Metody v klimatologii

III. Využití R v klimatologii

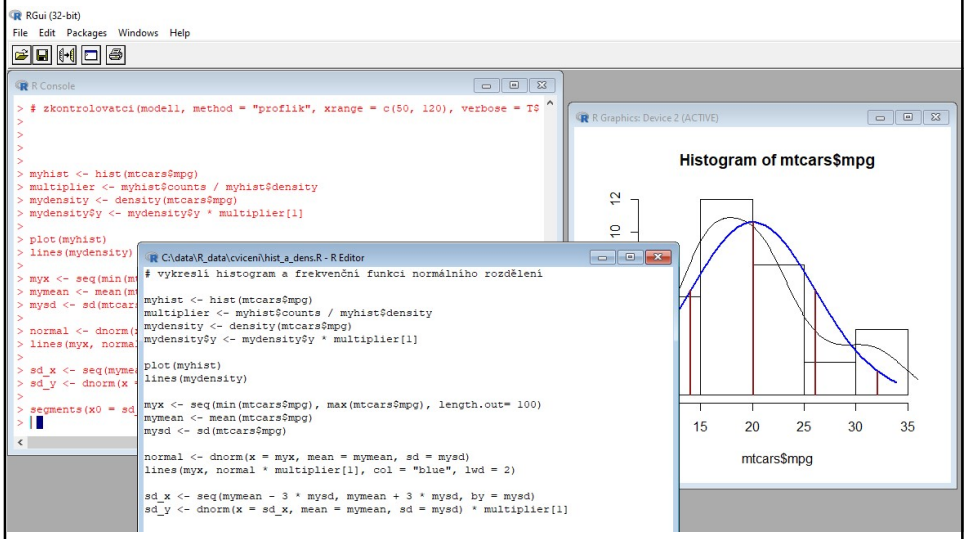
Úvod do R

https://cran.r-project.org/

- Interaktivní „výpočetní prostředí" pro statistické výpočty a grafiku
- Volně šiřitelný pod GNU GPL
- Nezávislý na platformě
- Řádkový interpret, objektově orientovaný
- Umožňuje jednoduché skriptování
- Snadno rozšiřitelný (> 2000 rozšiřujících knihoven na WWW)
- Nepřeberné množství výukových materiálů

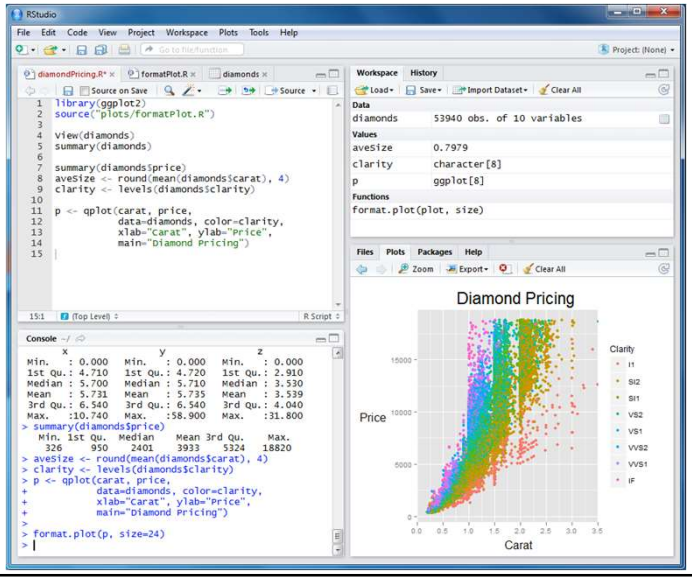
- Počáteční investice do ovládání jazyka R
- Nutnost často upgradovat R i některé knihovny

Úvod do R



Úvod do R

Grafické uživatelské rozhraní – R studio, Rcmd



R v klimatologii

1) Speciální knihovny (libraries)

