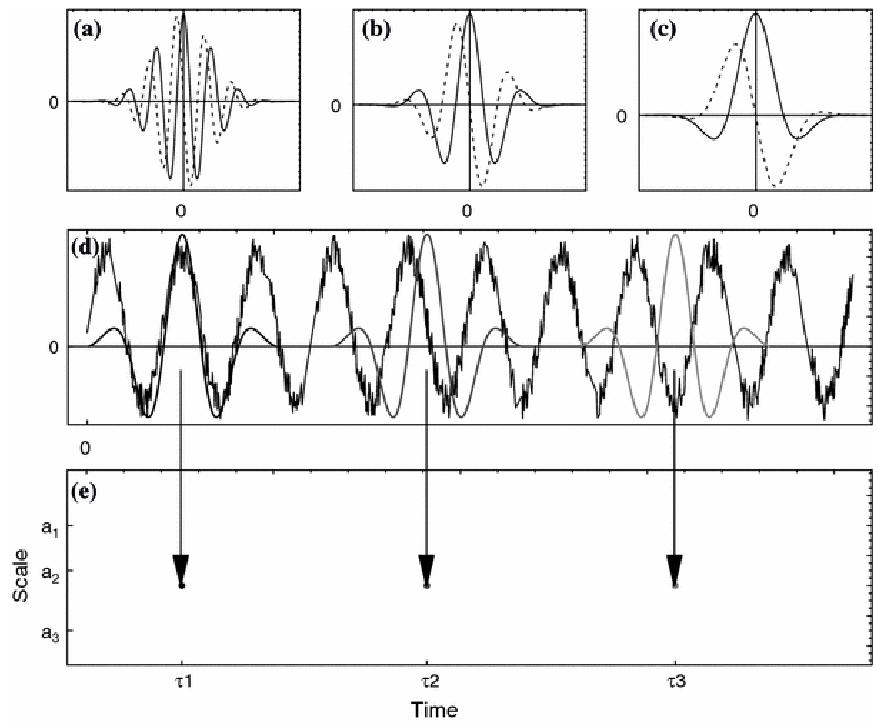
$$\varphi_{a,\tau}(t) = \frac{1}{\sqrt{a}}\varphi\left(\frac{t-\tau}{a}\right) .$$

The wavelet transform of a time series $x(t)$ with respect to a chosen mother wavelet is performed as follow:

$$W_x(a,\tau) = \frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} x(t)\varphi^*\left(\frac{t-\tau}{a}\right)dt = \int_{-\infty}^{+\infty} x(t)\varphi^*_{a,\tau}(t)dt$$

where * denotes the complex conjugate form. The wavelet coefficients, $W_x(a,\tau)$, represent the contribution of the scales (the a values) to the signal at different time positions (the τ values).

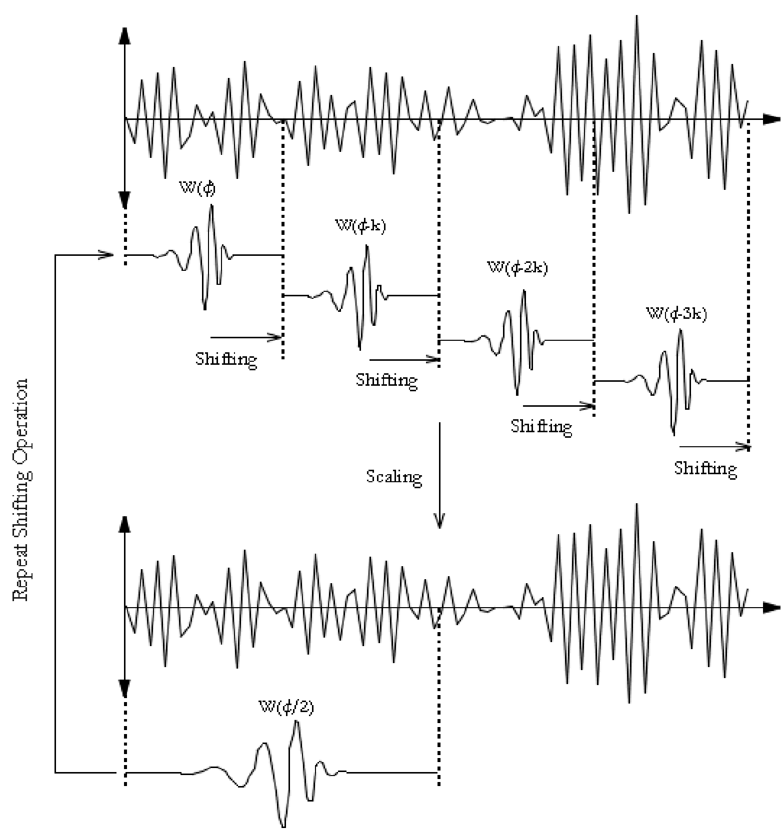
The wavelet transform can be thought as a cross-correlation of a signal $x(t)$ with a set of wavelets of various "widths" or "scales" a, at different time positions τ.



Wavelet analysis: (a-c) form of Morlet wavelet as a function of parameters 'a' for τ = 0: Real and imaginary parts of wavelet.
d) Morlet is superimposed and moved across the signal at different time positions (τ1τ2τ3) etc.
e) the fit of the morlet is plotted as a two dimensional plot.
τ1 = match of wavelet is high, thus high R(Wx) value.
τ 2 = weak match thus low R(Wx) value.
τ3 = perfect opposition, thus high negative R(Wx) value.
Where R is the Real part of wavelet.

Discrete Wavelet transform
The continuous wavelet transform was computed by changing the scale of the analysis window, shifting the window in time, multiplying by the signal, and integrating over all times.

Discrete Wavelets: filters of different cutoff frequencies are used to analyze the signal at different scales. The signal is passed through a series of high pass filters to analyze the high frequencies, and it is passed through a series of low pass filters to analyze the low frequencies.



Subsampling a signal

Corresponds to reducing the sampling rate, or removing some of the samples of the signal.

i.e. subsampling by two refers to dropping every other sample of the signal.

Subsampling by a factor *n* reduces the number of samples in the signal *n* times.

- A DWT is non-redundant.
- The number of blocks of wavelet power at each scale is a function of non-overlapping wavelet width.
- In a typical DWT, frequencies are spaced at unit powers of two and the count of blocks in time will increase by unit powers of two as these fixed frequencies increase.
- the DWT is fast and its time-frequency representation of a signal requires only modest memory, it is not practical for time-frequency spectral analysis

Choice of the mother wavelet:

There are several considerations in making the choice of a wavelet, for example;

1. **real versus complex wavelets**: Complex returns phase information, a real only power, but is useful in pinpointing peak frequency.
2. **continuous or discrete wavelets**: Continuous = redundant decomposition but more robust to noise. Discrete = fast implementation but number of scales at the time interval depend on data length.

- If information about phase interactions b/w 2 series- continuous and complex (Morlet, Mexican hat).

3. **Wide vs. narrow:** It's a trade off. A wide wavelet function will give good frequency resolution at the loss of time resolution, while a narrow wavelet function will yield good time resolution and poor frequency resolutions.
4. **Shape:** Reflect the type of feature in the time series. Records with sharp jumps of sets should use a box-car like Haar, while smooth use Morlet or cosine type function.

Wavelet Power spectrum:

- Allows quantification of the main periodic component of a given time series and its evolution through time.

Local Wavelet Power spectrum:

Computed by first taking a discrete Fourier transform of the time series

$$S_x(f,\tau) = ||W_x(f,\tau)||^2$$

The Fourier spectrum of a signal can be compared with the global wavelet power spectrum.

Global Wavelet Power Spectrum:

The averaged variance contained in all wavelet coefficients of the same frequency ***f:***

$$\bar{S}_x(f) = \frac{\sigma_x^2}{T}\int_0^T ||W_x(f,\tau)||^2\, d\tau$$

Where:
σ_x^2 is the variance of the time series *x* and *T* is the duration of the time series

Mean Variance at each time location:

Obtained by averaging the frequency components:

$$S_x(\tau) = \frac{\sigma_x^2\, \pi^{\frac{1}{4}} \tau^{\frac{1}{2}}}{C_g} \int_0^\infty \left(\frac{1}{f}\right)^{\frac{1}{2}} ||W_x(f,\tau)||^2\, df$$

Where $C_g = \int_0^\infty \frac{||\varphi(f)||^2}{f}\, df$

Wavelet Coherency and Phase Difference:

- Used to measure the direct correlation between the spectra of two non stationary time series. i.e . Quantify statistical relationships between 2 non stationary signals.
- In Fourier, the coherency is used to determine the association between x(t) and y(t).
- The wavelet coherence $R_{x,y}(f,\tau)$, is equal to 1 when there is a perfect linear relation at a particular time location and frequency between the two signals *x(t) and y (t)* respectively.

Zero padding and cone of influence:

- An artificial increase in the length of the time series to the next higher power of two by adding zero-value samples.
- helps to avoid false 'wrap around' periodic events.

Disadvantage: as wavelet gets closer to the edge of the time series, part of it will exceed the edge, and thus artificially decreasing the value of the wavelet transform.

Cone of influence = zone where the 'edge' effects are present.

Criteria for applying Wavelet analysis

1. Minimum time series with at least 30-40 data points with periodic components smaller than 20-25% of the series length.

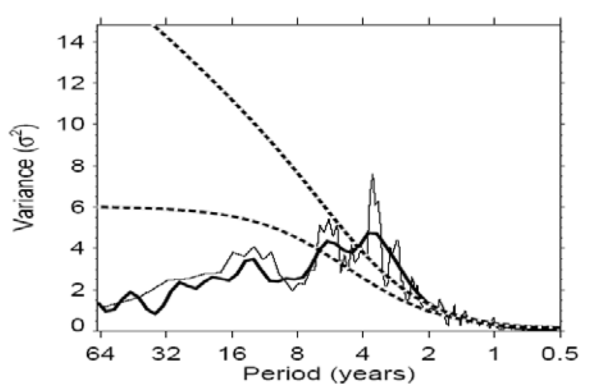


FIG. 6. Fourier power spectrum from Fig. 3, smoothed with a five-point running average (thin solid line). The thick solid line is the global wavelet spectrum for the Niño3 SST. The lower dashed line is the mean red-noise spectrum, while the upper dashed line is the 95% confidence level for the global wavelet spectrum, assuming $\alpha = 0.72$.

Statistical Significance:

1. Is the spectra observed at a particular position on the time scale due to random processes.
2. Determine background spectra – white noise or red noise(i.e. bootstrapping)

To compute the wavelet transform for a time series are thus:

1. Choose a mother wavelet,
2. Find the Fourier transform of the mother wavelet,
3. Find the Fourier transform of the time series,
4. Choose a minimum scale α_0,
5. For each scale, do:
   - Using the equation appropriate for your mother wavelet),
   - Compute the daughter wavelet at that scale;
   - Normalize the daughter wavelet by dividing by the square-root of the total wavelet variance (the total of $(\psi)^2$ should then be 1, thus preserving the variance of the time series);
   - Multiply by the FT of your time series;
   - Inverse transform back to real space;

6. Make a contour plot.
7. Define confidence limits based on auto-regressive red or white noise.

References:

Torrence, C. and .P Compo (1998). A practical guide to Wavelet Analysis. Bulletin of the American Meteorological Society.

Cazelles, B., Chavez, M., Berteaux, D., Menard, F., Olav, Vik., Jenourier, S., and N. Stenseth. (2008). Wavelet analysis of ecological time series. Oecologia. 156: 287 -304

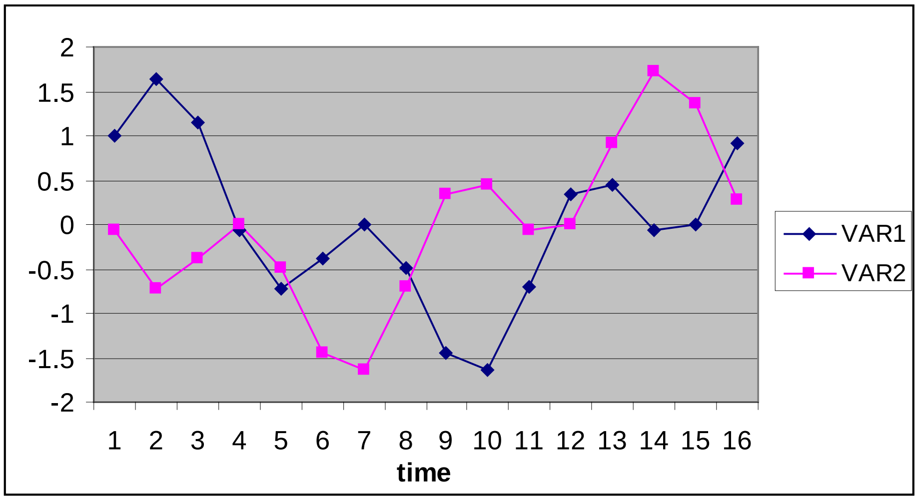
Anonymous. A Really Friendly Guide to Wavelets . (2004) http://pagesperso-orange.fr/polyvalens/clemens/wavelets/wavelets.html#eq2

PRIMER ON COHERENCE ANALYSIS

Gale Bravener

Let's start with a data set with two variables (2 different time series)

time	VAR1	VAR2
1	1.000	-.058
2	1.637	-.713
3	1.148	-.383
4	-.058	.006
5	-.713	-.483
6	-.383	-1.441
7	.006	-1.637
8	-.483	-.707
9	-1.441	.331
10	-1.637	.441
11	-.707	-.058
12	.331	-.006
13	.441	.924
14	-.058	1.713
15	-.006	1.365
16	.924	.266



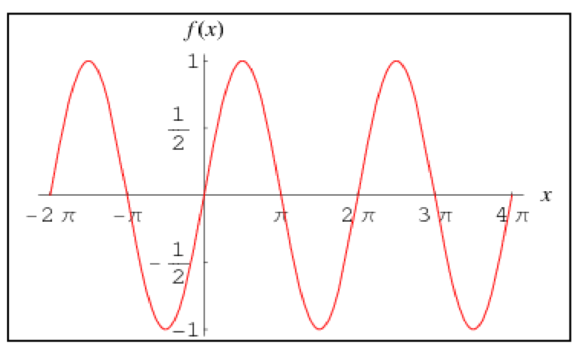
(Statsoft website)

Some analyses we have are now capable of doing with these time series:

Smoothing	Spectral Analysis
Correlation	Cross-Correlation
Regression	Cross-Spectral Analysis
Auto-Correlation	Wavelet Analysis

<u>Spectral Analyses:</u>

- Spectral analyses (Periodogram / Power spectrum) decompose a complex time series into a few underlying periodic (sine and cosine) functions, to uncover one or more recurring cycles of different lengths, which at first may have just looked like random noise.



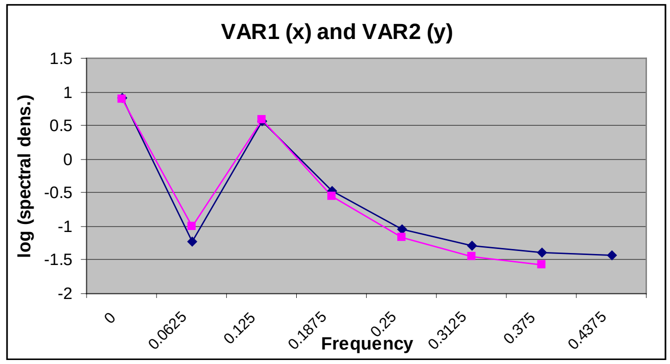
- The purpose of spectral analysis is to identify cycles of different lengths, rather than using the length of the seasonal component which is known *a priori* and then including it in some theoretical model of moving averages or autocorrelations (ARMA).

Spectral Analysis (univariate)

(Using our example data from page 1)

Spectral Analysis results:					
Frequency	Period	Cosine Effects	Sine Effects	X density	Y density
0		0	0	0	0.024
0.0625	16	1.006	0.028	8.095	7.798
0.125	8	0.033	0.079	0.059	0.101
0.1875	5.33	0.374	0.559	3.617	3.845
0.25	4	-0.144	-0.144	0.333	0.278
0.3125	3.2	-0.089	-0.06	0.092	0.067
0.375	2.67	-0.075	-0.031	0.053	0.036
0.4375	2.29	-0.07	-0.014	0.04	0.026
0.5	2	-0.068	0	0.037	0

Periodogram for each time series (var1 and var2)

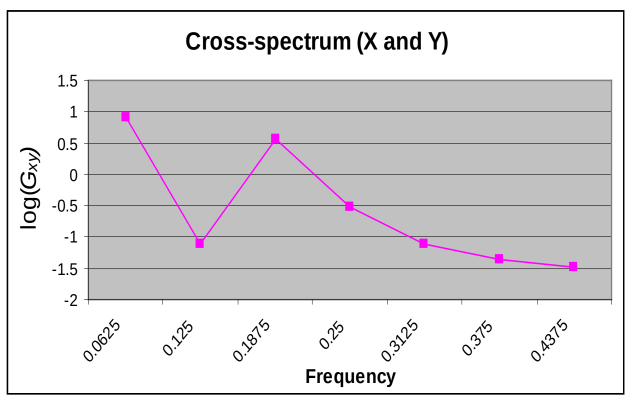


Cross-Spectral Analysis (multivariate)

- Used for two or more time series with concurrently measured variables
- Purpose is to uncover potential correlation, and lag, between two time series
- The cross-spectral density function of two sets of random data evolves directly from the cross-correlation function.

