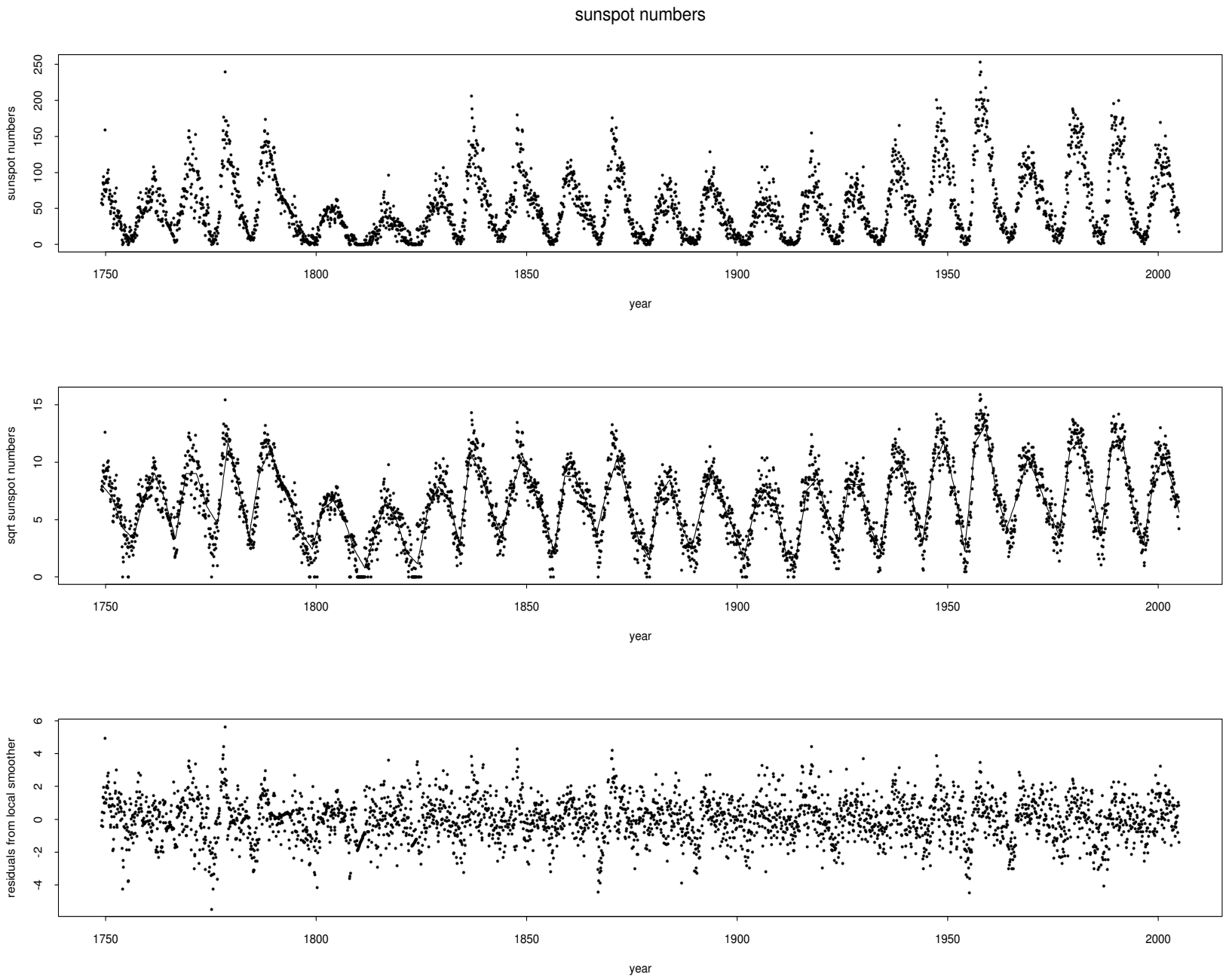
Statistical Modeling of Sunspot Cycles

Yaming Yu
Department of Statistics, UC Irvine

Suspot Number Data

- The longest directly observed index of solar activity
  - 1610: Galileo first viewed sunspots with his new telescope.
  - 1749: Daily observations were started at the Zurich Observatory.
  - 1849: Continuous (daily) observations were obtained with the addition of more observatories.
- Sunspot No. = No. of individual spots + 10 × No. of groups
- – Top: monthly averages of the International Sunspot Numbers.
  - Middle: local smoother fit to sqrt(SSN).
  - Bottom: residuals.