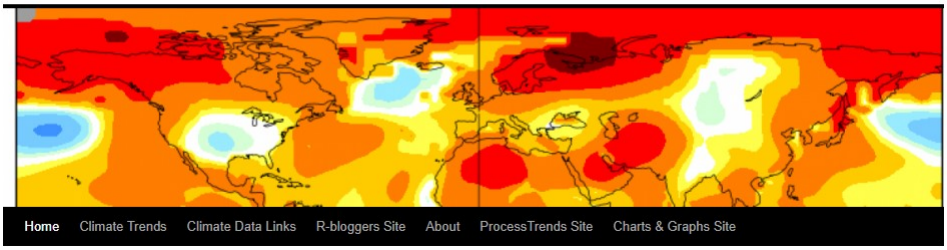
dplR, treeclim – dendroklimatologie
ncdf4 – analýza polí ve formátu NetCDF
climatol
…

2) www stránky

https://www.benjaminbell.co.uk/2018/01/getting-climate-data.html
https://rclimate.wordpress.com/

RClimate

Using R and Data to Understand Climate Change



Základy práce v R

S využitím dodaných skriptů obsahujících vybrané příkazy jazyka R a datových souborů si projděte základy práce v programovém prostředí R a zopakujte si základní úlohy popisné statistiky

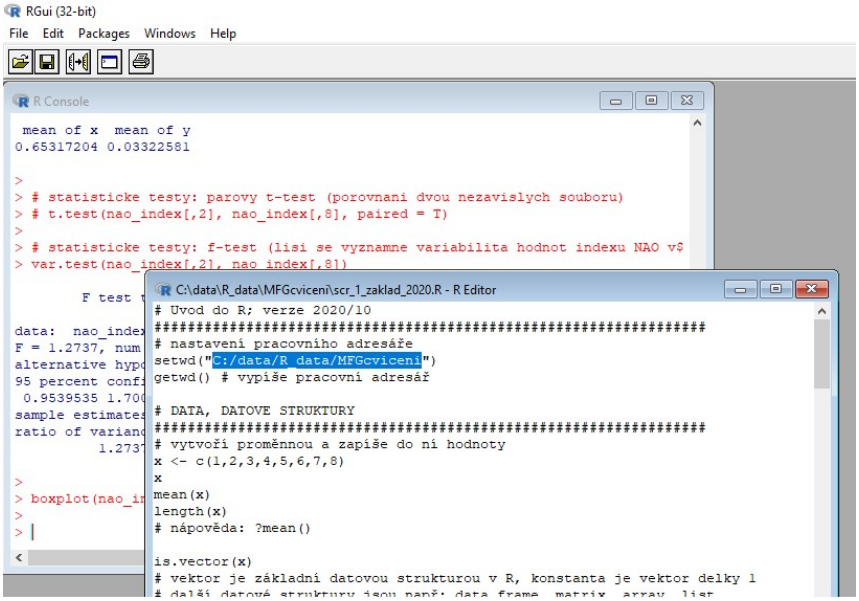
- Nainstalujte program R: https://cran.r-project.org/
- Vytvořte si pracovní adresář, např. C:/data/MFGcviceni
- **Do tohoto adresáře si z ISu zkopírujte všechny soubory, které najdete ve složce R_cviceni**
- Spusťte program R a otevřete si úvodní ukázkový skript:

  File -> Open script …

  scr_1_zaklad_2020.R
- Na 4. řádku upravte cestu k vašemu pracovnímu adresáři :

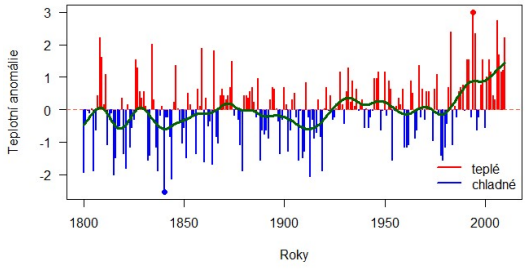
  setwd("C:/data/MFGcviceni")