Cross spectral analysis results :				
Frequency	Period	Cross-Density	Cross-Quadrature	Cross-Amplitude
		$C_{xy}(f)$	$Q_{xy}(f)$	$\vert G_{xy}(f)\vert$
0		0.000	0.000	0.000
0.0625	16	2.356	-7.588	7.945
0.125	8	-0.048	0.061	0.077
0.1875	5.33	-2.926	2.312	3.729
0.25	4	-0.269	0.142	0.305
0.3125	3.2	-0.074	0.026	0.079
0.375	2.66	-0.043	0.009	0.044
0.4375	2.285	-0.033	0.003	0.033
0.5	2	0.000	0.000	0.000

Cross-periodogram of the two time series



These are some of the values from the cross-spectral analysis *(from Statsoft website)*.

BACKGROUND
Cross Spectral Analysis

1. *Cross spectral density function* is calculated as:

$$G_{xy}(f) = C_{xy}(f) - i\, Q_{xy}(f)$$

Where *i (sometimes called j)* = $\sqrt{-1}$.

Unlike the power spectrum, the cross-spectrum is complex valued (consists of a real and an imaginary part) as it contains amplitude and phase information.

- The real part of the cross-spectrum, $C_{xy}(f)$, is known as the *cross-density* (*coincident spectrum, co-spectrum, or coincident spectral density function).* It gives the in-phase correlation at a given frequency between two series.

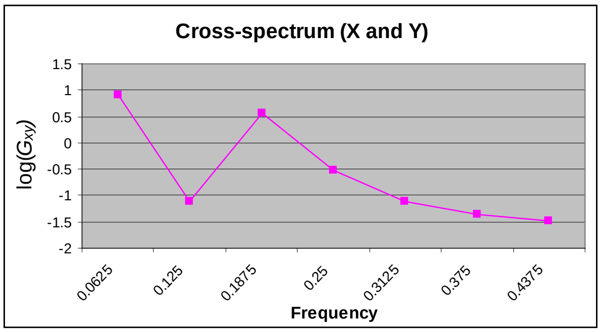
- The complex (or imaginary) part of the cross-spectrum is known as the *quadrature spectrum (quad-spectrum, or cross-quadrature spectral density)*. It gives the correlation at a given frequency between the two series where the time axis of one has been shifted by a quarter of a wavelength. As an example, the sine and cosine functions are in perfect quadrature (Platt and Denman 1975).

2. *Cross spectral density function* can also be expressed as:

$$G_{xy}(f) = |G_{xy}(f)|\, e^{-j\,\theta_{xy}(f)}$$

Where the *magnitude* (or *cross-amplitude*), $|G_{xy}(f)|$ and the *phase angle (or phase shift or phase spectrum)*, $\theta_{xy}(f)$ are related to $C_{xy}(f)$ and $Q_{xy}(f)$ by:

$$|G_{xy}(f)| = \sqrt{C_{xy}(f)^2 + Q_{xy}(f)^2} \qquad \textit{and} \qquad \theta_{xy}(f) = \tan^{-1}[Q_{xy}(f) / C_{xy}(f)]$$



Cross-amplitude, $|G_{xy}(f)|$ can be interpreted as a measure of covariance between the respective frequency components in the two series at *f*. **We can conclude from the results shown in the table above that the .0625 and .1875 frequency components in the two series covary.**

The *phase angle (phase shift)*, $\theta_{xy}(f)$ estimates are measures of the extent to which each frequency component of one series leads the other.

COHERENCE (aka: coherency, squared coherence, squared coherency, coherence function)

When applying cross-spectral density information to physical problems, it is often desirable to use a real-valued quantity, given by the *coherence.*

$$\gamma^2_{xy}(f) = \frac{|G_{xy}(f)|^2}{G_{xx}(f)\, G_{yy}(f)} \qquad 0 \geq \gamma^2_{xy}(f) \leq 1$$

Where f is the frequency, G_{xy} is the cross spectrum density of $x(t)$ and $y(t)$, G_{xx} is the power (or auto) spectrum of $x(t)$ and G_{yy} is the power (or auto) spectrum of $y(t)$ (Bendat and Piersol 1971).

What does it tell us?

Coherence indicates how well x corresponds to y at each frequency. It is analogous to the coefficient of determination (R^2) in simple correlation (Harris 1967). "The coherence function is a quantitative measure of the linear correlation between two random variables" (Loewen et al. 2007).

- When $\gamma^2_{xy}(f)$ is near 0 at a particular frequency, $x(t)$ and $y(t)$ are said to be incoherent at that frequency, which simply means they are uncorrelated.
- When $\gamma^2_{xy}(f)$ is near 1 $x(t)$ and $y(t)$ are said to be fully coherent.
- If $\gamma^2_{xy}(f)$ is between one and zero, there may be extraneous noise present in the measurements, the system relating $x(t)$ and $y(t)$ is not linear and/or $y(t)$ is an output due to an input $x(t)$ as well as other inputs.

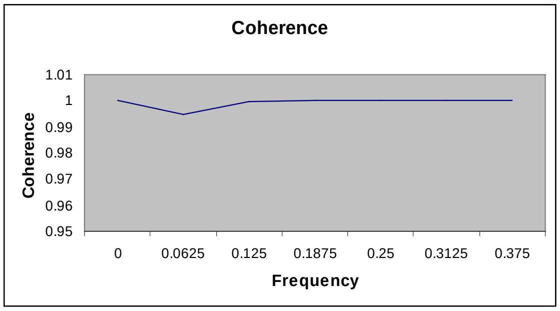
Important to note: The coherence is non-negative (between 0 and 1) because it measures the correlation between aligned frequency components.

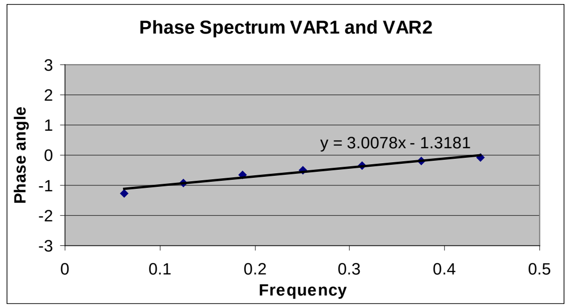
For example, if *x(t)* shows a strong pattern of alternating positive and negative values and *y(t) equals approximately -x(t),* then processes *x(t)* and *y(t)* will be strongly coherent but out of phase (Bendat & Piersol 1971).

Phase is measured by the phase spectrum, $\theta_{xy}(f)$

Time Delay Application:

It is the slope of the phase spectrum that corresponds to the delay between *x(t)* and *y(t).*



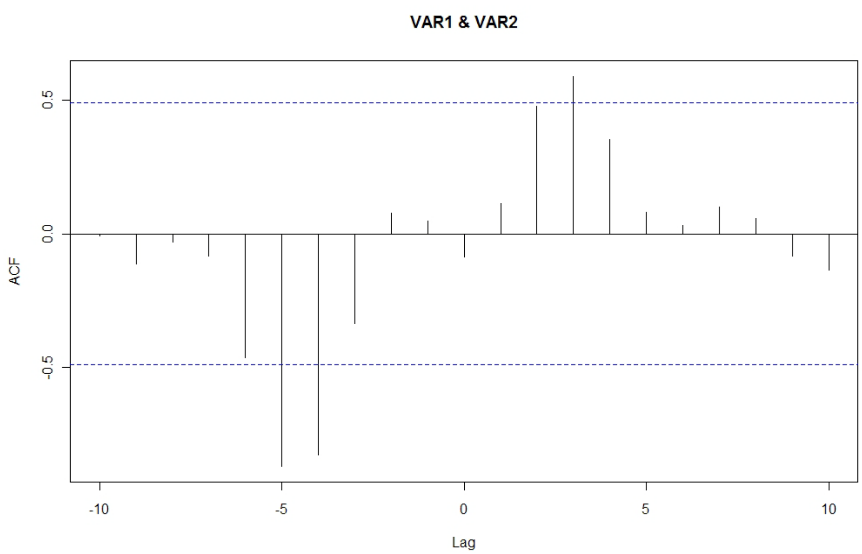


For example, a slope of 3.01 may suggest that series *x(t)* leads *y(t)* by 3.01 units (eg. 3 years). However, because of the inverse tangent ($\tan^{-1}$), the phase angle is between π and −π and it is liable to produce discontinuities in the phase spectrum. This could affect our interpretation of the slope.

Note: lag for example data should be 3 years. Slope of linear trendline for phase angle = 3.01. So in this case, the phase angle does very well at estimating the time delay.

Dr. Mark Loewen (U of Alberta) suggests that in some cases, a better method to determine time delays is to plot the cross-correlation function, or *cross-correlogram,* which is calculated to incorporate a lag of k, as:

$$r_{xy}(k) = g_{xy}(k)/\sqrt{(g_{xx}(0)g_{yy}(0))}$$



ADVANCED TOPICS

Wavelet Coherence:

- Along with the wavelet cross-spectrum and phase spectrum, coherence is calculated for the same applications and using the same equations used as those described above. Used to quantify the relationships between two non-stationary signals (Cazelles et al. 2008, Rouyer et al. 2008 – see Feb. 23 readings).

- Equal to 1.0 when there is a perfect linear relation at a particular time location and frequency between the two signals.

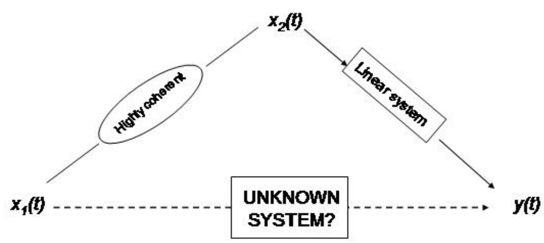
Partial Coherence Functions:

- For multiple input, single output systems
- Used to reveal the existence of a linear relationship between $\Delta x_1(t)$ and $\Delta y(t)$ even when such a relationship is not apparent from the ordinary coherence function between $x_1(t)$ and $y(t)$

Calculated as: $$\gamma^2_{1y}(f) = \frac{|S_{1y}(f)|^2}{G_{11}(f)\, G_{yy}(f)} \qquad 0 \geq \gamma^2_{1y}(f) \leq 1$$

Example:

If we assume that coherence between $x_1(t)$ and $y(t)$ =1, we may be inclined to believe there is a linear relationship between these two variable. But if there is a third variable, $x_2(t)$, which is highly coherent with $x_1(t)$. In this case, the high coherence between $x_1(t)$ and $y(t)$ might only be a reflection of the fact that $x_2(t)$ is highly coherent with $x_1(t)$ and $x_2(t)$ is related via a linear system to $y(t)$. If the partial coherence is computed between $x_1(t)$ and $y(t)$, it might turn out to be a very small number near zero.



Multiple Coherence Functions

- Combines ordinary coherence and partial coherence.
- Used for multiple inputs and one output.

For example, a two-input single-output linear model was used in Loewen et al. (2007) to investigate the coherence between water level and wind speed (two inputs) and water current (one output).

References:

Bendat, J.S. and Piersol, A.G. 1971. Random Data: Analysis and Measurement Procedures. John Wiley & Sons, Inc. New York.

Diggle, P.J. 1990. Time Series: A Biostatistical Introduction. Clarendon Press, Oxford.

Harris, B. 1967. Spectral Analysis of Time Series. John Wiley and Sons, Inc.

Loewen, M. R., Ackerman, J. D., Hamblin, P.F. 2007. Environmental implications of stratification and turbulent mixing in a shallow lake basin. Can. J. Fish. Aquat. Sci. 64: 43-57.

Platt, T. and Denman, K.L. 1975. Spectral Analysis in Ecology. Annu. Rev. Ecol. Syst. 6: 189-210

Priestley, M.B. 1981. Spectral Analysis and Time Series Volume 2. Multivariate Series, Prediction and Control. Academic Press, New York.

Time Series Analysis. 2008. Statsoft, Inc. Accessed February 22, 2009, URL http://www.statsoft.com/textbook/stathome.html

CHAPTER 8: PRIMER ON SPATIAL DISTRIBUTION

Justin Sheehy

A distribution (or set of geographic observations) representing a particular phenomenon or characteristic across a landscape or location.

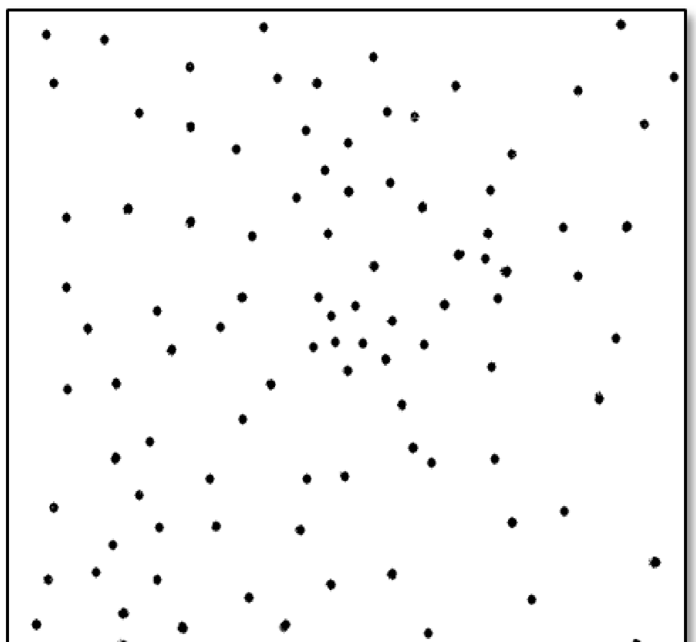
There are three spatial distributions that are typically considered:

1. Random
2. Uniform or Regular
3. Aggregated or Clumped

Random Distribution

For a distribution to be considered random, it must satisfy two conditions:

1. Condition of equal probability – any point has an equal probability of occurring at any position on the plane.

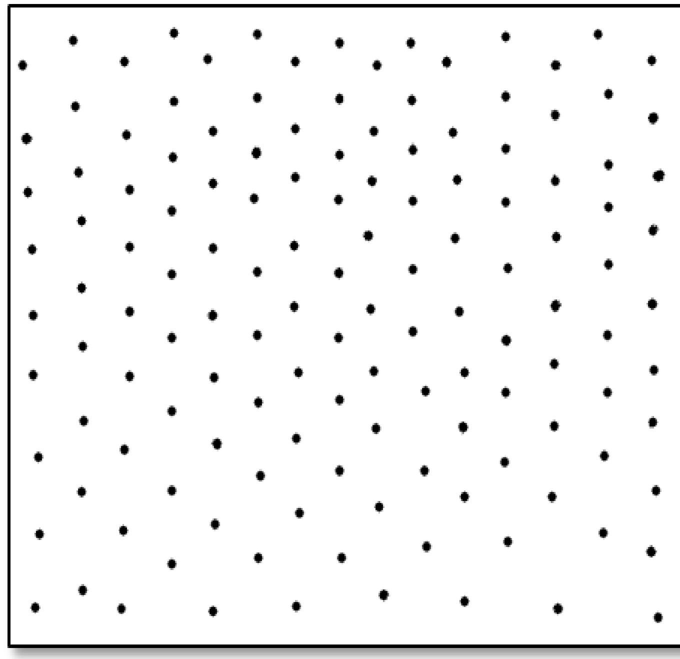


2. Condition of Independence – The position of any point on the plane is independent of the position of any other point

- Tend to be modeled/represented with a Poisson distribution, equal mean and variance.
- If *N* points are located randomly in a region, then the probability that a point falls within a particular subdivision of area *A* can be seen as an event which occurs with probability λA, where λ is the density.
- Not common pattern amongst animal populations
- Very common amongst wind dispersed plant populations (i.e. Dandelions)

Uniform/Regular Distribution

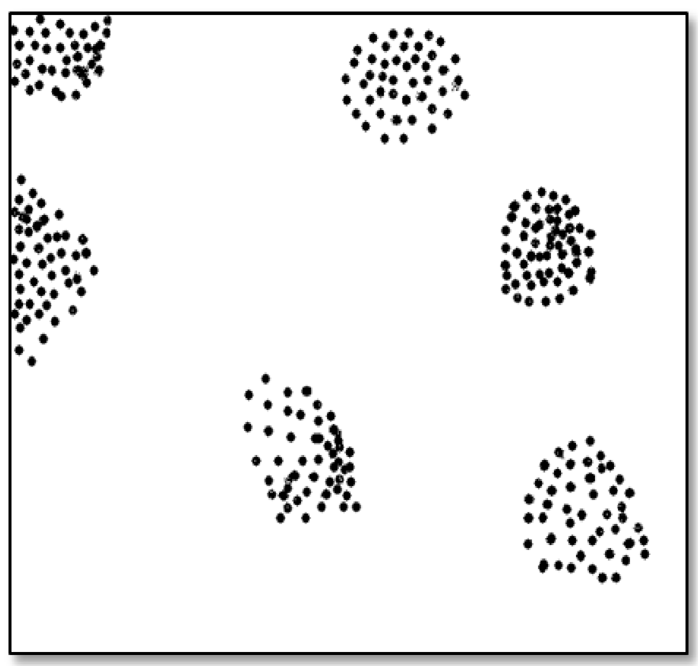
- A geographical distribution in which each member has its own space and all members are approximately the same distance away from their nearest neighbour
- Tend to be modeled/represented with a binomial distribution
- Very common amongst bird populations, especially during breeding season
- Attack patterns of Wood Beetle on Norway Spruce Logs (Byers 1984)
- I.e. Nesting birds on the beach will be at uniform distances from one another
    - o Distance tends to equal neck reach of bird
- In Areas of High Predation, can be the safest pattern strategy for nesting
    - o Upland thicket hens (Picman 1988)



