

Příklady korelace

Jiří Holčík

Nejdříve zopakování – definiční vztahy

Zůstaneme-li pro začátek v doméně diskrétního času, pak základní formulí definující korelační posloupnost je

$$R_{x_1x_2}(m) = \sum_n^N x_1(n)x_2(n+m) \quad (1)$$

kde x_1 a x_2 jsou posloupnosti, jejichž korelaci (vzájemný vztah) odhadujeme. Pokud jsou obě posloupnosti navzájem různé, tj. platí, že $x_1(n) \neq x_2(n)$, pak hovoříme o tzv. *vzájemné korelační posloupnosti*. Často se též používá překlad anglického názvu *cross correlation sequence*, tedy *křížová korelační posloupnost*. Jestliže chceme vyzkoumat vzájemné souvislosti mezi vzorky jedné a téže posloupnosti, tj. $x_1(n) = x_2(n)$, pak hovoříme o *autokorelační posloupnosti*, anglicky *autocorrelation sequence*. Zkoumáme-li posloupnosti centrovaných dat, tj. platí, že $x_1(n) = x_{1orig}(n) - m_{x1orig}$ a $x_2(n) = x_{2orig}(n) - m_{x2orig}$, kde m_{x1orig} , příp. m_{x2orig} jsou průměry obou posloupností, pak hovoříme o tzv. *vzájemné* či *auto kovarianční posloupnosti*, v závislosti na to, zda $x_1(n) \neq x_2(n)$ nebo $x_1(n) = x_2(n)$.

Zůstaneme-li u obecnějšího tvaru korelační posloupnosti, tj. u vzájemné korelační posloupnosti, pak se můžeme setkat s různými variantami definičního vztahu.

$$a) \quad R_{x_1x_2}(mT_{vz}) = \sum_{n=-\infty}^{\infty} x_1(nT_{vz})x_2(nT_{vz}+mT_{vz}) \quad (2)$$

Tento vztah má teoretický význam, protože předpokládá znalost obou posloupností v nekonečném časovém intervalu, jak do kladných i záporných časů. Vzkovovací periody T_{vz} jsou v tomto vztahu uvedeny, abychom připomenuli **časový** rozměr obou uvažovaných vstupních posloupností $x_1(n)$ a $x_2(n)$ i výsledné korelační posloupnosti $R_{x_1x_2}(mT_{vz})$. Protože v praxi známe obě posloupnosti pouze v konečném časovém intervalu, má výpočetní význam následující definiční vztah

$$b) \quad R_{x_1x_2}(mT_{vz}) = \sum_{n=0}^N x_1(nT_{vz})x_2(nT_{vz}+mT_{vz}) \quad (3)$$

Uvažujeme-li rovnoměrné vzorkování a současně netrváme na zdůraznění časového charakteru všech posloupností, pak lze za předpokladu konečných posloupností $x_1(n)$ a $x_2(n)$ použít definiční vztah (1). Jako alternativu této definice lze v odborné literatuře najít i následující tvar

$$c) \quad R_{x_1x_2}(m) = \sum_n^N x_1(n)x_2(n-m) \quad (4)$$

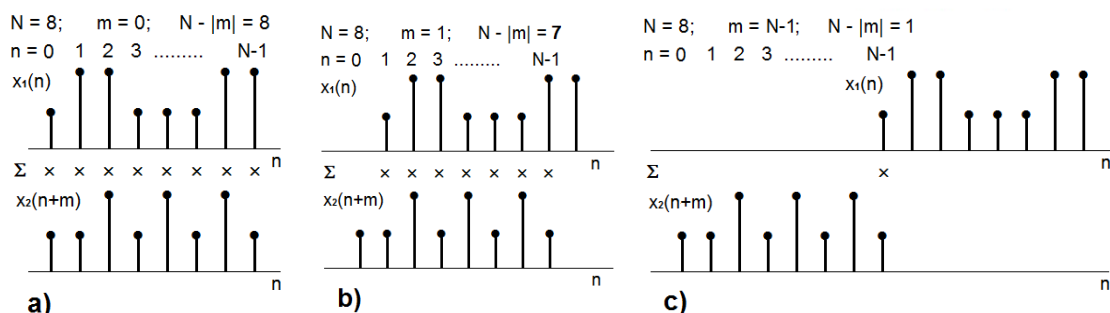
který se liší opačným znaménkem parametru posunu m u posloupnosti x_2 . Tato odlišnost vyjadřuje opačný časový posun posloupnosti x_2 při výpočtu. Pro výpočet autokorelační funkce tato varianta nemá, jak za chvíli uvidíme, žádný význam. Pro výpočet vzájemné korelace je třeba při interpretaci vypočtené korelační posloupnosti zvážit směr posunu x_2 při výpočtu.

Zatímco všechny až dosud uvedené vztahy mají tak říkajíc energetický charakter (hodnota autokorelační posloupnosti pro $m=0$ se rovná energii posloupnosti $x(n)$), v některých aplikacích může být užitečné použít tzv. výkonový vztah, kdy vypočtenou sumu součinů hodnot do výpočtu zahrnutých posloupností vztáhneme k časové jednotce, tj. podělíme ji délkou časových řad $x_1(n)$ a $x_2(n)$. Definiční vztah pak nabývá tvaru

$$d) \quad R_{x_1x_2}(m) = \frac{1}{N} \sum_n^N x_1(n)x_2(n+m) \quad \text{resp.} \quad R_{x_1x_2}(m) = \frac{1}{N} \sum_n^N x_1(n)x_2(n-m) \quad (5)$$

Tato varianta se s výhodou používá u periodických posloupností.

Poslední modifikací, kterou zde uvedeme, je tvar vztahu (5), který vyplývá praktického výpočtu korelační posloupnosti dvou konečných posloupností. Potřeba tohoto vztahu snad dostatečně jasně vyplývá z obr.1.



Obr.1 Výpočet korelační posloupnosti dvou konečných posloupností

Vzájemným posunem se snižuje počet párů, které lze mezi sebou vynásobit, a tedy se naskytá otázka, zda pro zachování rozumného průběhu korelační funkce je vhodnější vztáhnout celkovou sumu k neměnnému počtu vzorků N nebo pouze k počtu možných součinů pro danou hodnotu posunu m. Tedy na příklad ve tvaru

$$e) \quad R_{x_1x_2}(m) = \frac{1}{N - |m|} \sum_{n=0}^{N-|m|-1} x_1(n)x_2(n+m); \quad m = 0, 1, \dots, N-1 \quad (6)$$

Zatímco varianty podle vztahu (5) deformují průměrnou hodnotu odhadu korelační posloupnosti (hovoříme o *vychýleném* odhadu), výpočet podle vztahu (6) se chová příznivě k odhadu střední hodnoty (*nevychýlený* odhad), naopak má nepříznivý vliv na odhadnutou hodnotu rozptylu pro větší hodnoty posunu m.

V oblasti spojitého času, ve které byla korelační funkce definovaná historicky primárně, je sumační vztah nahrazen integrační formulí, tedy např.

$$f) \quad R_{x_1x_2}(\tau) = \frac{1}{T} \int_T x_1(t)x_2(t+\tau)dt \quad \text{resp.} \quad R_{x_1x_2}(\tau) = \frac{1}{T} \int_T x_1(t)x_2(t-\tau)dt \quad (7)$$

případně lze použít všechny ostatní varianty, jak vyplývají ze vztahů (1) až (5).

Ted' opakování vlastností odhadu korelačních posloupností

Vzájemná korelační posloupnost není komutativní, tj. $R_{x_1x_2}(n) \neq R_{x_2x_1}(n)$. Tedy její hodnoty pro kladnou a zápornou časovou poloosu se obecně liší.

Pro autokorelační posloupnost platí:

- na rozdíl od vzájemné korelační posloupnosti je sudá, tj. $R(-n) = R(n)$;
- pro všechna celá m je $R(0) \geq R(m)$, tj. její hodnota pro m=0 je větší nebo rovna jejím hodnotám pro všechny ostatní hodnoty posunu m. Hodnoty $R(0)$ teoreticky nabývá pro $m \neq 0$ pouze v případě periodických posloupností pro posun rovný periodě posloupnosti. U nevychýleného odhadu mohou hodnoty autokorelační posloupnosti i překročit velikost $R(0)$, což bývá způsobeno růstem rozptylu pro velké hodnoty posunu m.
- $R(0)$ je rovna energii, resp. výkonu posloupnosti $x(n)$ – viz vztahy (4), resp. (5) či (6); při nulové střední hodnotě je rovna rozptylu hodnot prvků posloupnosti;
- pokud je posloupnost $x(n)$ periodická, pak je její autokorelační posloupnost rovněž periodická s toutéž periodou – viz bod b).

