A Generalized Linear Mixed Model for Enumerated Sunspots

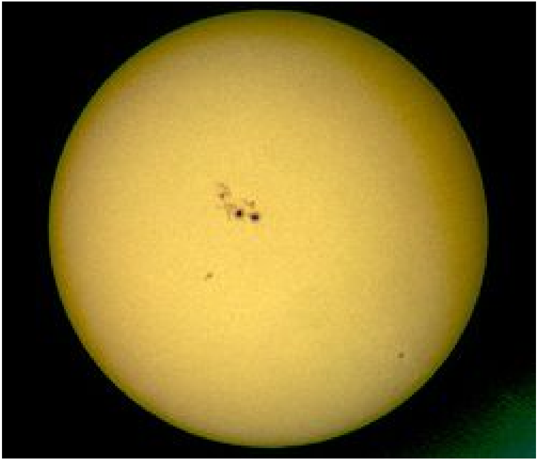
Jamie Riggs
Applied Statistics and Research Methods
Deep Space Exploration Society

AAVSO 100th Annual Meeting

October 8, 2011

UNIVERSITY of NORTHERN COLORADO

Solar Beauty Spots



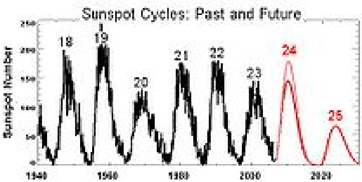
Presentation Outline

- Introduction
- Background
- Wald Approach
- Statistical Models for Counts Data
- Future Development
- Acknowledgments
  Rodney Howe, Solar Bulletin Editor, AAVSO
  Trent Lalonde, Applied Statistics, University of Northern Colorado

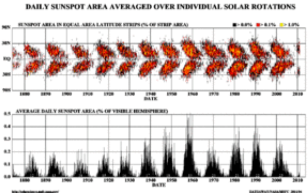
The Physics

- Sunspot generation a current research area
- Sunspots thought to be the visible counterparts of magnetic flux tubes in the Sun's convective zone
- Differential rotation (coriolis effect) stresses the tubes which puncture the Sun's surface
- Energy flux from the Sun's interior decreases and with it surface temperature
- Sunspot activity cycles about every eleven years
- Early in the cycle, sunspots appear in the higher latitudes and then move towards the equator as the cycle approaches maximum: this is called Spörer's law

Sunspot Cycle and Butterfly Plot



(a) Eleven-year cycle



(b) Spörer's Law

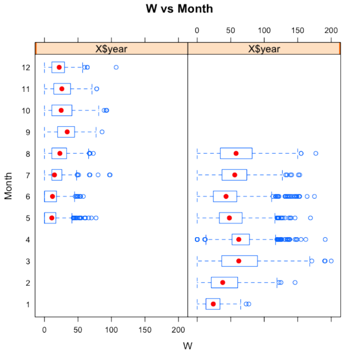
The Astronomy

- First noted sunspots in 364 BC by Chinese astronomer Gan De
- First telescopy in 1610 by English astronomer Thomas Harriot
- Rudolf Wolf established the Wolf Number in 1848
- AAVSO began the American Relative number in 1949
- Overall, weighted monthly count averages are assumed to be unbiased estimates of the true monthly sunspot numbers
- As sunspot cycle in the last 5 months is increasing from a minimum, monthly corrections are anticipated

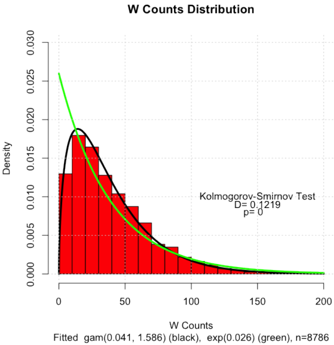
The Statistics

- Multiple observers ($\sim$ 80) worldwide
- Three random variables: sunspot counts, observers, and monthly sunspot numbers
- Sunspot numbers are known to follow an approximately 11-year sinusoidal cycle
- The statistical model needs to tie the average monthly sunspot numbers to the observer-reported counts
- The statistical model should predict sunspot numbers