Aggregated or Clumped Distribution

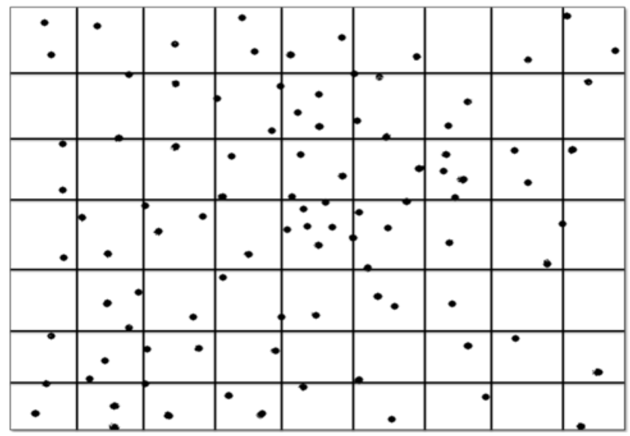
A geographical distribution where individuals occur in clusters

- Too dense to be explained by chance
- Correlated to environmental heterogeneity/ availability of resources
- Most common pattern amongst animals in stochastic environments
- Tends to have a negative binomial distribution
- I.e. Wolves travelling in packs
- Defense Mechanism against predators
- Tend to be modeled/represented with a Negative Binomial Distribution



Quadrat Analysis

- A planar study region is divided into a grid with cells of equal size (quadrats)
- Number of points in each cell, or in randomly selected cells, is recorded
- Quadrats are typically square in shape
- Tries to determine if the distribution of an area is random or nonrandom
- Null hypothesis, H_0, is typically random distribution



Therefore, if observed frequency distribution does not conform to one expected from a random point process:

- Reject null hypothesis (randomness)
- Accept alternative hypothesis
  - Either regular or clustered, depending on direction the observed values differ from those expected

Variance-Mean Ratio (VMR)

- In Poisson Distribution, variance = mean, variance/mean = 1
- Observed point distributions may be measured for their difference from the expected Poisson realizations by testing the significance of the difference between the observed ratio and 1.
- The difference has a standard error of $\sqrt{\frac{2}{N-1}}$, where N is number of observations and N-1 df
- The test statistic $(t^*) = \frac{Observed - Expected}{SE}$, *Expected* value is 1
- VMR > 1 = more clustered than random
- VMR < 1 = more regular than random

For Example, if you had an area with (N=50) animals, and measure the mean number of animals per quadrat (μ) and variance between quadrats for the area, you can determine the approximate distribution of the population.

Uniform/Regular	Random	Clustered
Mean (μ) = 0.500	μ = 0.500	μ = 0.500
Variance (δ^2) = 0.241	δ^2= 0.497	(δ^2) = 27.00
VMR = 0.482	VMR = 0.994	VMR = 54
$t^* = \frac{0.482-1}{0.2041} = -2.538$	$t^* = \frac{0.994-1}{0.2041} = -0.030$	$t^* = \frac{54-1}{0.2041} = 259.65$

Following calculations, use a one-sided T-test with *N*-1 df to determine if significant

Chi Square (Or Goodness of Fit) Test

- Another method for measuring the difference between observed and expected (typically random) distributions.
- Similar to VMR, but uses only expected and observed values

Steps for Chi square test

1. Estimate population mean of the Poisson distribution is: $m = \lambda a$, where λ is the population density, and *a* is the quadrat area, with a sample mean: $\hat{m} = u$
2. Calculate the expected frequencies: $NP(r) = N\exp(-\hat{m})\frac{\hat{m}^r}{r!}$ (*r* = 0,1,2,3…), where *r* is the number of points per quadrat.
3. Test for difference using: $$X^2 = \sum_{i=0}^{N} \frac{[Observed - Expected]^2}{Expected}$$

$$X^2 = \sum_{r=0}^{w} \frac{[f_r - NP(r)]^2}{NP(r)}$$

4. Check chi square table for significance, using *N*-1 df. If significant, reject null hypothesis.

Number of points per quadrat	Observed Frequencies: Uniform/ Regular	Observed Frequencies: Random	Observed Frequencies: Clustered	Expected frequency with Poisson model (λa = 0.500)
0	25	30	44	30.32
1	25	16	2	15.16
2	0	3	0	3.79
3+	0	1	4	0.730
N=	50	50	50	50.00
X^2_{49}	7.3203	0.3145	36.039	
$P_{0.05}$	0.0003	0.76	0.000002	

Nearest Neighbour Analysis

- Measures the difference between an observed spatial point pattern and randomness
- Uses the distribution, in a random point pattern, of the distance between a point and its closest neighbouring point

Steps for Nearest Neighbour

1. Measure the distances between points and nearest neighbours
2. Calculate mean distance, using: $\bar{d} = \frac{1}{N}\sum_{i=1}^{N} d_i$
3. Calculate the mean of the normal distribution, using: $E(\bar{d}) = \frac{1}{2\lambda^{1/2}}$
4. Calculate variance of distribution, using: $\mathrm{var}(\bar{d}) = \frac{4-\pi}{4\lambda\pi N}$
5. Calculate test statistic (ϕ), using: $\phi = \frac{\bar{d} - E(\bar{d})}{\sqrt{\mathrm{var}(\bar{d})}}$

Note: There is another index suggested: $D^* = \frac{\bar{d}}{E(\bar{d})}$

In this index, a perfectly uniform pattern leads to a D^* value of 2.14191, a random pattern leads to a D^* value of 1, and a clustered patter leads to a D^* value of 0

In the following table,

$$E(\bar{d}) = \frac{1}{2\sqrt{(50/10000)}} = 7.071$$

$$\mathrm{var}(\bar{d}) = \frac{4-\pi}{4(50/10000)\pi(50)} = 0.273$$

	$\bar{d}$	D^*	ϕ
Perfectly regular (uniform)	15.00	2.12	$\frac{15.00-7.071}{0.5227}=15.17$
Random	7.59	1.07	$\frac{7.59-7.071}{0.5227}=0.993$
Perfectly clustered	0.00	0.00	$\frac{0-7.071}{0.5227}=-13.53$

References

Arocena, J.M., and Ackerman, J.D. (1998) The use of Statistical Tests to Describe the Basic Distribution Pattern of Iron Oxide Nodules in Soil Thin Sections. Soil Science America Journal. 62:1346–1350.

Byers, J. A. (1984). Nearest Neighbor Analysis and Simulation of Distribution Patterns Indicates an Attack Spacing Mechanism in the Bark Beetle, Ips typographus (Coleoptera: Scolytidae). Environmental Entomology. 13 (5): 1191- 1200.

Picman, J. (1988). Experimental Study of Predation on Eggs of Ground-Nesting birds: Effects of habitat and Nest Distribution. The Condor. 90: 124-131.

Ripley, B. D. (1981). Spatial Statistics. Wiley, 252pp

Rogers, A. (1974). Statistical analysis of spatial dispersion: the quadrat method. London: Pion, pp 1-30

CHAPTER 9: PRIMER ON USE OF INDICES TO DETERMINE SPATIAL PATTERNS

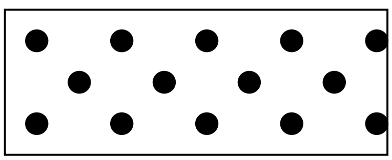
Michael Janssen

Introduction
Spatial indices are used to provide an indication of spatial pattern.
Note: An "index" does not represent a real quantity, and is therefore different from a "measure"

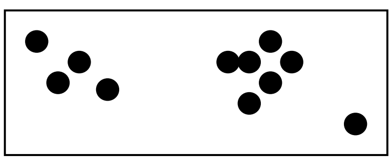
Spatial indices can be used to:

1. Describe a spatial pattern at a given location
2. Indicate relationships within spatial data e.g. autocorrelation
3. Indicate spatial association between individuals or groups

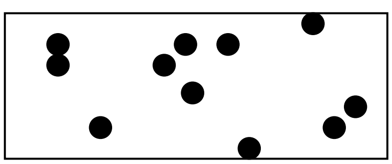
Review – Basic spatial patterns:



Uniform
Synonyms: regular, even, negative contagion, under-dispersed



Aggregated
Synonyms: clumped, patchy, contagious, positive contagion, over-dispersed



Random

1. Statistics describing spatial patterns at a given time and place- *Indices of Dispersion*

- Indices of dispersion are used to measure the distribution of organisms across a landscape.
- The type if index used depends on how the spatial data were collected

Spatial patterns can be inferred from data that:

- was collected using sampling quadrats
- includes the entire population accurately plotted on a map
- includes only a sample of individuals plotted over 2 dimensional space

Indices of dispersion for quadrat counts
Quadrat count data: A series of counts of individual organisms in quadrats of constant size and shape are taken. Quadrats can be randomly sampled from the area of interest, sampled side by side ("contiguous quadrats"), or placed so the entire area is sampled.

Consider the quadrats:

2	6	6
2	6	6
2	4	4

$\bar{x} = 4$
n = 9
$s^2 = 4$
$\sum (x) = 38$

A good index of dispersion:
- moves in a smooth manner from uniform-random-aggregated
- is not affected by n or mean frequencies
- has a known sampling distribution so CI's can applied and significance tested.

Variance-to-mean ratio (VMR):

$VMR = s^2 / \bar{x}$ For details, and hypothesis testing, see Justin's primer

Green's coefficient (C_x):

$$C_x = \frac{\frac{s^2}{\bar{x}} - 1}{\Sigma(x) - 1}$$

*Unfortunately the sampling distribution has not been worked out, so it is difficult to assign confidence intervals

Morisita's Index of dispersion (I_d):

$$I_d = n\left[\frac{\Sigma(x^2) - \Sigma x}{(\Sigma x)^2 - \Sigma x}\right]$$

H_o = Random distribution $\chi^2 = I_d(\sum x - 1) + n - \sum x$ d.f. = $n - 1$

Standardized Morisita Index (I_p):
- standardized so it fits on a scale of -1 to +1

- 1st calculate I_d
- Then calculate 2 critical values from the Morisita Index

Uniform Index = $M_u = \dfrac{\chi^2_{.975} - n + \Sigma x}{(\Sigma x) - 1}$

where $\chi^2_{.975}$ = *Chi Square* from table with d.f. = $n - 1$ that has 97.5% of area to the right

Clumped Index $M_c = \dfrac{\chi^2_{.025} - n + \Sigma x}{(\Sigma x) - 1} =$

where $\chi^2_{.025}$ = *Chi Square* from table with d.f. = $n - 1$ that has 2.5% of area to the right

Standardized Morisita Index (I_p) cont…
Then calculate the standardized Morisita Index using one of the following:

When $I_d \geq M_c > 1.0$, $$I_p = 0.5 + 0.5\left(\frac{I_d - M_c}{n - M_c}\right)$$

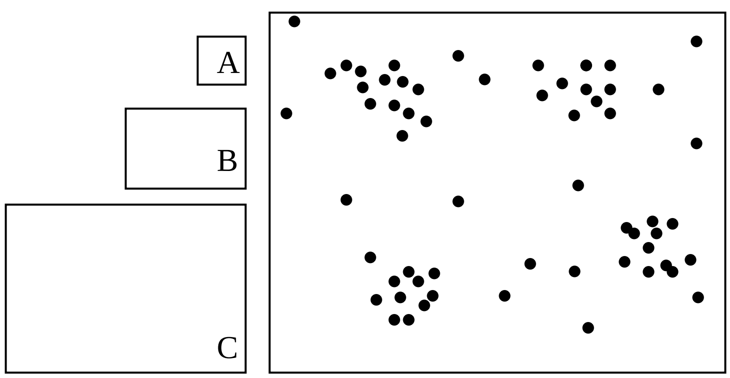
When $M_c > I_d \geq 1.0$, $$I_p = 0.5\left(\frac{I_d - 1}{M_u - 1}\right)$$

When $1.0 > I_d > M_u$, $$I_p = -0.5\left(\frac{I_d - 1}{M_u - 1}\right)$$

When $1.0 > M_u > I_d$, $$I_p = -0.5 + 0.5\left(\frac{I_d - M_u}{M_u}\right)$$

The standardized Morisita Index (I_p) ranges from -1.0 to +1.0 with 95% confidence intervals at +0.5 and -0.5

The Scale Problem:



Adapted from Krebs 1999

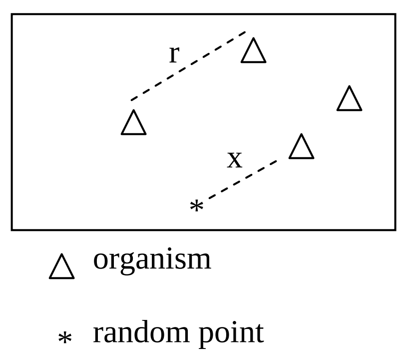
Quadrat size can influence the indices of dispersion

In this hypothetical clumped population with regularly distributed clumps:

quadrat A will suggest a random distribution,
quadrat B will suggest a clumped distribution,
quadrat C will suggest a uniform distribution.

Spatial Pattern from Distance Methods:
Data is often collected such that we have 2 kinds of measurements:

- Distance from random points to the nearest organism (x_i)
- Distance from a random organism to it's nearest neighbor (r_i)



$$h = \frac{\Sigma(x_i^2)}{\Sigma(r_i^2)}$$

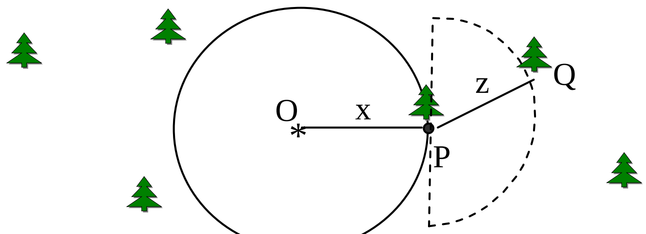
***h* = Hopkin's test for randomness**

under H_o of randomness *h is* F distributed with d.f. = 2*n* in numerator and denominator

An index of of this pattern is $I_H = h / (1+h)$
I_H is near 1 when there is clumping, near 0 when uniform, and near 0.5 when the data are random

Spatial Pattern from Distance Methods cont…

T-square sampling procedure:



1. Random point O is located
2. Distance x is measured from O to nearest organism P
3. A second distance z is measured to it's nearest neighbor, constrained to be in the hemisphere to the right of the dashed line, Q. i.e. the angle must be greater than 90˚.

Hine's statistic = $h_T = \dfrac{2n[2\,\Sigma\,(x_i^2) + \Sigma\,(z_i^2)]}{[(\sqrt{2}\;\Sigma x_i) + \Sigma z_i]^2}$

- H_0 is a random distribution such that $h_T = 1.27$
- H_0 is evaluated by comparing the calculated h_T to critical values from a table for this statistic.
- h_T smaller than 1.27 indicates a uniform pattern, larger than 1.27 indicates a clumped pattern.

Spatial Pattern from mapped data

If we have the entire population of interest mapped, we can use the Clark and Evans test:

$\bar{r}_A$ = mean distance to the nearest neighbor $= \dfrac{\Sigma r_i}{n}$

where r_i = distance to nearest neighbor for individual i
n = number of individuals in study area
ρ = density of organisms = number in area / size of area

Under Ho of a random spatial pattern:

$\bar{r}_E$ = expected distance to nearest neighbor $= \dfrac{1}{2\sqrt{\rho}}$

$$z = \frac{\bar{r}_A - \bar{r}_E}{s_r} \qquad s_r = \frac{0.2616}{\sqrt{n\rho}}$$

Index of Aggregation = $R = \dfrac{\bar{r}_A}{\bar{r}_E}$

If pattern is random, R =1
If uniform, R approaches 2.15
If clumped, R approaches 0

2. Indices of spatial autocorrelation:

Recall:

The correlation coefficient: $$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}}$$

And the temporal autocorrelation coefficient: (with a lag of k) $$r_k = \frac{\sum_{t=1}^{n-k}(y_t - \overline{y})(y_{t+k} - \overline{y})}{\sum_{t=1}^{n}(y_t - \overline{y})^2}$$

With some spatial datasets we might logically expect that observations that are close together may be more similar than observations that are far apart. As an index of **spatial autocorrelation** we can use Moran's I:

$$I = \left(\frac{n}{S_0}\right)\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \overline{x})(x_j - \overline{x})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

Where x_i is the value of the observation at point i, *n* is the total number of observations, w_{ij} is the weight between observation i and j, and S_o is the sum of all w_{ij}'s **:**

$$S_O = \sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}$$

W_{ij} is chosen by the researcher according to the expected nature of the correlation:

- If you predict that values in adjacent quadrats are correlated you can set w_{ij} =1 when i and j share a boundary, and w_{ij} = 0 if they don't

- If you predict that the strength of the correlation depends on the distance between i and j, you can set $w_{ij} = 1/d_{ij}$, where d_{ij} is some measure of distance between i and j

Testing for spatial autocorrelation

Under the **Null Hypothesis** of no correlation, $I_o = -1/(n-1)$.
Since we can calculate variance of I, we can test whether observed I (denoted Î)is significantly different from I_o:

$$Z = \frac{\hat{I} - I_0}{\sqrt{\left(Var\left(\hat{I}\right)\right)}}$$

If $z \geq 1.96$, we reject Ho in favor of positive autocorrelation
If $z \leq 1.96$, we reject Ho in favor of negative spatial autocorrelation

3. Indices of Spatial Association:

If two types of organisms often occur together in the same place, they are spatially associated

Organisms that don't move*: e.g. Plants*

If we have presence/absence data for quadrats, for 2 species, we can create a 2x2 contingency table:

		Species A		
		present	absent	
Species B	present	a	b	
	absent	c	d	
	Total			n

a = # quadrats where both species are present, etc…

H_o is no association between species A and B, so

$a = b = c = d$

we use H_o to calculate expected values of a,b,c,d

test whether the expected values are significantly different from observed using Chi-Square with d.f. = (R-1) (C-1)

If $p < 0.05$ and ad>cb the data suggest positive association

If $p < 0.05$ and ad<cb the data suggest negative association

Organisms that move:

When an organism is capable of moving it's distribution can change rapidly and so we are necessarily concerned with how organisms are distributed in both space and time.