Hodnota autokovarianční posloupnosti pro $m=0$ je rovna disperzi vstupní posloupnosti $x(n)$ pokud je určena pomocí vztahu

$$R_{xx}(m) = \frac{1}{N} \sum_n^N [x(n) - m_x][x(n+m) - m_x] \quad (8)$$

Korelační posloupnost a Pearsonův korelační koeficient

Připomeňme, jak získáme Pearsonův korelační koeficient. Pro jeho odhad lze použít vztah

$$r = \frac{\frac{1}{N} \sum_{n=1}^N (x_n - \bar{x})(y_n - \bar{y})}{\sigma_x \sigma_y} = \frac{1}{N} \sum_{n=1}^N \frac{(x_n - \bar{x})}{\sigma_x} \cdot \frac{(y_n - \bar{y})}{\sigma_y} \quad (9)$$

Když tento vztah srovnáme s výpočtem hodnoty vzájemné korelační posloupnosti pro $m=0$, tj.

$$R_{x_1x_2}(0) = \frac{1}{N} \sum_{n=0}^{N-1} x_1(n)x_2(n+0) = \frac{1}{N} \sum_{n=0}^{N-1} x_1(n)x_2(n) \quad (10)$$

pak můžeme konstatovat, že pokud by srovnávané posloupnosti $x_1(n)$ a $x_2(n)$ byly standardizovány, pak každý vzorek jejich vzájemné korelační posloupnosti je roven odpovídající hodnotě Pearsonova korelačního koeficientu. To znamená, že každá hodnota korelační posloupnosti vyjadřuje míru lineární závislosti vzájemně posunutých hodnot korelační posloupnosti. Ale protože hodnoty posloupností $x_1(n)$ a $x_2(n)$ zpravidla nejsou standardizovány, pak se hodnoty korelační posloupnosti nepohybují v rozsahu hodnot Pearsonova korelačního koeficientu, tj. v intervalu $\langle -1; 1 \rangle$, nýbrž vyjadřují pouze jistou relativní vazbu obou posloupností.

A konečně vztah korelační posloupnosti a konvoluce dvou posloupností

Nechť je vzájemná korelační posloupnost definována vztahem (4), tj.

$$R_{x_1x_2}(m) = \sum_n^N x_1(n)x_2(n-m)$$

a konvoluce vztahem

$$y(n) = x_1(n) * x_2(n) = \sum_m x_1(m) \cdot x_2(n-m) = \sum_m x_1(n-m) \cdot x_2(m). \quad (11)$$

Po záměně symbolů n a m v konvolučním vztahu dostáváme

$$y(m) = x_1(m) * x_2(m) = \sum_m x_1(n) * x_2(m-n) \quad (12)$$

Po srovnání vztahu (4) a vztahu (12) lze konstatovat, že oba vztahy by byly ekvivalentní, pokud by posloupnost x_2 byla invertována v čase, tj. když by ve formuli (12) bylo $x_2(n-m)$.

Když jsme si všechno potřebné zopakovali tak se dejme do počítání (a taky přemýšlení o tom co počítání přineslo).

Začneme případem se spojitým časem. To nám umožní spočítat výsledek analyticky, a tak si ukázat některé obecné zákonitosti.

Příklad 1

Vypočítejte vzájemnou korelační funkci funkcí $x_1(t) = \cos(\pi t)$ a $x_2(t) = 4\sin(\pi t)$. Ověřte, jaký vliv na průběh výsledné korelační funkce má alternace znaménka v druhém členu definičního vztahu (vztah (7)).

Řešení:

Argumenty obou zadaných harmonických funkcí jsou $\pi t = \omega t = 2\pi f t = 2\pi t/T = 2\pi t/2$, obě harmonické funkce tedy mají tutéž periodu $T = 2$ [časové jednotky]. Z toho by nám mohlo vyplynout, že i výsledná korelační funkce by mohla být periodická a vzhledem k tomu, že obě zadané funkce jsou harmonické, i výsledná korelační funkce by mohla být harmonická.

Nejdříve se ale zamysleme nad tím, jaké hodnoty bude nabývat korelační funkce pro nulový posun, tj. pro $\tau = 0$. Pro výsledné řešení nemá tato úvaha rozhodující význam, nicméně umožní srovnat výsledek této rozvahy se získaným analytickým řešením. Takže

$$R_{x_1x_2}(0) = \frac{1}{T} \int_T x_1(t)x_2(t+\tau)dt = \frac{1}{2} \int_0^2 \cos(\pi t) \cdot 4 \sin(\pi t) dt = 2 \int_0^2 \cos(\pi t) \cdot \sin(\pi t) dt \quad (13)$$

Odtud bychom mohli pokračovat výpočtem integrálu, což zase nemusí být příliš složité, když víme, že $2\sin\alpha \cdot \cos\alpha = \sin(2\alpha)$. V tom případě lze pokračovat

$$R_{x_1x_2}(0) = \int_0^2 \sin(2\pi t) dt = \left| \int \sin(ax) dx = -\frac{1}{a} \cos(ax) \right|_0^2 = -\frac{1}{2\pi} [\cos(2\pi t)]_0^2 = -\frac{1}{2\pi} [1 - 1] = 0 \quad (14)$$

Tím víme, že korelační funkce zadaných goniometrických funkcí začíná pro $\tau = 0$ v nule. K takovému výsledku by bylo možné dojít i z úvahy, že $\cos(\alpha)$ a $\sin(\alpha)$ jsou ortogonální (pravoúhlé, kolmé) funkce (to už víme ze souvislosti obou funkcí s kruhovým pohybem i z jejich vyjádření v Gaussově komplexní rovině) a integrál takových funkcí přes periodu je nulový.

S tím, co už o úloze víme, můžeme spočítat průběh korelační funkce. Začneme s kladným posunem, tedy $+\tau$.

$$\begin{aligned} R_{x_1x_2}(\tau) &= \frac{1}{T} \int_T x_1(t)x_2(t+\tau)dt = \frac{1}{2} \int_0^2 \cos(\pi t) \cdot 4 \sin(\pi(t+\tau)) dt = \\ &= 2 \int_0^2 \cos(\pi t) \cdot \sin(\pi(t+\tau)) dt = \frac{2}{2} \int_0^2 [\sin(\pi\tau) + \sin(2\pi t + \pi\tau)] dt = \\ &= \int_0^2 \sin(\pi\tau) dt + \int_0^2 \sin(2\pi t + \pi\tau) dt = \sin(\pi\tau)[t]_0^2 + 0 = 2 \sin(\pi\tau). \end{aligned} \quad (15)$$