Monthly Submissions and Histogram



(c) Monthly counts



(d) Histogram with fitted pdfs

Wolf, Wald, and Shapley

The Framers

- Wolf, R, 1848.
  - Developed the Wolf number (a International sunspot number, relative sunspot number, or Zürich number)
  - A quantity measuring the number of sunspots and groups of sunspots on the Sun's surface
  - The relative sunspot number R is computed as

$$R = k(10g + s)$$

  where

    - s is the number of individual spots
    - g is the number of sunspot groups
    - k is a factor that varies with location and instrumentation

The Framers

- Wald, A., The Fitting of Straight Lines if Both Variables are Subject to Error, *Annals Mathematical Statistics*, 1940, Vol. 11, No. 3, pp. 284-300.
  - Response, Y, and predictor, X are random variables
  - Method of least squares (SLR) usually used
  - Fit parameters different for $Y \sim f(X)$ and $X \sim f(Y)$

The Framers

- Shapley, A.H., Reduction of Sunspot-Number Observations, *Publication of the Astronomical Society of the Pacific*, 1949, Vol. 61, No. 358, pp 13-21.
  - Adapted Wald's method to correct observations from many observers to the American Relative sunspot number
  - Correction factor accounts for variations in equipment and seeing conditions
  - A "statistical weight" per observer is also used

The American Relative Sunspot Number

Shapley via Wald

$$R_i = k_i(10g_i + s_i) \tag{1}$$

Shapley via Wald

$$R_i = k_i(10g_i + s_i) \tag{1}$$

$$R_a = \frac{\sum_{i=1}^{N} w_i k_i R_i}{\sum_{i=1}^{N} w_i} \tag{2}$$

Shapley via Wald

$$R_i = k_i(10g_i + s_i) \tag{1}$$

$$R_a = \frac{\sum_{i=1}^{N} w_i k_i R_i}{\sum_{i=1}^{N} w_i} \tag{2}$$

$$R_{sm} = \frac{1}{24}\left(R_{a,i-6} + R_{a,i+6} + 2\sum_{j=i-5}^{5} R_{a,j}\right) \tag{3}$$

Poisson Models

Poisson Distribution

Poisson probability distribution function

$$f(y_i; \mu_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} = e^{-\mu_i} \frac{1}{y_i!} e^{y_i \log(\mu_i)}, \quad i = 1, 2, \ldots, N \qquad (4)$$

Poisson Distribution

Poisson probability distribution function

$$f(y_i; \mu_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} = e^{-\mu_i} \frac{1}{y_i!} e^{y_i \log(\mu_i)}, \quad i = 1, 2, \ldots, N \tag{4}$$

GLM canonical link to a monotone function of μ_i

$$\log(\mu_i) = \sum_{i,j} \beta_i x_{ij}, \quad i = 1, \ldots, N, \; j = 1, 2, \ldots, n_i \tag{5}$$

Poisson Distribution

Poisson probability distribution function

$$f(y_i;\mu_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!} = e^{-\mu_i}\frac{1}{y_i!}e^{y_i \log(\mu_i)}, \quad i = 1, 2, \ldots, N \qquad (4)$$

GLM canonical link to a monotone function of μ_i

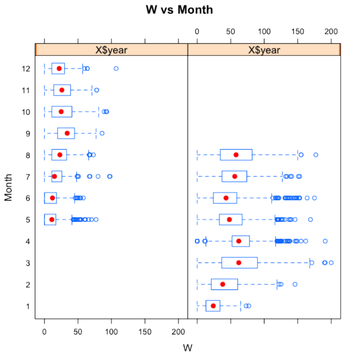
$$\log(\mu_i) = \sum_{i,j} \beta_i x_{ij}, \quad i = 1, \ldots, N, \; j = 1, 2, \ldots, n_i \qquad (5)$$

