

3. Metody analýzy časových řad v klimatologii

3.1 Periodicita a cykličnost

Klima je vyjádřeno různými prvky (např. teplota vzduchu, srážky, indexy), kolísajícími v prostoru a čase:

$$\{a, b, c, \dots\} = f(x, y, z, t),$$

a, b, c, ... jsou prvky klimatu, x, y, z jsou prostorové souřadnice a t je čas.

Při studiu časových změn platí pro dané místo

$$\{a, b, c, \dots\} = f(t).$$

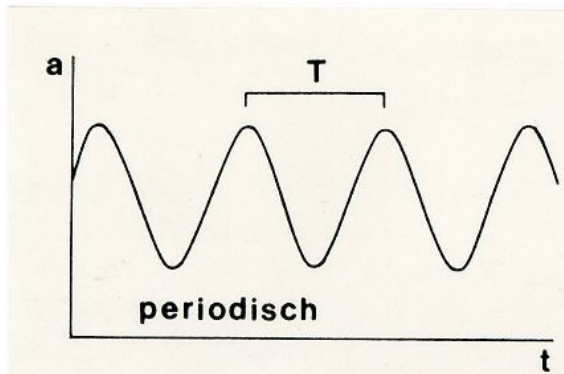
Informace o časových změnách je získávána v podobě časové řady

$$\{a_i, b_i, c_i, \dots\} = f(t_i) \text{ s } t_{i+1} - t_i = \Delta t = \text{konst.}, i = 1, 2, \dots, n$$

Δt je časový interval (např. den, měsíc, rok) a n je délka řady.

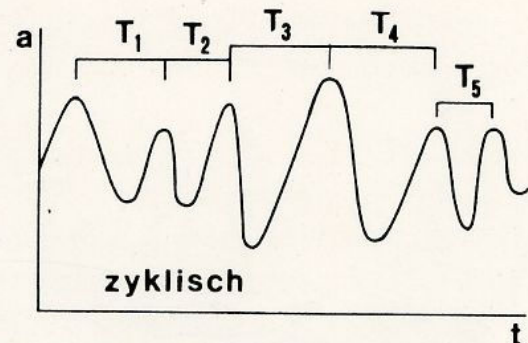
Typy kolísání v časových řadách

Periodická: s periodou T



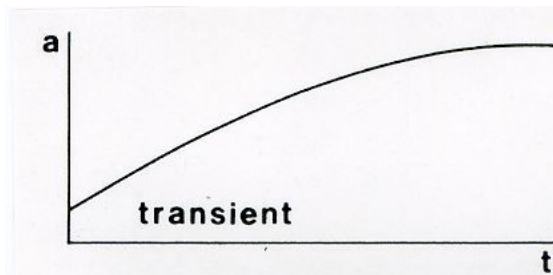
$$a(t_i) = a(t_i + T)$$

Cyklická (kvaziperiodická, rytmická): s periodou T

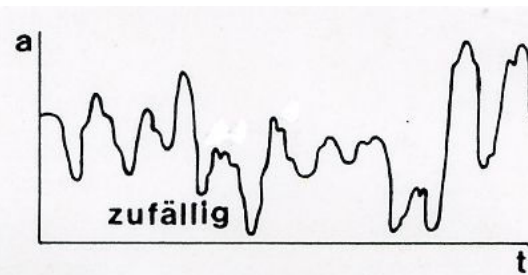


$$a(t_i) \approx a(t_i + \bar{T})$$

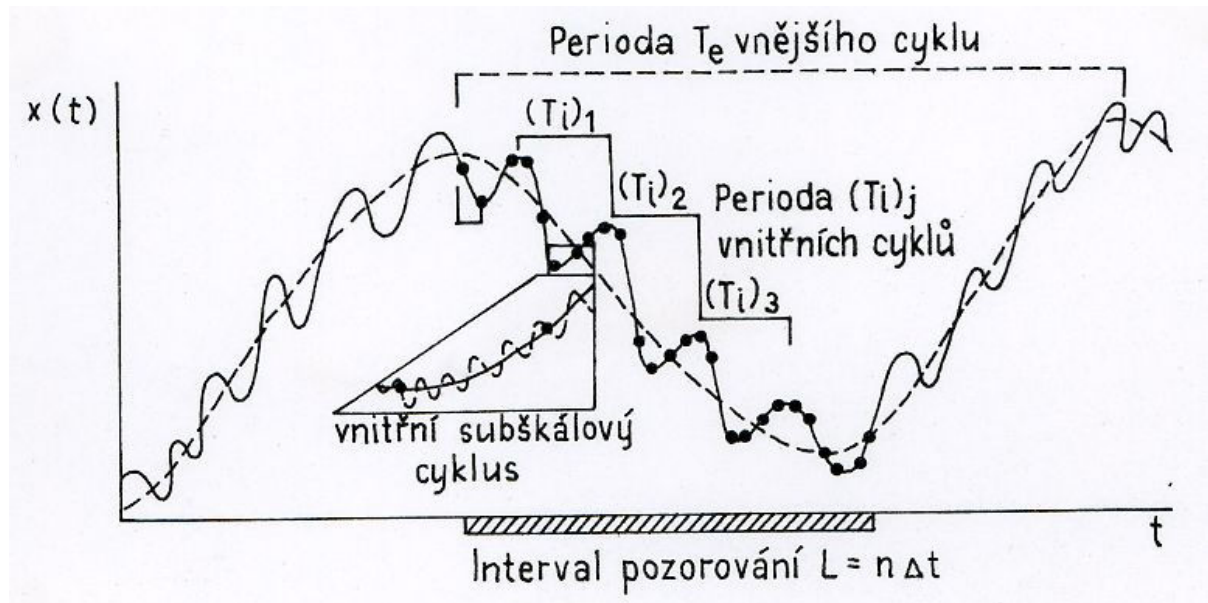
Transientní: lineární nebo nelineární trend



Náhodná (stochastická)



Perioda T ← Cyklus T → Náhodnost



Cyklus s periodou T_p : $\Delta t \ll T_p \ll n \Delta t$

Vnitřní cykly: $T_p < \Delta t$

Vnější cykly: $T_p > n \Delta t$

Podmínky aplikovatelnosti statistických metod

- **nezávislost** hodnot časové řady
- **stacionarita** řady, tj. statistické charakteristiky řady S_j ($j = 1, 2, \dots, J$) jsou časově nezávislé, tedy $\partial S_j / \partial t = \text{konst.}$
- **ergodicita**, tj. průměr dané řady (výběrový soubor) s jejím prodlužováním v časovém intervalu L se blíží střední hodnotě celé řady (základního souboru)

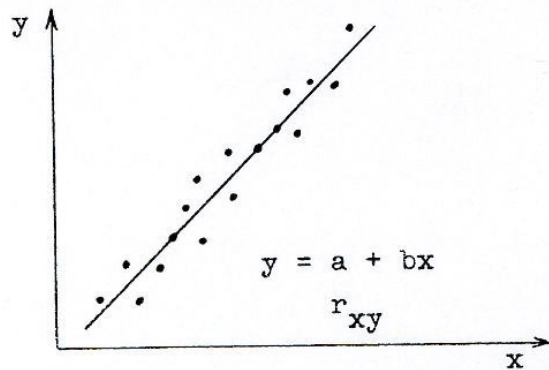
Nesplnění podmínek: výsledky analýzy mají platnost pouze pro danou realizaci časové řady (tj. daný výběr), je třeba testovat výsledky se zřetelem na vlastnosti základního souboru.

Nestacionarita – přirozená vlastnost klimatických řad (změny a kolísání klimatu).

3.2 Korelační analýza

Korelační koeficient

Podmínka použití: hodnoty obou řad jsou nezávislé a mají přibližně normální rozdělení.



$$r_{xy} = \frac{1}{n-1} \frac{\sum_{i=1}^n x'_i \cdot y'_i}{s_x \cdot s_y}$$

$$x'_i = x_i - \bar{x}, \quad y'_i = y_i - \bar{y}$$

s_x, s_y - standard deviation

r_{xy}^2 = explained variance / total variance

Test criterion of significance of r_{xy} :

$$t = \frac{r_{xy} \sqrt{n-2}}{\sqrt{1-r_{xy}^2}} \quad \text{with t-distribution and } \nu = n-2.$$

\hat{r}_{xy} -transformation (Fisher) for non-normally distributed series:

$$\hat{r}_{xy} = \frac{1}{2} \ln \frac{1+r_{xy}}{1-r_{xy}}$$

$$u_p = (\hat{r}_{xy} - z/\sqrt{n-3}; \hat{r}_{xy} + z/\sqrt{n-3})$$

$$z = \hat{r}_{xy} \sqrt{n-3}$$

Autokorelační koeficient

Autokovariance

$$V_x(\tau) = \frac{1}{n-\tau-1} \sum_{i=1}^{n-\tau} x'_i \cdot x'_{i+\tau}$$

Autokorelace

$$r_x(\tau) = V_x(\tau) / s_1 \cdot s_2$$

$\tau = 0, 1, 2, \dots, M < n/2$ - time lag, M - maximum lag

A diagram of the computation:

I.	x_1	x_2	x_3	x_4	x_5	x_6	$\tau = 0 \cdot \Delta t$	$r_x(0) = 1$
II.	x_2	x_3	x_4	x_5	x_6	x_7	$\tau = 1 \cdot \Delta t$	$r_x(1)$
III.	x_3	x_4	x_5	x_6	x_7	x_8	$\tau = 2 \cdot \Delta t$	$r_x(2)$
IV.	x_4	x_5	x_6	x_7	x_8	x_9	$\tau = 3 \cdot \Delta t$	$r_x(3)$ etc.