Tady jsme si na druhém řádku výpočtu pomohli vzorcem

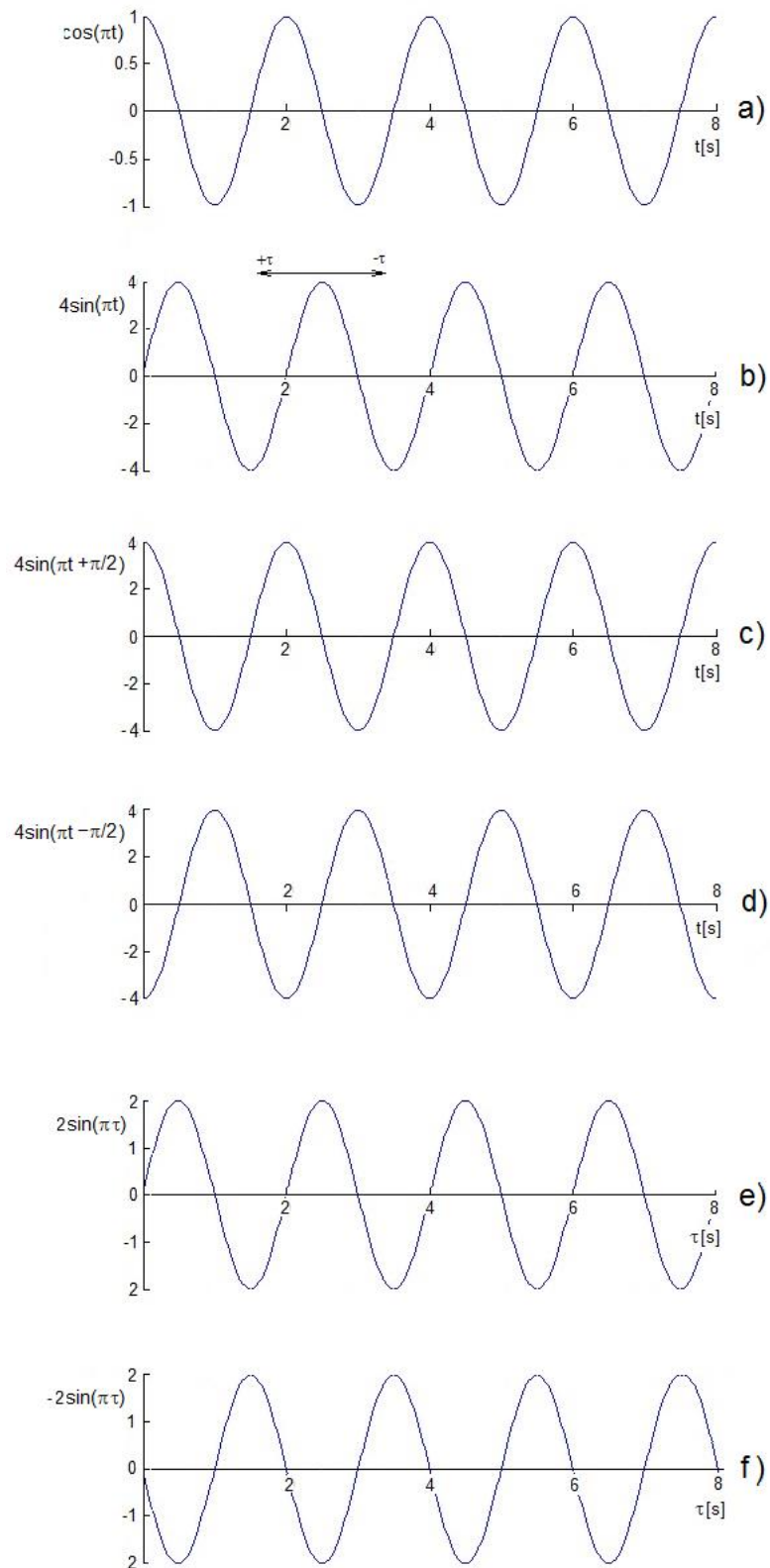
$$\sin \alpha + \sin \beta = 2 \cdot \sin \frac{\alpha+\beta}{2} \cdot \cos \frac{\alpha-\beta}{2} \quad (16)$$

z kterého po srovnání s úhly πt a $\pi(t+\tau)$ plyne třeba, že $\alpha = 2\pi t + \pi\tau$ a $\beta = \pi\tau$. (Ještě existuje druhá možnost, kdy $\alpha = 2\pi t + \pi\tau$ a $\beta = -\pi\tau$, ale i s touto variantou bychom dospěli k témuž výsledku, jen ten výpočet integrálů by byl o něco malinko zamotanější.)

V prvním integrálu na posledním řádku výpočtu je funkce sinus nezávislá na integrační proměnné t , proto není předmětem integrace a lze ji vytknout před integrál. Druhý integrál představuje integraci funkce sinus (v tomto případě je již funkcí integrační proměnné t , byť fázově posunutou o $\pi\tau$) právě přes jednu periodu. Proto je hodnota integrálu nulová,

Jak rozumět tomu co vyšlo? Především se potvrdila počáteční úvaha, že vypočítaná korelační funkce bude periodická, potvrdilo se i to, že je harmonická. Navíc, protože $\sin(0) = 0$, platí i to co jsme už dříve spočítali a je vyjádřeno výrazem (14) a v podstatě se znovu potvrdila ortogonalita funkcí \sin a \cos . Dál pomůže obr.2. Pro daný konkrétní případ se vypočtená vzájemná korelační funkce jeví jako lichá (pozor – nelze zobecnit). Průběhy a) a b) reprezentují obě zadané funkce v poloze, kdy $\tau = 0$. Při výpočtu korelační funkce pomocí vztahu používajícího $+\tau$ dochází při $\tau > 0$ k posunu funkce \sin směrem k menším hodnotám na časové ose (vlevo). To znamená, že podobnost obou křivek posunem z výchozího postavení nejdříve roste a maxima dosahuje, když se posune o čtvrt periody, tj. o $\pi/2$, jak je naznačeno v řádku c). Protože funkce

$\sin(\pi\tau)$ s nárůstem hodnoty τ také nejdříve roste a svého maxima dosahuje pro $\tau = 1/2$, odpovídá to očekávanému nárůstu hodnoty korelace. Pokud se τ zvětšuje, tj. funkce na druhém řádku se posunuje vlevo, podobnost obou funkcí se zvětšuje. Posléze se podobnost zase zmenšuje, atd. jak ukazuje průběh korelační funkce v řádku e).



Obr.2 Vzájemná poloha zadaných harmonických funkcí a průběhy vypočtených korelačních funkcí

Pokud by byla korelační funkce definována pomocí vztahu

$$R_{x_1x_2}(\tau) = \frac{1}{T} \int x_1(t)x_2(t-\tau)dt,$$

pak je možné zopakovat výpočet ekvivalentně postupu z předchozího výpočtu podle vztahu (15) a realizuje se podle formule (17) nebo lze tvar korelační funkce zjistit elegantněji dosazením hodnoty $-\tau$ do výsledné formulky, tj. $R_{x_1x_2}(\tau) = 2\sin(\pi(-\tau))$. Protože \sin je funkce lichá dostáváme konečně $R_{x_1x_2}(\tau) = -2\sin(\pi\tau)$

$$\begin{aligned} R_{x_1x_2}^-(\tau) &= \frac{1}{T} \int x_1(t)x_2(t-\tau)dt = \frac{1}{2} \int_0^2 \cos(\pi t) \cdot 4 \sin(\pi(t-\tau))dt = \\ &= 2 \int_0^2 \cos(\pi t) \cdot \sin(\pi(t-\tau))dt = \frac{2}{2} \int_0^2 [\sin(-\pi\tau) + \sin(2\pi t - \pi\tau)]dt = \\ &= -\int_0^2 \sin(\pi\tau)dt + \int_0^2 \sin(2\pi t - \pi\tau)dt = -\sin(\pi\tau)[t]_0^2 + 0 = -2\sin(\pi\tau). \end{aligned} \quad (17)$$

Při výpočtu korelační funkce pomocí vztahu s $-\tau$ dochází při $\tau > 0$ k posunu funkce \sin směrem k větším hodnotám na časové ose (vpravo). To znamená, že podobnost obou křivek posunem z výchozího postavení nejdříve klesá a minima dosáhne pro posun o $\tau = -1/2$ (obr.2d). Protože funkce $\sin(\pi\tau)$ s nárůstem hodnoty τ od nuly nejdříve roste, odpovídá průběh vypočtené korelační funkce očekávanému poklesu funkce $-2\sin(\pi\tau)$ – viz obr.2f. □□□

A od této chvíle se už budeme zabývat pouze případy s diskrétním časem.

Příklad 2

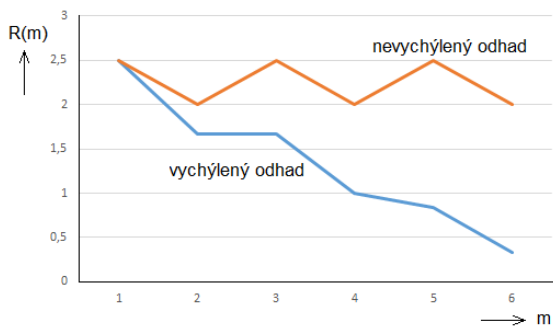
Vypočítejte autokorelační posloupnost posloupnosti $\{1; 2; 1; 2; 1; 2\}$.

Řešení:

Než začneme počítat, shrňme, co o úloze víme. Zadaná posloupnost má 6 prvků/vzorků. Je periodická s periodou 2 vzorky a délka posloupnosti je určena celočíselným násobkem v ní obsažených period. Protože je periodická, měla by být periodická i její autokorelační posloupnost. Protože autokorelační posloupnost je sudá, nezáleží na směru posunu posloupnosti x_2 , tj. zda pro výpočet použijeme vztah (1) nebo (4). Vztah (4) je použit, snad jen proto, že nabízí lepší formální znázornění výpočtu, a hlavně umožní srovnání s výpočty dalších příkladů. Abychom dokázali srovnat kvalitu odhadů autokorelační posloupnosti bude výpočet proveden pomocí vychýleného (vztah (5)) i nevychýleného (vztah (6)) odhadu.

