

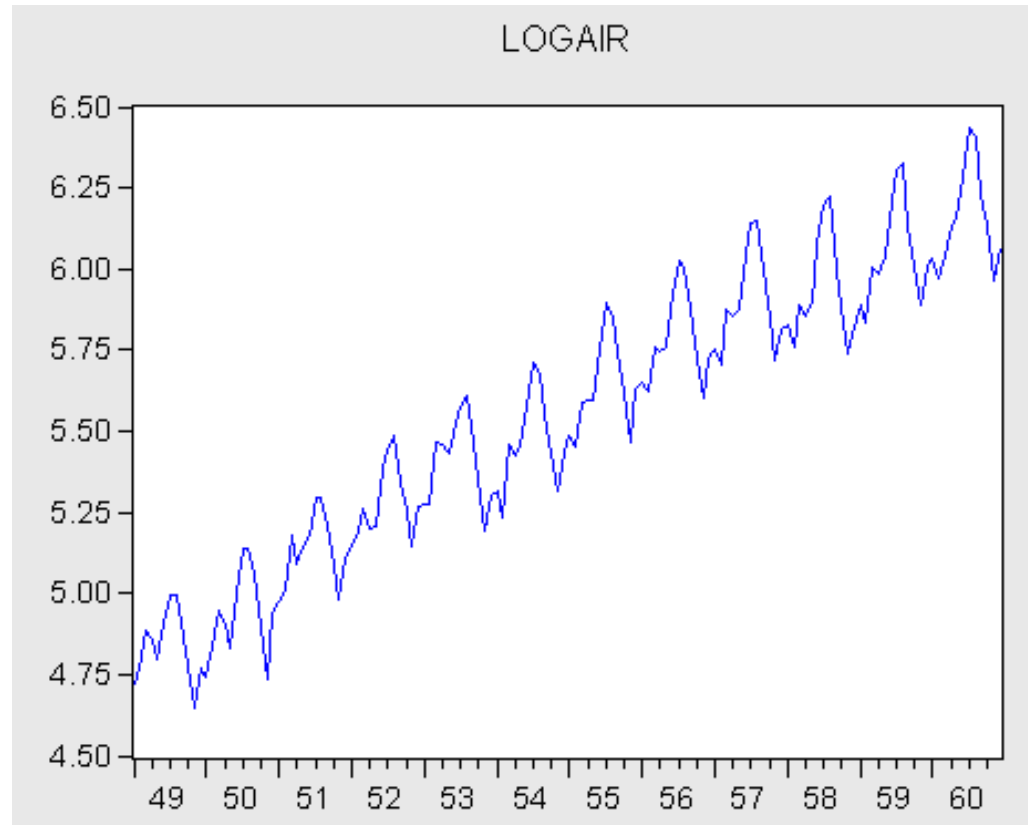
Spektrálna analýza

Beáta Stehlíková

Časové rady, FMFI UK, 2011/2012

Motivácia

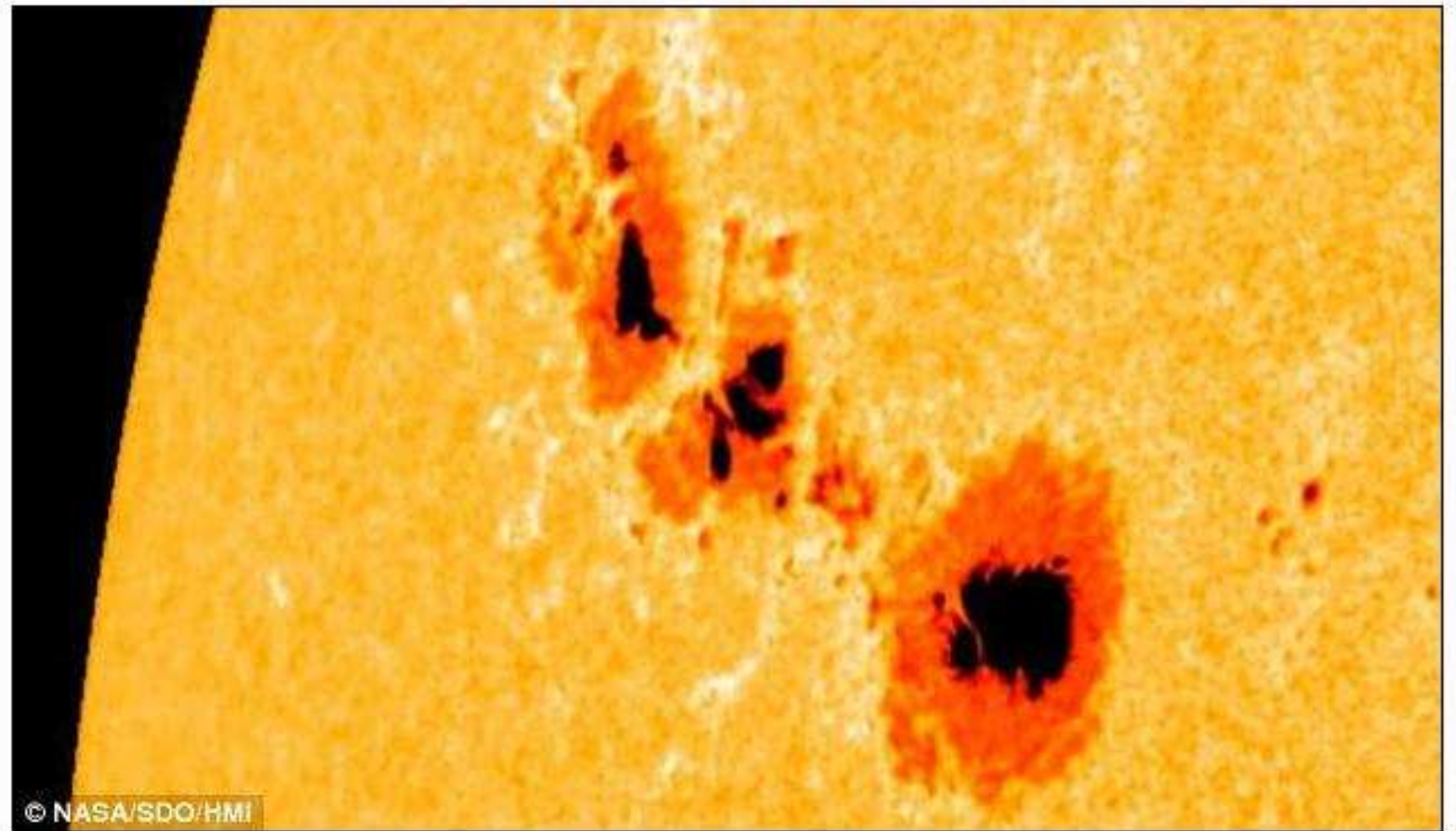
- Príklad zo začiatku semestra - počet cestujúcich aerolinkami:



- Podobne napr. mesačné dáta nezamestnanosti, kvartálne dáta HDP, ...
- Nie vždy to však musí byť takto jasné

Motivácia

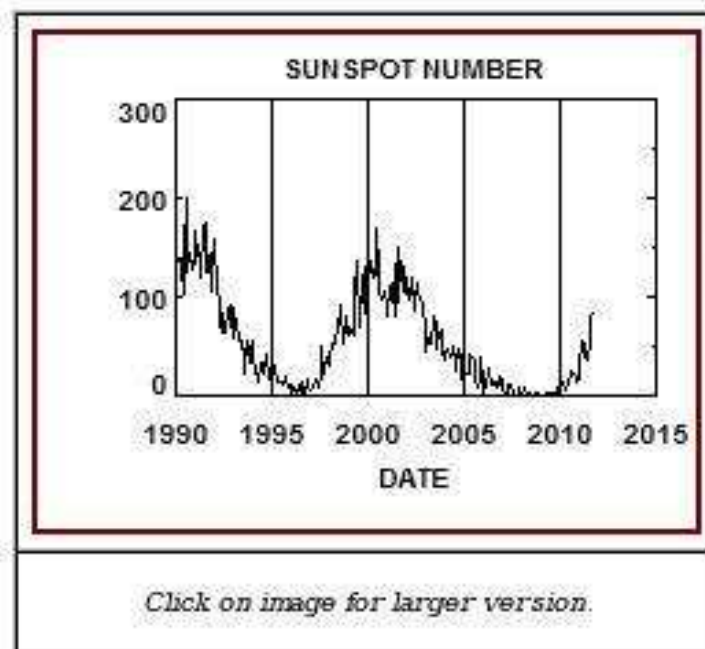
- Slné škvvrny:



<http://www.dailymail.co.uk/sciencetech/article-2042428/Best-auroras-seen-Britain-thanks-huge-solar-flares.html>

Motivácia

- Slné škvvrny:



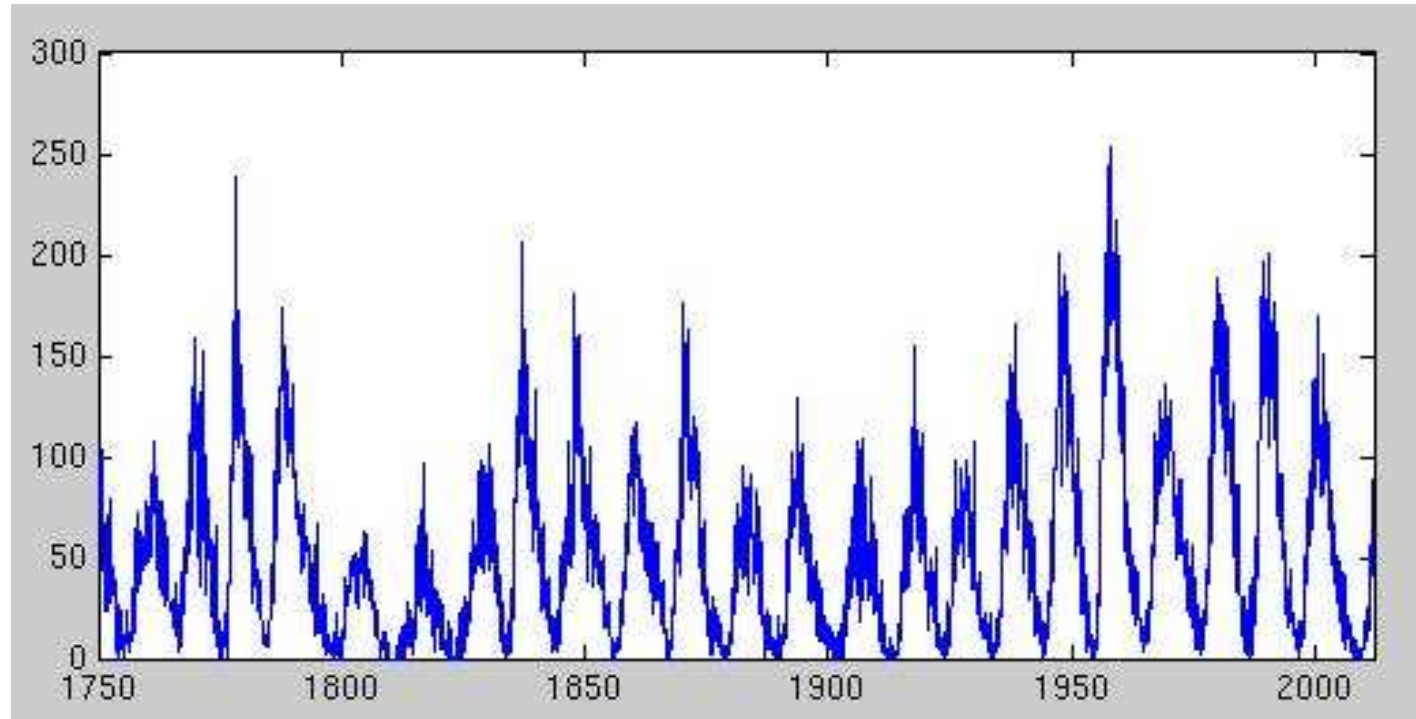
Sunspot Numbers

In 1610, shortly after viewing the sun with his new telescope, Galileo Galilei (or was it Thomas Harriot?) made the first European observations of Sunspots. Continuous daily observations were started at the Zurich Observatory in 1849 and earlier observations have been used to extend the records back to 1610. The sunspot number is calculated by first counting the number of sunspot groups and then the number of individual sunspots.