Confidence levels of r_{xy} by Andersson:

$$r_x(\tau) = \frac{-1 \pm z(p) \sqrt{n - \tau - 2}}{n - \tau - 1}$$

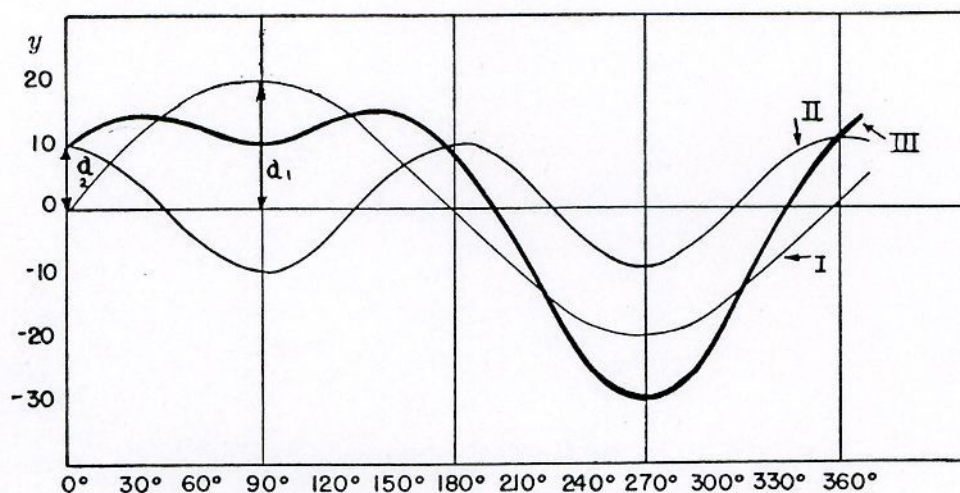
Křížově-korelační koeficient

series X: $x_1 \ x_2 \ x_3 \ x_4 \ x_5 \ x_6 \ x_7 \ \dots$

series Y: $y_1 \ y_2 \ y_3 \ y_4 \ y_5 \ y_6 \ y_7 \ \dots$ with possible shift τ

$$r_{Cxy} = \frac{\sum_{i=1}^{n-\tau} x'_i \cdot y'_{i+\tau}}{\sqrt{\sum_{i=1+\tau}^n x_i'^2 \sum_{i=1}^{n-\tau} y_i'^2}} \quad -1 \leq r_{Cxy} \leq 1$$

3.3 Harmonická analýza



Example of superposition of two waves.

- I: $y_1 = a_1 \sin(x + A_1), \quad A_1 = 0^\circ;$
 II: $y_2 = a_2 \sin(2x + A_2), \quad A_2 = 90^\circ;$
 III: $y = y_1 + y_2.$

Řada spojitá, nekonečná a periodická – superpozice konečného počtu sinových a kosinových funkcí – Fourierova analýza – Fourierova řada:

$$x(t) = A_0/2 + \sum_{i=1}^n A_i \sin \omega t + \sum_{i=1}^n B_i \cos \omega t, \quad \omega = 2\pi/T = 2\pi f$$

$$A_i = \frac{2}{T} \int_0^T x(t) \sin i\omega t dt \quad B_i = \frac{2}{T} \int_0^T x(t) \cos i\omega t dt$$

Pro klimatologické řady, nesplňující uvedené podmínky, lze nahradit Fourierovy koeficienty Besselovou formulí, aplikovatelnou pro přibližně periodickou řadu:

$$x_j(t_j) = \bar{x} + \sum_{i=1}^{N/2} \left[A_i \sin\left(\frac{360^\circ}{P} it\right) + B_i \cos\left(\frac{360^\circ}{P} it\right) \right]$$

$$A_i = \frac{2}{N} \sum_{j=1}^N x_j \sin\left(\frac{360^\circ}{P} it_j\right) \quad i = 1, 2, \dots, N/2 - 1$$

$$B_i = \frac{2}{N} \sum_{j=1}^N x_j \cos\left(\frac{360^\circ}{P} it_j\right) \quad i = 1, 2, \dots, N/2$$

Počet harmonik $h_i = P/i$ je $N/2$, $h_{\min} = 2\Delta t$, $h_{\max} = P$.

$1/P_{\min}$ = Nyquist's frequency

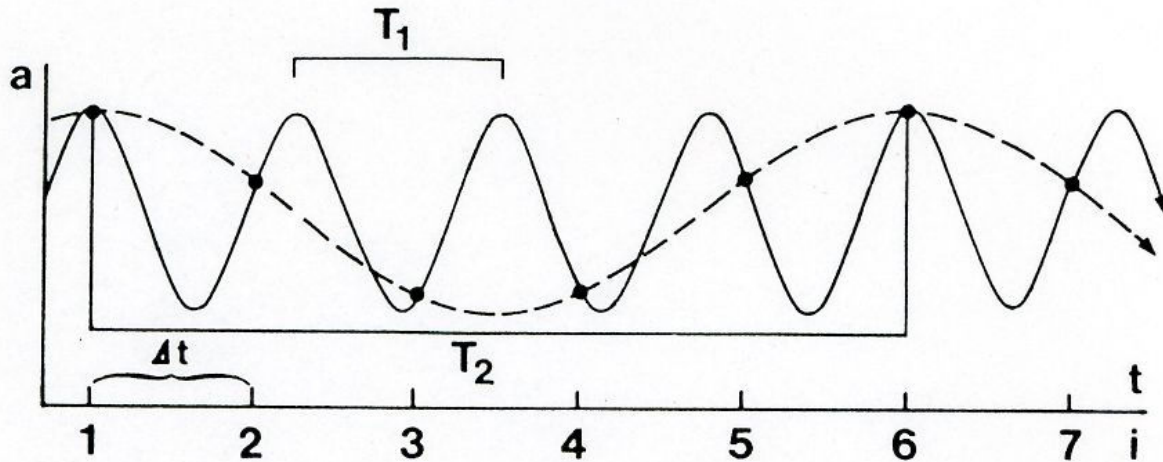
Explained variance through h_i :

for $i < N/2$ $C_i^2/2s^2$ with $C_i = \sqrt{A_i^2 + B_i^2}$

for $i = N/2$ $C_{N/2}^2/s^2$ with $C_{N/2} = \sqrt{B_{N/2}^2}$

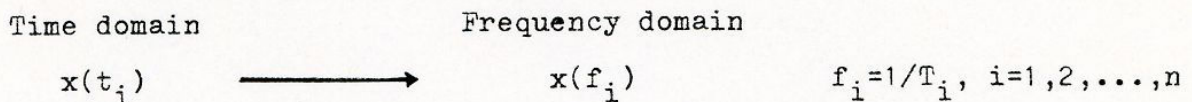
Time of maximum of h_i :

$$t_i = \frac{P}{360^\circ i} \arctan(A_i/B_i)$$



Falešný efekt (aliasing effect) – skutečná kolísání s periodou T_1 ($5/4\Delta T$) mohou být se zřetelem na interval měření Δt chybně interpretována periodou T_2 ($5\Delta T$), kde $T_2 > T_1$

3.4 Spektrální analýza



Spectral analysis according to Blackman and Tukey:

Raw spectral estimates of $S_x(h)$:

$$S_x(0) = 1/2M \left[v_x^2(0) + \sum_{k=1}^{M-1} D(k) v_x^2(k) \right] \quad \text{for } h = 0$$

$$S_x(h) = 1/2M \left[v_x^2(0) + \sum_{k=1}^{M-1} D(k) v_x^2(k) \cos \frac{\pi h k}{M} \right] \quad \text{for } 0 < h < M$$

$$S_x(M) = 1/2M \left[v_x^2(0) + \sum_{k=1}^{M-1} D(k) v_x^2(k) (-1)^k \right] \quad \text{for } h = M.$$

$$k = 0, 1, 2, \dots, M; \quad D(k) = \begin{cases} 1/2(1 + \cos \pi k/M) & \text{for } 0 < k < M \\ 0 & \text{for } -\infty < k < 0 \text{ and } M < k < +\infty \end{cases}$$

Smoothed spectral estimates:

Hamming window:

$$S_x(0) = 0.54S_x(0) + 0.46S_x(1)$$

$$S_x(h) = 0.23S_x(h-1) + 0.54S_x(h) + 0.23S_x(h+1) \text{ for } 0 < h < M$$

$$S_x(M) = 0.54S_x(M) + 0.46S_x(M-1)$$

Hanning window:

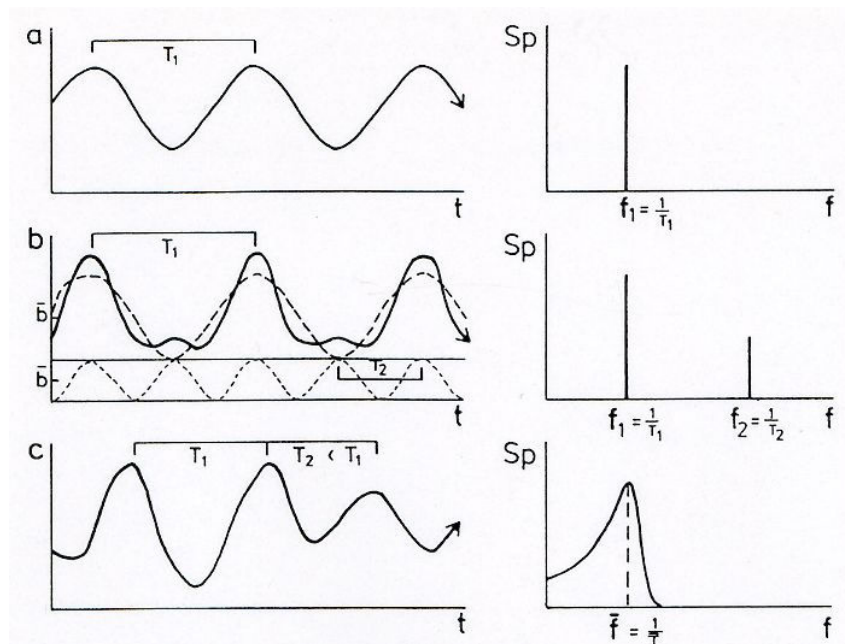
$$S_x(0) = 1/2 (S_x(0) + S_x(1))$$

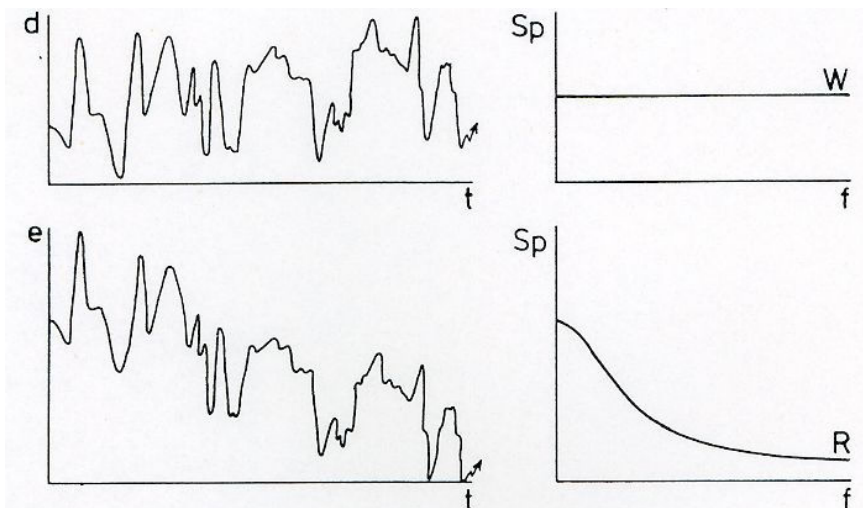
$$S_x(h) = 1/4 (S_x(h-1) + 2S_x(h) + S_x(h+1)) \text{ for } 0 < h < M$$

$$S_x(M) = 1/2 (S_x(M-1) + S_x(M))$$

Interpretace spektra

- Periodická funkce s periodou T_1
- Periodická funkce s periodami T_1 a T_2
- Cyklická kolísání se střední periodou T
- Náhodná čísla (tzv. bílý šum nebo bílé spektrum)
- Náhodná čísla s trendem (tzv. rudý šum nebo rudé spektrum)





Markov's time series ($r_x(1)$ differs significantly from zero and further $r_x(2) \approx [r_x(1)]^2$, $r_x(3) \approx [r_x(1)]^3$, etc. has a red spectrum:

$$RN(h) = \frac{1 - [r_x(1)]^2}{1 + [r_x(1)]^2 - 2r_x(1) \cos \frac{h\pi}{M}} \cdot WN$$

where white spectrum WN

$$WN = \frac{1}{M} \left[(1/2 S_x(0) + S_x(M)) + \sum_{h=1}^{M-1} S_x(h) \right]$$

Confidence limits:

$$G_\alpha = RN(h) \frac{\chi_{\nu, \alpha}^2}{\nu} \quad \nu = \frac{2n - M/2}{M}$$

Schéma spektrální analýzy maximální entropie (Maximum Entropy Spectral Analysis – MESA)

Ausgangsreihe
 $x(j\Delta t) = x_j$

ME-Analyse
 datenadaptive Filterung

$$\sum_{k=0}^M a_M(k) x_{j-k} = x_{jF}$$

mit $a_M(0) = 1$

Zufallszahlenreihe
 $x_{jF} = z_j$ mit $E(Z_j) = 0$
 und $E(Z_j Z_{j-k}) = \delta_{k,0} Q_M$

Sukzessive Berechnung der Autoregressionskoeffizienten $a_M(M)$
 nach BURG mit $M = 1$ beginnend aus

$$a_M(M) = -2 \frac{\sum_{k=1}^{N-M} y_{kM} y'_{kM}}{\sum_{k=1}^{N-M} (y_{kM}^2 + y'_{kM}{}^2)}$$

mit $y_{kM} = x_k + \sum_{j=1}^{M-1} a_{M-1}(j) x_{k+j}$, wobei $y_{k1} = x_k$,

und $y'_{kM} = x_{k+M} + \sum_{j=1}^{M-1} a_{M-1}(j) x_{k+M-j}$, wobei $y'_{k1} = x_{k+1}$,

sowie der $a_M(k)$ aus der Rekursionsgleichung

$$a_M(k) = a_{M-1}(k) + a_M(M) a_{M-1}(M-k) \quad \text{für } k = 1, 2, \dots, M-1.$$

Q_M ergibt sich aus der Rekursionsgleichung

$$Q_M = Q_{M-1} (1 - a_{M-1}^2(M-1)) \quad \text{mit } Q_0 = \frac{1}{N} \sum_{k=1}^N x_k^2.$$

Zeitbereich

 Frequenzbereich

$$D(f) = 1 + \sum_{k=1}^M a_M(k) e^{i2\pi f k}$$

Filterdurchlaß-

Charakteristik -----

Energie-Spektrum $P(f)$
 der Ausgangsreihe

$$|D(f)|^2 P(f) = P_F(f)$$

Energie-Spektrum der

Zufallszahlenreihe

$$P_F(f) = \text{const} = Q_M$$

Maximum-Entropie-Spektrum

$$P(f) = Q_M |D(f)|^{-2} = \frac{Q_M}{\left| 1 + \sum_{k=1}^M a_M(k) e^{i2\pi f k} \right|^2}$$

3.5 Křížově-spektrální a koherenční analýza

Cross-spectrum function: $P_{xy}(h) = C_{xy}(h) + i Q_{xy}(h)$

Real part of cross-spectrum $C_{xy}(h)$ (Kospektrum, Wirksamspektrum):

$$C_{xy}(h) = \Delta t / \pi \sum_{p=0}^M \epsilon_p (R_{xy}(p) + R_{yx}(p)) \cos hp\pi/M$$

$$\text{with } \epsilon_p = \begin{cases} 1/2 & \text{for } p = 0, M \\ 1 & \text{for } 0 < p < M \end{cases}$$

Imaginary part of cross-spectrum $Q_{xy}(h)$ (Quadraturspektrum, Blindspektrum):

$$Q_{xy}(h) = \Delta t / \pi \sum_{p=0}^M \epsilon_p (R_{xy}(p) - R_{yx}(p)) \sin hp\pi/M$$

$$\text{with } \epsilon_p = \begin{cases} 1/2 & \text{for } p = 0, M \\ 1 & \text{for } 0 < p < M \end{cases}$$

$C_{xy}(h)$ vyjadřuje podíl oscilací různých frekvencí na celkové kovarianci při nulovém posunu.