Základy práce v R

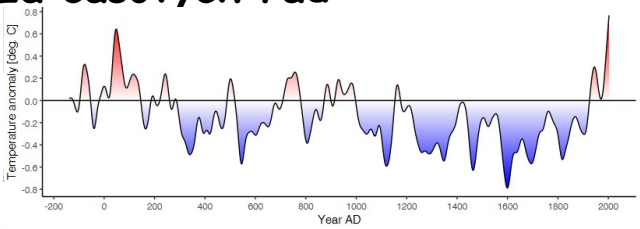


Základy práce v R

- Klikněte na začátek prvního řádku. Postupně tiskněte CTRL-R. Program bude vykonávat jednotlivé příkazy kódu
- Sledujte výstupy příkazů na konzole a později také v grafickém okně
- Příkazy můžete provádět opakovaně v závislosti, kde umístíte kurzor či jakou část kódu vyberete
- U jednotlivých příkazů můžete měnit nastavení parametrů či dodávat parametry další s pomocí helpu: ? nazev_funkce()
- Projdete-li všechny příkazy, můžete vyzkoušet script **scr_2_graf.R**, který vytvoří přiložený graf. Část jeho kódu využijete ve cvičení



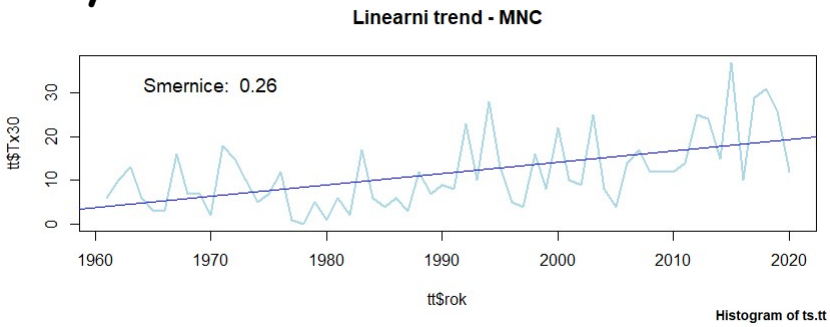
Analýza časových řad



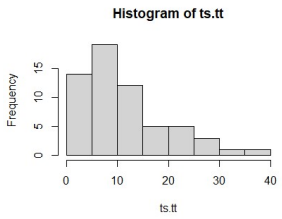
- Explorační (průzkumová) analýza s cílem otestovat vlastnosti časové řady (homogenizace, autokorelace, přítomnost odlehlých hodnot, normalita, homoskedascita, …)
- **Analýza trendové složky**
- **Analýza cyklického chování**

- Modelování časových řad (např. ARIMA modely),
- Analýza extrémních hodnot (viz **část IV**)
- Hledání změn v chování časové řady (change point detection) – nejen jako součást procesu homogenizace, ale též jako indikátor změny režimu chování (structural changes)
- Superposed Epoch Analysis, …

Analýza trendu

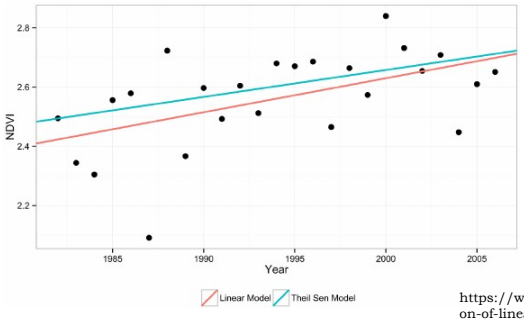


- Neměli bychom si zkontrolovat, zda je MNČ vůbec použitelná?
- Vypočtený trend nevystihuje dobře kolísání v celé délce časové řady



Neparametrický odhad lineárního trendu

Může existovat řada důvodů, proč je MNČ k odhadu lineárního trendu nevhodná



https://www.researchgate.net/figure/Comparison-of-linear-and-Theil-Sen-regression-slopes-for-a-pixel-in-the-Sahel_fig3_266705601

- Hledáme metody, která je využitelná z hlediska vlastností analyzovaných dat
- Nezávisí na rozdělení hodnot
- Není ovlivněna přítomností odlehlých hodnot,
- …