The Half-Weight Index (*HWI*) is commonly used as an index of association:

$$HWI = x/[x + y_{ab} + 0.5(y_a + y_b)]$$

a and b can represent two different individuals, groups, or species
x = number of instances a and b were observed together
y_{ab} = number of instances a and b were observed apart, at the same time (often $y_{ab} = 0$)
y_a = number of instances a was observed and b was not
y_b = number of instances b was observed and a was not

To test H_o of random association:

1st randomly generate many alternative data sets with an equal number of individuals and observations as the original data set
2nd calculate *HWI* for each pair of individuals from the observed data set
3rd calculate *HWI* for each pair of individuals from each of the randomly generated data sets

4th calculate S

$$S = \sum_{i=1}^{D}\sum_{j=1}^{D}\frac{(O_{ij} - E_{ij})^2}{D^2}$$

D = total number of individuals observed
O_{ij} = *HWI* for individuals i and j
E_{ij} = mean *HWI* for individuals i and j from the randomly generated data sets

5th Compare S from the observed data set to the distribution of S values from the random data sets. If S from the observed data set is larger than 95% of the S values from the random data sets, we reject H_o

References/further readings:

Spatial pattern and indices of dispersion:

Krebs, C.J. 1999. Ecological methodology, 2nd Edition. Addison Wesley Longman, New York.

Hurlbert, S.H. 1990. Spatial distribution of the montane unicorn. Oikos **58**: 257 -271.

Spatial autocorrelation:

Cliff, A.D. & J.K. Ord. 1981. Spatial processes: models and applications. Pion, London.

Cliff, A.D. and J.K. Ord. 1973. Spatial autocorrelation. Pion, London

Waldor, T. 1996. The spatial autocorrelation coefficient Moran's *I* under heterscedasticity. Statistics in Medicine **15**: 887-892.

Spatial Association:

Dale, M.R.T. 1999. Spatial pattern analysis in plant ecology. Cambridge University Press, New York.

Ginsberg, J.R. and T.P. Young. 1992. Measuring association between individuals or groups in behavioral studies. Animal Behavior **44**: 377-379.

Bejder, L, Fletcher, D., and S. Brager. 1998. A method for testing association patterns of social animals. Animal Behavior **56**: 719-725.

General:

Gustafson, E.J. 1998. Quantifying landscape spatial pattern: What is the state of the art? Ecosystesm **1**: 143-156.

Perry, J. N., Liebhold, A. M., Rosenberg, M. S., Dungan, J., Miriti, M., Jakomulska, A. and Citron-Pousty, S. 2002. Illustrations and guidelines for selecting statistical methods for quantifying spatial pattern in ecological data. Ecography **25**: 578–600.

Appendix: Summary of dispersion indices.
All references were either cited in the primer, or handed out as assigned readings.

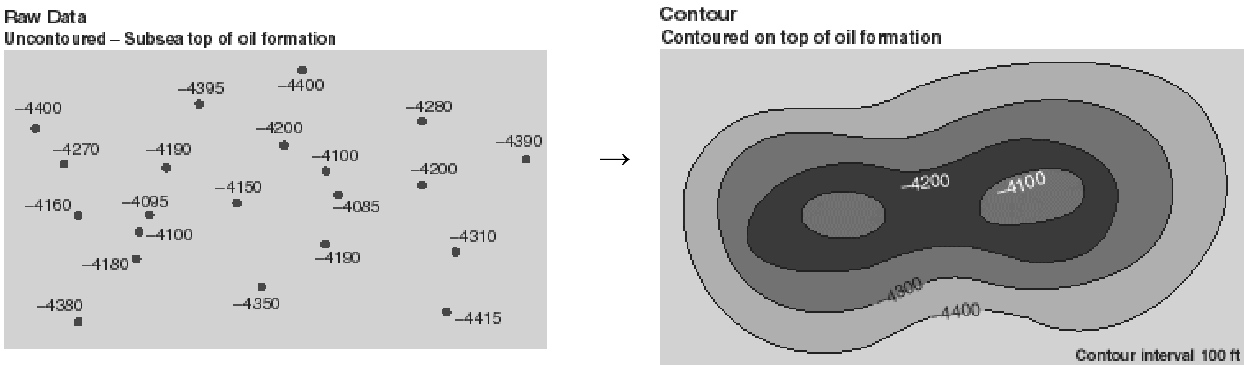
			Values expected if distribution is			
Name	**Symbol**	**Data**	**Uniform**	**Random**	**Aggregated**	**Reference**
Variance to mean ratio	VMR	Quadrat	<1	1	>1	Krebs 1999 (eg Morales et al. 2008, Hurlbert 1990)
Green's coefficient	C_x	Quadrat	<0	0	>0	Krebs 1999 (eg. Hurlbert 1990)
Morisita Index	I_d	Quadrat	<0	0	>0	Krebs 1999 (eg Hurlbert 1990)
Standardized Morisita Index	I_p	Quadrat	-1 - 0	0	0 - 1	Krebs 1999 (eg Hurlbert 1990)
Inverse clustering coefficient	1/k	Quadrat	<0	0	>0	Krebs 1999 (eg Arocena and Ackerman 1998)
Hopkin's Index	I_H	mapped points	0	0.5	1	Krebs 1999
Hine's statistic	h_T	mapped points	< 1.27	1.27	>1.27	Krebs 1999
Clark and Evans	R	mapped points	0	1	2.15	Krebs 1999 (eg Frohlich and Quenau 1995)
Goodness of fit	X^2	Quadrat or mapped points	-	-	-	Krebs 1999 (eg:Arocena and Ackerman 1998)

CHAPTER 10: PRIMER ON PATIAL SMOOTHING

Gale Bravener

"Smoothing is a statistical technique…to capture important patterns in a set of data, while leaving out noise or other fine-scale structures…" (Sheehy 2009)

- Like smoothing of time series data:
  - spatial smoothing is used to remove distracting noise present in spatial data in order to reveal patterns that are not evident visually.
  - spatial smoothing calculates a value at a specific location as a function (e.g. average) of its neighbours



I. WHEN TO USE SMOOTHING?

- Detecting patterns in spatial data, especially point data
- Provides insight but not precise estimates of location, spread or trends.
- Useful where data are known to be of low precision or small sample sizes
- To convert discrete point data to a contour map or continuous density map

II. WHAT MAKES A GOOD SMOOTHER?

Some objectives of smoothing:

1) reduce the variance so that underlying trends can be seen
2) reduce attention to outliers or transients
3) examine patterns in residuals that can be revealed once the smoothed trend has been removed
4) minimize the effect of aggregation in what may be a summary data point across an entire region

(Kafadar 1999)

Some characteristics of a good smoother:

1) displays the true pattern as accurately as possible
2) performance should not be impaired if the data are not evenly spaced
3) its output should be 'smooth' where *F* (the true function of x_i) is, without attempting to smooth over obvious breaks
4) unusual values, unsupported by neighbours, should stand out clearly in the residuals, not in the smooth

(Kafadar 1999)

III. TYPES OF SMOOTHERS

Recall temporal smoothers: *Moving Averages, Weighted Average, Kernel, Spline, etc.*

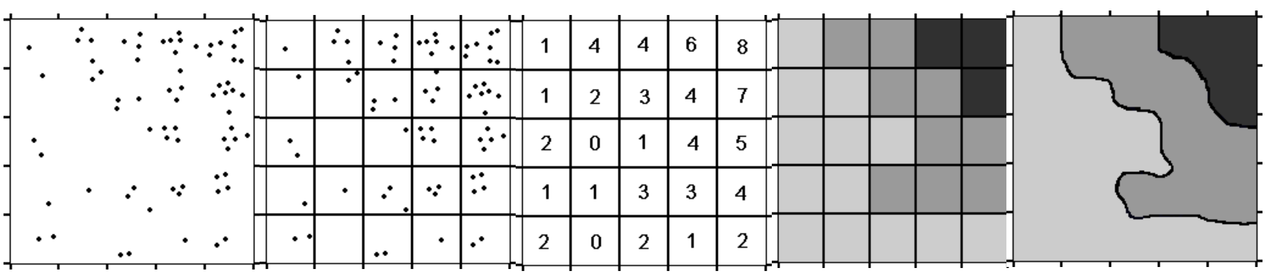
Spatial smoothers:

- Linear**:** ***Trend Surface Analysis, Moving Averages, Kernel, Spline, IDW***
- Non-Linear: ***Median filters, Head-banging***

➢ Linear smoothers expose extremely broad, non-specific trends, and non-linear smoothers identify sharp distinctions between regions

Choropleth maps

- simple smoother – converts point data to choropleth to isopleths (= contour map)
- degree of smoothing depends on the number of classes the values are put into
- smooths by assigning data values to categories (interval to ordinal data).



raw quadrats densities choropleth contours

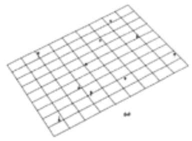
1. Trend surface analysis

- entire surface is approximated by a polynomial equation
- surface is estimated by an ordinary **least squares regression**.
- peculiar in that the two independent variables represent two perpendicular spatial dimensions, and the dependent variable represents a variable (e.g., elevation)

- ***Trend surface analysis (cont'd)***

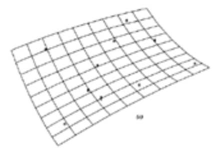
- a linear equation (first order polynomial) describes a tilted plane surface:

  $z = a + bx + cy$



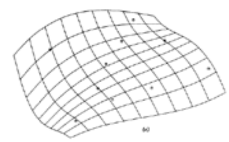
- quadratic equation (second order polynomial) describes a simple hill or valley

  $z = a + bx + cy + dx^2 + exy + fy^2$

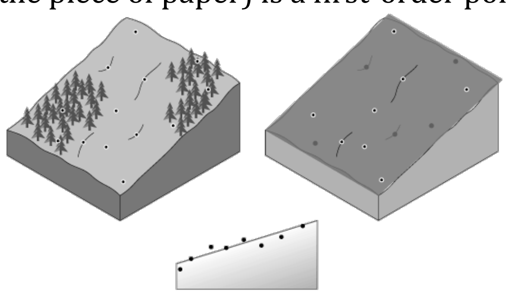


- cubic equation adds more complexity

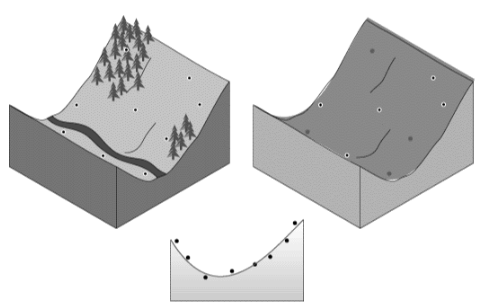
  $z = a + bx + cy + dx^2 + exy + fy^2 + gx^3 + hx^2y + ixy^2 + jy^3$



- By analogy, it's like taking a piece of paper and fitting it to a landscape (e.g. a slope). A flat plane (no bend in the piece of paper) is a first-order polynomial (linear).

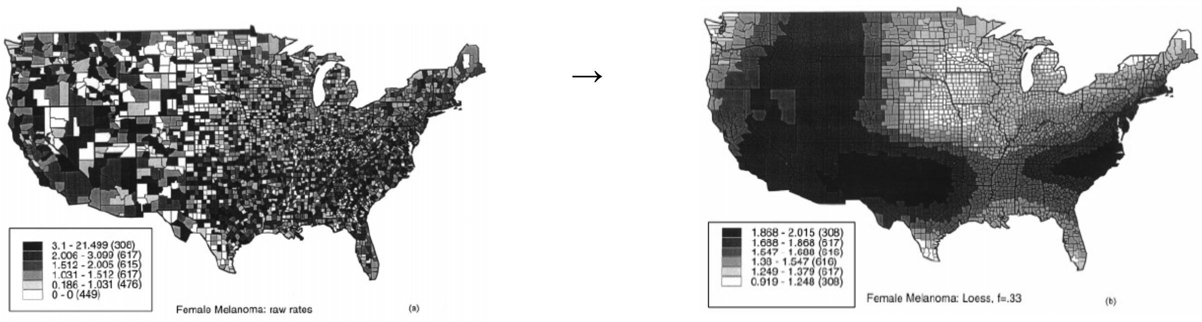


- But a flat piece of paper will not accurately capture a landscape containing a valley unless you bend it. Allowing for one bend is like a second-order polynomial (quadratic), two bends a third-order (cubic), and so forth



LOESS (locally weighted regression) is a type of trend surface analysis:

- for fitting a regression surface to data through multivariate smoothing.
- The dependent variable is smoothed as a function of the independent variables in a moving fashion analagous to moving averages in time series (Cleveland 1988)

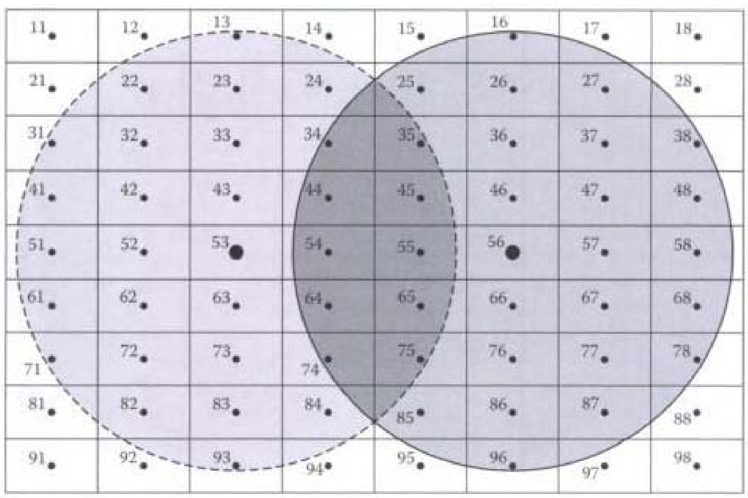


(*Kafadar 1999)*

2. Moving Average Smoothers *(aka Moving Windows)*

- uses circular or square filtering window
- uses the average value of data points within the window to calculate the value
- averaged values have less variability and are thus spatially smoothed.

example:



- circle around quadrat 53 defines the window, average of the 33 quadrats within window gives the smoothed value for 53
- The choice of window size is very important (large windows reveal better regional patterns than local patterns)

A more common approach is to use **weighted** moving averages for smoothing. This allows points nearby to have more influence than points far away. Common weighted methods include:

3. Kernel estimation

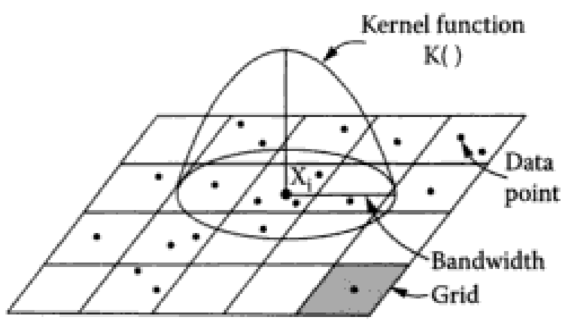
- similar to moving averages in that it uses a moving window, but the kernel method weighs nearby objects more than far ones.

Kernel estimation (cont'd)

- For a 2-D surface, a common kernel is:

$$\hat{f}(x) = \frac{1}{nh^2\pi}\sum_{i=1}^{n}[1 - \frac{(x-x_i)^2 + (y-y_i)^2}{h^2}]^2$$

- Where *h* is bandwidth, *n* is the number of points within the bandwidth, $(x-x_i)^2 + (y-y_i)^2$ measures the deviation in x-y coordinates between points (x_i, y_i) and (x, y).
- A kernel function looks like a bump centered at each point x_i, and tapering off to 0 over a window or "bandwidth"



4. Inverse Distance Weighted (IDW)

- another moving window method
- estimates unknown values as the weighted average of its surrounding points, in which the weight is the inverse of the distance raised to a power
- it is an exact smoother, so the exact known data values are honoured

$$z_u = \frac{\sum_{i=1}^{s} z_i d_{iu}^{-k}}{\sum_{i=1}^{s} d_{iu}^{-k}}$$

where z_u is the unknown value to be estimated at u, z_i is the attribute value at control point i, d_{iu} is the distance between points i and u, s is the number of control points used in estimation, and k is the power. The higher the power, the stronger (faster) the effect of distance decay is (i.e., nearby points are weighted much higher than remote ones). In other words, distance raised to a higher power implies stronger localized effects.