Teď už se pusťme do výpočtu.

		součet realizovatelných součinů vzorků posloupnosti	nevychýlený odhad $1/(N- m)$	vychýlený odhad $1/N$
$x_1(n)$	1 2 1 2 1 2			
$x_2(n)$	1 2 1 2 1 2	$1+4+1+4+1+4 = 15$	$15/6 = 2,5$	$15/6 = 2,500$
$x_2(n-1)$	1 2 1 2 1 2	$2+2+2+2+2 = 10$	$10/5 = 2,0$	$10/6 = 1,667$
$x_2(n-2)$	1 2 1 2 1 2	$1+4+1+4 = 10$	$10/4 = 2,5$	$10/6 = 1,667$
$x_2(n-3)$	1 2 1 2 1 2	$2+2+2 = 6$	$6/3 = 2,0$	$6/6 = 1,000$
$x_2(n-4)$	1 2 1 2 1 ...	$1+4 = 5$	$5/2 = 2,5$	$5/6 = 0,833$
$x_2(n-5)$	1 2 1 2 ...	$2 = 2$	$2/1 = 2,0$	$2/6 = 0,333$



Obr.3 Průběhy autokorelačních posloupností z Příkladu 2 vypočítaných oběma způsoby odhadu

Průběhy obou odhadů autokorelační posloupnosti jsou zobrazeny na obr.3.

U nevychýleného odhadu je první hodnota vskutku největší a opakuje se tak, jak jsme očekávali s periodou dva vzorky. Hodnoty mezi extrémami nejsou zas až o tolik menší. Střední hodnota zadané posloupnosti je 1,5 a všechny hodnoty posloupnosti se od ní příliš neliší, jsou si tedy poměrně podobné, korelované. Není tedy důvod, aby se hodnoty autokorelační posloupnosti velice zmenšovaly.

Průběh vychýleného odhadu je viditelně deformován dělením konstantním celkovým počtem vzorků při snižování počtu spočítatelných součinnů. Přestože z výčtu hodnot v tabulce není patrná periodicita autokorelačního odhadu, z grafického vyjádření je náznak periodicity trochu zřejmý. □□□

Příklad 3

Jak se změní výsledek, použije-li pro posloupnost z předchozího příkladu odhad autokorelační posloupnosti pomocí kruhového algoritmu, při kterém se vzorky, dostávající se při posunu řady mimo interval možného násobení, přesouvají na druhou stranu posloupnosti.

Řešení:

		součet součinnů vzorků posloupnosti	kruhový odhad 1/N
$x_1(n)$	1 2 1 2 1 2		
$x_2(n)$	1 2 1 2 1 2	$1+4+1+4+1+4 = 15$	$15/6 = 2,5$
$x_2(n-1)$	2 1 2 1 2 1	$2+2+2+2+2+2 = 12$	$12/6 = 2,0$
$x_2(n-2)$	1 2 1 2 1 2	$1+4+1+4+1+4 = 15$	$15/6 = 2,5$
$x_2(n-3)$	2 1 2 1 2 1	$2+2+2+2+2+2 = 12$	$12/6 = 2,0$
$x_2(n-4)$	1 2 1 2 1 2	$1+4+1+4+1+4 = 15$	$15/6 = 2,5$
$x_2(n-5)$	2 1 2 1 2 1	$2+2+2+2+2+2 = 12$	$12/6 = 2,0$

Při tomto způsobu výpočtu zůstává počet součinnů zachován, tedy nevádí dělení konstantním počtem vzorků a výsledné hodnoty se rovnají hodnotám nevychýleného odhadu. To všechno ale platí jen za situace, kdy vstupní posloupnost obsahuje celočíselný násobek period. Pokud tento případ neplatí, pak se výsledek deformuje zavedením periodicity, jak uvidíme hnedle v následujícím příkladu. □□□

Příklad 4

Určete všemi třemi algoritmy odhadu autokorelační posloupnosti následujících posloupností – a) {1; 2; 1; 2; 1}; b) {2; 1; 2; 1; 2}; c) {1; 2; 1; 2; 1; 2; 1}; d) {2; 1; 2; 1; 2; 1; 2} a konečně e) {1; 1; 2; 1; 2; 1; 2}. Průběhy vypočítaných autokorelačních posloupností navzájem srovnejte, příp. srovnejte s průběhy autokorelačních posloupností vypočítaných v předchozím příkladu.

Řešení:

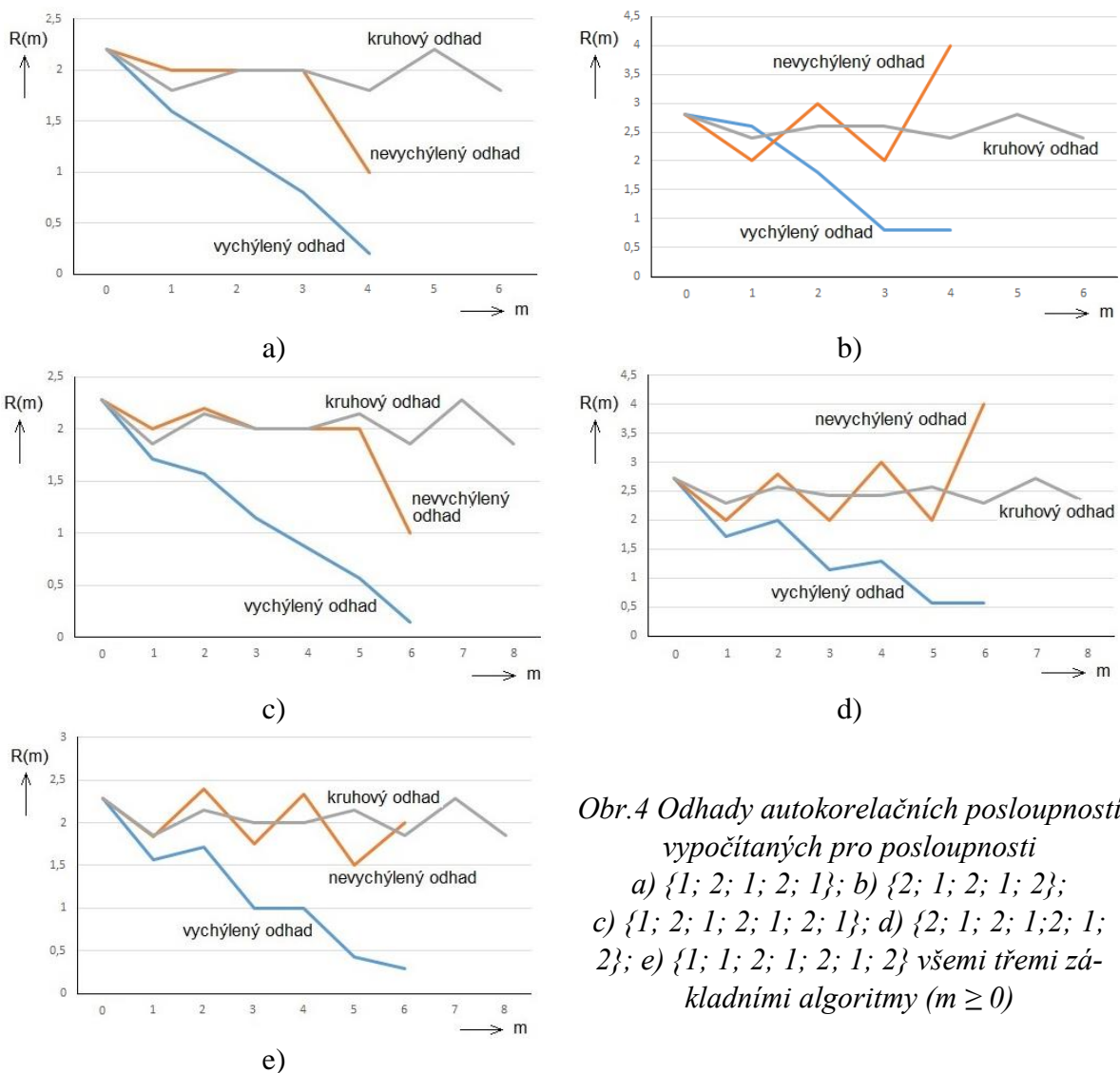
Všechny čtyři zadané posloupnosti nějakým způsobem porušují tu pěknou periodicitu, kterou měla posloupnost {1; 2; 1; 2; 1; 2} v předchozích dvou příkladech. Zatímco posloupnosti {1; 2; 1; 2; 1} a {2; 1; 2; 1; 2} jsou o jeden vzorek kratší, zbylé tři posloupnosti jsou naopak o jeden vzorek delší. V případě posloupnosti {1; 2; 1; 2; 1; 2; 1} je přidána jednička na konec

posloupnosti, u posloupnosti $\{2; 1; 2; 1; 2; 1; 2\}$ je přidána na začátek dvojka a u posloupnosti $\{1; 1; 2; 1; 2; 1; 2\}$ jednička. Přestože absolutní změny délek jednotlivých posloupností nejsou příliš velké (\pm jeden vzorek), relativní hodnoty zas až tak nevýznamné nejsou. Je to plus mínus necelých sedmáct procent.