sunspot numbers
sunspot numbers
0
50
100
150
200
250
1750
1800
1850
1900
1950
2000
year
sqrt sunspot numbers
0
5
10
15
1750
1800
1850
1900
1950
2000
year
residuals from local smoother
-4
-2
0
2
4
6
1750
1800
1850
1900
1950
2000
year

Features of Sunspot Cycles

- **A lot of noise.**
- **Quasi-periodicity: average cycle length is 11 years (Wolf 1852).**
- **Asymmetry: rise to maximum is faster than fall to minimum (Waldmeier 1935, 1939).**
- **Waldmeier effect: stronger cycles tend to take less time to rise to maximum amplitude.**
- **Long-term (8–9 cycles) periodicity (?)**
- **...**

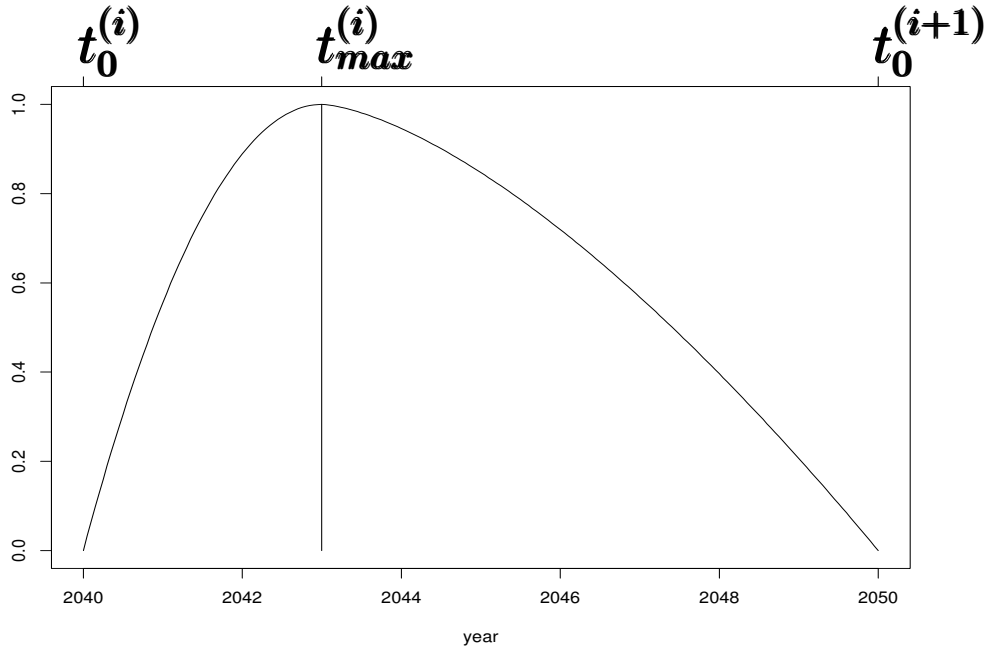
How to quantify the statistical significance?

Physical models of the solar dynamo are unfortunately lacking...
But we can build statistical models.

Modeling Each Cycle by Simple Functions

Notation for cycle i

- $t_0^{(i)}$: start of cycle i
- $t_{max}^{(i)}$: time at cycle maximum
- $t_0^{(i+1)}$: end of cycle i



- R_t: "average solar activity level" at time t
  - For the rising phase $t < t_{max}^{(i)}$

$$R_t = c_i \left(1 - \left(\frac{t_{max}^{(i)} - t}{t_{max}^{(i)} - t_0^{(i)}} \right)^{\alpha_1} \right);$$

  - For the declining phase $t > t_{max}^{(i)}$

$$R_t = c_i \left(1 - \left(\frac{t - t_{max}^{(i)}}{t_0^{(i+1)} - t_{max}^{(i)}} \right)^{\alpha_2} \right).$$

- cycle length $= t_0^{(i+1)} - t_0^{(i)}$;
  time to rise to maximum $= t_{max}^{(i)} - t_0^{(i)}$;
  amplitude $= c_i$.