Problems with IDW:

- when points are clustered, or more dense in some areas than others, steps appear
- spikes or pits occur around data points because they are "honoured"

5. Splining

- polynomial regression is "global", while moving averages, kernel, IDW are "local" – splining offer a compromise by using a "piecewise polynomial"
- creates a surface that predicts the values exactly at all control points and has the least change in slope at all points
- cubic splines are commonly used

$$z(x,y) = \sum_{i=1}^{n} A_i d_i^2 \ln d_i + a + bx + cy$$

where x and y are the coordinates of the point to be interpolated, $d_i = \sqrt{(x-x_i)^2 + (y-y_i)^2}$ is the distance from the control point (x_i, y_i), and A_i, a, b, and c are the $n + 3$ parameters to be estimated. These parameters are estimated by solving a system of $n + 3$ linear equations (see Chapter 11), such as

$$\sum_{i=1}^{n} A_i d_i^2 \ln d_i + a + bx_i + cy_i = z_i;$$

$$\sum_{i=1}^{n} A_i = 0 \; ; \; \sum_{i=1}^{n} A_i x_i = 0 \; ; \text{ and } \sum_{i=1}^{n} A_i y_i = 0$$

Note that the first equation above represents n equations for $i = 1, 2, \ldots, n$, and z_i is the known attribute value at point i.

Problems:

- Splines tend to generate steep gradients in areas where data is sparse.
- poor for surfaces which show marked fluctuations - can cause wild oscillations
- are popular in general surface interpolation packages but are not common in GISs

6. Kriging

- a common method for interpolation.
- models spatial variation as three components
  - a spatially correlated component (representing the regionalized variable)
  - a drift or structure (representing the trend)
  - a random error
- The heart of kriging is the ***semivariogram***
- This is the *a priori* information that you must supply in order to interpolate
- The idea is to get an estimate of the distance one would need to travel before data points become uncorrelated.

First, remember the definition of variance:

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{N-1}$$

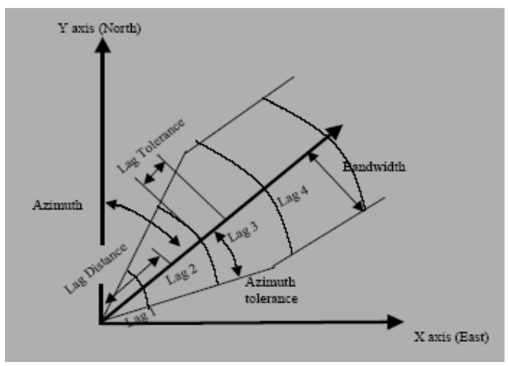
The variance of a data set is a number, but the semivariance is a curve derived from the data according to:

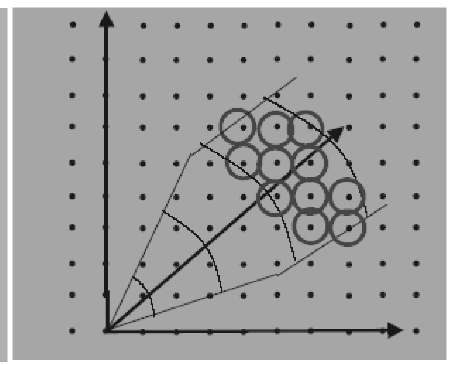
$$\gamma^*(h) = \frac{\sum (y(x) - y(x+h))^2}{2N}$$

where *h* is the lag distance between data points. This is known as the *experimental variogram*, computed from the data.

A model is chosen by matching the shape of the curve of the experimental variogram to the shape of the curve of the mathematical function.

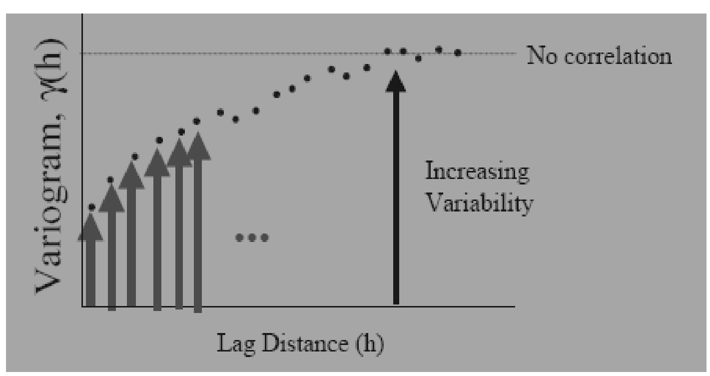
1. Calculate variogram using all points that fall in the lag and angle tolerance



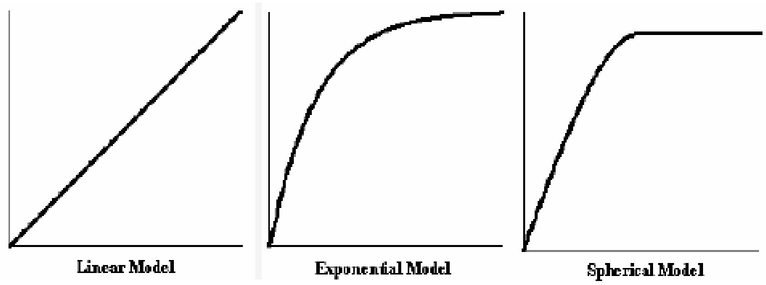


- If data is isotropic, an omni-directional semivariogram is used (angle is ignored)
- If data is anisotropic, the directional semivariogram is indicated (eg. vertical or horizontal)

2. Repeat for all points and all lags, and plot



- **There are several models of semivariance top pick from (linear, exponential, spherical, etc.)…….The trick is to pick the one that best fits your data!**

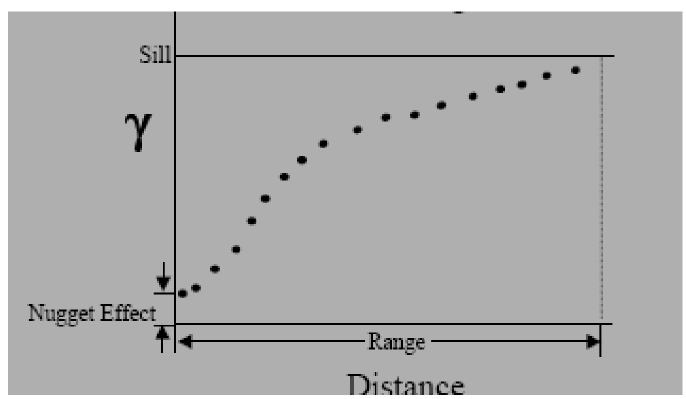


- **The model is then used to determine weights applied to neighbouring points.**

For example, the exponential model:

$$\gamma(h) = c_0 + c\left(1 - e^{-\frac{h}{a}}\right)$$

- *Nugget (c_0):* variation or measurement error – seen as y-intercept
- *Range (a):* controls the degree of correlation between data points, usually represented as a distance
- *Sill (c):* The value of the semivariance model as the lag (h) goes to infinity - it is equal to the total variance of the data set.



The *linear model* is the simplest and one of the most common

In the ***linear model***:
- the semivariance increases as the lag increases
- no indication of a sill or range
- concerned with the slope and intercept

$$\gamma(h) = c_0 + bh$$

and the slope (b) is nothing more than the ratio of the sill (c) to the range (a).

"… linear smoothers tend to be somewhat unsatisfying in situations with abrupt features, extreme values, or outliers. Because non-linear smoothers use medians, abrupt features will be retained far better than with linear smoothers" (Kafadar 1999).

7. Head-banging

- median-based smoother intended for use on non-gridded spatial data
- designed to remove small-scale local variations within a data-set while preserving regional trends
- tends to not over-smooth zones of sharp transition, which can be good or bad
- The procedure is as follows:

1. For each point or area whose value, y_i, is to be smoothed, determine the NN nearest neighbours to $\mathbf{x}_i$.
2. From among these NN neighbours, define a set of pairs around the point/area, such that the 'triple' (pair plus target point at $\mathbf{x}_i$) are roughly collinear. (Formally, the angle formed by the two segments with $\mathbf{x}_i$ in the centre should not exceed, say $\pm 45^\circ$ from 180°. Denote this threshold ϕ, for example, $\phi = 45^\circ$.) Let NTRIP be the maximum number of such triples. If there are more than NTRIP pairs that satisfy the ϕ criterion, choose those whose angles are closest to 180°.
3. Let (a_k, b_k) denote the (higher, lower) of the two y-values in the kth pair, and let A = median $\{a_k\}$, B = median $\{b_k\}$.
4. The smoothed value is $\tilde{y}_i$ = median $\{A, y_i, B\}$ = median{median$\{a_k\}$, y_i, median$\{b_k\}$}.

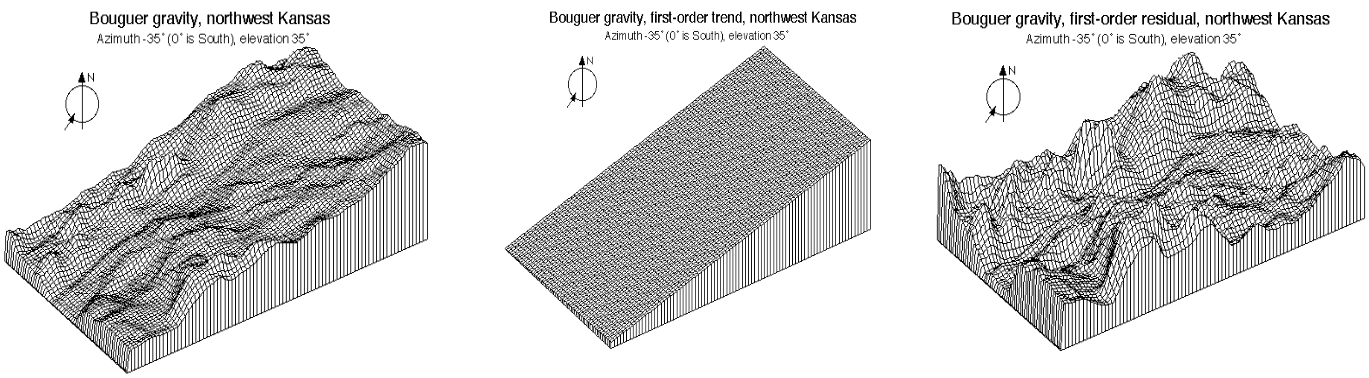
(Kafadar 1999)

Problems:

- Smooths out spikes, which may not be good, depending on the data.
- Tends to produce high rates on the boundaries of study areas.

IV. HOW TO USE SMOOTHERS?

- Different smoothers are better for different purposes and different data
- The best strategy is to use several smoothers and compare the results
- Use an iterative approach (smooth, plot, adjust; re-smooth, re-plot, readjust)
- As with time series data, the residuals (noise) can be evaluated, or removed (detrended) to identify patterns in the data.



Removing the first order trend (a tilted plane) reveals clearer patterns in the residuals

V. INTRODUCTION TO SPATIAL INTERPOLATION

- ***The goal of interpolation is to derive a value at some intermediate location other than where data are taken***

- Many methods used for exploration (smoothing) are also used for interpolating
    - Inverse Distance Weighting (IDW), Splining, Kriging

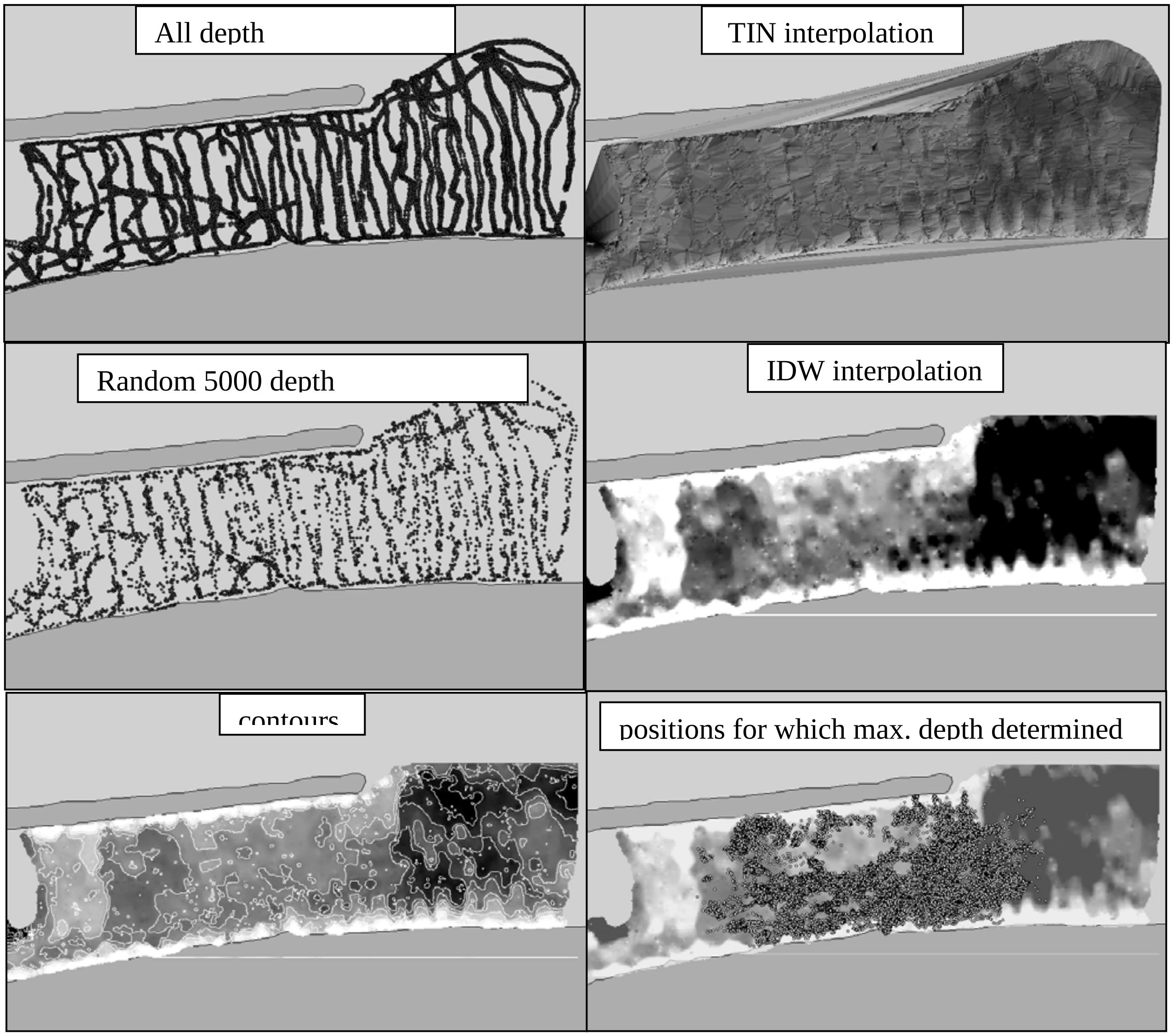
- As with smoothing, different interpolators are better for certain types of data - (See ArcGIS Spatial Analyst information at end of primer for details).

<u>An example from my own data:</u>

The goal: Use measured depth data to interpolate bathymetry, and estimate water depth at many (~75,000) x-y locations where fish were positioned.

1. Mapped x-y depth points in the 100m x 400m study area using ArcGIS
2. Interpolated depths across entire area
    a. TIN (triangular irregular network) attempted first
    b. Splining, Kriging, IDW all attempted
3. Realized transects and clumped data having too much effect
4. Took random sample of 5000 points, and re-evaluated interpolation methods
5. Decided on IDW as best method – fewest spikes in data, least influence of transects on data
6. used "extract values to points" tool in ESRI Spatial Analyst to obtain depths

See next page for graphics…



REFERENCES

Borradaille, G.J. 2004. Statistics of earth science data: their distribution in time, space, and orientation. Springer. 351 pages.

Cleveland, W.S. and Devlin, S.J. 1988. Weighted Regression: An approach to Regression Analysis by Local Fitting. Journal of the American Statistical Association. 83(403), 596-610.

Haining, Robert, P. 2003. Spatial Data Analysis: Theory and Practice. Cambridge Univ.. 432 pp.

Hastie, T. and Tibshirani, R. 1990. Generalized Additive Models. CRC Press. 335 pages.

Kafadar, K. 1999. Simultaneous smoothing and adjusting mortality rates in u.s. counties: melanoma in white females and white males. *Statist. Med*. 18, 3167-3188

Lloyd, Christopher D. 2006. Local Models for Spatial Analysis. CRC Press. 244 pp.

Ripley, Brian D. 1981. Spatial Statistics. John Wiley & Sons. Inc. 252 pp.

Sheehy, Justin. 2009. Temporal Smoothing Primer: IBIO*6000.

Wang, F. 2006. Quantitative Methods and Applications in GIS. CRC Press. 265 pp.

CHAPTER 11: PRIMER ON SPATIAL AUTOCORRELATION

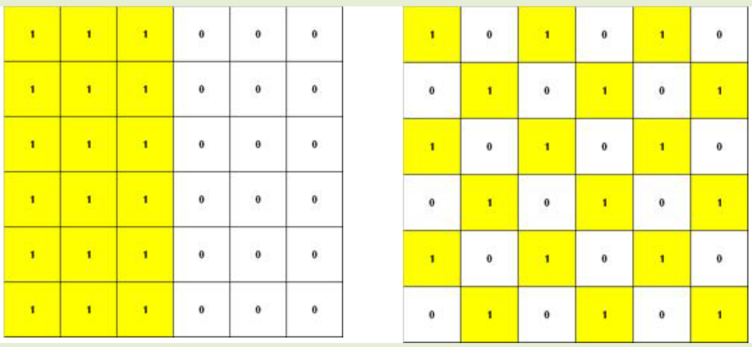
Mark D'Aguiar

If there is systematic spatial variation in the variate, then the phenomenon being studied is said to exhibit spatial auto correlation.