Neparametrický odhad lineárního trendu

- Nejprve testujeme, zda řada obsahuje významný trend
  Mann- Kendallův test
- Následně odhadujeme parametry přímky –
  Theill-Senova metoda odhadu parametrů

- M-K test je testem na přítomnost monotonního trendu v časové řadě
- Lze ho využít, pokud nejsou splněny podmínky pro použití MNČ
- Neparametrický odhad – nevyžaduje splnění předpokladu normality
- Není citlivý k odlehlým hodnotám
- Lze ho aplikovat i na časovou řadu s chybějícími hodnotami

- ☹ Vyžaduje, aby hodnoty členů časové řady byly nekorelované

Neparametrický odhad lineárního trendu

Mann-Kendallův test analyzuje rozdíly ve znaménkách mezi dřívějšími a pozdějšími datovými body. Pokud je v řadě přítomen trend, hodnoty znamének budou mít tendenci neustále růst resp. neustále klesat.

The MK test tests whether to reject the null hypothesis (H_0) and accept the alternative hypothesis (H_a), where

H_0: No monotonic trend

H_a: Monotonic trend is present

The initial assumption of the MK test is that the H_0 is true and that the data must be convincing beyond a reasonable doubt before H_0 is rejected and H_a is accepted.

The MK test is conducted as follows (from Gilbert 1987, pp. 209-213) :

1. List the data in the order in which they were collected over time,$x_1, x_2, \ldots, x_n$, which denote the measurements obtained at times $1, 2, \ldots, n$, respectively.

2. Determine the sign of all $n(n-1)/2$ possible differences $x_j - x_k$, where $j > k$. These differences are

$x_2 - x_1, x_3 - x_1, \ldots, x_n - x_1, x_3 - x_2, x_4 - x_2, \ldots, x_n - x_{n-2}, x_n - x_{n-1}$

3. Let $\operatorname{sgn}(x_j - x_k)$ be an indicator function that takes on the values 1, 0, or -1 according to the sign of $x_j - x_k$, that is,

$\operatorname{sgn}(x_j - x_k) = 1$ if $x_j - x_k > 0$

$= 0$ if $x_j - x_k = 0$, or if the sign of $x_j - x_k$ cannot be determined due to non-detects

$= -1$ if $x_j - x_k < 0$

4. Compute

$$S = \sum_{k-1}^{n-1} \sum_{j-k+1}^{n} \operatorname{sgn}(x_j - x_k) \qquad (1)$$

which is the number of positive differences minus the number of negative differences. If is a positive number, observations obtained later in time tend to be larger than observations made earlier. If is a negative number, then observations made later in time tend to be smaller than observations made earlier.

Neparametrický odhad lineárního trendu Mann-Kendallův test

5. If $n \leq 10$, follow the procedure described in Gilbert (1987, page 209, Section 16.4.1) by looking up S in a table of probabilities (Gilbert 1987, Table A18, page 272) . If this probability is less than α (the probability of concluding a trend exists when there is none), then reject the null hypothesis and conclude the trend exists. If n cannot be found in the table of probabilities (which can happen if there are tied data values), the next value farther from zero in the table is used. For example, if $S = 12$ and there is no value for $S = 12$ in the table, it is handled the same as $S = 13$.
If $n > 10$, continue with steps 6 through 10 to determine whether a trend exists. This follows the procedure described in Gilbert (1987, page 211, Section 16.4.2).
6. Compute the variance of S as follows:

$$\operatorname{VAR}(S) = \frac{1}{18}\left[n(n-1)(2n+5) - \sum_{p-1}^{g} t_p(t_p-1)(2t_p+5)\right]$$

where g is the number of tied groups and tp is the number of observations in the p th group.

7. Compute the MK test statistic, Z_{MK}, as follows:

$Z_{MK} = \frac{S-1}{\sqrt{VAR(S)}}$ if $S > 0$

$= 0$ if $S = 0$

$= \frac{S+1}{\sqrt{VAR(S)}}$ if $S < 0$

A positive (negative) value of Z_{MK} indicates that the data tend to increase (decrease) with time.