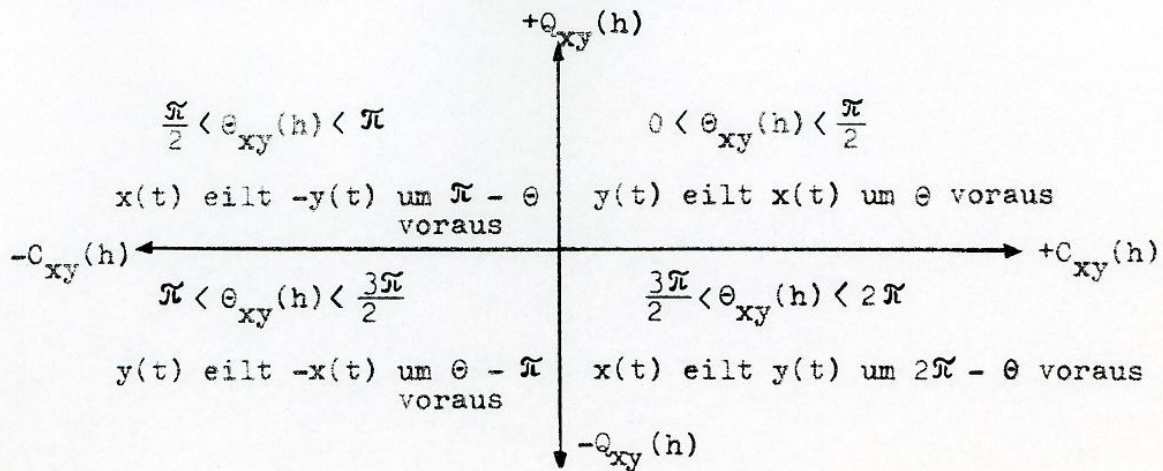
$Q_{xy}(h)$ vyjadřuje podíl různých harmonik na celkové kovarianci (harmoniky řady x jsou posunuty o $1/4$ periody zpět a řada y je beze změny).

Other possible expression of the cross-spectrum function:

$$P_{xy}(h) = |A_{xy}(h)| e^{-i\theta_{xy}(h)}$$

Amplitude: $A_{xy}(h) = \sqrt{SC_{xy}^2(h) + SQ_{xy}^2(h)} \quad h=0,1,\dots,M$

Phase angle: $\theta_{xy}(h) = \tan^{-1}(SQ_{xy}(h)/SC_{xy}(h)) \quad h=2,3,\dots,M-2$



Coherency:

$$K_{xy}^2(h) = \frac{SC_{xy}^2(h) + SQ_{xy}^2(h)}{SP_x(h) + SP_y(h)} \quad h = 1, 2, \dots, M$$

$$0 \leq K_{xy}^2(h) \leq 1$$

Confidence level:

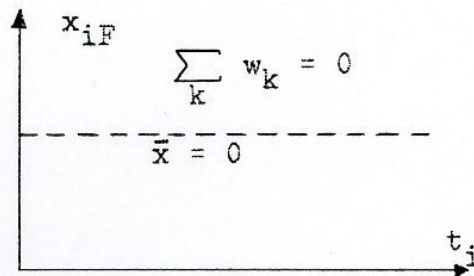
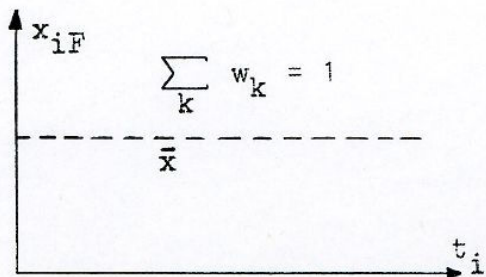
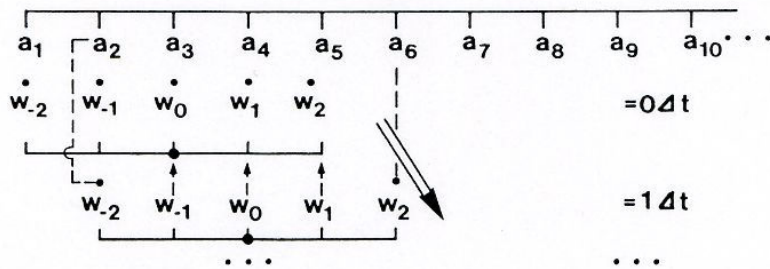
$$\beta = \sqrt{1 - \alpha^{\frac{1}{\nu-1}}} \quad \nu = \frac{2n - M/2}{M}$$

3.6. Filtrace

Schéma filtrace časové řady

$$x_{iF} = \sum_{k=-n}^m w_k \cdot x_{i+k}$$

$w_k = w_{-n}, w_{-n+1}, \dots, w_{-1}, w_0, w_1, \dots, w_{m-1}, w_m$ - filter weights



$n = m, w_k = w_{-k}, k = 0, 1, 2, \dots$ - symmetrical filter weights

Response function:

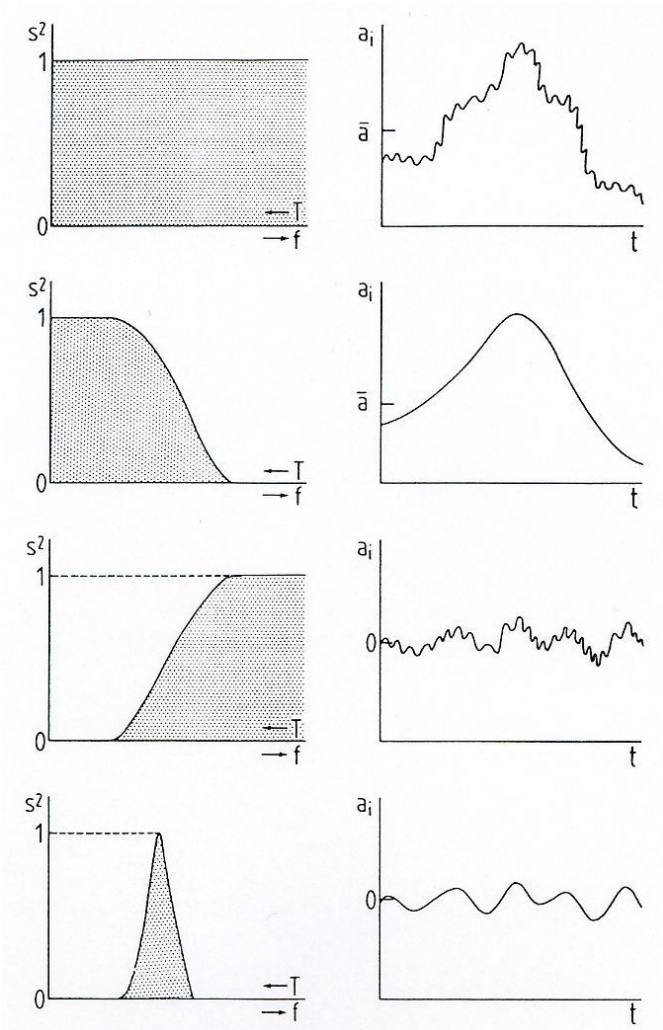
$$R(f) = w_0 + 2 \sum_{k=1}^n w_k \cos 2\pi f k \quad 0 \leq R(f) \leq 1$$

$R(f) = 0$ amplitude completely suppressed

$R(f) = 1$ amplitude without change

Filtrační funkce a časová řada

- Bez filtrace – výchozí údaje
- Nízkofrekvenční filtrace
- Vysokofrekvenční filtrace
- Pásmová filtrace



Pásmový filtr

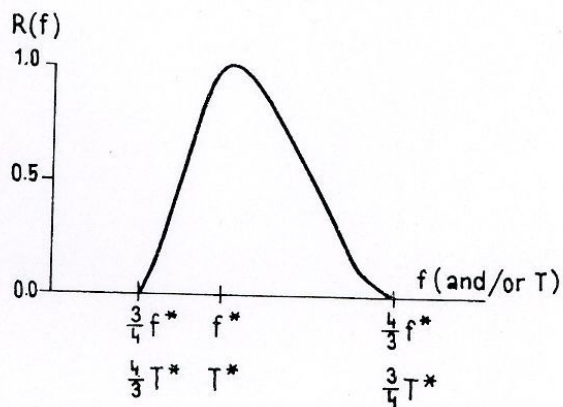
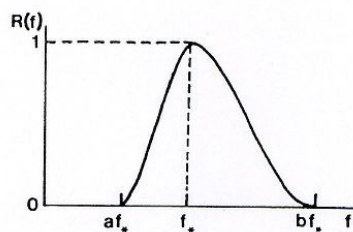
$$R(f) = 0 \quad \text{for } \frac{1}{T_n} < f < af^* \quad \text{and/or} \quad T_n > T > bT^*$$

$$2R(f) = 1 + \cos(2T^* \pi f) \quad \text{for } af^* < f < f^* \quad \text{and/or} \quad bT^* > T > T^*$$

$$2R(f) = 1 - \cos(T^* \pi f) \quad \text{for } f^* < f < bf^* \quad \text{and/or} \quad T^* > T > aT^*$$

$$R(f) = 0 \quad \text{for } bf^* < f < \frac{1}{2\Delta t} \quad \text{and/or} \quad aT^* > T > 2\Delta t$$