Z obecného hlediska také není příliš důležité, kolik je kterých vzorků (v tomto případě jedniček a dvojek) v jednotlivých posloupnostech. Ovšem pro výklad tvaru vypočítaných autokorelačních posloupností u těchto krátkých a jednoduchých posloupností mohou tyto počty roli hrát.

Pro výpočet opět použijme modifikace definičního vztahu (4). Zřejmě už není nezbytné uvádět celý detailní výpočet tak, jako v předchozích dvou příkladech. Uveďme pouze výsledné posloupnosti.

- a) posloupnost $\{1; 2; 1; 2; 1\}$
 nevychýlený odhad: $R_n(m) = \{2,2; 2,0; 2,0; 2,0; 1,0\}; m = 0, 1, \dots, 4;$
 vychýlený odhad: $R_v(m) = \{2,2; 1,6; 1,2; 0,8; 0,2\}; m = 0, 1, \dots, 4;$
 kruhový odhad: $R_k(m) = \{2,2; 1,8; 2,0; 2,0; 1,8 \mid 2,2; 1,8; \dots\}; m = 0, 1, \dots, 6;$
- b) posloupnost $\{2; 1; 2; 1; 2\}$
 nevychýlený odhad: $R_n(m) = \{2,8; 2,0; 3,0; 2,0; 4,0\}; m = 0, 1, \dots, 4;$
 vychýlený odhad: $R_v(m) = \{2,8; 2,6; 1,8; 0,8; 0,8\}; m = 0, 1, \dots, 4;$
 kruhový odhad: $R_k(m) = \{2,8; 2,4; 2,6; 2,6; 2,4 \mid 2,8; 2,4; \dots\}; m = 0, 1, \dots, 6;$



Obr.4 Odhady autokorelačních posloupností vypočítaných pro posloupnosti a) $\{1; 2; 1; 2; 1\}$; b) $\{2; 1; 2; 1; 2\}$; c) $\{1; 2; 1; 2; 1; 2; 1\}$; d) $\{2; 1; 2; 1; 2; 1; 2\}$; e) $\{1; 1; 2; 1; 2; 1; 2\}$ všemi třemi základními algoritmy ($m \geq 0$)

c) posloupnost $\{1; 2; 1; 2; 1; 2; 1\}$

nevychýlený odhad:

$$R_n(m) = \{2,286; 2,000; 2,200; 2,000; 2,000; 2,000; 1,000\}; m = 0, 1, \dots, 6;$$

vychýlený odhad:

$$R_v(m) = \{2,286; 1,714; 1,571; 1,143; 0,857; 0,571; 0,143\}; m = 0, 1, \dots, 6;$$

kruhový odhad:

$$R_k(m) = \{2,286; 1,857; 2,143; 2,000; 2,000; 2,143; 1,857 \mid 2,286; 1,857; \dots\}; \\ m = 0, 1, \dots, 8;$$

d) posloupnost $\{2; 1; 2; 1; 2; 1; 2\}$

nevychýlený odhad:

$$R_n(m) = \{2,714; 2,000; 2,800; 2,000; 3,000; 2,000; 4,000\}; m = 0, 1, \dots, 6;$$

vychýlený odhad:

$$R_v(m) = \{2,714; 1,714; 2,000; 1,143; 1,286; 0,571; 0,571\}; m = 0, 1, \dots, 6;$$

kruhový odhad:

$$R_k(m) = \{2,714; 2,286; 2,571; 2,429; 2,429; 2,571; 2,286 \mid 2,714; 2,286; \dots\}; \\ m = 0, 1, \dots, 8;$$

e) posloupnost $\{1; 1; 2; 1; 2; 1; 2\}$

nevychýlený odhad:

$$R_n(m) = \{2,286; 1,833; 2,400; 1,750; 2,333; 1,500; 2,000\}; m = 0, 1, \dots, 6;$$

vychýlený odhad:

$$R_v(m) = \{2,286; 1,571; 1,714; 1,000; 1,000; 0,429; 0,286\}; m = 0, 1, \dots, 6;$$

kruhový odhad:

$$R_k(m) = \{2,286; 1,857; 2,143; 2,000; 2,000; 2,143; 1,857 \mid 2,286; 1,857; \dots\}; \\ m = 0, 1, \dots, 8;$$

Co ta jednoduchá ztráta periodičnosti dokázala napáchat na autokorelačních odhadech je i graficky vyjádřeno na obr.4. Hodnota pro nulový posun, tj. $m = 0$ je vždy pro všechny tři algoritmy odhadu stejná. Ovšem je různá pro jednotlivé posloupnosti. Její hodnota je logicky závislá na délce jednotlivých posloupností, je závislá i na relativních četnostech hodnot (jedniček a dvojek) v posloupnosti. Proto největší a nejmenší hodnoty nabývá první hodnota autokorelační posloupnosti pro nejkratší vstupní posloupnosti a) a b). Táž je první hodnota u posloupností c) a e), protože mají stejnou délku i stejné poměrné zastoupení obou použitých hodnot.

Bohužel už není vždy první hodnota odhadu autokorelační posloupnosti největší tak, jak očekává teorie. Platí to jen u vychýlených a kruhových odhadů. V případě nevychýlených odhadů, které vedou k zvyšování rozptylu hodnot autokorelační posloupnosti se zvětšujícím se přesahem analyzovaných posloupností, dochází k výraznému nárůstu či poklesu odhadovaných hodnot. Zda nárůst či pokles, závisí opět na relativních četnostech jedniček a dvojek v posloupnostech. Je-li více dvojek a dvojka je i na konci posloupnosti dochází ke zvětšení odhadu. Je-li více jedniček a jednička je i na konci posloupnosti dochází k poklesu odhadované hodnoty.

Vychýlený i nevychýlený odhad jsou délkou omezeny na délku vstupní posloupnosti, z logiky principu výpočtu zavádí kruhový odhad do průběhu autokorelační posloupnosti periodicitu s periodou danou délkou vstupní posloupnosti (v numerických výsledcích je vyznačena svislou čarou).

Vychýlené odhady (opět v závislosti na principu výpočtu) mají klesající trend, oproti nevychýleným odhadům je díky tomu i méně výrazné vyjádření periodicity. $\square\square$

Příklad 5

Určete nevychýlený i vychýlený odhad vzájemné korelační posloupnosti posloupností:

$$x_1 = \{1; 2; 3; 2; 1\} \text{ a } x_2 = \{1; 2; 1; 0; 1\}.$$

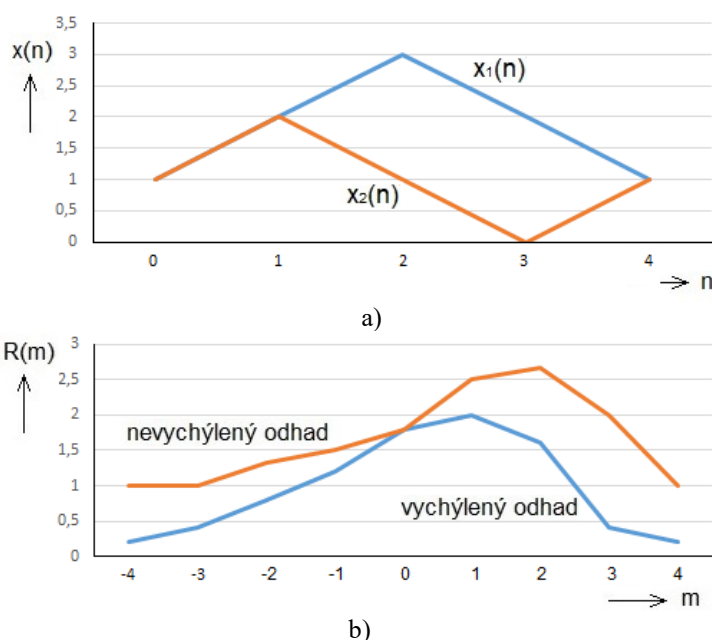
Řešení:

Obě zadané posloupnosti mají tutéž délku. Protože vzájemná korelační posloupnost nemusí být obecně symetrická vzhledem k počátku časové osy, tj. $m = 0$, spočítáme její hodnoty pro kladné i záporné hodnoty posunu. Stále počítáme podle variant vztahu (4).