- $\alpha_1, \alpha_2 > 1$: the same shape parameters for all cycles.

A Nonlinear Regression Model

- Model sqrt of sunspot numbers to stablize the variance:

$$\sqrt{Y_t} \overset{ind}{\sim} N(\beta_0 + \beta_1 t + R_t,\ \sigma^2)$$

- Cycle-specific parameters
  - $T_0 = (t_0^{(i)},\ i = 0, 1, \ldots, k)$;
  - $T_{max} = (t_{max}^{(i)},\ i = 0, \ldots, k-1)$;
  - $C = (c_i,\ i = 0, \ldots, k-1)$.

  Total number of available cycles $k = 24$.

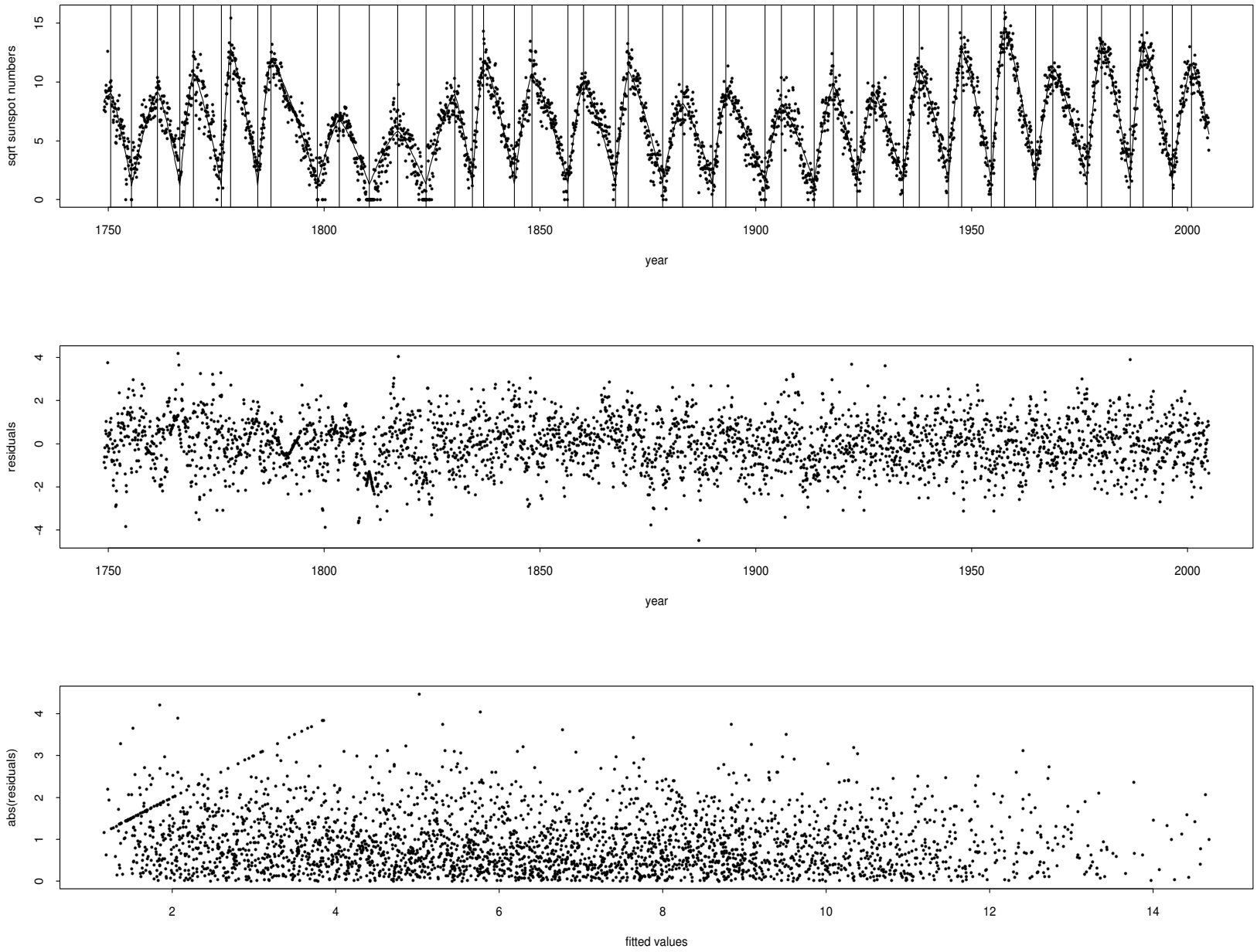
- Specify non-informative priors.

