

A Generalized Linear Mixed Model for Enumerated Sunspots

Jamie Riggs

Applied Statistics and Research Methods
Deep Space Exploration Society

**Deep Space Exploration Society
Board of Directors Meeting**

July 14, 2012

UNIVERSITY of
NORTHERN COLORADO



Solar Beauty Spots



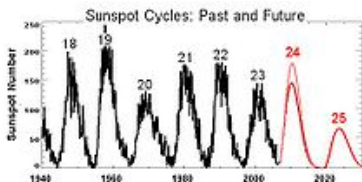
Presentation Outline

- Introduction
- Background
- Wald Approach
- General Linear Mixed Models
- Future Development
- Acknowledgments
 - Rodney Howe, Solar Bulletin Editor, AAVSO
 - Trent Lalonde, Applied Statistics, University of Northern Colorado

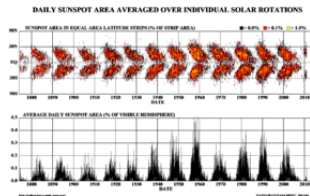
The Physics

- 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

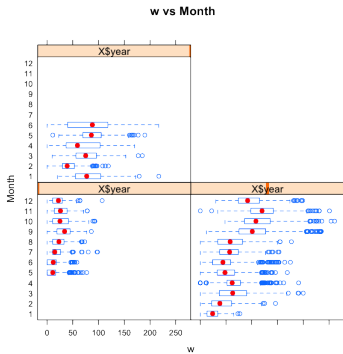
The Astronomy

- First noted sunspots in 364 BC by Chinese astronomer Gan De
- First telescope 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
- Sunspot cycle since 2010 is increasing from a minimum

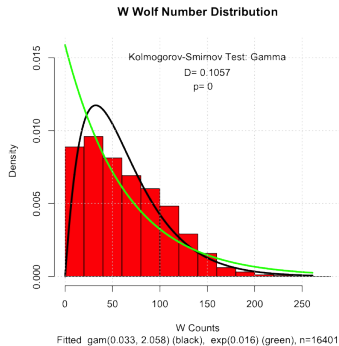
The Statistics

- Multiple observers (~ 60) 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 tie historical numbers and predict future numbers

Monthly Sunspot Counts Summary



(c) Monthly counts



(d) Histogram with fitted pdfs

Wolf, Wald, and Shapley

The Framers

- Wolf, R, 1848.
 - Developed the Wolf number (an 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 American Relative Sunspot Number

$$R_i = k_i(10g_i + s_i), \text{ } i^{\text{th}} \text{ observer's Wolf number} \quad (1)$$

$$R_a = \frac{\sum_{i=1}^N w_i k_i R_i}{\sum_{i=1}^N w_i}, \text{ monthly sunspot number} \quad (2)$$

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

smoothed monthly number

Sunspot Counts from Poisson Models

Poisson Distribution

Poisson probability distribution function models counts

$$\begin{aligned} f(y_i; \mu_i) &= \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} && (4) \\ &= \exp\{y_i \log(\mu_i) - e^{\log \mu_i} - \log(y_i!)\}, \quad i = 1, 2, \dots, N \end{aligned}$$

Poisson Distribution

Poisson probability distribution function models counts

$$\begin{aligned} f(y_i; \mu_i) &= \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} & (4) \\ &= \exp\{y_i \log(\mu_i) - e^{\log \mu_i} - \log(y_i!)\}, \quad i = 1, 2, \dots, N \end{aligned}$$

GLM canonical link to a monotone function of mean counts μ_i

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

Poisson Distribution

Poisson probability distribution function models counts

$$f(y_i; \mu_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} \quad (4)$$

$$= \exp\{y_i \log(\mu_i) - e^{\log \mu_i} - \log(y_i!)\}, \quad i = 1, 2, \dots, N$$

GLM canonical link to a monotone function of mean counts μ_i

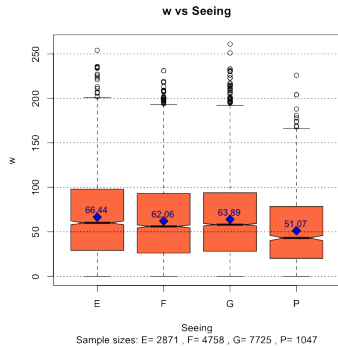
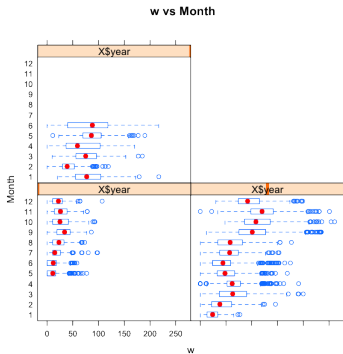
$$\log(\mu_i) = \sum_{i,j} \beta_j x_{ij}, \quad i = 1, \dots, N, \quad j = 1, 2, \dots, n_i \quad (5)$$

Matrix form with observer, period, experience, seeing, etc.

$$\log(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (6)$$

Estimation of the American Relative Sunspot Number R_a

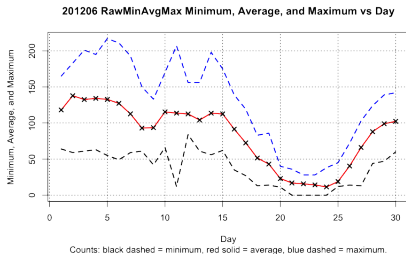
Estimation of R_a



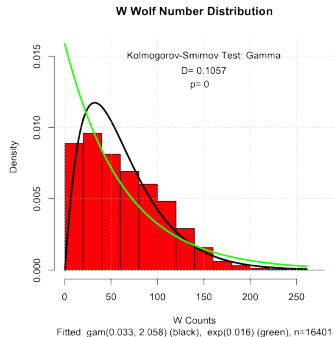
(e) Boxplots of Wolf numbers by Month

(f) Boxplots of Wolf numbers by Experience

Estimation of R_a



(g) Range of daily sunspot counts.