		součet realizovatelných součinnů vzorků posloupností	nevychýlený odhad	vychýlený odhad
$x_1(n)$	1 2 3 2 1			
$x_2(n+4)$... 0 1	1 = 1	$1/1 = 1,000$	$1/5 = 0,2$
$x_2(n+3)$... 1 0 1	$0+2 = 2$	$2/2 = 1,000$	$2/5 = 0,4$
$x_2(n+2)$... 2 1 0 1	$1+0+3 = 4$	$4/3 = 1,333$	$4/5 = 0,8$
$x_2(n+1)$	1 2 1 0 1	$2+2+0+2 = 6$	$6/4 = 1,500$	$6/5 = 1,2$
$x_2(n)$	1 2 1 0 1	$1+4+3+0+1 = 9$	$9/5 = 1,800$	$9/5 = 1,8$
$x_2(n-1)$	1 2 1 0 1	$2+6+2+0 = 10$	$10/4 = 2,500$	$10/5 = 2,0$
$x_2(n-2)$	1 2 1 0 ...	$3+4+1 = 8$	$8/3 = 2,667$	$8/5 = 1,6$
$x_2(n-3)$	1 2 1 ...	$2+2 = 4$	$4/2 = 2,000$	$2/5 = 0,4$
$x_2(n-4)$	1 2 ...	$1 = 1$	$1/1 = 1,000$	$1/5 = 0,2$

Na zobrazených průbězích (obr.5b) obou odhadů je zcela zřejmá jejich nesymetričnost.

Nevážený odhad (první výsledkový sloupec ve výše uvedené tabulce) korelační posloupnosti má maximum pro $m = 1$. Tuto polohu ctí též vychýlený odhad, což lze snadno očekávat, když jeho hodnoty jsou jen konstantně poníženy. Tato poloha odpovídá slícování maxim obou zadaných posloupností. Na druhé straně, nevychýlený odhad nabývá též poměrně vysokých hodnot, ale své maximum má lehce posunuté pro $m = 2$. To je způsobeno samozřejmě odlišným váhovááním, ale i nepřímým vlivem těch částí obou posloupností, které se bezprostředně nepodílí na výpočtu (jsou posunuty mimo intervaly možného pronásobení).



Obr.5 (a) Zadané posloupnosti a (b) odhady jejich vzájemné korelační posloupnosti podle zadání Příkladu 5

Příklad 6

Určete odhady vzájemné korelační posloupnosti dvou následně zadaných posloupností:

$$x_1 = \{-1; 1; -1; 1; -1; 1\} \text{ a } x_2 = \{-1; 1; -1\}.$$

Řešení:

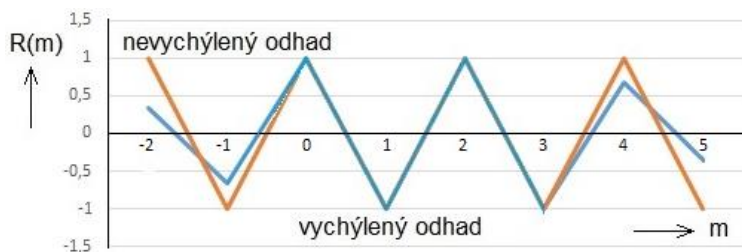
Obě zadané posloupnosti mají tentokrát různou délku. Začátek výpočtu v USTÁLENÉM stavu by měl nastat ve chvíli, kdy VŠECHNY vzorky kratší posloupnosti UŽ dokážeme vynásobit vzorky delší posloupnosti ($m = 0$) a konec výpočtu ustáleného stavu nastane tehdy, když

dokážeme JEŠTĚ všechny vzorky kratší posloupnosti vynásobit se vzorky delší posloupnosti. Výpočet by ale měl začínat ($m = -2$), pokud UŽ můžeme mezi sebou vynásobit ALESPŇ JEDEN prvek z každé ze zadaných posloupností a končit, pokud JEŠTĚ můžeme mezi sebou vynásobit alespoň jeden prvek z každé posloupnosti. Tedy počet sečítaných součinů, který by mohl ovlivnit výběr algoritmu vychýleného či nevychýleného odhadu, může nastat pouze na začátku a konci výpočtu.

Pokud budeme posloupnost $\{-1; 1; -1; 1; -1; 1\}$ považovat za posloupnost x_1 a posloupnost $\{-1; 1; -1\}$ za x_2 (jak jsou uvedeny ve vztazích (1) a (4)), bude vhodnější/názornější použít pro odhad vzájemné korelace vztah (4).

		součet realizovatelných součinů vzorků posloupnosti	nevychýlený odhad	vychýlený odhad
$x_1(n)$	-1 1 -1 1 -1 1			
$x_2(n+2)$	-1 1 -1	$1 = 1$	$1/1 = 1$	$1/3 = 0,333$
$x_2(n+1)$	-1 1 -1	$(-1)+(-1) = -2$	$-2/2 = -1$	$-2/3 = -0,667$
$x_2(n)$	-1 1 -1	$1+1+1 = 3$	$3/3 = 1$	$3/3 = 1,000$
$x_2(n-1)$	-1 1 -1	$(-1)+(-1)+(-1) = -3$	$-3/3 = -1$	$-3/3 = -1,000$
$x_2(n-2)$	-1 1 -1	$1+1+1 = 3$	$3/3 = 1$	$3/3 = 1,000$
$x_2(n-3)$	-1 1 -1	$(-1)+(-1)+(-1) = -3$	$-3/3 = -1$	$-3/3 = -1,000$
$x_2(n-4)$	-1 1 -1	$1+1 = 2$	$2/2 = 1$	$2/3 = 0,667$
$x_2(n-5)$	-1 1 -1	$-1 = -1$	$-1/1 = -1$	$-1/3 = -0,333$

Grafické znázornění výsledných korelačních posloupností je na obr.6.



Obr.6 Vzájemné korelační posloupnosti podle zadání Příkladu 6 (modrá – vychýlený odhad; žlutá – nevychýlený odhad)

Je jasně zřejmé, že vliv nevychýleného, resp. vychýleného odhadu se uplatňuje pouze v přechodných neustálených odhadech na začátku a na konci výsledné vzájemné korelační posloupnosti.

Jak plyne z definice korelační posloupnosti, její hodnoty vyjadřují vzájemný vztah/podobnost mezi dvěma posloup-

nostmi, příp. jak se tento vztah mění se změnou vzájemné polohy srovnávaných posloupností. Z toho plyne, že hodnota by měla být vysoká, když se oba srovnávané průběhy navzájem podobají a naopak nízká, pokud se srovnávané posloupnosti nepochybují.

Z této logiky lze usoudit, že vysoká hodnota odhadu korelační posloupnosti bude tehdy, když se bude posloupnost $\{-1; 1; -1\}$ vyskytovat v místě, kde se v posloupnosti $\{-1; 1; -1; 1; -1; 1\}$ právě vyskytuje tato sekvence. To nastává v ustáleném stavu pro $m = 0$ a $m = 2$. V přechodných intervalech lze částečnou tvarovou podobnost mezi oběma posloupnostmi nalézt i pro $m = -2$ a $m = 4$. A naopak. Tam, kde se v posloupnosti $\{-1; 1; -1; 1; -1; 1\}$ vyskytuje dílčí sekvence s hodnotami s opačnou polaritou, tj. $\{1; -1; 1\}$, tam bude hodnota korelační posloupnosti minimální. Že jsou tyto extrémní hodnoty právě $+1$ a -1 , plyne z hodnot posloupností x_1 a x_2 ; obecně to není hodnota korelačních koeficientů. V tomto konkrétním případě ano, protože posloupnost x_1 má nulovou střední hodnotu a jednotkový rozptyl a posloupnost x_2 se těmito vlastnostem dostatečně blíží. □□□

Příklad 7

Určete odhady vzájemné korelační posloupnosti dvou následujících posloupností: $x_1 = \{-1; 1; -1; 1; -1; 1; -1; 1\}$ a $x_2 = \{1; -1; 1\}$.

Řešení:

Na rozdíl od předešlého příkladu 6 se změnila polarita hodnot posloupnosti x_2 . Tím pádem, na základě rozboru řešení v předešlém příkladu, lze snad bez výpočtu rovnou psát:

nevychýlený odhad: $R_n(m) = \{-1; 1; -1; 1; -1; 1; -1; 1\}$; $m = -2, -1, \dots, 5$

vychýlený odhad: $R_n(m) = \{-0,333; 0,667; -1; 1; -1; 1; -0,667; 0,333\}$; $m = -2, -1, \dots, 5$.

Kdo nevěří, ať si spočítá.