- Use MCMC (Gibbs sampler with Metropolis-Hastings steps) to fit the model.

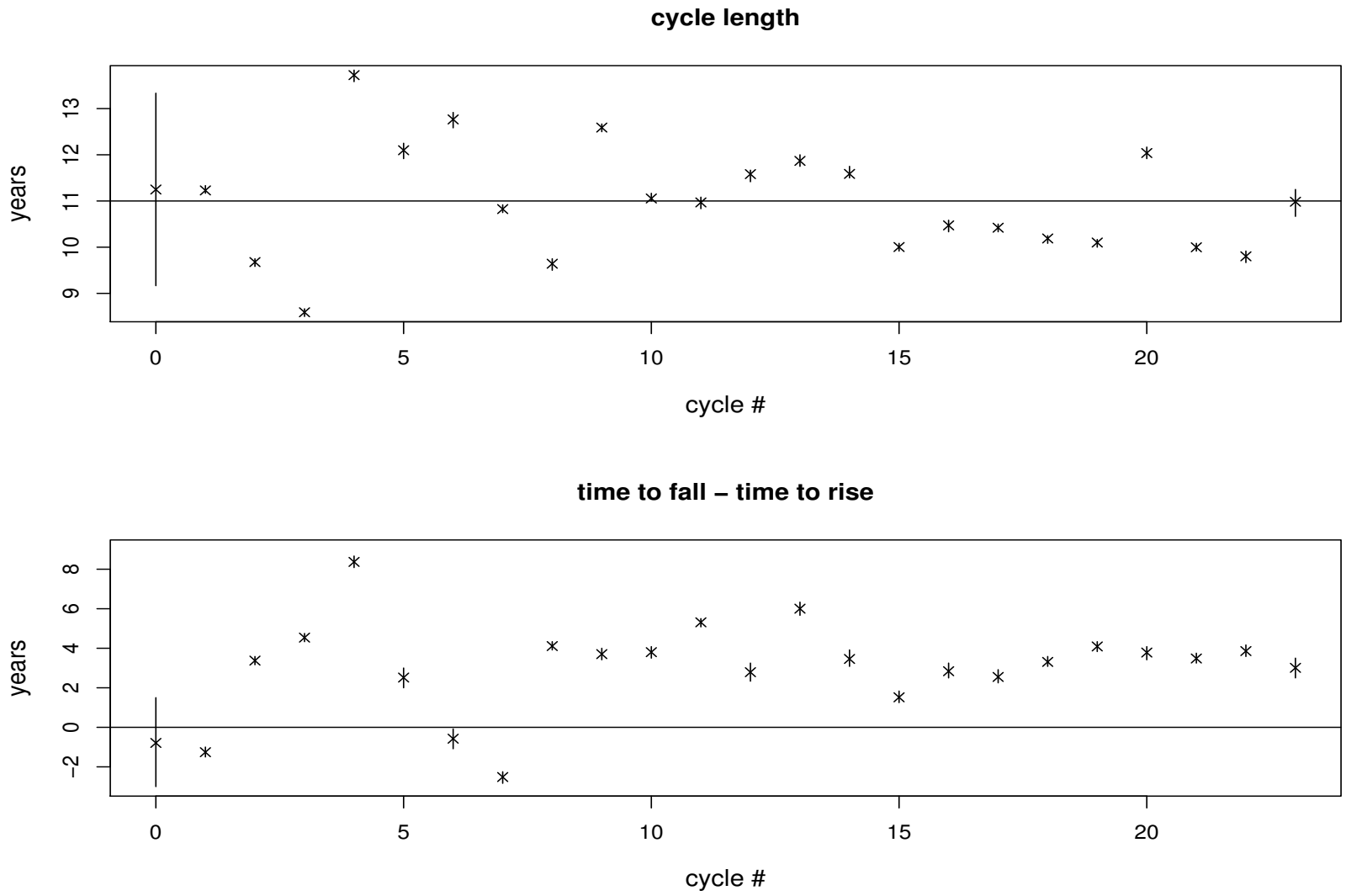
Posterior Inference

Fitted model and residuals

- – **Top: sqrt(SSN) with fitted values.**
  Vertical lines represent one posterior draw of $(T_0,\ T_{max})$.
  – **Middle: residuals vs. time (year).**
  – **Bottom: residuals vs. fitted values.**