Neparametrický odhad lineárního trendu Mann-Kendallův test

8. Suppose we want to test the null hypothesis

H_0: No monotonic trend

versus the alternative hypothesis

H_a: Upward monotonic trend

at the Type I error rate α , where $0 < \alpha < 0.5$. (Note that α is the tolerable probability that the MK test will falsely reject the null hypothesis.) Then H_0 is rejected and H_a is accepted if $Z_{MK} \geq Z_{1-\alpha}$, where $Z_{1-\alpha}$ is the $100(1-\alpha)^{th}$ percentile of the standard normal distribution. These percentiles are provided in many statistics book (for example Gilbert 1987, Table A1, page 254) and statistical software packages.

9. To test H_0 above versus

H_a: Downward monotonic trend

at the Type I error rate α, H_0 is rejected and H_a is accepted if $Z_{MK} \leq -Z_{1-\alpha}$.

10. To test the H_0 above versus

H a : Upward or downward monotonic trend

at the Type I error rate α, H_0 is rejected and H_a is accepted if $|Z_{MK}| \geq Z_{1-\alpha/2}$, where the vertical bars denote absolute value.

Počítačové programy běžně poskytují p-hodnoty příslušející vypočtenému testovacímu kritériu, pode které lze výsledek testu jednoznačně interpretovat

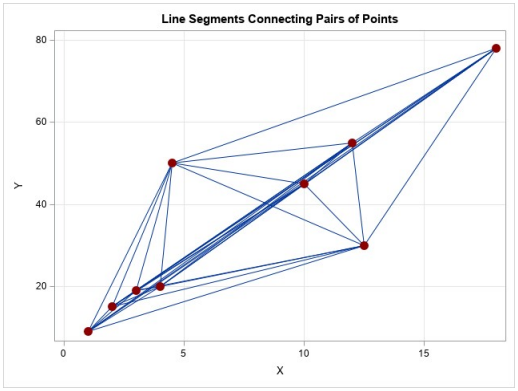
Mann-Kendallův test v R

```
# vyžaduje knihovnu
if(!require(Kendall)){install.packages("Kendall")}
library(Kendall)

MKD <- MannKendall(tt$Tx30)
summary(MKD)

     Score =  652 , Var(Score) = 24489.33
     denominator =  1738.724
     tau = 0.375, 2-sided pvalue =3.1829e-05
```

V naší v časové řadě se vyskytuje rostoucí statisticky významný monotónní trend. Jeho parametry zjistíme a trend vykreslíme **Theil-Senovou metodou**

Metoda odhadu parametrů lineárního trendu (Theil-Sen estimate)

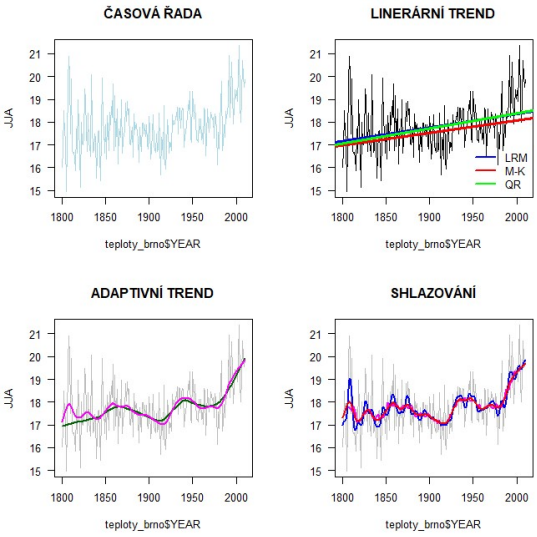
Opět se jedná o neparametrickou metodu



1) Spojíme úsečkami všechny body časové řady
2) Pro každou úsečku vypočteme její směrnici
3) Směrnici časové řady vypočteme jako **medián** všech směrnic jednotlivých úseček
4) Stejným způsobem vypočteme i absolutní člen jako druhý parametr přímky