□□□

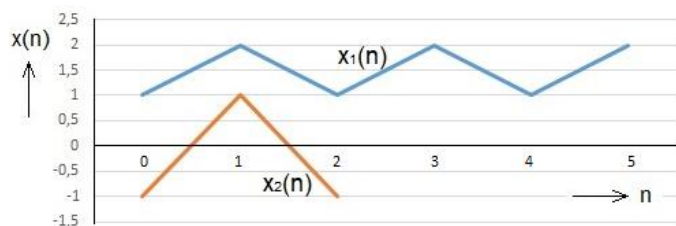
Příklad 8

Určete odhady vzájemné korelační posloupnosti dvou následujících posloupností: $x_1 = \{1; 2; 1; 2; 1; 2\}$ a $x_2 = \{-1; 1; -1\}$.

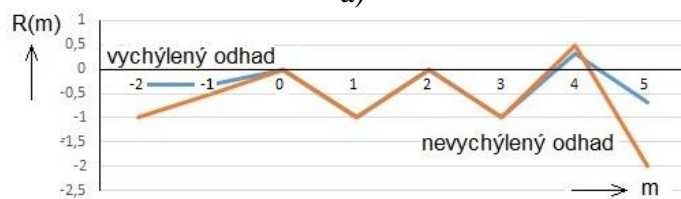
Řešení:

Posloupnost x_2 je tentokrát stejná jako v příkladu 6, posloupnost x_1 je rovněž oscilující s toutéž periodou ($T = 2$), ale obsahuje pouze kladné hodnoty a střední hodnota je 1,5. Způsob výpočtu se nemění, jen pro kontrolu jej uvedme:

		součet realizovatelných součinů vzorků posloupnosti	nevychýlený odhad	vychýlený odhad
$x_1(n)$	1 2 1 2 1 2			
$x_2(n+2)$	-1 1 -1	$-1 = -1$	$(-1)/1 = -1$	$(-1)/3 = -0,33$
$x_2(n+1)$	-1 1 -1	$1+(-2) = -1$	$(-1)/2 = -0,5$	$(-1)/3 = -0,33$
$x_2(n)$	-1 1 -1	$(-1)+2+(-1) = 0$	$0/3 = 0$	$0/3 = 0,00$
$x_2(n-1)$	-1 1 -1	$(-2)+1+(-2) = -3$	$(-3)/3 = -1$	$(-3)/3 = -1,00$
$x_2(n-2)$	-1 1 -1	$(-1)+2+(-1) = 0$	$0/3 = 0$	$0/3 = 0,00$
$x_2(n-3)$	-1 1 -1	$(-2)+1+(-2) = -3$	$(-3)/3 = -1$	$(-3)/3 = -1,00$
$x_2(n-4)$	-1 1 -1	$(-1)+2 = 1$	$1/2 = 0,5$	$1/3 = 0,33$
$x_2(n-5)$	-1 1 -1	$(-2) = -2$	$(-2)/1 = -2$	$(-2)/3 = -0,67$



a)



b)

Obr.7 (a) Zadané posloupnosti a (b) odhady jejich vzájemné korelační posloupnosti (modrá – vychýlený odhad; žlutá – nevychýlený odhad) podle zadání Příkladu 8

V ustáleném intervalu nabývá odhad korelační funkce hodnot 0 a -1 . Nula znamená srovnání bez korelace, jak je zvykem u korelačních koeficientů, je to v tomto intervalu maximální hodnota, a tedy to znamená, že pro $m = 0$ a $m = 2$ se obě posloupnosti nejvíce podobají.

Hodnoty obou odhadů pro $m = 4$ větší než nula jsou dány výpočtem v přechodném intervalu. Větší vychylky (v absolutní hodnotě) na konci odhadů, než na jejich začátku vyplývají ze

skutečnosti, že posloupnost x_1 začíná jedničkou a končí dvojkou – poslední vzorek odhadu se tedy bere s větší vahou než první. □□□

Příklad 9

Určete nevychýlený odhad vzájemné korelační posloupnosti dvou následujících posloupností: $x_1 = \{1; 2; 1; 1; 2; 1\}$ a $x_2 = \{-1; 1; -1\}$.

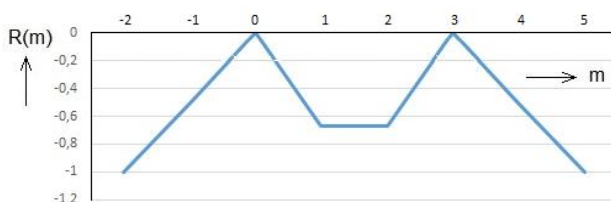
Řešení:

Perioda posloupnosti x_1 je v tomto případě $T = 3$ a je v ní obsažena dvakrát. Současně je rovna i délce posloupnosti x_2 .

Výsledný odhad korelační posloupnosti je roven

$$R_n(m) = \{-1,000; -0,500; 0,000; -0,667; -0,667; 0,000; -0,500; -1,000\}, m = -2, -1, \dots, 5.$$

Zobrazen je tento průběh na obr.8. Nulová maxima korelační posloupnosti se vyskytují pro $m = 0$ a $m = 3$, tj. pozice, kde se v posloupnosti x_1 vyskytuje největší hodnota nebo jinak, kdy je posloupnost $\{-1; 1; -1\}$ nejpodobnější dílčímu trojúhelníku $\{1; 2; 1\}$ v posloupnosti x_1 . Na rozdíl od všech předchozích příkladů je vypočítaná korelační posloupnost symetrická, což je způsobeno symetrií obou vstupních posloupností.



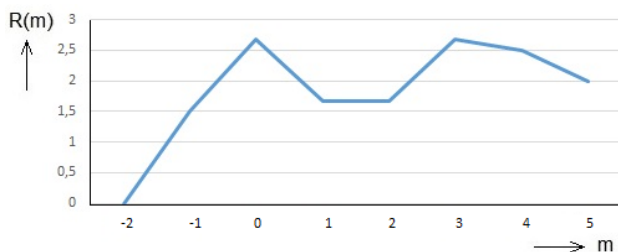
Obr.8 Nevychýlený odhad korelační posloupnosti dle zadání Příkladu 9 □□□

Příklad 10

Určete nevychýlený odhad vzájemné korelační posloupnosti dvou následujících posloupností: $x_1 = \{0; 1; 2; 0; 1; 2\}$ a $x_2 = \{1; 2; 3\}$.

Řešení:

Výsledný odhad korelační posloupnosti je $R_n(m) = \{0; 1,5; 2,667; 1,667; 1,667; 2,667; 2,5; 2\}$, $m = -2, -1, \dots, 5$.



Obr.9 Nevychýlený odhad korelační posloupnosti dle zadání Příkladu 10

Jsou-li dva první vzorky součástí přechodného děje s překryvem jednoho a dvou vzorků, pak logicky největší shoda/korelace mezi x_1 a x_2 nastává s prvním úplným překryvem mezi vzorky x_1 a x_2 , tj. pro $m = 0$. Je to tehdy, když nastává úplná shoda v lineárním nárůstu hodnot vzorků jak posloupnosti x_1 , tak i x_2 . □□□

Příklad 11

Určete nevychýlený i vychýlený odhad vzájemné korelační posloupnosti dvou následujících posloupností: $x_1 = \{1; 2; 1; 2; 1\}$ a $x_2 = \{0; 1; 0\}$.

Řešení:

Pokud přistoupíme k řešení tohoto zadání bez hlubší rozvahy tímž postupem, jak v předcházejících příkladech a budeme předpokládat, že délka posloupnosti x_1 je 5 a x_2 je 3 vzorky, pak je:

nevychýlený odhad: $R_n(m) = \{0; 0,5; 0,667; 0,333; 0,667; 0,5; 0\}$; $m = -2, -1, \dots, 4$;

vychýlený odhad: $R_v(m) = \{0; 0,333; 0,667; 0,333; 0,667; 0,333; 0\}$; $m = -2, -1, \dots, 4$.

Vidíme, že oba odhady se liší pouze v jednom vzorku v přechodné fázi výpočtu na začátku a na konci vypočtených posloupností.

K výpočtu lze ale přistoupit i z jiného pohledu. Posloupnost x_2 obsahuje pouze jeden nenulový vzorek, který je z obou stran obklopen nulami. Tedy faktická délka posloupnosti x_2 je pouze jeden vzorek a nemá smysl vztahovat vypočtené hodnoty k dříve uvažované délce posloupnosti x_2 , nýbrž pouze k délce jeden vzorek. Pak jsou obě výsledné posloupnosti (nevychýlený i vychýlený odhad) stejné, a to $\{0; 1; 2; 1; 2; 1; 0\}$. A to je posloupnost x_1 , pouze ohraničená nulami, představující přechodný počáteční a koncový děj.