The matrix form including observer, period, and seeing conditions

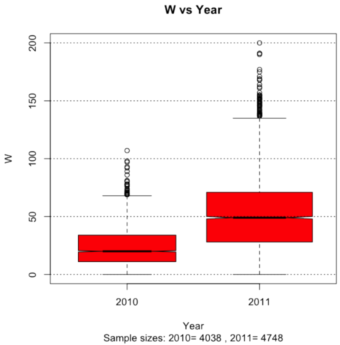
$$\log(\boldsymbol{\mu}_{\mathbf{f}}) = \mathbf{X}\boldsymbol{\beta}, \qquad (6)$$

Estimation of R_a

Estimation of R_a

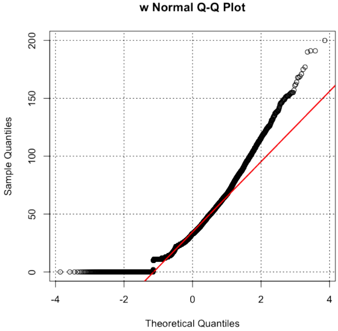


(e) Boxplots of Wolf numbers by Month

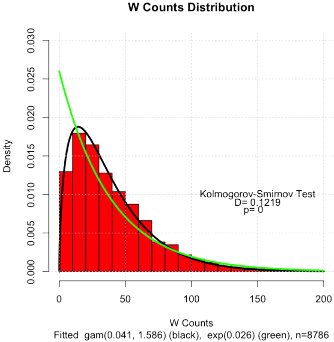


(f) Boxplots of Wolf numbers by Year

Estimation of R_a



(g) The normal Q-Q plot for the Wolf number.



(h) Wolf number distribution.

GLM with Poisson Error Structure

- Several models were fitted using the independent variables observer, seeing conditions, and time sequence
- Two error structures were used: Poisson and negative binomial

GLM with Poisson Error Structure

- Several models were fitted using the independent variables observer, seeing conditions, and time sequence
- Two error structures were used: Poisson and negative binomial
- The final model is

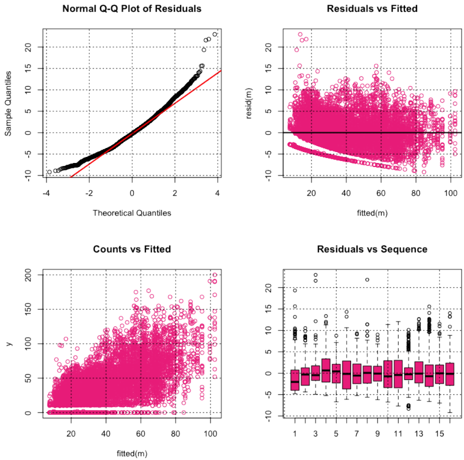
$$\log(\mu_{ij}) = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{1ij} + \boldsymbol{\beta_3 x_{3ij}} + \eta_{ij},$$

$$x_{1ij} = j^{th} \text{ appearance of the } i^{th} \text{ observer}$$

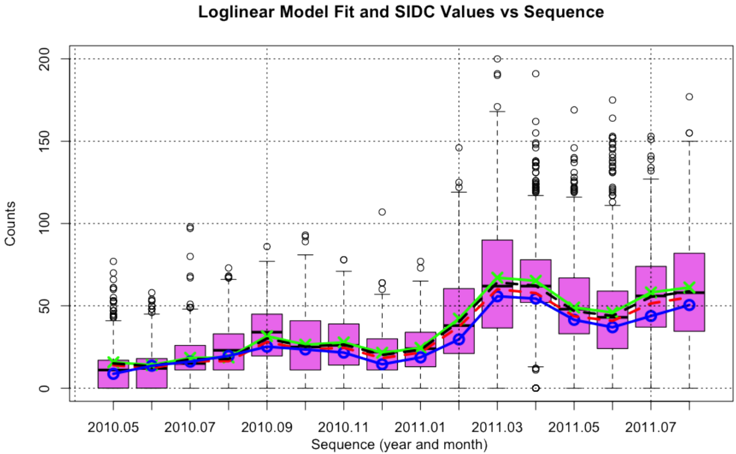
$$x_{2ij} = j^{th} \text{ occurrence of the } i^{th}\text{observer's seeing condition}$$

$$\mathbf{x_{3ij}} = j^{th} \text{ occurrence of the } i^{th}\text{time sequence}$$