(h) Wolf number distribution.

R_a from GLMM with Poisson Error Structure

- Model mixes random observer and fixed seeing conditions, experience, time sequence, etc.
- Two error structures: Poisson (counts) and normal (observer)

R_a from GLMM with Poisson Error Structure

- Model mixes random observer and fixed seeing conditions, experience, time sequence, etc.
- Two error structures: Poisson (counts) and normal (observer)
- The model is

$$\log(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\nu}, \quad (7)$$

$\boldsymbol{\mu}$ = the vector of mean sunspot numbers by observer

\mathbf{X} = the fixed effects matrix

$\boldsymbol{\beta}$ = the vector of fixed effects parameters

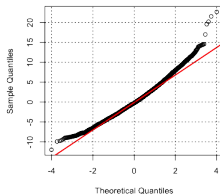
\mathbf{Z} = the random effects matrix of observer identifiers

$\boldsymbol{\nu}$ = the random effects parameter vector

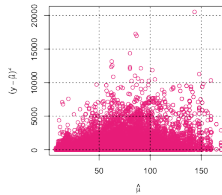
(8)

R_a GLMM with Poisson Error Structure Diagnostics

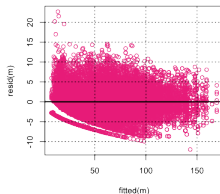
Normal Q-Q Plot of Residuals



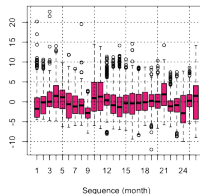
Poisson Variance vs Mean



Residuals vs Fitted



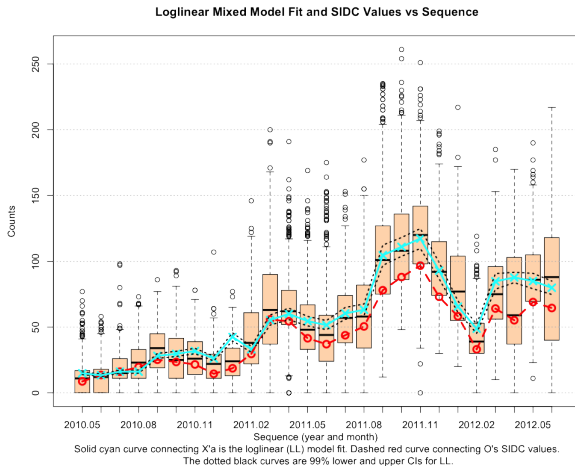
Residuals vs Sequence



R_a GLMM with Poisson Error Structure Diagnostics

- $s^2/\bar{x} = 18.54494 \gg 1$ indicates overdispersion
- Poisson error structure with normal random effect insufficient to account for all the R_a
- Concave up Normal Q-Q plot indicates right-skewed residuals
- Residuals vs. Fitted plot pattern indicates missing or misspecified predictors
- Preliminary use of Gamma error structure for observer random effect reduces the mean-variance ratio
- Poisson error structure with normal random effect insufficient to account for all the R_a GLMM deviation

R_a GLMM with Poisson Error Structure Predictions



R_a GLMM Overdispersion

Table: Improvements from Error Structure Changes

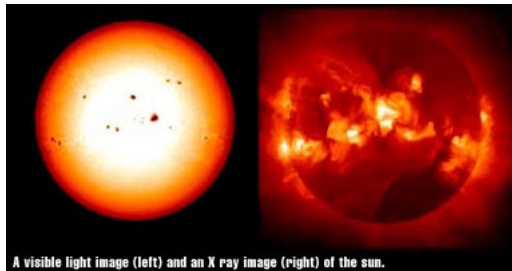
$y u$ Dist	Link $g(\mu)$	u Dist	Link $v(u)$	Method	s^2/\bar{x}
Poisson	log	fixed	NA	GLS	22.87
Poisson	log	Normal	identity	log-likelihood	21.66
Poisson	log	Poisson	log	h-likelihood	18.54
Poisson	log	Gamma	identity	h-likelihood	?
Gamma	log	Gamma	identity	h-likelihood	?
Gamma	inverse	inverse Gamma	inversey	h-likelihood	?

Future Development

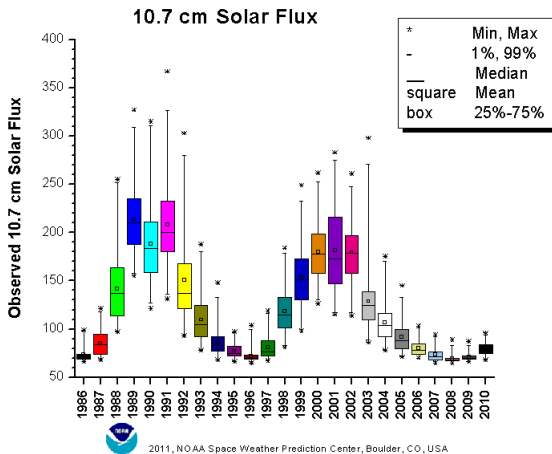
Future Development

- Introduce observer equipment factor (fixed)
- Test for the effect of the Solar hemisphere
- Test different error structures for counts and for observer random variables
- Braid in
 - Optical observations from Europe
 - X-ray
 - 10.7cm radio

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

**Deep Space Exploration Society
Board of Directors Meeting**

July 14, 2012

UNIVERSITY of
NORTHERN COLORADO