<http://solarscience.msfc.nasa.gov/SunspotCycle.shtml>

Motivácia

- Slné škvvrny - graf z dlhšieho obdobia:



Dáta zo stránky <http://solarscience.msfc.nasa.gov/SunspotCycle.shtml>

- Otázka: Ako určiť periódu?

Spektrum

- Postupnosť $\{\gamma_j\}_{j=-\infty}^{\infty} \rightarrow$ vytvárajúca funkcia

$$g(z) = \sum_{j=-\infty}^{\infty} \gamma_j z^j$$

- Stacionárny proces Y s autokovarianciami $\{\gamma_j\}_{j=-\infty}^{\infty} \rightarrow$ spektrum

$$s_Y(\omega) = \frac{1}{2\pi} g(e^{-i\omega}) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_j e^{-i\omega j},$$

kde i je imaginárna jednotka.

Vlastnosti spektra

- Spektrum $s_Y(\omega)$:
 - ◇ dá sa napísať v tvare:

$$s_Y(\omega) = \frac{1}{2\pi} \left[\gamma_0 + 2 \sum_{j=1}^{\infty} \gamma_j \cos(\omega j) \right]$$

- ◇ nadobúda reálne hodnoty
 - ◇ je párna funkcia
 - ◇ je periodická funkcia s periódou 2π
- stačí nám poznať hodnoty spektra medzi 0 a π
- Dá sa dokázať, že $s_Y(\omega) \geq 0$ [Fuller, 1976]

Spektrum a disperzia

- Výpočet autokovariancií zo spektra [Hamilton, Proposition 6.1]:

$$\gamma_k = \int_{-\pi}^{\pi} s_Y(\omega) e^{i\omega k} d\omega$$

- Pre $k = 0$ dostaneme: $\gamma_0 = \int_{-\pi}^{\pi} s_Y(\omega) d\omega$, a keď že spektrum je párna funkcia:

$$\gamma_0 = 2 \int_0^{\pi} s_Y(\omega) d\omega,$$

čiže disperzia γ_0 je dvojnásobkom plochy pod spektrom na intervale $[0, \pi]$

- Z priebehu spektra teda vidíme, ktoré frekvencie najviac prispievajú k disperzii procesu - tie, kde má spektrum vysokú funkčnú hodnotu

Odhad spektra - výberový periodogram

- Máme dáta y_1, \dots, y_T \rightarrow chceme z nich odhadnúť spektrum časového radu
- Prvá myšlienka: nahradíme autokovariancie v definícii spektra výberovými autokovarianciami, takto dostaneme - výberový periodogram:

$$\hat{s}_y(\omega) = \frac{1}{2\pi} \sum_{j=-T+1}^{T-1} \hat{\gamma}_j e^{-i\omega j} = \frac{1}{2\pi} \left[\hat{\gamma}_0 + 2 \sum_{j=1}^{T-1} \hat{\gamma}_j \cos(\omega j) \right]$$

- Nevýhody:
 - ◇ odhady majú veľkú disperziu
 - ◇ presnosť sa nezlepšuje so zvyšujúcim sa počtom dát (lebo odhadujeme stále viac autokovariancií) \rightarrow potrebujeme iný odhad spektra