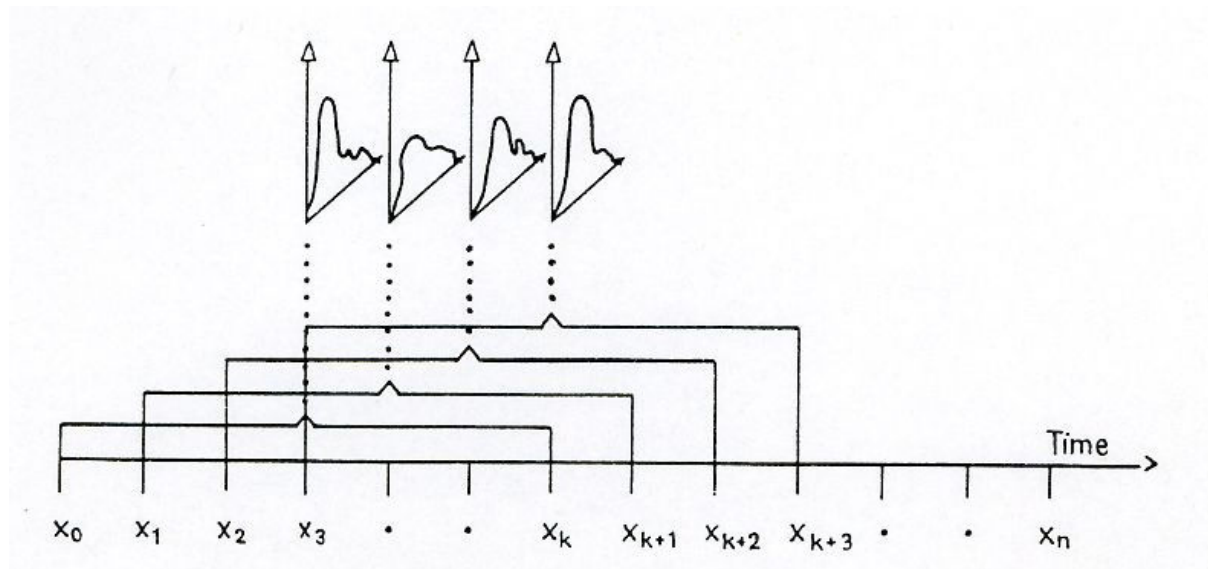
with $b > a$, $T_n = n \cdot \Delta t$

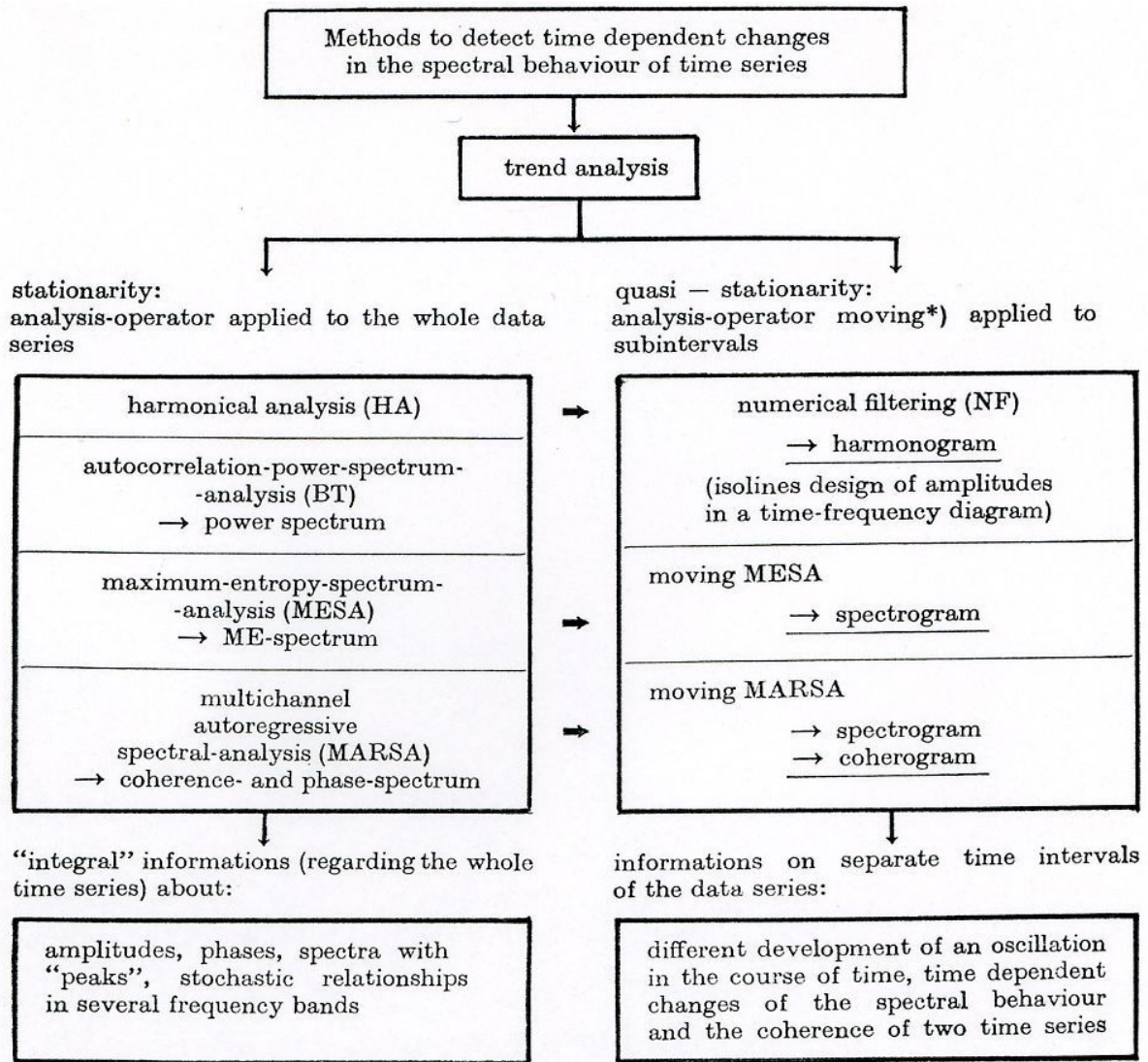


3.7 Metody klouzavé analýzy

- podmínka stacionarity časové řady může být snadněji splněna v jejích kratších realizacích → aplikace metod klouzavé analýzy

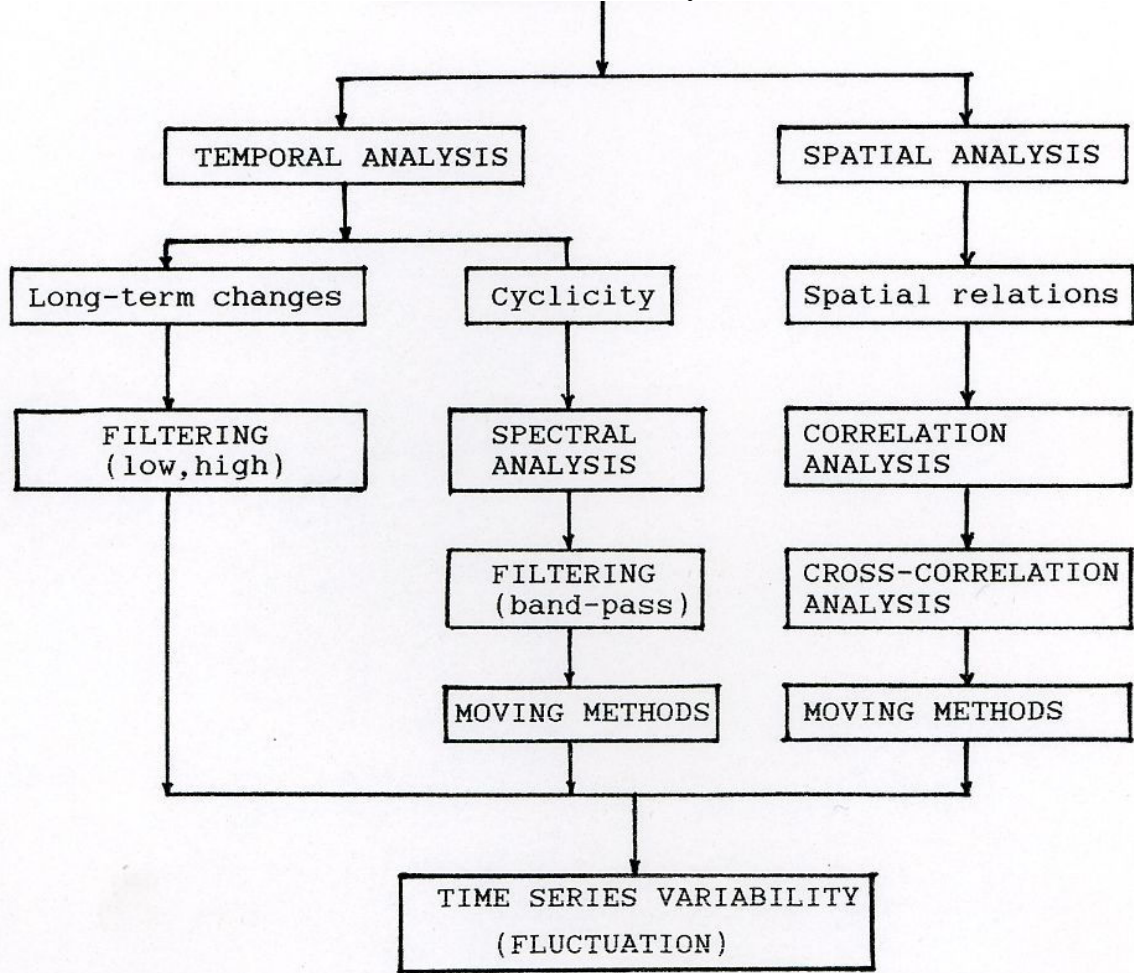
Schéma výpočtu dynamické (klouzavé) MESA





*) moving, or sliding, or adaptive, or dynamic method.