- **The fit is better for recent data ($year > 1850$) than for the less reliable data in the past.**

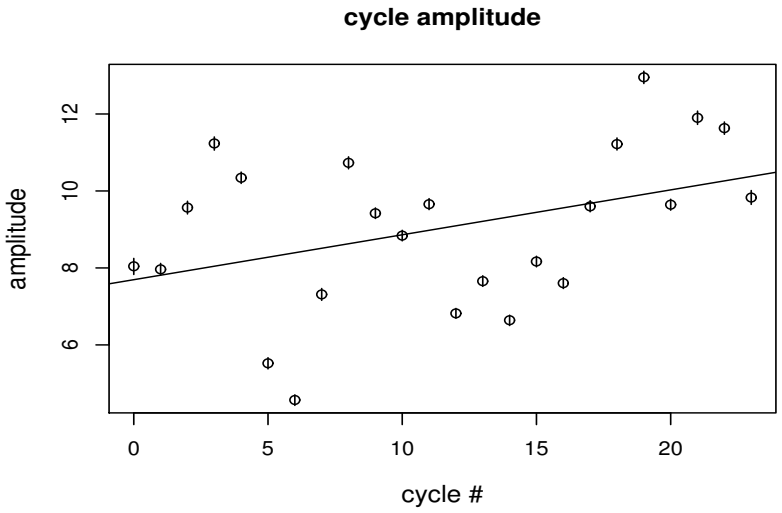

sqrt sunspot numbers
0
5
10
15
1750
1800
1850
1900
1950
2000
year
residuals
-4
-2
0
2
4
1750
1800
1850
1900
1950
2000
year
abs(residuals)
0
1
2
3
4
2
4
6
8
10
12
14
fitted values

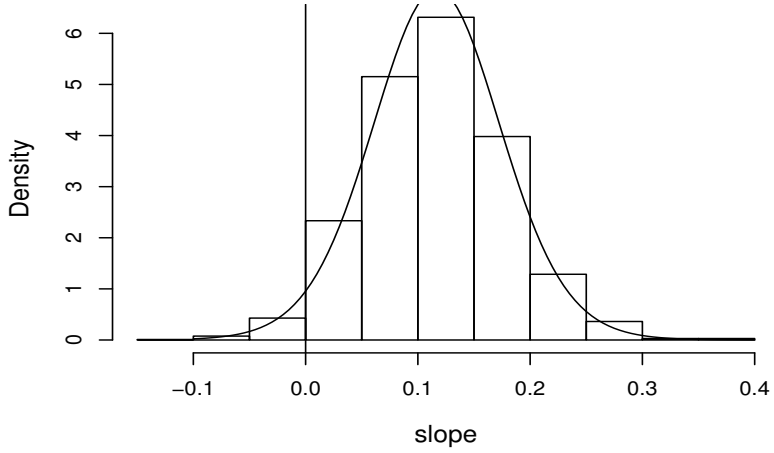
Posterior Inference: Cycle Length Patterns