Odhad spektra - - vý berový periodogram

- Nevýhody - podrobnejšie:
 - ◇ [Fuller, 1976]: pre veľký rozsah výberu má podiel $\frac{2\hat{s}_y(\omega)}{s_Y(\omega)}$ približne $\chi^2(2)$ rozdelenie a tieto podiely sú pre rôzne ω približne nezávislé
 - ◇ $E[\chi^2(2)] = 2$, a teda

$$E[\hat{s}_y(\omega)] \sim s_Y(\omega)$$

- to je OK

- ◇ Ale 95 percentný interval spoľahlivosti pre $\chi^2(2)$ je $(0.05, 7.4)$ a teda IS pre spektrum je

$$(0.025\hat{s}_y(\omega), 3.7\hat{s}_y(\omega))$$

- príliš široký

Odhad spektra - neparametrický odhad

- Idea: ak sú blízke hodnoty frekvencií, majú podobné hodnoty spektra → za odhad spektra pre danú frekvenciu zoberieme vážený priemer hodnôt výberového periodogramu \hat{s}_y z okolitých frekvencií:

$$(1) \quad \hat{s}_Y(\omega_j) = \sum_{m=-h}^h \kappa(\omega_{j+m}, \omega_j) \hat{s}_y(\omega_{j+m})$$

kde

- ◇ $\omega_j = 2\pi j/T$
- ◇ konštanta h udáva počet okolitých frekvencií, ktoré berieme do úvahy pri výpočte odhadu (nazýva sa **bandwidth**)
- ◇ funkcia κ určuje **váhy**, s ktorými tieto frekvencie vstupujú do odhadu (v súčte dávajú 1)

Odhad spektra - neparametrický odhad

- Iný prístup k na priemerovaniu:

$$\hat{s}_Y(\omega) = \int_{\omega-\nu}^{\omega+\nu} w(\omega, \lambda) \hat{s}_y(\omega) d\lambda,$$

kde $\int_{\omega-\nu}^{\omega+\nu} w(\omega, \lambda) d\lambda = 1$

- Zoberme obyčajný aritmetický priemer (nevážený), t.j. $w = 1/(2\nu)$:

$$\begin{aligned} \hat{s}_Y(\omega) &= \int_{\omega-\nu}^{\omega+\nu} \frac{1}{2\nu} \hat{s}_y(\omega) d\lambda \\ &= \frac{1}{2\pi} \left[\hat{\gamma}_0 + 2 \sum_{j=1}^{T-1} \hat{\gamma}_j \frac{\sin(\nu j)}{\nu j} \cos(\omega j) \right] \end{aligned}$$

Odhad spektra - neparametrický odhad

- Všeobecnejšie:

$$\hat{s}_Y(\omega) = \frac{1}{2\pi} \left[\hat{\gamma}_0 + 2 \sum_{j=1}^{T-1} \hat{\gamma}_j \kappa_j^* \cos(\omega j) \right]$$

- Funkcia $\{\kappa_j^*\}$ sa nazýva **kernel**:

- ◇ pre nevážený priemer $\kappa_j^* = \frac{\sin(\nu j)}{\nu j}$
- ◇ obyčajný výberový periodogram: $\kappa_j^* = 1$
- ◇ často sa používa **Bartlettov odhad**

Bartlettov odhad

- Kernel:

$$(2) \quad \kappa_j^* = \begin{cases} 1 - \frac{j}{q+1} & \text{pre } j = 1, \dots, q, \\ 0 & \text{pre } j > q. \end{cases}$$

čím dostaneme odhad

$$s_Y(\omega) = \frac{1}{2\pi} \left[\hat{\gamma}_0 + 2 \sum_{j=1}^q \left(1 - \frac{j}{q+1} \right) \hat{\gamma}_j \cos(\omega j) \right].$$

Bartlettov odhad

- Ako zvolit' parameter q vo vzt'ahu (2), resp. parameter h v (1)?

[Hamilton, str. 167]:

One practical guide is to plot an estimate of the spectrum using several different bandwidths and rely on subjective judgement to choose the bandwidth that produces the most plausible estimate.