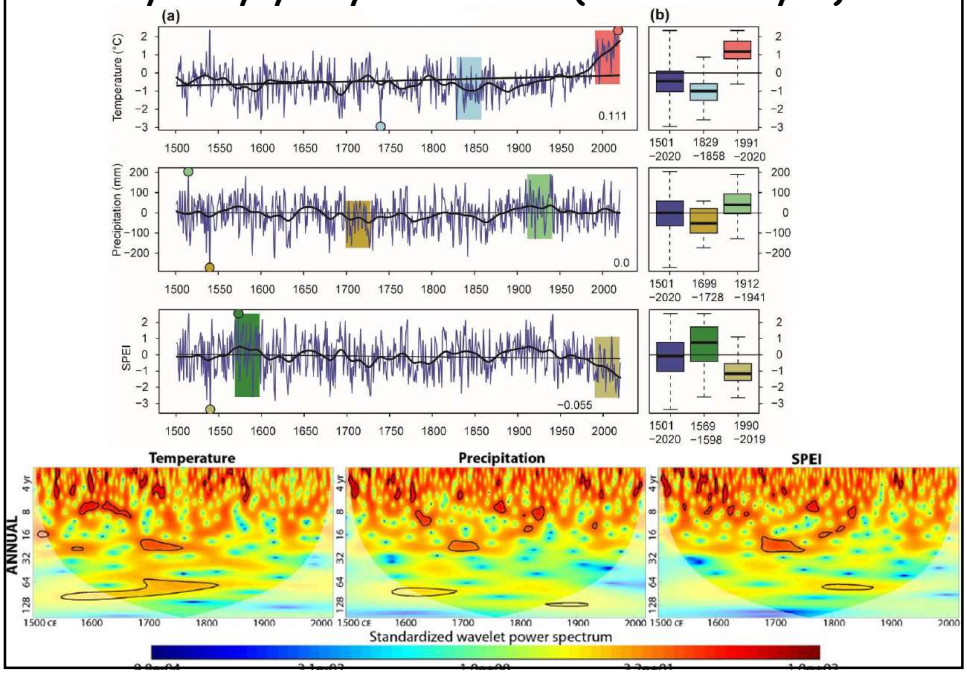
Různé metody odhadu trendu

- Běžný způsob odhadu lineárního trendu (LRM) může být ovlivněn výskytem odlehlých hodnot.
- V takovém případě je vhodnější odhad **lineárního** trendu neparametrickými metodami (např. M-K či QR)
- **Adaptivní** trend lze sestavit pomocí lokální regrese či pomocí splinových funkcí
- Krátkodobé tendence v časové řadě lze vystihnout jejím **shlazením** klouzavými průměry, Gaussovým filtrem či řadou dalších metod.

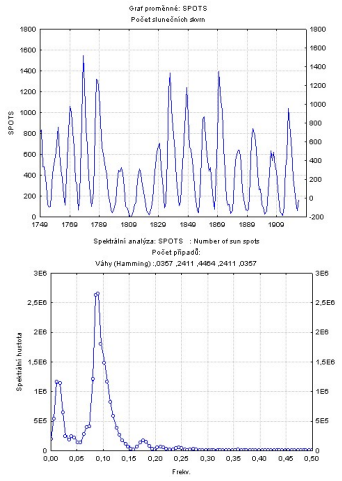


V programu R byl sestaven skript pro aplikaci různých metod analýzy časové řady (**scr_3_casove_rady.R**), vstupní data obsahuje soubor **brno_t.csv**

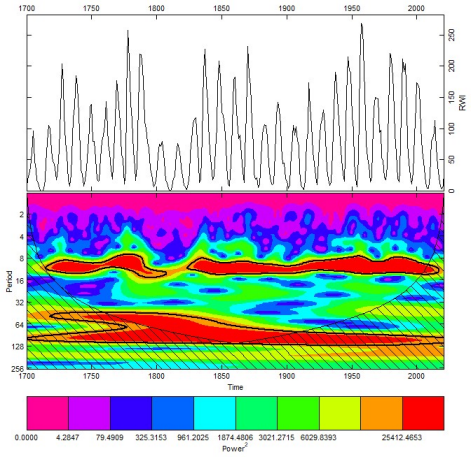
Příklady analýzy – cyklická složka (wavelet analysis)