Fig. 1 Spatial Autocorrelation – correlation of a variable with itself through space.

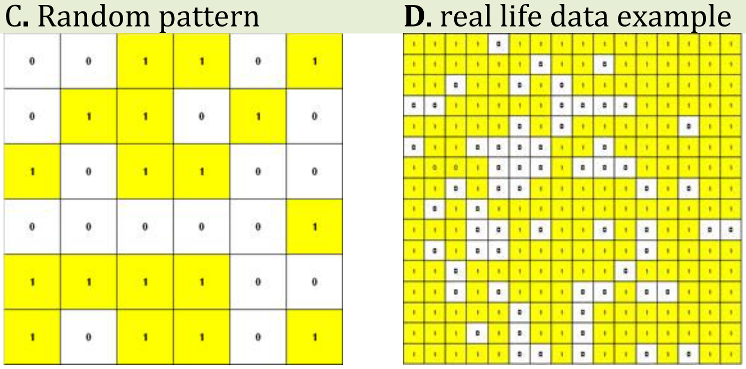
A. Completely separated pattern (+ve)

B. Evenly spaced pattern (-ve)



C. Random pattern

D. real life data example



- If there is any systematic pattern in the spatial distribution of a variable, it is said to be spatially autocorrelated
- If nearby or neighboring areas are more alike, this is *positive spatial autocorrelation*
- *Negative autocorrelation* describes patterns in which neighboring areas are unlike
- Random patterns exhibit *no spatial autocorrelation*

Why spatial autocorrelation is important

Recall: assumption that observations are independent

- Positive spatial autocorrelation may violate this.
- Measures the extent to which the occurrence of an event in an area, makes the occurrence on an event in a neighboring area / unit more probable
- Goals of spatial autocorrelation:
    - Measure the strength of spatial autocorrelation in a map
    - test the assumption of independence or randomness

SPATIAL AUTOCORRELATION

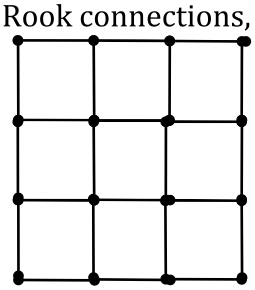
- Non-spatial independence suggests many statistical tools and inferences are inappropriate.
  - Correlation coefficients or ordinary least squares regressions (OLS) to predict a consequence assumes that the observations have been selected randomly.
  - If the observations, however, are spatially clustered in some way, the estimates obtained from the correlation coefficient or OLS estimator will be biased and overly precise.
  - *biased* : areas with higher concentration of events will have a greater impact on the model estimate and they will overestimate precision .

General Considerations:

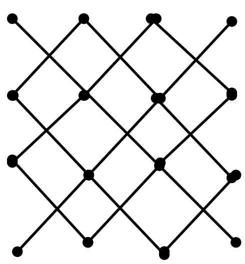
- most based on a sample of localities from an area, or points on a plane.
- Points can be regularly distributed (i.e. grid) or lattice like typical ecological sampling.
- Variables can be nominal (categorical e.g. colour morphs, BW, genotypes), Ordinaal (ranked e.g. n localities ranked in order of population density for a species occurring there), or interval (continuous e.g. gene frequencies, morphological measurements etc.)
- The single value of a variable at each point may be based of single observation (e.g. species i.d. of an individual found at a point in an area), or based on a sample of individuals taken from a locality
- Not all pairs of points will be correlated. **Investigator chooses** the criteria for nearest neighbor connections.

Simple adjacency / join structure / connectedness.

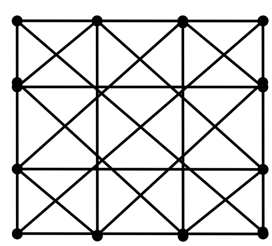
Rook connections,



Bishop connections,

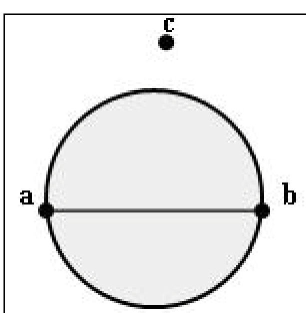
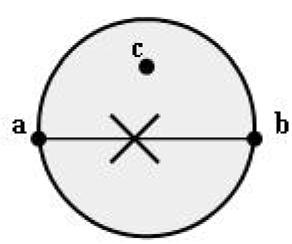


Queen Connections



More complex or irregularly distributed points:

- Gabriel-Connection graph- Any two localities A and B are considered connected if no other locality lies on or within the circle who's diameter is in the line AB.

	**Points a and b are Gabriel neighbours, as c is outside their diameter circle**		**The presence of point c within the circle prevents points a and b from being Gabriel neighbors.**

Indices of Spatial Autocorrelation
Join Count Analysis (Nominal Data)

- Join = edge connecting two point or localities
- points that are like or unlike with respect to the nominal variable (i.e. categorical variable such as spatial distribution of colour morphs.)
- Used to determine whether the classes of points in a regular grid or other spatial structure are random or patchy in their distribution
- Comparing the **observed** number of time that members of the same class are found at **adjacent** grid points with the number expected if the classes are randomly arranged.
- **Free sampling** vs. **non free sampling** (equations in APPENDIX A)

Example: refer to figure 1 A – D.

Join count computation

2	3	3	3	3	2	16
3	4	4	4	4	3	22
3	4	4	4	4	3	22
3	4	4	4	4	3	22
3	4	4	4	4	3	22
2	3	3	3	3	2	16
16	22	22	22	22	16	120

1. Select connection (i.e rook, queen)
2. Specify H_0 (probability that a cell is a particular colour, independent of all colors
3. Apply equations to determine observed joins [(BB,BW,WW) APPENDIX A]
4. Apply equations to determine Expected Values μ_1' , and expected variance .
5. Make decision

****more complex joins with weights use matrix and equations (See Cliff and Ord 1973)**

A. Positive autocorrelation

	BW Joins	BB Joins	WW Joins
# of Joins	6	27	27
z-Rand. statistic	-6.58	5.80	5.80
Variance Rand.	14.26	4.59	4.59
Expected # of Joins	30.86	14.57	14.57
Number of Observations	# of B's	# of W's	Total Joins
36	18	18	60

B. Negative autocorrelation

	BW Joins	BB Joins	WW Joins
# of Joins	60	0	0
z-Rand. statistic	7.72	-6.80	-6.80
Variance Rand.	14.26	4.59	4.59
Expected # of Joins	30.86	14.57	14.57
Number of Observations	# of B's	# of W's	Total Joins
36	18	18	60

C. Random model — no discernable autocorrelation

	BW Joins	BB Joins	WW Joins
# of Joins	35	13	12
z-Rand. statistic	1.10	-0.73	-1.20
Variance Rand.	14.26	4.59	4.59
Expected # of Joins	30.86	14.57	14.57
Number of Observations	# of B's	# of W's	Total Joins
36	18	18	60

D. Atriplex hymeneltrya — positive BB autocorrelation

	BW Joins	BB Joins	WW Joins
# of Joins	173	39	268
z-Rand. statistic	-1.14	2.00	0.24
Variance Rand.	70.67	17.69	22.97
Expected # of Joins	182.57	30.59	266.84
Number of Observations	# of B's	# of W's	Total Joins
256	65	191	480

Moran's I (Interval Ordinal)

- One of the oldest indicators of spatial autocorrelation (Moran, 1950). Still a defacto standard for determining spatial autocorrelation
- Applied to zones or points with continuous variables associated with them.
- achieved by division of the **spatial covariation** by the **total variation**.

$$I = \frac{N\sum_i \sum_j W_{i,j}(X_i - \bar{X})(X_j - \bar{X})}{(\sum_i \sum_j W_{i,j})\sum_i (X_i - \bar{X})^2}$$

N is the number of cases
X_i is the variable value at a particular location
X_j is the variable value at another location
X is the mean of the variable
W_{ij} is a weight applied to the comparison between location *i* and location *j*
*denotes the effect of i on j by the weight of w_{ij}

W_{ij} is chosen by the researcher according to the expected nature of the correlation:

- If you predict that values in adjacent quadrats are correlated you can set

w_{ij} =1 when i and j share a boundary, and w_{ij} = 0 if they don't

-i.e. If zone *j* is adjacent to zone *i*, the interaction receives a weight of 1

- Another option is to make W_{ij} a distance-based weight which is the inverse distance between locations I and j ($1/d_{ij}$), thus $w_{ij} = 1 / d_{ij;}$ where d_{ij} is some measure of distance between i and j
-
- Compares the sum of the cross-products of values at different locations, two at a time weighted by the inverse of the distance between the locations
- Similar to correlation coefficient, it varies between –1.0 and + 1.0

When autocorrelation is high, the coefficient is high
A high I value indicates positive autocorrelation

Testing for spatial autocorrelation / significance

- Empirical distribution can be compared to the theoretical distribution by dividing by an estimate of the theoretical standard deviation

$$Z(I) = \frac{I - E(I)}{S_{E(I)}}$$

where:

$$S_{E(I)} = SQRT[\frac{N^2\sum_{ij} w_{ij}^2 + 3(\sum_{ij} w_{ij})^2 - N\sum_i(\sum_j w_{ij})^2}{(N^2-1)(\sum_{ij} w_{ij})^2}]$$

Under the **Null Hypothesis** of no correlation, *E(I)* = -1/(*n*-1).

Since we can calculate variance of I, we can test whether observed I (denoted I)is significantly different from E(I):

If z ≥ 1.96, we reject Ho in favor of positive autocorrelation *If z ≤ 1.96, we reject Ho in favor of negative spatial autocorrelation*

Geary's C (Interval/ Ordinal)

$$C = \frac{[(N-1)[\sum_i \sum_j W_{ij}(X_i - X_j)^2]}{2(\sum_i \sum_j W_{ij}(X_i - \overline{X})^2}$$

- Similar to Moran's I (Geary, 1954)
- Interaction is not the cross-product of the deviations from the mean, but the **deviations in intensities** of each observation location with one another
- Value typically range between 0 and 2
- If value of any one zone are spatially unrelated to any other zone, the expected value of C will be 1
- Values less than 1 (between 1 and 2) indicate **negative** spatial autocorrelation
- Inversely related to Moran's I
- Does not provide identical inference because it emphasizes the differences in values between pairs of observations, rather than the covariation between the pairs.
- Moran's I gives a more global indicator, whereas the Geary coefficient is **more sensitive to differences in small neighborhoods.**

Testing the Significance

$$Z(C) = \frac{C - E(C)}{S_{E(C)}}$$

SUMARRY

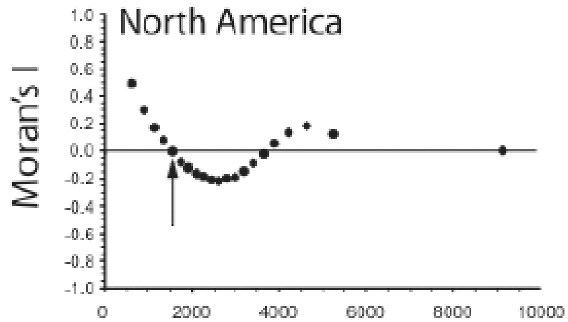
HOW TO COMPUTE SPATIAL AUTO CORRELATION (MORANS I and GEARY's C)

1. Specification of Ho: probablility that a quadrat received a particular x_i, is the same for each quadrat, and the level of X_i observed is fixed independently
2. Significance level: Examine H_o at 0.05
3. Sampling distribution assume normality of I and C
4. Determine region of rejection, indicated by H_1

5. The measurement scale dictates the type of measure,
6. assign weights to the cases
7. create a matrix representing the relationships between variables into software which will compute measure of the spatial autocorrelation between the input data matrix (**APPENDIX A).**
8. **When building a contiguity matrix be cognizant of the relationships between cases, are neighbors determined on a eight directional Queens case or non-diagonal, four directional Rooks case.**

OTHER TESTS (point wise distance data)

- Creating a correllogram using the coefficients vs. distance.



Mantel Test (Mantel z)

- Moran's I is a parametric test while Mantel's test is **semi-parametric**
- Both test against **the null that there is no spatial autocorrelation.**
- Moran's I does this with a correlation that is weighted by inverse distances;
- Mantel test examines the correlation between two distance matrices and generating a null distribution for this correlation by randomly permuting one of the matrices:
    - o distance matrix': consists of the distance between all pairs of sites
    - o Correlation matrix: consists of the **similarilty** between the values across all pairs of sites
- Visualize using correlation coefficients.
- Mantel process tests patterns using a randomization test in which one of the matricies is shuffled, and the resulting resulting coefficient compared with the observed (unshuffled) regression. The end product = Mantel z value which indicates whether or not autocorrelation changes with distance. (Koenig 1998)
- Plot mantel correlogram –this tests for autocorrelation relative to the overall data set.

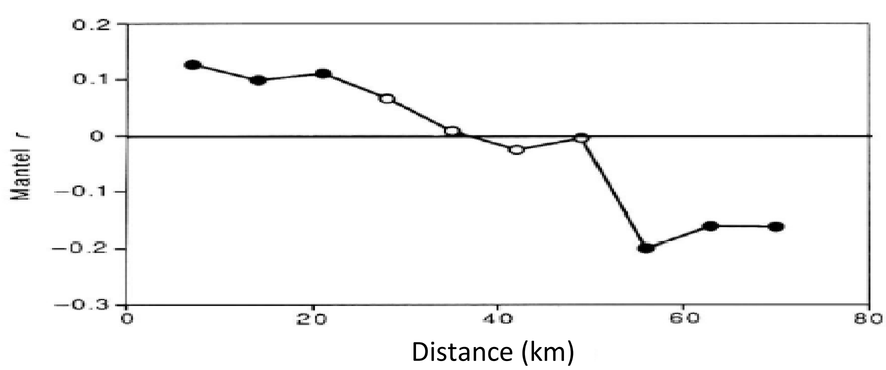


Fig. 1. Mantel correlogram for Bray-Curtis dissimilarities and geographic distances using 7 km distances classes. The solid circles are significantly different from zero at the 95% confidence level, using an error rate of α/n where n is the number of distance classes tested (= 10).

The analysis of regression residuals

- On many occasions regression analysis is carried out to look for autocorrelation in the residuals from a regression. Detection of autocorrelation among residuals can imply:
    1) The presence of nonlinear relationship b/w dependant and non dependant variables;
    2) The omission of one or more regressor variables
    3) That the regression should have an autoregressive structure.

If '1' is important, different models can be specified and interaction terms among independant variable included.

If '2' is the main cause of the auto correlation, additional variables may be suggested by plotting residuals on a map and looking for regular patterns.

If '3' is thought to be the main cause, some kind of transformation needs to be carried out (refer to chapter 5 Cliff and Ord, 1973)

Uses of assessment of spatial autocorrelation:

- identification of patterns which may reveal an underlying process,
- describe a spatial pattern and use as evidence, such as a diagnostic tool for the nature of residuals in a regression analysis,
- as an inferential statistic to buttress assumptions about the data,
- data interpolation technique.

How to Correct for Spatial Autocorrelation in Regression:

- indicates incomplete model, there may be a missing variable. Therefore add an additional variable which may change data pattern.
- incorrect model specification. The data may not be appropriate for a linear fit, or a non-spatial effect may be manifest in the residuals, nuisance spatial autocorrelation. Substantive spatial autocorrelation occurs when there is missing values.
- dominant or extreme cases, outliers, which should have been found at data screening stage.
- systematic measurement error in response variable (non-random). A case in which error increases as values increase, or *vice versa*.
- regression model is inappropriate, reflects the need for an explicitly spatial model. A spatially autoregressive model which incorporates a spatial lag operator into the regression computation. The approach for the implementation of spatial autoregressive models is as follows:
    - establish nature of spatial dependency,
    - use information to choose appropriate model form,
    - fit model using maximum likelihood operators,
    - calculate residuals from model,
    - test residuals, and
    - adjust model based upon residuals. (after Haining, 1990)

References

Cliff, A.D. & J.K. Ord. 1981. Spatial processes: models and applications. Pion, London.

Cliff, A.D. and J.K. Ord. 1973. Spatial autocorrelation. Pion, London

Koenig, W. 1998. Spatial autocorrelation of ecological phenomena. Tree. 14: 22-26

Longly, P., Smith, M. and M. Goodchild. 2008. Geospatial analysis: a comprehensive guide to principles, techniques and software tools. www.spatialanalysisonline.com.

Sokal R.R. 1978. Spatial autocorrelation in biology. Biol. J. Linn. Soc. 10: 199-228

Waldor, T. 1996. The spatial autocorrelation coefficient Moran's *I* under heterscedasticity. Statistics in Medicine **15**: 887-892.