Príklad 1 - simulované dáta

- Pripomeňme si AR(2) proces s periodickým charakterom:

$$x_t = 1.4x_{t-1} - 0.85x_{t-2} + u_t$$

Vieme:

- ◇ korelácie spĺňajú diferenčnú rovnicu

$$\rho(t) - 1.4\rho(t-1) + 0.85\rho(t-2) = 0,$$

ktorá má všeobecné riešenie

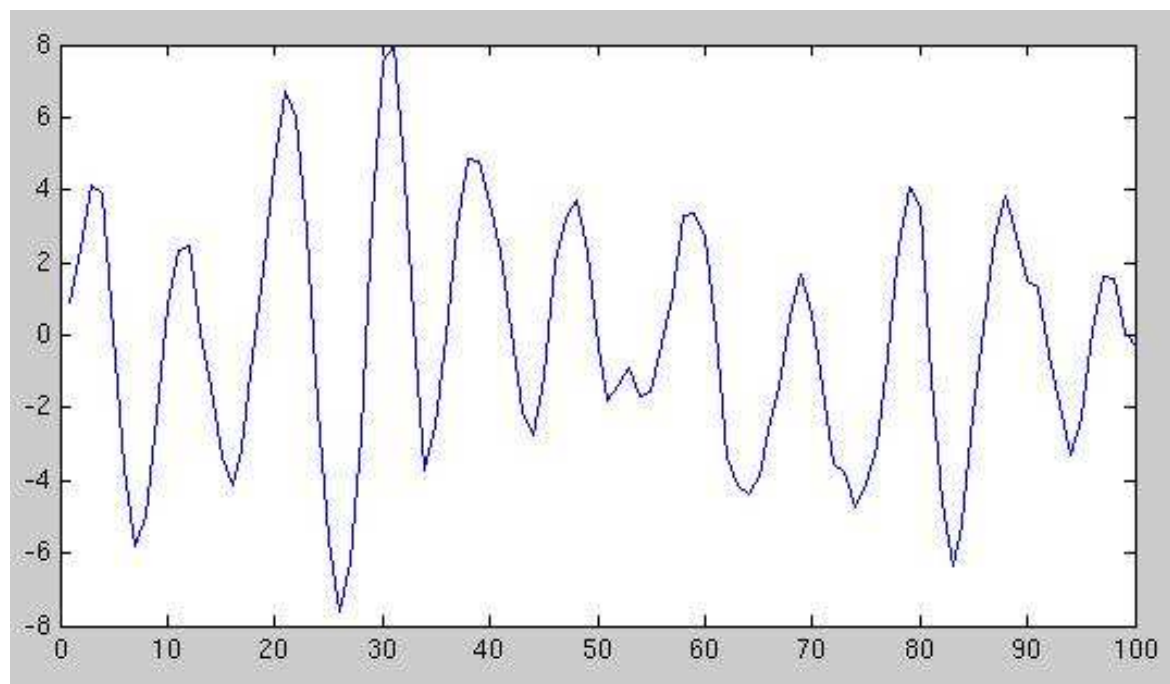
$$\rho(t) = 0.922^t (c_1 \cos(0.709t) + c_2 \sin(0.709t))$$

- ◇ sínus a kosínus vo všeobecnom riešení:

$$\cos(kt), \sin(kt) \rightarrow \text{perióda } \frac{2\pi}{k} = 8.862 \approx 9$$

Príklad 1 - simulované dáta

- Vygenerovaný priebeh:



Príklad 1 - simulované dáta

- Odhad spektra:

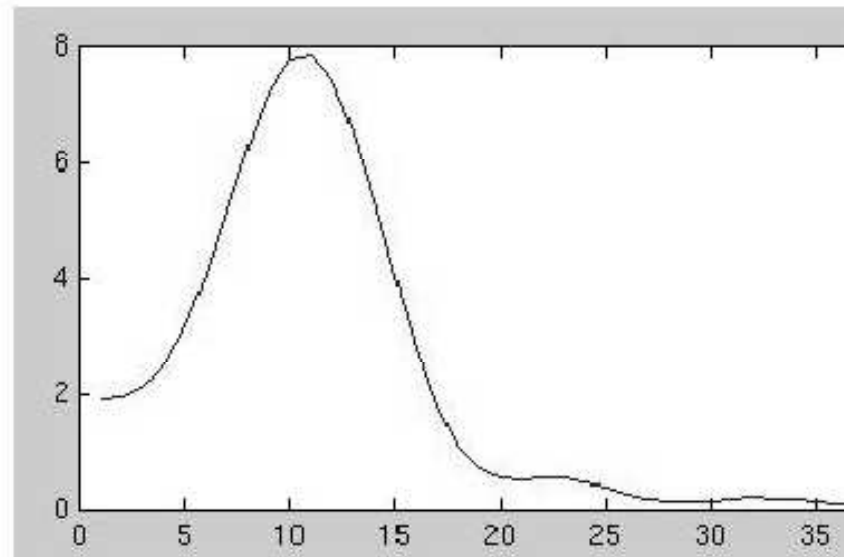
```
>> plot(sy)
>> [smax,imax]=max(sy)

smax =
    7.8658

imax =
    11

>> T/imax
ans =
    9.0909

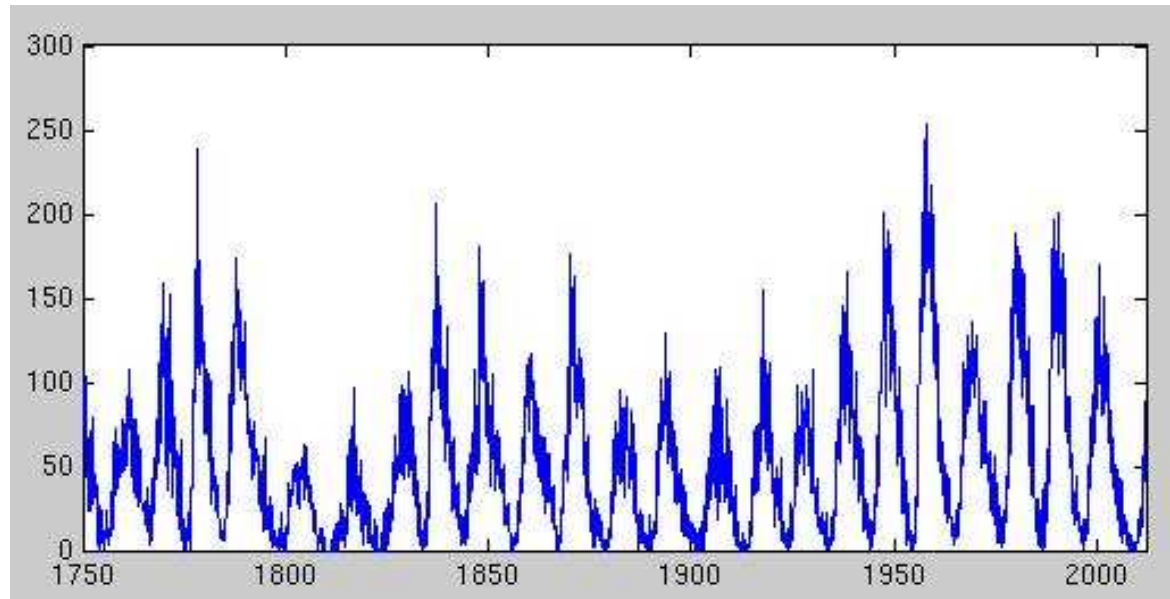
>>
```



- Hodnota i na x-ovej osi zodpovedá frekvencii $\omega_i = 2\pi i/T$, kde T je počet dát