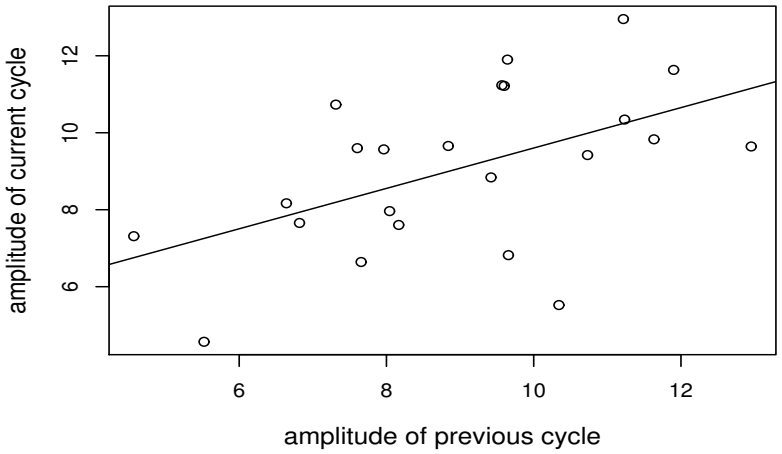


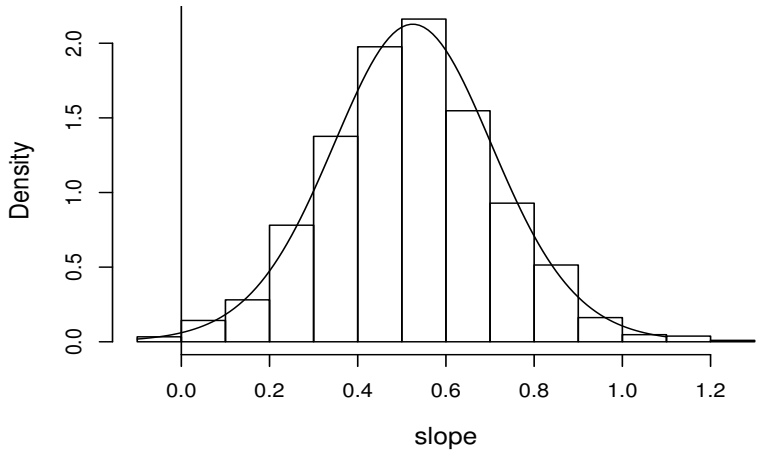
- **Average cycle length is around 11 years**
  (×'s mark posterior means)
- **Error bars are small**
  (Vertical bars represent 50% marginal credible intervals)
- **The cycle length has no apparent upward or downward trend.**
- **With few exceptions, cycles take more time to decline than to rise.**
- **Only about half of Cycle # 0 is observed, hence the large error bars.**