Time series analysis



3.8 Wavelet analýza

- časově-frekvenční analýza signálu v nestacionárních řadách

Wavelet transformace W_x^ψ časové řady $x(t)$ je definována výrazem (1):

$$W_x^\psi(\tau, \lambda) = \frac{1}{\sqrt{|\lambda|}} \int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t-\tau}{\lambda} \right) dt, \quad (1)$$

kde $\psi^{(\tau, \lambda)}(t)$ je tzv. wavelet (wavelet – česky vlnka). Wavelet si můžeme představit jako okénko, pomocí kterého se časová řada $x(t)$ postupně zkoumá. Parametr τ (translace) určuje část signálu $x(t)$, kde je právě wavelet lokalizován a druhý parametr λ (tzv. škála) vystupuje v definici (1) jako frekvenční parametr, hvězdička * zde značí proměnnou komplexně sdruženou. Wavelet transformace je v podstatě konvolucí původního signálu $x(t)$ s novou funkcí $\psi^{(\tau, \lambda)}(t)$.

Wavelet je funkce oscilující kolem x -ové osy, která dále musí splňovat některé podmínky, jako například, že její integrál je roven nule (2) anebo, že integrál kvadrátu je roven jedné (3) (podrobněji viz například [4]).

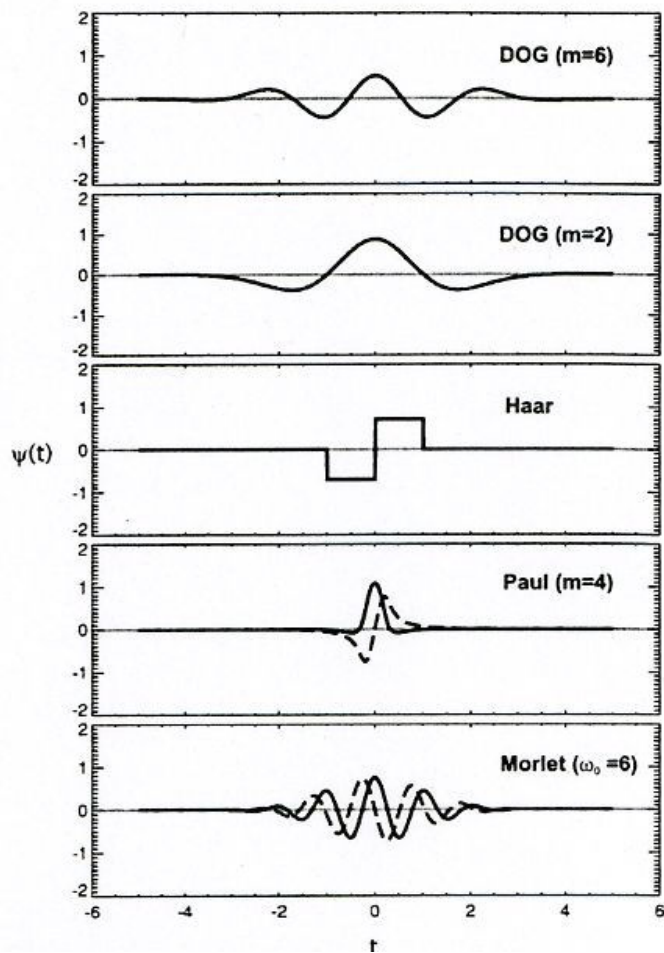
$$\int_{-\infty}^{\infty} \psi(t) dt = 0 \quad (2)$$

$$\int_{-\infty}^{\infty} \psi^2(t) dt = 1 \quad (3)$$

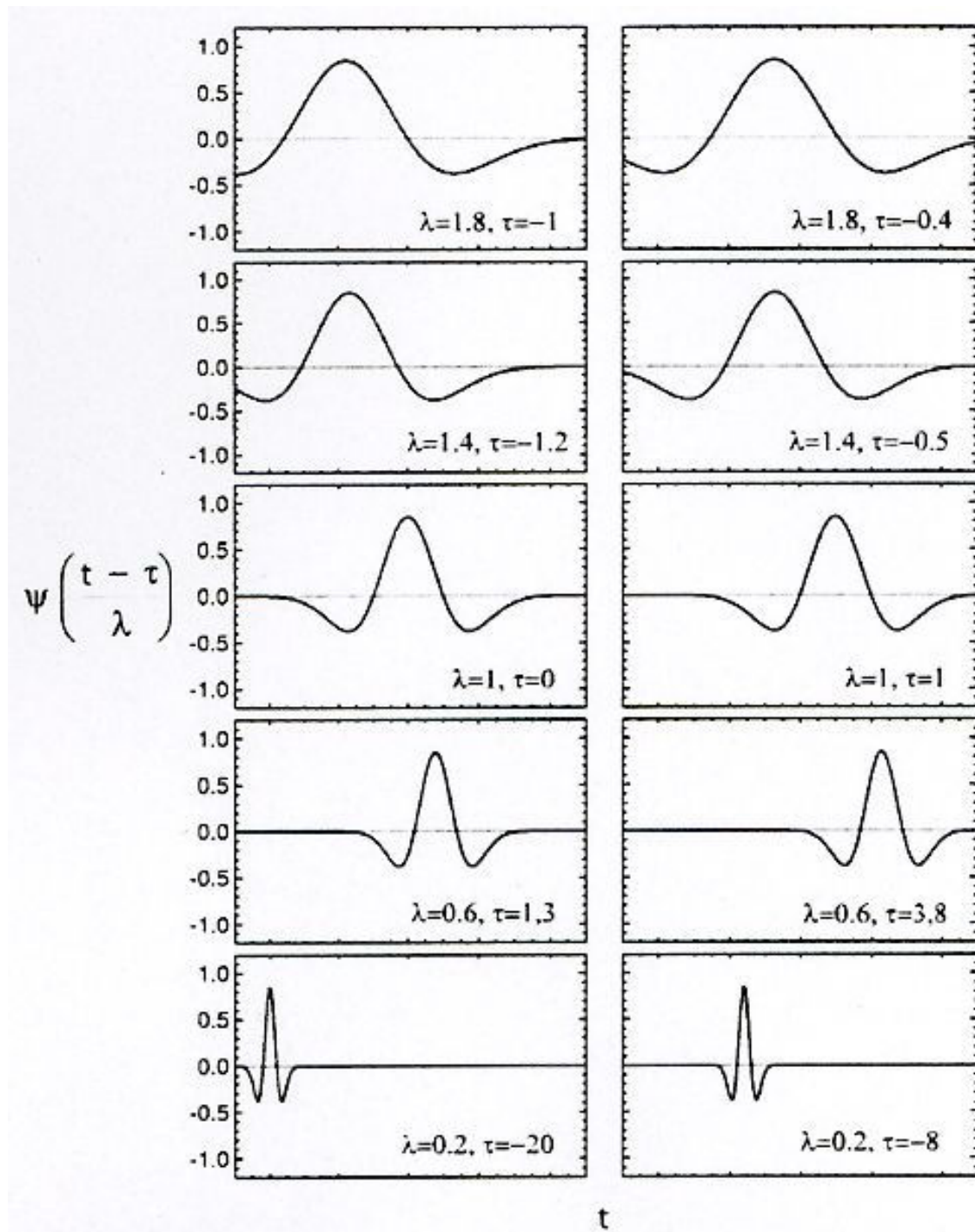
Všechny wavelety závisí na tzv. mateřském waveletu (mother wavelet), ze kterého se generují další tzv. dceřiné wavelety (daughter wavelets). V tab. 1 jsou uvedeny definice tří často používaných waveletů a grafy na obr. 1 pak ukazují, jak tyto mateřské wavelety vypadají. Symboly v tab. 1 označují parametry mateřských waveletů (m, ω_0), gamma funkci (Γ), komplexní jednotku (i) a početní konstantu (C). Grafy na obr. 2 ilustrují způsob, jakým jsou dále jednotlivé wavelety generovány – jedná se o postupnou translaci a kontrakci daného waveletu podél signálu v závislosti na parametrech τ a λ .