APPENDIX A

Computational Formulas, expected values, and variances of autocorrelation statistics

Nominal Data (for K ≥ 2 classes or types)

Join counts

Total joins **A = ½** $\sum_{i=1}^{n} li$

For localities of the same type:

Number of rr joins = $\frac{1}{2}\sum_{ij} \mathbf{w_{ij}}\,(\mathbf{rr})_{ij}$

Expected values or rr joins ($\mu_1'(BB) = \frac{\mathbf{Wn_r^2}}{\mathbf{2n^2}}$,

Expected variance:

$\mu_2 = [1/4]^{}[(S_1 n_2 2)/n^2 + ((S_2 - 2S_1)\, n_r^3)/n^3 + (([W^2]_2 + S_1 - S_2)\, n_r^4)/n^4 - W^2\, [(n_r^2)/n^2]]^2$

For localities of different types

Number of rs joins = $\frac{1}{2}\sum_{ij} \mathbf{w_{ij}}\,(\mathbf{rs})_{ij}$

Expected Values of rs joins: $\mu_1'(BB) = \frac{W n_r n_s}{n^2}$,

Morans statistic

Expected values: μ_1' = -(n-1)$^{-1}$

Expected variance:

$$\mu_2 = \frac{\mathbf{n}\ [(\mathbf{n}^2 - \mathbf{3}_1 + \mathbf{3})S_1 - nS_2 + \mathbf{3}W^2] - b_2\ \left[\left(n^2 - n\right)S_1 - 2nS_2 + 6W^2\right]}{(n-\mathbf{1})^3 W^2} \frac{(\mathbf{S}_2 - 2S_1)n_r^3}{n^3}$$

Adjacency and Weighted Matrices

A. Source data

	+4.55	+5.54
+2.24	-5.15	+9.02
+3.10	-4.39	-2.09
	+0.46	-3.06

B. Adjacency matrix, **W**

	1	2	3	4	5	6	7	8	9	10
1	0	1	0	1	0	0	0	0	0	0
2	1	0	0	0	1	0	0	0	0	0
3	0	0	0	1	0	0	1	0	0	0
4	1	0	1	0	1	0	0	1	0	0
5	0	1	0	1	0	1	0	0	1	0
6	0	0	0	0	1	0	0	0	0	1
7	0	0	1	0	0	0	0	1	0	0
8	0	0	0	1	0	0	1	0	1	0
9	0	0	0	0	1	0	0	1	0	1
10	0	0	0	0	0	1	0	0	1	0

A. Computation of variance/covariance-like quantities, matrix **C**

Var/Covar matrix			1	2	3	4	5	6	7	8	9	10	Diagonal values:
	z(i)	z(i)-mean	1.22	2.08	3.53	-6.17	-5.41	-0.56	4.52	8.00	-3.11	-4.08	(z(i)-mean)^2
1	2.24	1.22	1.48	2.53	4.30	-7.52	-6.59	-0.68	5.50	9.74	-3.79	-4.97	1.48
2	3.10	2.08	2.53	4.32	7.33	-12.83	-11.25	-1.17	9.39	16.62	-6.47	-8.48	4.32
3	4.55	3.53	4.30	7.33	12.45	-21.77	-19.09	-1.98	15.94	28.22	-10.98	-14.40	12.45
4	-5.15	-6.17	-7.52	-12.83	-21.77	38.09	33.40	3.47	-27.89	-49.36	19.21	25.19	38.09
5	-4.39	-5.41	-6.59	-11.25	-19.09	33.40	29.29	3.04	-24.45	-43.29	16.84	22.09	29.29
6	0.46	-0.56	-0.68	-1.17	-1.98	3.47	3.04	0.32	-2.54	-4.49	1.75	2.29	0.32
7	5.54	4.52	5.50	9.39	15.94	-27.89	-24.45	-2.54	20.41	36.13	-14.06	-18.44	20.41
8	9.02	8.00	9.74	16.62	28.22	-49.36	-43.29	-4.49	36.13	63.97	-24.89	-32.65	63.97
9	-2.09	-3.11	-3.79	-6.47	-10.98	19.21	16.84	1.75	-14.06	-24.89	9.68	12.70	9.68
10	-3.06	-4.08	-4.97	-8.48	-14.40	25.19	22.09	2.29	-18.44	-32.65	12.70	16.66	16.66
Mean	1.02											SSD	196.68
SD	4.67												4.67

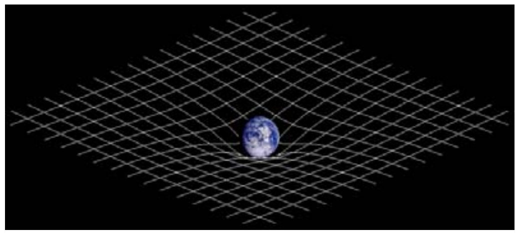
B. **C*W**: Adjustment by multiplication of the weighting matrix, **W**

1	0.00	2.53	0.00	-7.52	0.00	0.00	0.00	0.00	0.00	0.00	-4.99
2	2.53	0.00	0.00	0.00	-11.25	0.00	0.00	0.00	0.00	0.00	-8.72
3	0.00	0.00	0.00	-21.77	0.00	0.00	15.94	0.00	0.00	0.00	-5.84
4	-7.52	0.00	-21.77	0.00	33.40	0.00	0.00	-49.36	0.00	0.00	-45.25
5	0.00	-11.25	0.00	33.40	0.00	3.04	0.00	0.00	16.84	0.00	42.04
6	0.00	0.00	0.00	0.00	3.04	0.00	0.00	0.00	0.00	2.29	5.34
7	0.00	0.00	15.94	0.00	0.00	0.00	0.00	36.13	0.00	0.00	52.07
8	0.00	0.00	0.00	-49.36	0.00	0.00	36.13	0.00	-24.89	0.00	-38.12
9	0.00	0.00	0.00	0.00	16.84	0.00	0.00	-24.89	0.00	12.70	4.66
10	0.00	0.00	0.00	0.00	0.00	2.29	0.00	0.00	12.70	0.00	15.00
										Covar SS	16.19

CHAPTER 12 PRIMER ON SPATIAL-TEMPORAL ANALYSIS

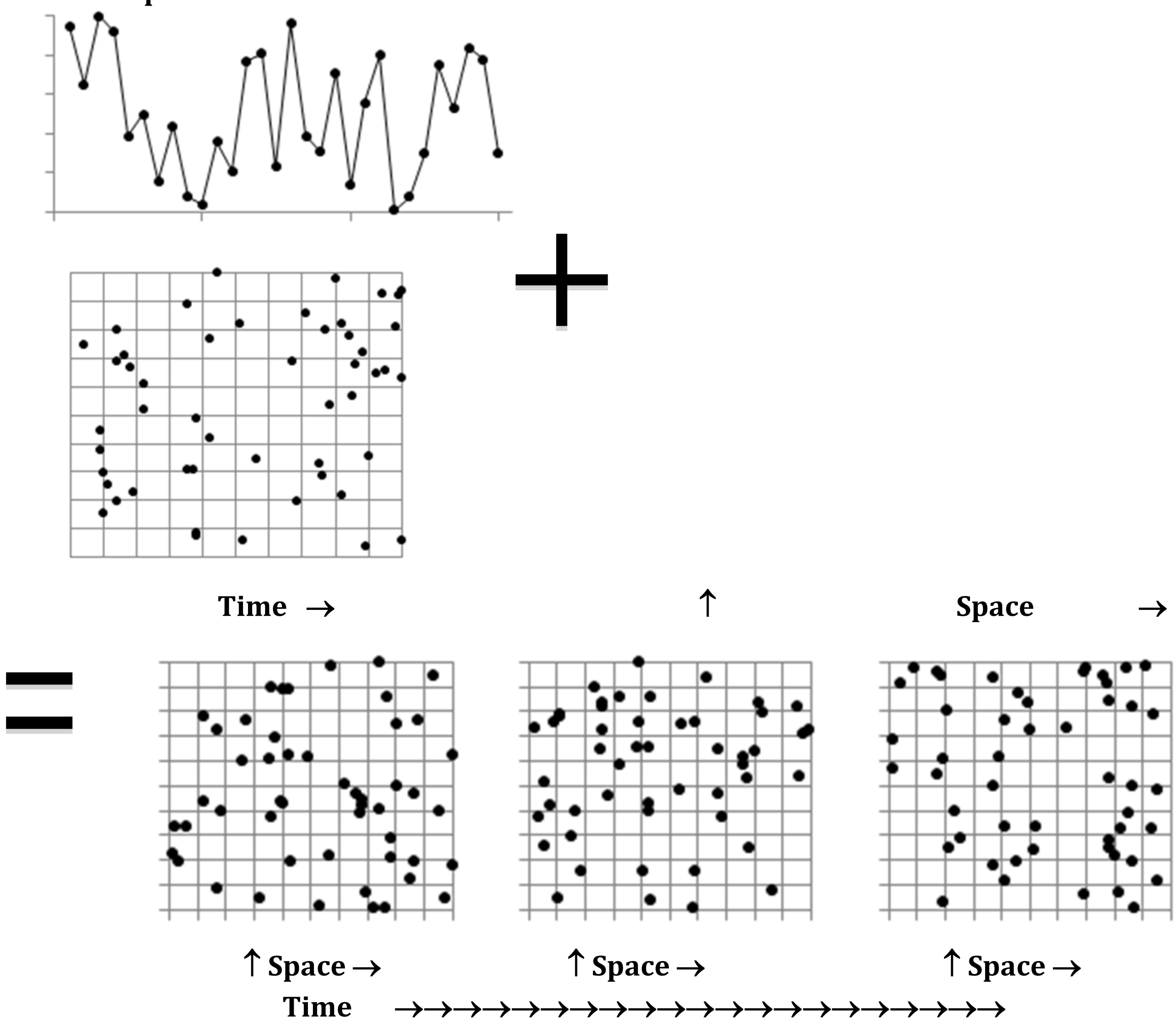
Timothy J. Bartley

Spatial-Temporal Analysis



from <http://en.wikipedia.org/wiki/Spacetime>

I. Space and Time



II. Autologistic Regression in Space and Time

a. Linear regression
- ⇒ used to examine the relationship between one or more independent variables (x_j) and a continuous dependent variable (Y_i)
- ⇒ least squares function

$$Y_i = \beta_0 + \beta_1 x_{1,i} + \beta_1 x_{2,i} + \ldots + \beta_j x_{j,i} + \varepsilon_i \qquad \text{where} \qquad i = 1, \ldots, n$$

b. Logistic regression
- ⇒ used to predict the probability of occurrence of an event
- ⇒ fit data to a logistic curve using a logit model
- ⇒ the dependent variable (Y) is binary (0,1 or present/absent)

$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \beta_1 x_2 + \ldots + \beta_j x_j + \varepsilon \quad \text{OR} \quad \pi_i = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_j x_j}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_j x_j}}$$

π = probability that Y = 1 for a given x_j
β_0 = intercept
β_j = coefficent that relates observed occurrence Y to covariates
x_j = covariates (explanatory variable)
ε = binomially distributed error term

c. Maximum likelihood
- ⇒ used to evaluate logistic regression
- ⇒ least squares cannot be used for model with binary dependent variable
- ⇒ fits values for the parameters which maximize the probability of obtaining the observed data
- ⇒ the set of values which give the greatest likelihood are used as estimators of the parameters of interest

$$\lambda = \ln L = \prod_{n=1}^{N} \ln f(g_n; \beta_0, \beta_1, \ldots, \beta_j)$$

L = the likelihood function
g_n = each of N total observations
β_j = parameters which are to be estimated

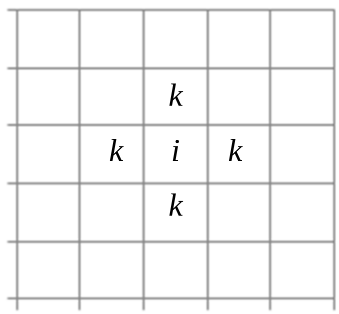
- ⇒ significance can be tested by taking the ratio of each parameter estimate to its standard error (the z-ratio), with values of magnitude 1.96 or greater significant at the α = 0.05 level
- ⇒ significance can also be tested using Monte Carlo maximum likelihood simulations

⇒ best model fit can be determined in several ways including though the use of information criteria, such as Akaike's information criterion (AIC), deviance information criterion (DIC), or Bayesian information criterion (BIC)

d. Autologistic regression

⇒ logistic regression model with extra explanatory variables representing spatial effects

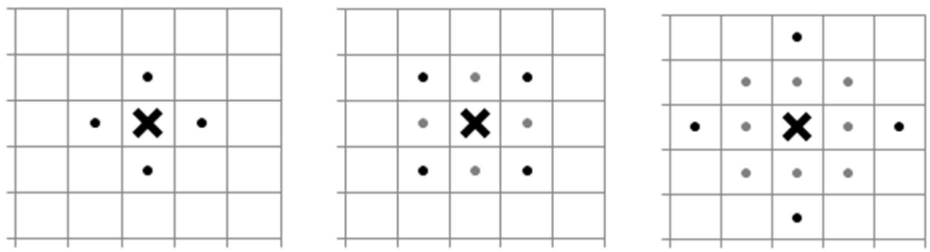
⇒ incorporates spatial dependence to account for potential autocorrelation



$$\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \sum_j \beta_j x_{j,i} + \sum_k \alpha_k Y_{k,i} + \varepsilon_i$$

π = probability that $Y = 1$ at position i for a given $x_{j,i}$
Y =binary (0,1) occurence
β_0= intercept
β_j = coefficent that relates observed occurrence Y to covariates
x_j = covariate (explanatory variable)
α_k= coefficient that relates observed occurrence Y at position k to predicted probability of occurrence π at position i
ε = binomially distributed error term

⇒ Neighbour order (l)



First Order (l=1) Second Order (l=2) Third Order (l=3)

⇒ Spatial terms can be averaged into a single spatial autocovariate (assuming isotropy)

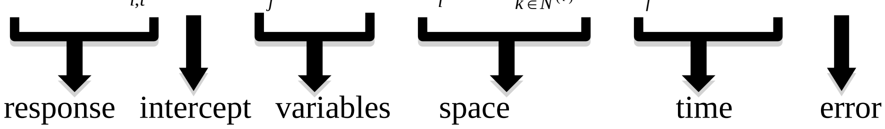
$$\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \sum_j \beta_j x_{j,i} + s(Y_i) + \varepsilon_i \qquad \text{where}$$

$$s(Y_i) = \frac{\sum_{k \neq i} w_k Y_{k,i}}{\sum w_k} \qquad \text{and} \qquad w_k \propto \frac{1}{\text{distance}_{k,i}}$$

e. Spatial-temporal autologistic regression model (STARM)

⇒ logistic regression model with extra explanatory variables representing both spatial and temporal effects

$$\ln\left(\frac{\pi_{i,t}}{1-\pi_{i,t}}\right) = \beta_0 + \sum_j^J \beta_j x_{j,i,t} + \frac{1}{2}\sum_l^L \alpha_{J+l} \sum_{k \in N^{(l)}}^K Y_{k,t} + \sum_f^F \delta_{J+K+s} Y_{i,t-f} + \varepsilon_{i,t}$$



π = probability that $Y = 1$ at position i for a given $x_{j,i}$

β_0 = intercept

β_j = coefficent that relates observed occurrence Y to covariates

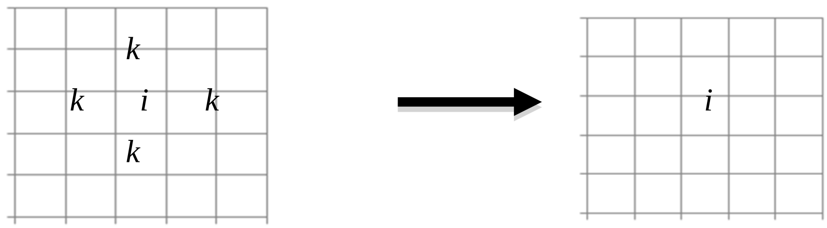
x_j = covariate (explanatory variable)

α = coefficient that relates observed occurrence Y at position k from neighbourhood l to predicted probability of occurrence π at position i

ε = binomially distributed error term

f = each of F total time lags

δ = coefficient that relates observed occurrence Y at position k and time lag f to predicted probability of occurrence π at position i



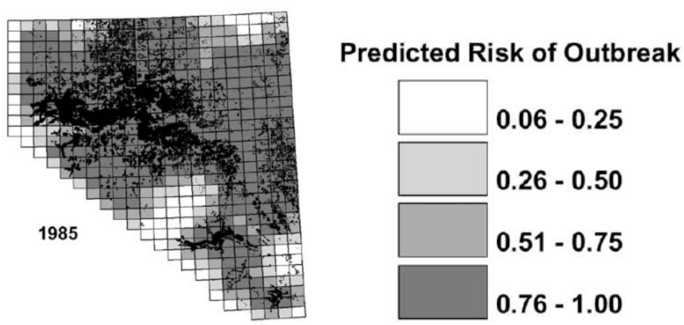
Y for each k at time t - f $\qquad$ π that Y = 1 at i at time t

⇒ achieved using 'Markov random fields'
⇒ define the probability of an event based on a number of variables

$$p(Y_{i,t} = 1 | Y_{k,t'} : k \in N_i, Y_{t'} : t' = t-1, \ldots, t-F)$$

⇒ probability that Y at position i and time t is equal to 1 given the neighbours k in the neighbourhood N_i and given Y at time lags from $t-1$ to $t-F$

Example: Mountain pine beetle outbreaks in British Columbia.



black spots represent actual outbreaks

Variable type	Variable	Explanation and rationale
Temporal	lag1	Presence/absence of mountain pine beetle in a cell the previous year.
	lag 2	Same, two years previous.
	lag 3	Same, three years previous.
Spatial	lst nbhd	First-order neighborhood (nearest four neighbours).
	2nd nbhd	Second-order neighborhood (nearest eight neighbours).
	infestations	Number of discrete infestations in each cell. This differs from the response variable, the presence/absence of red attack in each cell.
Environmental	tmin	Minimum temperature over calendar year.
	tmax	Maximum temperature over calendar year.
	tmean	Mean temperature over calendar year.
	cold[a]	Number of days cold enough to cause lethal mortality to overwintering larvae, after Wygant (1940). Temperatures less extreme than −37°C can be lethal early and late in the year, and complete mortality occurs when larvae are exposed to temperatures < −37°C for short periods (Wygant 1940, Somme 1964, Stahl et al. 2006b).
	warm[a]	Mean August temperature. Development and emergence of new mountain pine beetle adults is closely governed by temperature. Peak flight occurs in a narrow window in summer (McCambridge 1971, Safranyik 1978, Safranyik and Carroll 2006).
	ddegg[a]	Accumulated degree days above 5.5°C from August to end of growing season.
	dd[a]	Accumulated degree days above 5.5°C from August in previous year to current July.
	Pla[a]	0/1 indicator variable: sufficient heat accumulation to hatch 50% of eggs before winter (306°C degree days).
	Plb[a]	0/1 indicator variable: sufficient heat accumulation to develop and emerge on a univoltine life cycle (833°C degree days).
	P2[a]	0/1 indicator variable if minimum winter temperatures were higher than −40°C.
	elevation	Mean elevation of cell, based on 25 sampled points (regular design) within cell. This may be a useful proxy for host tree distribution, as lodgepole pine do not grow at high elevations over our study area.