GLM with Poisson Error Structure Diagnostics

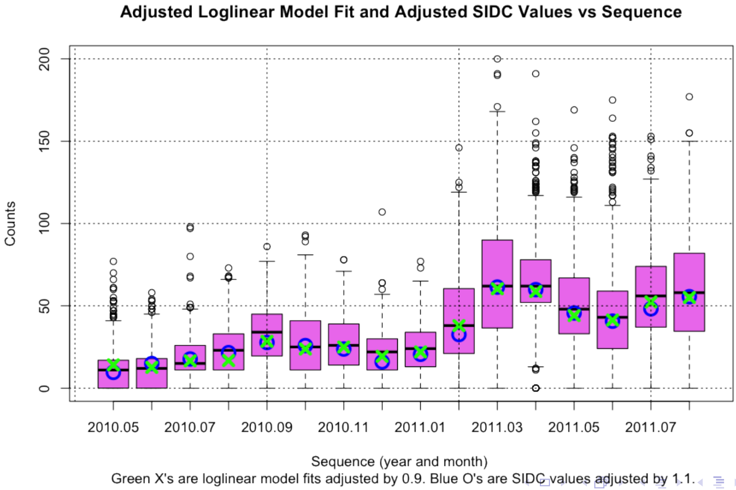


GLM with Poisson Error Structure Diagnostics



Green X's connected by the curve is the loglinear (LL) model fit. Solid blue curve is SIDC values.
The dashed red curve is a 99% lower CI for LL. The dashed black curve is a 99% upper CI for SIDC.

GLM with Poisson Error Structure Diagnostics

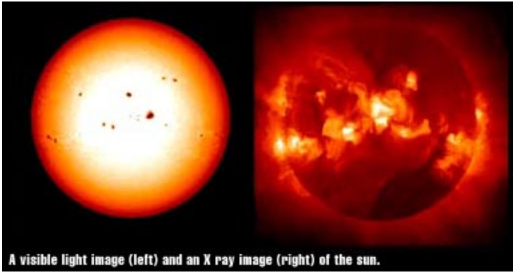


Future Development

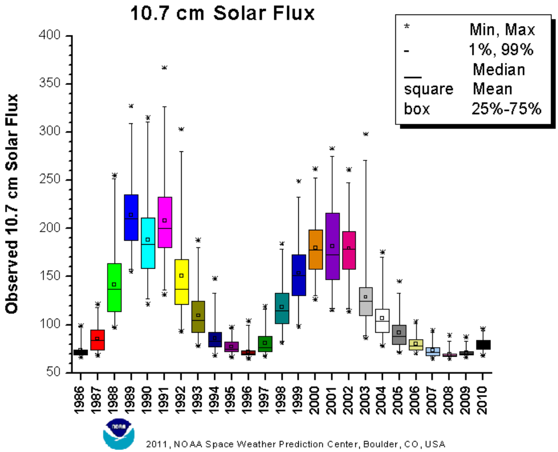
Future Development

- Introduce an observer's equipment factor (fixed)
- Test for the effect of the Solar hemisphere
- Braid in
  - Optical observations from Europe
  - X-ray observations from GOES-15 satellite
  - 10.7cm radio from Deep Space Exploration Society, Canada, and Australia

Soft X-rays



10.7 cm (2800 MHz) Radio



A Generalized Linear Mixed Model for Enumerated Sunspots

Jamie Riggs
Applied Statistics and Research Methods
Deep Space Exploration Society

AAVSO 100th Annual Meeting

October 8, 2011

UNIVERSITY of NORTHERN COLORADO