DOG	$\frac{(-1)^{m+1}}{\sqrt{\Gamma\left(m + \frac{1}{2}\right)}} \frac{d^m}{dt^m} e^{-\frac{t^2}{2}}$
Paul	$\frac{2^m i^m m!}{\sqrt{\pi} (2m)!} (1-it)^{-(m+1)}$
Haar	$\psi(t) = \begin{cases} -1/\sqrt{2}, & t \in (-1; 0) \\ 0, & t \in (-\infty; -1) \cup (1; \infty) \\ 1/\sqrt{2}, & t \in (0; 1) \end{cases}$
Morlet	$C e^{-i\omega_0 t} \left(e^{-\frac{t^2}{2}} - \sqrt{2} e^{-\frac{\omega_0^2}{4}} e^{-t^2} \right)$

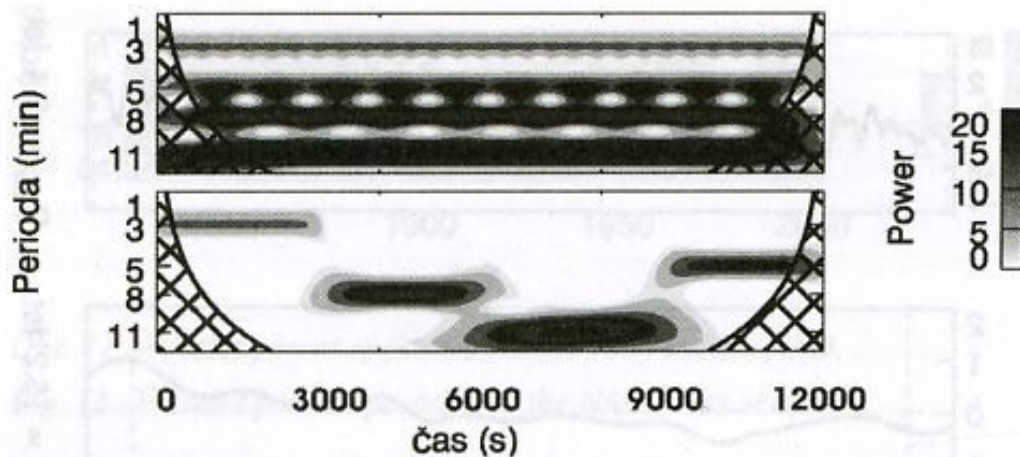
Tab. 1. Mateřské wavelety DOG, Paul, Haar, Kotlet



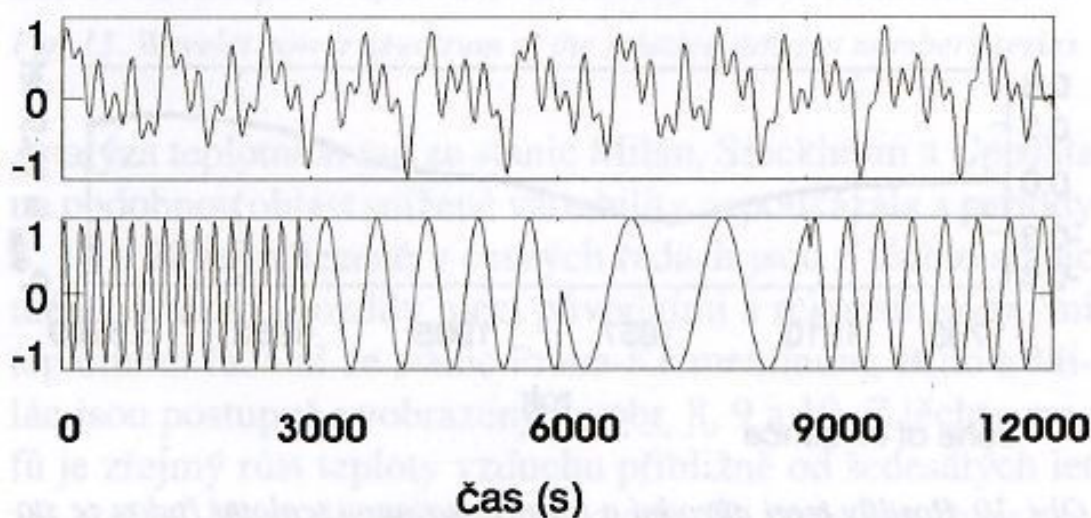
Obr. 1. Reálná (plně) a imaginární (přerušovaně) části mateřských waveletů



Obr. 2. Generace waveletů – posun a kontrakce mateřského waveletu pro různé σ a λ



Obr. 3 Spojitá wavelet transformace stacionárního (nahore) a nestacionárního (dole) signálu.



Obr. 4 Stacionární (nahore) a nestacionární (dole) testovací signál.

Výběr mateřského waveletu

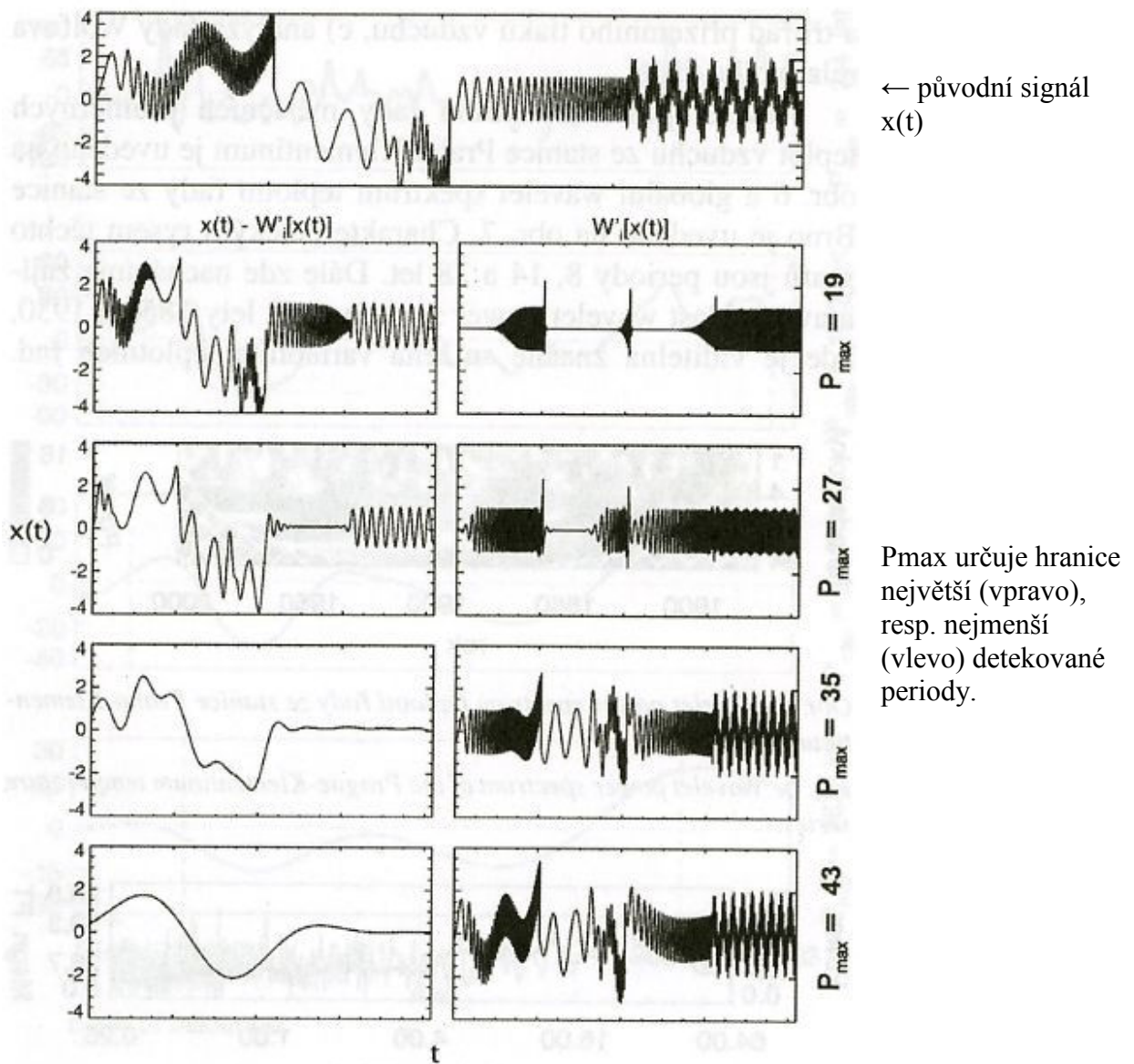
- Morlet – dobré časové a frekvenční rozlišení
- Paul – malý „cone of influence“ (nekorektní výsledky)

Zobrazení výsledků

- wavelet power spektrum – kvadrát absolutní hodnoty wavelet transformace $|W_x^\psi|^2$
- globální wavelet spektrum – integrál wavelet power spektra „podél časové osy“

$$\frac{1}{t} \int_0^t |W_x^\psi|^2 dt$$

Rekonstrukce testovacího signálu

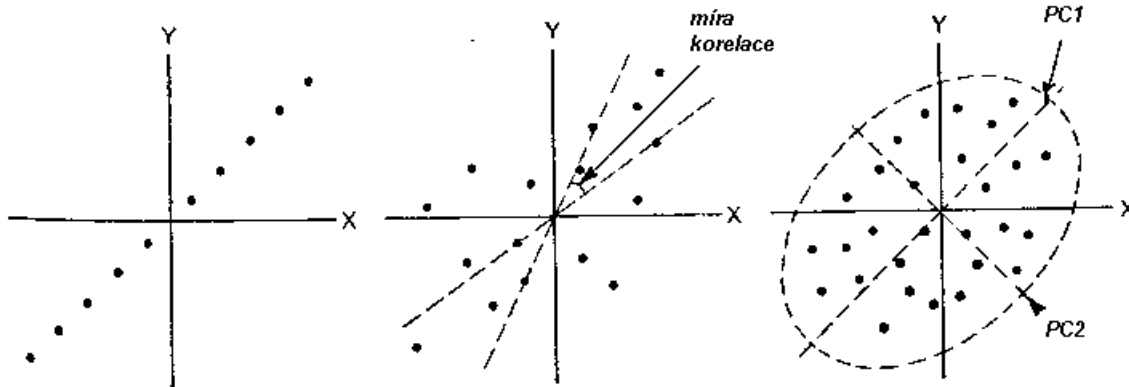


rozdíl původního a
rekonstruovaného signálu

rekonstrukce signálu
(inverzní wavelet transformace)

3.9 Analýza hlavních komponent (Principal Component Analysis)

Cílem metody je eliminování duplicit a zhuštění informace obsažené v původních proměnných do menšího počtu vzájemně nekorelovaných proměnných. Tyto nové proměnné (hlavní komponenty) popisují soubor syntetičtěji a úsporněji.