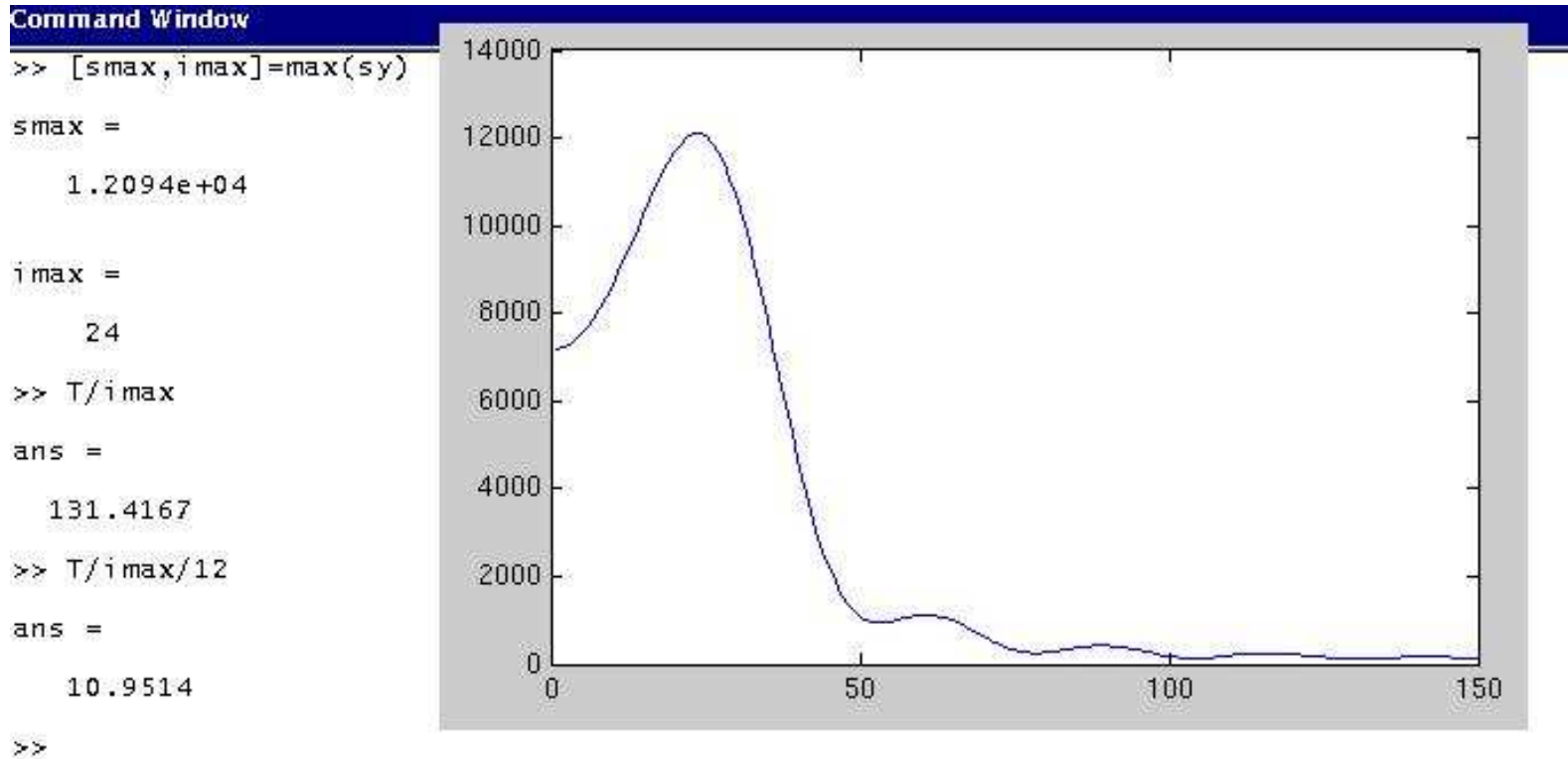
Príklad 2 - slnečné škvrny

- Priebeh dát:



Príklad 2 - slnečné škvrny

- Odhad spektra:



Príklad 2 - slnečné škvrny

- Známa perióda:

The "sunspot number" is then given by the sum of the number of individual sunspots and ten times the number of groups. Since most sunspot groups have, on average, about ten spots, this formula for counting sunspots gives reliable numbers even when the observing conditions are less than ideal and small spots are hard to see. Monthly averages (updated monthly) of the sunspot numbers ([181 kb JPEG image](#)), ([307 kb pdf-file](#)), ([62 kb text file](#)) show that the number of sunspots visible on the sun waxes and wanes with an approximate 11-year cycle.

<http://solarscience.msfc.nasa.gov/SunspotCycle.shtml>