Příklady analýzy – cyklická složka (wavelet analysis)



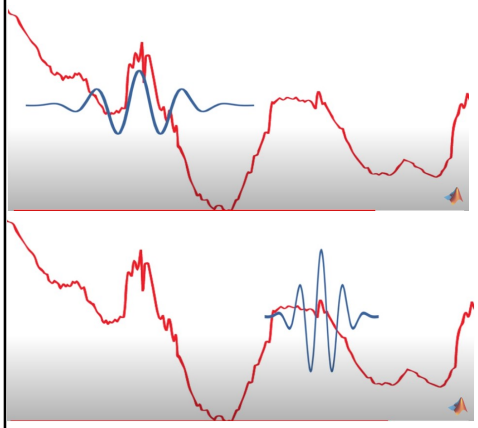
Odhad **spektrální hustoty** – předpokládá, že zastoupení cyklů v čase se nemění



Vlnková (wavelet) transformace

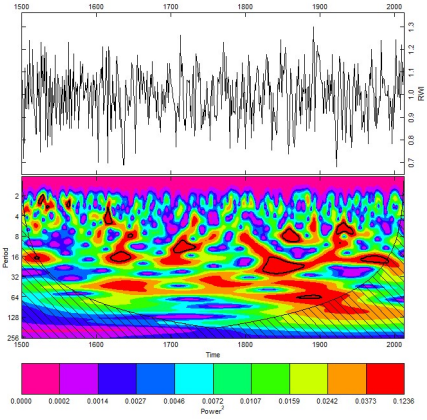
Příklady analýzy – cyklická složka (wavelet analysis)

„Mateřská" vlnka provádí po časové řadě dva druhy pohybu:



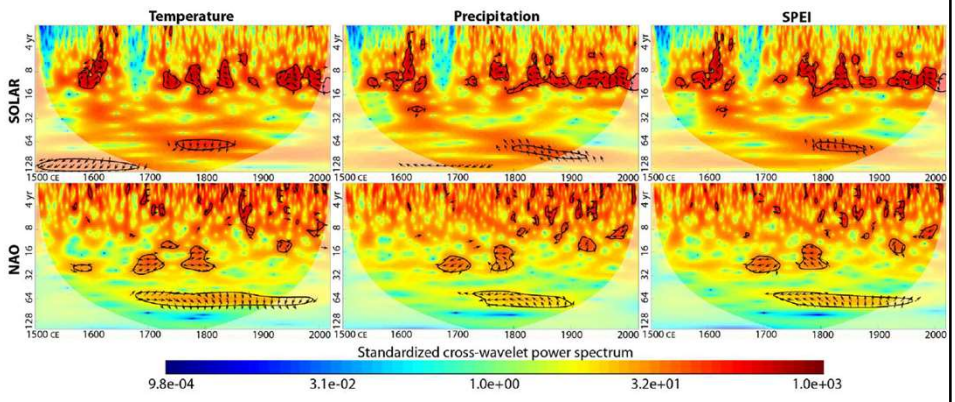
- posun
- škálování (roztažení a smrštění)

https://www.youtube.com/watch?v=QX1-xGVFqmw



Řada šířky letokruhů dubu, ve které se v některých obdobích objevuje statisticky významný cyklus délky 16 roků