Redukce dimenzionality – příliš mnoho proměnných

Příklady:

- jedna meteorologická proměnná měřená v n případech na p místech
- p meteorologických proměnných měřených na n stanicích v jednom případě
- p meteorologických proměnných měřených na jedné stanici v n případech.

Je-li p proměnných vysoce korelováno, lze vybrat m proměnných ($m \ll p$), které poskytují veškerou originální informaci o p proměnných.

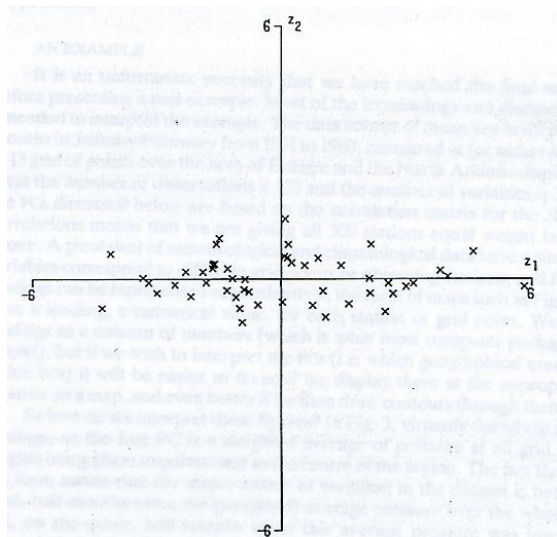
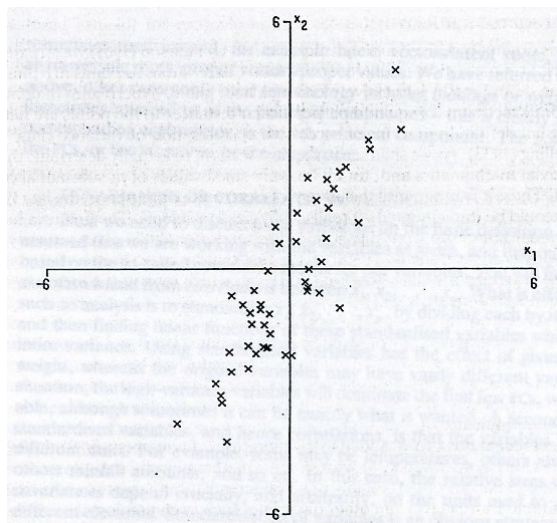
Strategie výběru m proměnných:

- podsoubor z originálních p proměnných
- vytvoření nových proměnných z p původních, které jsou od nich typově odlišné (při téže ztrátě informace lze získat větší redukci dimenzionality)

Analýza hlavních komponent

- p proměnných x_1, x_2, \dots, x_p
- lineární funkce p proměnných: $z = a_1x_1 + a_2x_2 + \dots + a_px_p$, kde a_1, a_2, \dots, a_p jsou konstanty
- změni-li se konstanty, dostaneme odlišnou lineární funkci a lze spočítat její rozptyl
- první hlavní komponenta (PC1) je lineární funkcí, která má maximální možný rozptyl, druhá hlavní komponenta (PC2) je lineární funkcí s maximálně možným rozptylem nekorelovatelná s PC1, třetí hlavní komponenta PC3 je lineární funkcí s maximálně možným rozptylem nekorelovatelná s PC1 a PC2, atd.
- PCA nabízí m -dimenzionální prezentaci údajů pro $m = 1, 2, \dots, p - 1$, tedy pro každé $m = 1, 2, \dots, p - 1$ dostáváme m lineárních funkcí proměnných x_1, x_2, \dots, x_p ukazující maximálně možný podíl originálních variací

Příklad (dvourozměrný):



- 50 měření teploty půdy (x_1) a teploty vzduchu (x_2)
- lineární funkce x_1 a x_2 objasňuje více kolísání než každá proměnná samostatně

- lineární funkce, která maximalizuje rozptyl, je první PC – z_1 , tj. všechna kolísání mohou být nyní vyjádřena s ohledem na z_1
- obrázky se shodují s tím, že došlo k rotaci os – druhá PC (z_2) jde pak kolmo na z_1
- PC minimalizuje sumu čtverců podle vzdáleností kolmých k z_1 (předchozí podle vertikálních vzdáleností)

Vlastní hodnoty a vlastní vektory (eigenvalues and eigenvectors)

- k -tá hlavní komponenta je dána vztahem $z_1 = a_{k1}x_1 + a_{k2}x_2 + \dots + a_{kpx}x_p$ pro $k = 1, 2, \dots, p$
- vlastní vektory: první vlastní vektor – koeficienty $a_{11}, a_{12}, \dots, a_{1p}$; další vlastní vektory jsou tvořeny koeficienty u proměnných x_1, x_2, \dots, x_p
- vlastní hodnota – rozptyl první hlavní komponenty (míra významnosti objasněných kolísání)

Vstupní matice výpočtu a standardizace hodnot

		Variables			
		1	2	...	m
Cases	1				
	2				
	3				
	⋮				
	n				

- standardizací se přisuzuje proměnným stejná váha (např. proměnné v různých jednotkách)

$$t_i = \frac{x_i - \mu}{\sigma}$$

Příklad – typický výstup PCA (I)

No.	Eigenvalue	Individual Cumulative		Scree Plot
		Percent	Percent	
1	3.487151	38.75	38.75	
2	2.130173	23.67	62.41	
3	1.098958	12.21	74.63	
4	0.994483	11.05	85.68	
5	0.543218	6.04	91.71	
6	0.383428	4.26	95.97	
7	0.225754	2.51	98.48	
8	0.136790	1.52	100.00	
9	0.000046	0.00	100.00	

- pořadové číslo nové proměnné (PC - hlavní komponenty)
- vlastní hodnota – část z celkového rozptylu původních dat vysvětlená každou z nových komponent
- procentuální vyjádření množství rozptylu vysvětleného komponentou
- kumulativní hodnota procentuálního podílu vysvětleného příslušnými komponentami (např. první 4 komponenty vysvětlují 85,68 % celkové variability původních dat)
- tzv. sutinový graf sloužící k určení počtu významných komponent

Příklad – typický výstup PCA (II)

Variables	Factor1	Factor2	Factor3	Factor4
AGR	0.523791	0.053594	0.048674	0.028793
MIN	0.001323	0.617807	-0.201100	0.064085
MAN	-0.347495	0.355054	-0.150463	-0.346088
PS	-0.255716	0.261096	-0.561083	0.393309
CON	-0.325179	0.051288	0.153321	-0.668324
SER	-0.378920	-0.350172	-0.115096	-0.050157
FIN	-0.074374	-0.453698	-0.587361	-0.051567
SPS	-0.387409	-0.221521	0.311904	0.412230
TC	-0.366823	0.202592	0.375106	0.314372

Tzv. zátěže (loadings) - představují míru korelace mezi původními a novými proměnnými.

Doporučená literatura:

- Brázdil, R. (1991): Kolísání vybraných meteorologických prvků ve střední Evropě v období přístrojových pozorování. Národní klimatický program ČSFR, sv. 2, Praha, 56 s.
- Pišoft, P., Kalvová, J. (2005): Wavelet analýza v meteorologii: teorie a přehled dosavadních výsledků. Meteorologické zprávy, 58, č. 1, s. 1-6.
- Pišoft, P., Kalvová, J., Brázdil, R. (2004): Cycles and trends in the Czech temperature series using wavelet transforms. International Journal of Climatology, 24, 1661-1670.
- Schönwiese, C.-D. (1985): Praktische Statistik für Meteorologen und Geowissenschaftler. Gebrüder Borntraeger, Berlin, Stuttgart, 231 s.
- Storch, H. von, Zwiers, F. W. (1999): Statistical Analysis in Climate Research. Cambridge University Press, Cambridge, 484 s.