Cycle Amplitude Patterns





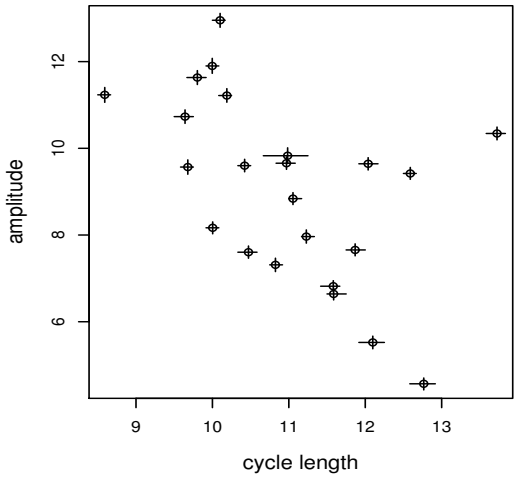


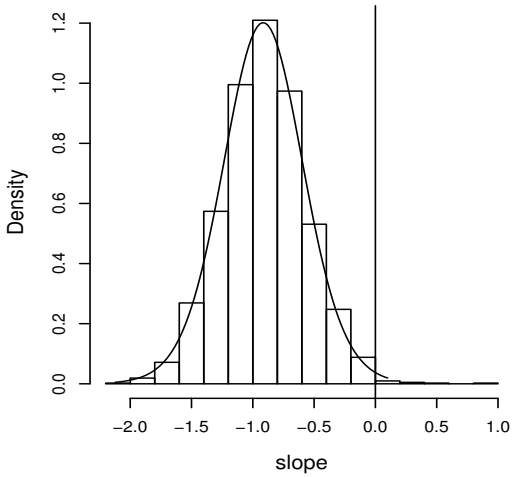


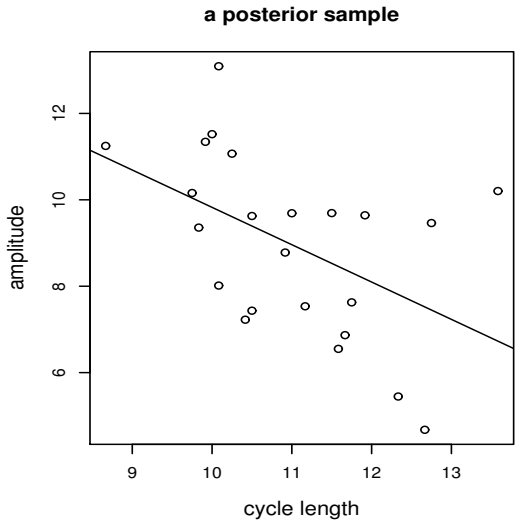
Evaluating Statistical Significance

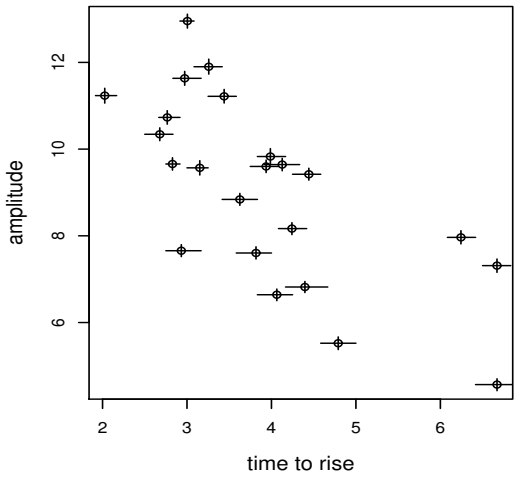
- **Wrong procedure: simple linear regression using the posterior mean as the true amplitudes.**
- **Ideally we should fit a hierarchical model.**
- **A two-stage simulation procedure to start with:**
  - **Draw posterior samples of the cycle amplitudes (done).**
  - **For each sample, fit the regression model of amplitude vs. cycle #, and then draw from the posterior of the regression coefficient.**
- **Because error bars are small, results (histogram) are nearly identical to those of simple linear regression (solid curve).**

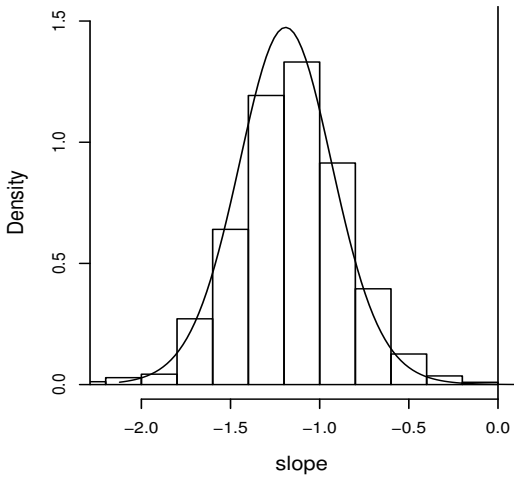
Relationship Between Cycle Length and Amplitude