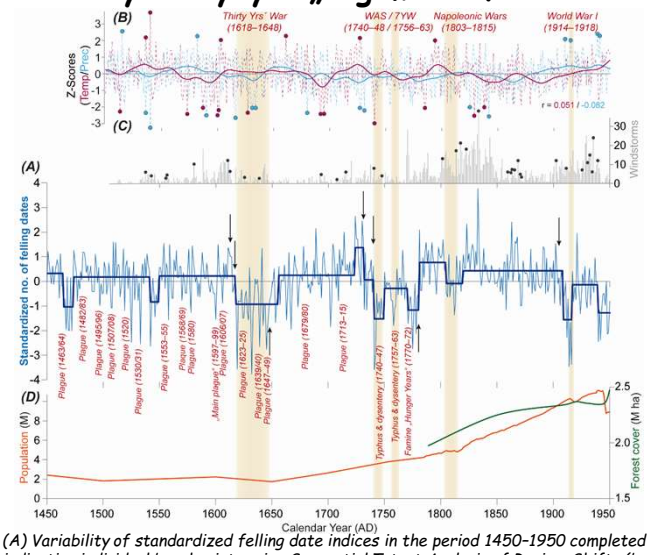
Příklady analýzy – cyklická složka (wavelet analysis)



„Cross-wavelet" spektrum mezi řadami teploty, srážek a SPEI a řadami vysvětlujících proměnných (sluneční aktivita a NAO index).

Šipky indikují fázový posun mezi dvěma řadami (ukazuje-li šipka vpřed, jsou obě řady ve fázi)

Příklady analýzy – „regime shift detection"



Kolář et al. (2022) Effects of social and climatic factors on building activity in the Czech lands between 1450 and 1950

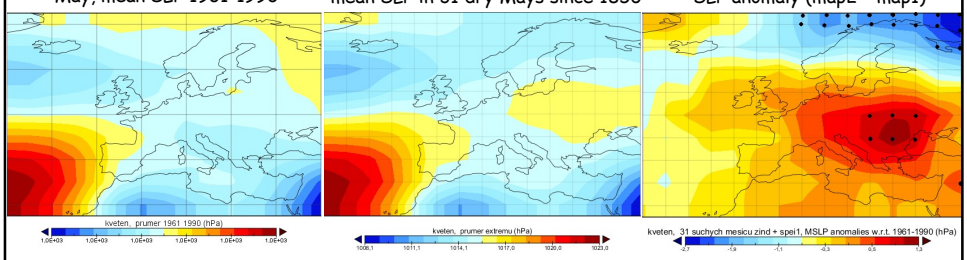
(A) Variability of standardized felling date indices in the period 1450–1950 completed with the model of regime shifts indicating individual breakpoints using Sequential T-test Analysis of Regime Shifts (bold line) and Pettitt test (arrows). The most serious wars, epidemics and famines are indicated. (B) Reconstruction of annual temperature (Dobrovolný et al. 2010) and precipitation (Dobrovolný et al. 2015) anomalies (w.r.t. 1961–1990) smoothed by Gaussian low pass filter and Pearson´s correlation coefficients between these climate parameters and standardized number of felling dates for the common period 1501–1950. (C) Number of recorded local windstorms with highlighted widespread windstorms (black dots) based on Brázdil et al. (2004). (D) Evolution of population and forest cover in the Czech lands (CSO). Note: WAS = War of the Austrian Succession, 7YW = Seven Years' War.

Příklad analýzy prostorové informace

Jaký je rozdíl mezi tlakovým polem v měsících, kdy je ve střední Evropě výrazné sucho (obrázek vlevo), v porovnání s normálními podmínkami (průměrné tlakové pole daného měsíce v období 1961-1990 – obrázek uprostřed).

Vpravo je výsledek v podobě rozdílu (normál – sucho). Vyznačeny jsou gridové body, ve kterých je rozdíl statisticky významný

May, mean SLP 1961–1990 | mean SLP in 31 dry Mays since 1850 | SLP anomaly (map2 – map1)



SLP at extremely dry Mays defined in the 1850–2010 period and SLP differences in extremely dry months compared to MSLP of the 1961–1990 reference period; black points mark statistically significant (a=0.05) SLP decrease/incerese w.r.t the 1961–1990 reference period

V programu R byl sestaven skript pro vizualizaci sestavených polí (**scr_4_slp_mapy.R**), vstupní data obsahuje soubor **au_sucho.csv**