(Aukema et al. 2008)

Synopsis: Spatial-temporal autologistic regression model (STARM) can be used to examine the probability the occurrence of an event across a spatial lattice over discrete time points.

III. Travelling Waves in Space and Time

⇒ "Travelling waves in cyclic populations refer to temporal shifts in peak densities moving across space in a wave-like fashion" (Johnson et al 2004).

⇒ used to examine the partial temporal synchrony or asynchrony of populations which are spatially separated

Time →→→→→→→→→→→→→→→→→→→→→→

(from Kaitala and Ranta 1998)

⇒ Data needed are estimates of population density (or other population factors) at different sites over time

time series for several spatial locations (from Moss et al 2000)

⇒ First, must determine a relationship between distance and dynamics (using Mantel correlation between two matrices, for example)

⇒ Second, fit a model to the data

⇒ Typically based on generalized additive models (GAMs)

⇒ Take a temporal pattern and add time lags based on location

⇒ For a distribution (normal, binomial, etc.), represents a function g that links the expected value $E(Y)$ of the distribution to the parameters $f_m(x_m)$

$$g(E(Y)) = \beta_0 + f_1(x_1) + f_2(x_2) + \ldots + f_m(x_m)$$

⇒ Models for travelling waves typically links a demographic variable (usually density) to spatial, temporal and other factors

$$\log D_{i,t} = m_i + tb_i + s(t + r(\cos\theta y_i + \sin\theta x_i)) + \varepsilon_{i,t}$$

response — $\log D_{i,t}$

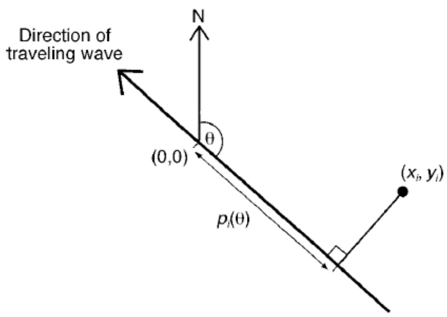
mean — m_i

long term trend — tb_i

spatial and temporal effects — $s(t + r(\cos\theta y_i + \sin\theta x_i))$

error — $\varepsilon_{i,t}$

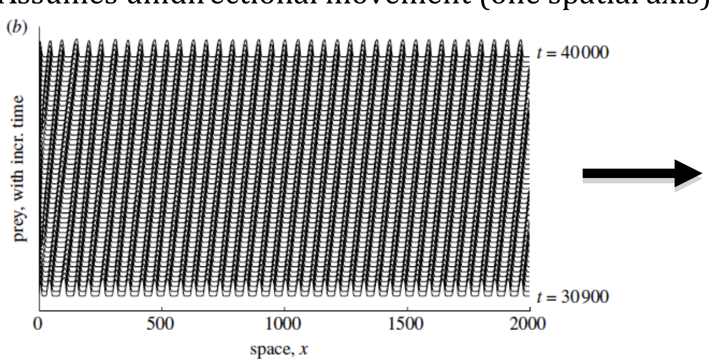
$D_{i,t}$ = demographic variable of interest
m = long term mean of D at i
b_i =long term trend at i
s = function representing temporal and spatial pattern
$r \propto \frac{1}{\text{distance}}$
θ = angle between direction of wave and North



(from Moss et al. 2000)

where $p_i(\theta) = r(\cos\theta y_i + \sin\theta x_i)$

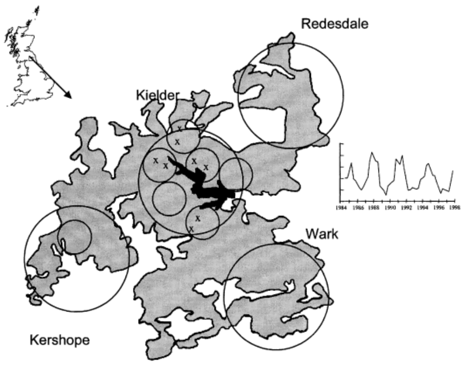
⇒ The *s* function can be any number of potential equations
⇒ Parameters r and θ are estimated by iterative analysis of deviance, with values for r and θ with a minimum deviance considered the best fit
⇒ Assumes unidirectional movement (one spatial axis)



(from Sherratt and Smith, 2008)

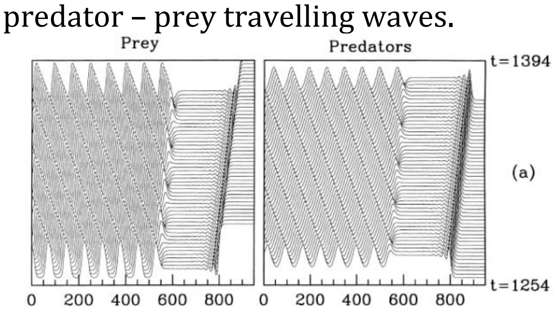
⇒ Can also fit a model for radiating rings from a central point
⇒ Models can also use second order partial differential equations
⇒ Third, test for significance

⇒ Significance can be tested by comparing the deviance of models iteratively randomly re-assigned to locations to the deviance for models of the observed data
⇒ Can be considered at a variety of scales



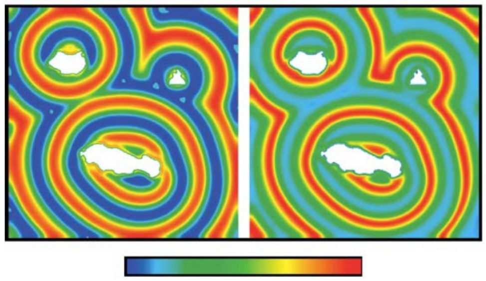
(MacKinnon et al. 2001)

Example: Modelling predator – prey travelling waves.



(From Sherratt 2001)

⇒ observed in larch budmoth, red grouse, voles, lemmings, lynx (see Sherratt and Smith 2008)



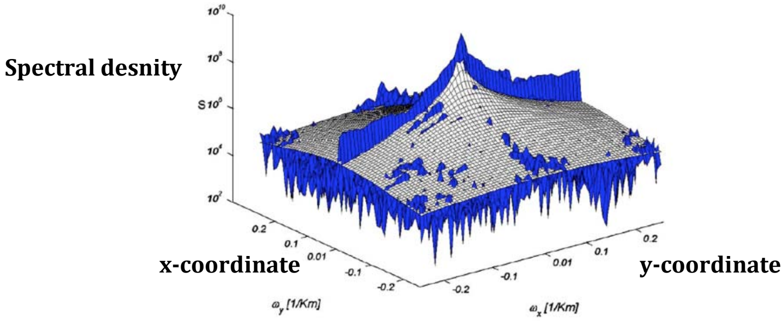
(Sherratt and Smith 2008)

Synopsis: Travelling wave models can be used to test for and estimate patterns in spatial asynchrony in temporal population dynamics.

III. Spectral Analysis in Space and Time

⇒ Can be used to deconstruct time series distributed in space into component frequencies
⇒ Requires that data are collected at even intervals in both space and time

Examples: Spatio-temporal patterns of rainfall.



(from De Michele and Bernardara 2005)

Mapping brain activity in space and time.

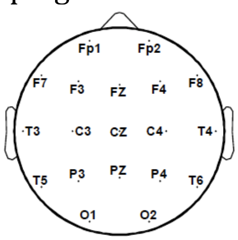
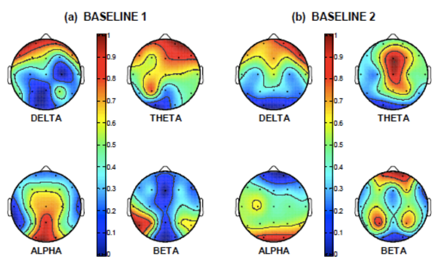


Fig 1 Electrode montage

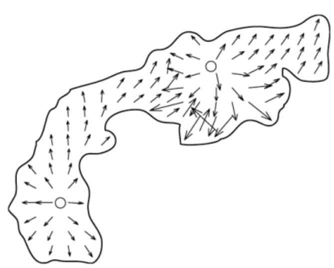


(from Santhosh et al. 2008)

IV. Wavelet Analysis in Space and Time

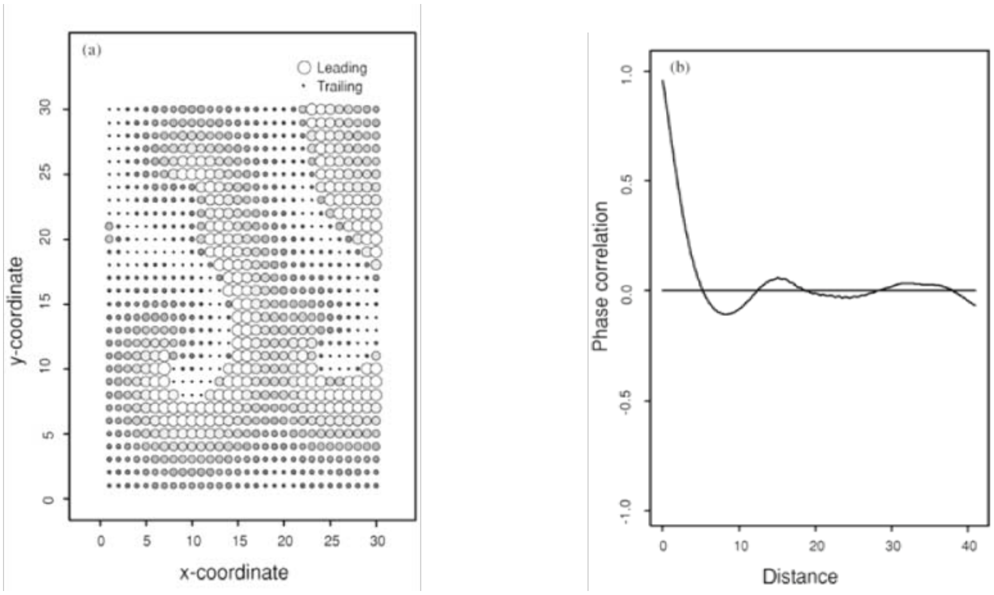
⇒ Recently have been applied to examine travelling waves in space and time

Examples: Travelling waves in the larch budmoth.



(from Johnson et al 2004)

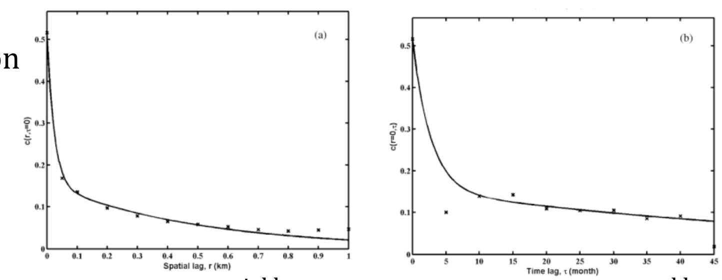
Travelling waves in host-parasitoid interactions.



(from Liebold et al. 2004)

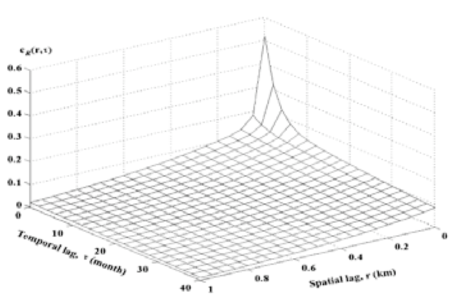
V. Kriging in Space and Time

correlation



spatial lag temporal lag

correlation



temporal lag spatial lag

(from Douaik 2005)

References

(1) Autologistic Regression in Space and Time

Augustin, N. H., M. A. Mugglestone and S. T. Buckland. 1996. An Autologistic Model for the Spatial Distribution of Wildlife. Journal of Applied Ecology 33(2): 339-347.

Aukema, B. H., A. L. Carroll, Y. Zheng, J. Zhu, K. F. Raffa, R. D. Moore, K. Stahl and S. W. Taylor. 2008. Movement of outbreak populations of mountain pine beetle: influences of spatiotemporal patterns and climate. Ecography 31: 348-358.

Anonymous. Logistic regression. Wikipedia. Accessed: March 28, 2009. URL: <http://en.wikipedia.org/wiki/Logistic_regression>

Anonymous. Maximum likelihood. Wikipedia. Accessed: March 28, 2009. URL: <http://en.wikipedia.org/wiki/Maximum_likelihood>

Chatterjee, S and A. S. Hadi. 2006. Regression Analysis by Example. Fourth Edition. Hoboken, New Jersey: John Wiley & Sons, Inc. 370 p.

Dormann, C. 2007. Assessing the validity of autologistic regression. Ecological Modelling 207: 234-242.

Hosmer, D. W. and S. Lemeshow. 1989. Applied Logistic Regression. First Edition. Toronto, Ontario: John Wiley & Sons, Inc. 307 p.

Sinclair, A. R. E., J. M. Fryxell and G. Caughley. 2006. Wildlife ecology, conservation and management. Second Edition. Malden, Massachusetts: Blackwell Publishing. 469 p.

Zhu, J., Y. Zheng, A. L Carroll and B. H. Aukema. 2008. Autologistic Regression Analysis of Spatio-Temporal Binary Data via Monte Carlo Maximum Likelihood. Journal of Agricultural, Biological and Environmental Statistics 13(1): 84-98.

(2) Travelling Waves in Space and Time

Johnson, D. M., O. N. Bjørnstad and A. M. Liebold. 2004. Landscape geometry and travelling waves in the larch budmoth. Ecology Letters 7: 967-974.

Kaitala, V. and E. Ranta. 1998. Travelling wave dynamics and self-organization in a spatio-temporally structured population. Ecology Letters 1: 186-192.

Lambin, X., D. A Elston, S. J. Petty and J. L. MacKinnon. 1998. Spatial asynchrony and periodic travelling waves in cyclic populations of field voles. Proceedings of the Royal Society B: Biological Sciences 265: 1491-1496.

MacKinnon, J. L., S. J. Petty, D. A. Elston, C. J. Thomas, T. N. Sherratt and X. Lambin. 2001. Scale Invariant Spatio-Temporal Patters of Field Vole Density. Journal of Animal Ecology 70(1): 101-111.

Moss, R., D. A. Elston and A. Watson. 2000. Spatial asynchrony and demographic traveling waves during red grouse population cycles. Ecology 81(4): 981-989.

Sherratt, J. A. 2001. Periodic travelling waves in cyclic predator-prey systems. Ecology Letters 4: 30-37.

Sherratt, J. A and M. J. Smith. 2008. Periodic travelling waves in cyclic populations: field studies and reaction-diffusion models. Journal of the Royal Society Interface 5: 483-505.

(3) Spectral Analysis in Space and Time

De Michele, C. and P. Bernardara. 2005. Spectral analysis and modeling of space-time rainfall fields. Atmospheric Research 77: 124-136.

Santhosh, J., G. Agrawal, B. Manvir, S. B. Nandeeshwara and S. Anand. 2008. Spatio-Temporal EEG Spectral Analysis of Shambhavi Maha Mudra Practice in Isha Yoga. <http://www.ishafoundation.org/news/happenings/Shambavi.pdf>

(4) Wavelet Analysis in Space and Time

Bjørnstad, O. N. and J. Bascompte. 2001. Syncrony and Second-Order Spatial Correlation in Host-Parasitoid Systems. Journal of Animal Ecology 70(6): 924-933.

Johnson, D. M., O. N. Bjørnstad and A. M. Liebold. 2004. Landscape geometry and travelling waves in the larch budmoth. Ecology Letters 7: 967-974.

Liebhold, A., W. D. Koenig and O. N. Bjornstad. 2004. Spatial Synchrony in Population Dynamics. Annual Review of Ecology, Evolution and Systematics 35: 467-490.

(5) Kriging in Space and Time

Douaik, A., M. Van Meirvenne, and T. Toth. 2005. Soil salinity mapping using spatio-temporal kriging and Bayesian maximum entropy with interval soft data. Geoderma 128: 234-248.