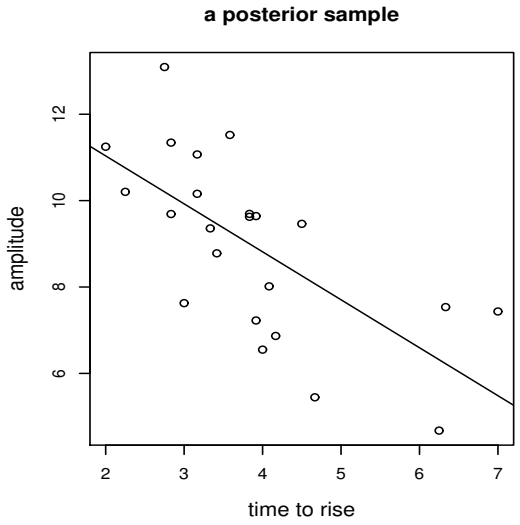










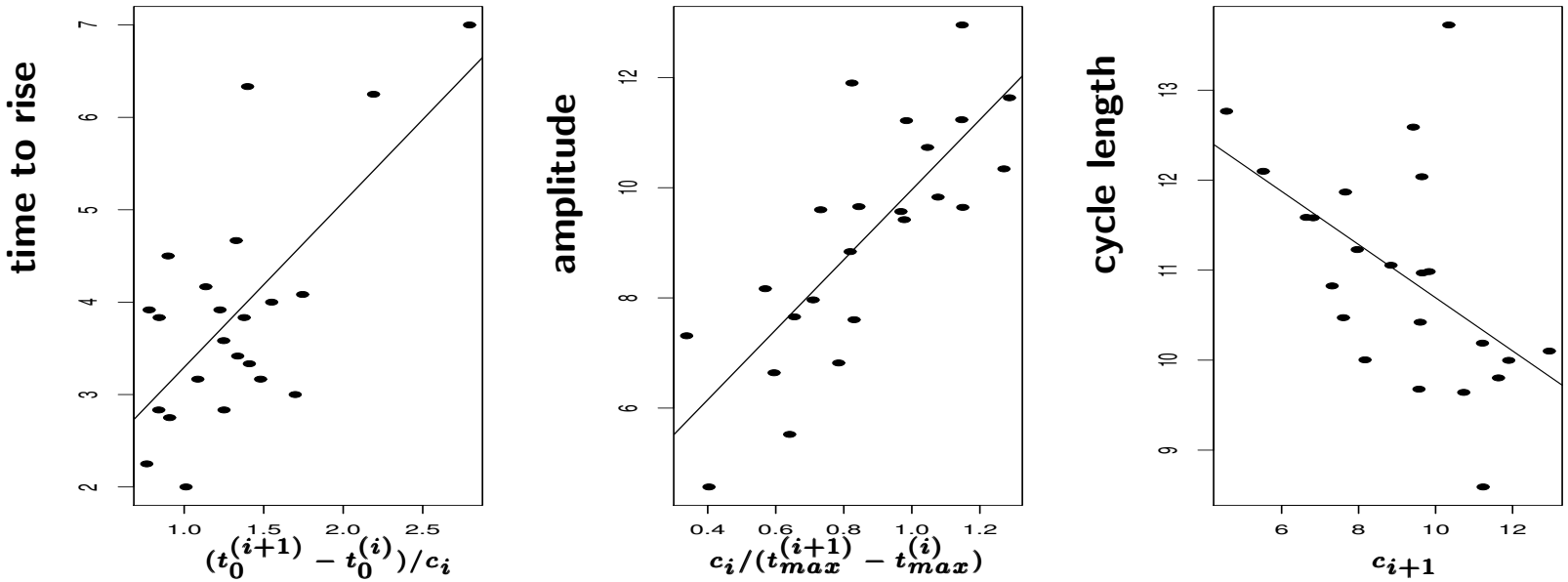


- **Row 1: amplitude vs. cycle length**
  - **Left: posterior means with 50% credible intervals.**
  - **Middle: Statistical significance of the regression slope. Little difference between simple linear regression and two-stage simulation.**
  - **Right: A posterior sample and its regression line.**
- **Row 2: amplitude vs. time to rise to cycle maximum**
  - **Middle: the error bars are large enough to make a (very small) difference.**

Some New Relations

An empirical model to generate cycle $i+1$ from cycle i:

- **Time of maximum:** $t_{max}^{(i+1)} - t_0^{(i+1)} \sim \theta_1 + \gamma_1 \frac{t_0^{(i+1)} - t_0^{(i)}}{c_i} + N(0, \sigma_1^2)$
- **The amplitude:** $c_{i+1} \sim \theta_2 + \gamma_2 \frac{c_i}{t_{max}^{(i+1)} - t_{max}^{(i)}} + N(0, \sigma_2^2)$
- **The length:** $t_0^{(i+2)} - t_0^{(i+1)} \sim \theta_3 + \gamma_3 c_{i+1} + N(0, \sigma_3^2)$



Forecasts

- Based on these new relations, cycle 24 is estimated to rise to maximum around year 2011.1 ± 1.1, with a maximum smoothed monthly sunspot number of 128 ± 36.
- Work in progress:
  - Data quality problems.
  - Comparison with similar models in the literature.
  - A more elaborate model to link cycle length, time to rise, and amplitude through hyperparameters.
  - Incorporating additional information, e.g., spatial location of sunspots, magnetic polarity information; joint modeling with 10.7cm flux, etc.
  - Better algorithms. More efficient computer code.
  - ...