A ještě jinak. Posloupnost x_2 vlastně představuje jednotkový impulz. Protože ale x_2 jednotkou nezačíná, ale je až na druhé pozici, reprezentuje x_2 jednotkový impulz posunutý o jeden vzorek. Teď si připomeňme vztah mezi korelací a konvolucí, jak jej vyjadřují vztahy (11) a (12) a k tomu definiční vlastnost jednotkového impulzu, která říká, že konvoluce jednotkového impulzu s danou posloupností je rovna této dané posloupnosti. Abychom spočítali konvoluci posloupností x_1 a x_2 , není třeba vyvíjet speciální úsilí, protože posloupnost x_2 je osově symetrická a není třeba ji proto v čase invertovat. Z toho důvodu je neváhovaná posloupnost $\{0; 1; 2; 1; 2; 1; 0\}$ současně i výsledkem konvoluce obou posloupností. A to, že je na začátku nula, je způsobeno posunem jednotkového impulzu v posloupnosti x_2 . $\square\square$

Příklady pro samostatné procvičení

Určete odhady (nevychýlený, vychýlený, neváhovaný) korelačních posloupností pro následující posloupnosti. Výsledek svého snažení můžete srovnat s uvedenými neváhovanými korelačními posloupnostmi. Důležité určitě bude zamyslet se nad spočítanými výsledky a zapřemýšlet nad tím, co znamenají, co říkají o vlastnostech obou zadaných posloupností a proč jsou takové, jaké jsou. Případně srovnat zadání a výsledky z jednotlivých zadání a zamyslet se, jak navzájem souvisí.

- 1) $x_1 = \{1; 2; 1; 2; 1\}$ a $x_2 = \{-1; 0; -1\}$;
neváhovaný výsledek: $R(m) = \{-1; -2; -2; -4; -2; -2; -1\}$; $m = -2, -1, \dots, 4$;
- 2) $x_1 = \{1; 2; 1; 2; 1\}$ a $x_2 = \{1; 0; 1\}$;
neváhovaný výsledek: $R(m) = \{1; 2; 2; 4; 2; 2; 1\}$; $m = -2, -1, \dots, 4$;
- 3) $x_1 = \{1; 2; 3; 2; 1, 2; 3\}$ a $x_2 = \{1; 2; 3\}$;
neváhovaný výsledek: $R(m) = \{3; 8; 14; 14; 10; 10; 14; 8; 3\}$; $m = -2, -1, \dots, 4$;
- 4) $x_1 = \{2; 3; 4; 3; 2; 3; 4\}$ a $x_2 = \{1; 2; 3\}$;
neváhovaný výsledek: $R(m) = \{6; 13; 20; 20; 16; 16; 20; 11; 4\}$; $m = -2, -1, \dots, 4$;
- 5) $x_1 = \{2; 3; 4; 3; 2; 3; 4\}$ a $x_2 = \{3; 2; 1\}$;
neváhovaný výsledek: $R(m) = \{2; 7; 16; 20; 20; 16; 16; 17; 12\}$; $m = -2, -1, \dots, 4$.

Průběhy vychýlených, resp. nevychýlených odhadů vzájemné korelační posloupnosti získáme podělením hodnot korelační posloupnosti $R(m)$ odpovídajícími vahami, tj. délkou kratší posloupnosti N_2 , resp. $N_2 - |m|$. $\square\square$

Příklad 12

Určete nevychýlený i vychýlený odhad autokorelační posloupnosti konstantní posloupnosti. Pro zjednodušení, nechť tato posloupnost obsahuje pouze jednotkové vzorky. Tedy např. posloupnosti $x = \{1; 1; 1; 1; 1; 1; 1; 1\}$.

Řešení:

Autokorelační posloupnost, která je výsledkem výpočtu korelace/podobnosti dané posloupnosti s posouváním průběhem téže posloupnosti tak vyjadřuje i souvislost mezi hodnotami sousedních vzorků/prvků této posloupnosti.

Na základě této úvahy lze předpokládat (hodnoty všech prvků konstantní posloupnosti jsou stejné, tedy jsou spolu hodnotově jednoznačně vázány), že velikost prvků autokorelační posloupnosti posloupnosti s konstantními prvky bude též konstantní.

Ověřme výpočtem.

		součet realizovatelných součinů vzorků posloupnosti	nevychýlený odhad	vychýlený odhad
x(n)	1 1 1 1 1 1 1 1			
x(n-0)	1 1 1 1 1 1 1 1	1+1+1+1+1+1+1+1 = 8	8/8 = 1	8/8 = 1,000
x(n-1)	1 1 1 1 1 1 1 1	1+1+1+1+1+1+1 = 7	7/7 = 1	7/8 = 0,875
x(n-2)	1 1 1 1 1 1 1...	1+1+1+1+1+1 = 6	6/6 = 1	6/8 = 0,750
x(n-3)	1 1 1 1 1 1...	1+1+1+1+1 = 5	5/5 = 1	5/8 = 0,625
x(n-4)	1 1 1 1 1...	1+1+1+1 = 4	4/4 = 1	4/8 = 0,500
x(n-5)	1 1 1 1...	1+1+1 = 3	3/3 = 1	3/8 = 0,375
x(n-6)	1 1 1...	1+1 = 2	2/2 = 1	2/8 = 0,250
x(n-7)	1 1...	1 = 1	1/1 = 1	1/8 = 0,125

Nevychýlený odhad, který zachovává střední hodnotu, opravdu potvrzuje výše uvedenou rozvahu; vychýlený odhad díky konstantnímu váhovému koeficientu logicky vede k lineárně klesajícím hodnotám.

Odhady spočítané pro kladný i záporný posun v prostředí MATLAB[®] pro posloupnost s 500 prvky je na obr.10. Protože víme, že autokorelační posloupnost je sudá, pak je symetricky lineárně klesající vychýlený odhad očekávaný, podobně jako konstantně jednotkový nevychýlený odhad. U nevychýleného odhadu může být zajímavé všimnout si oscilací hodnot odhadu pro velké hodnoty posunu tak, jak předpokládá teorie – zvětšením rozptylu odhadu pro velké m. Nicméně, srovnáme-li velikost kmitů s měřítkem na svislé ose, není třeba se kmitáním této velikosti nijak znervózňovat. □□□

Příklad 13

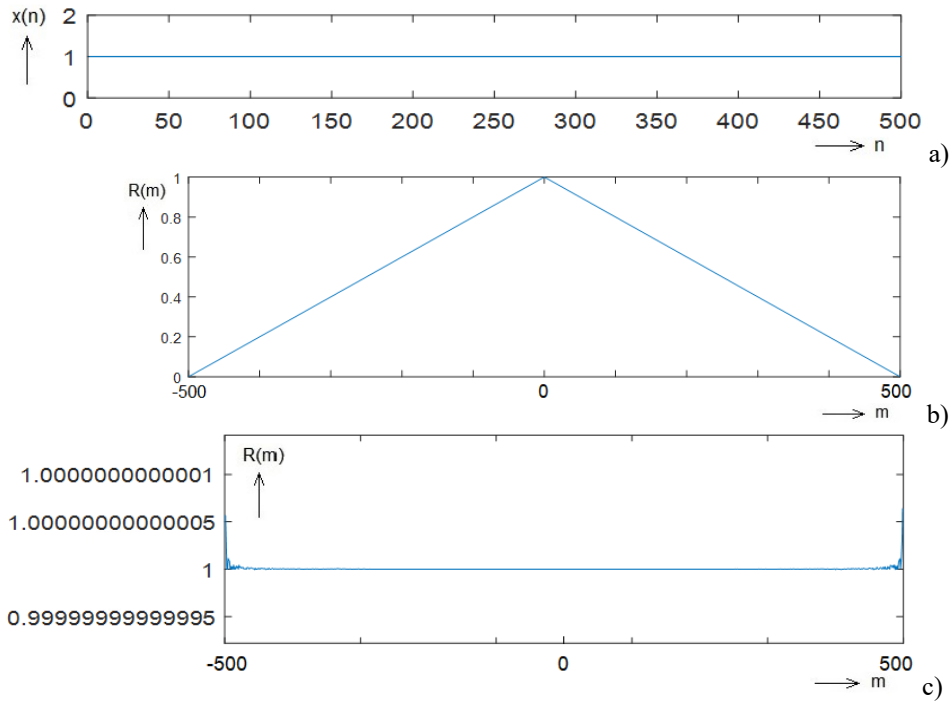
Rozmyslete, jak vypadá autokorelační posloupnost náhodné posloupnosti s nulovou střední hodnotou a jednotkovým rozptylem.

Řešení:

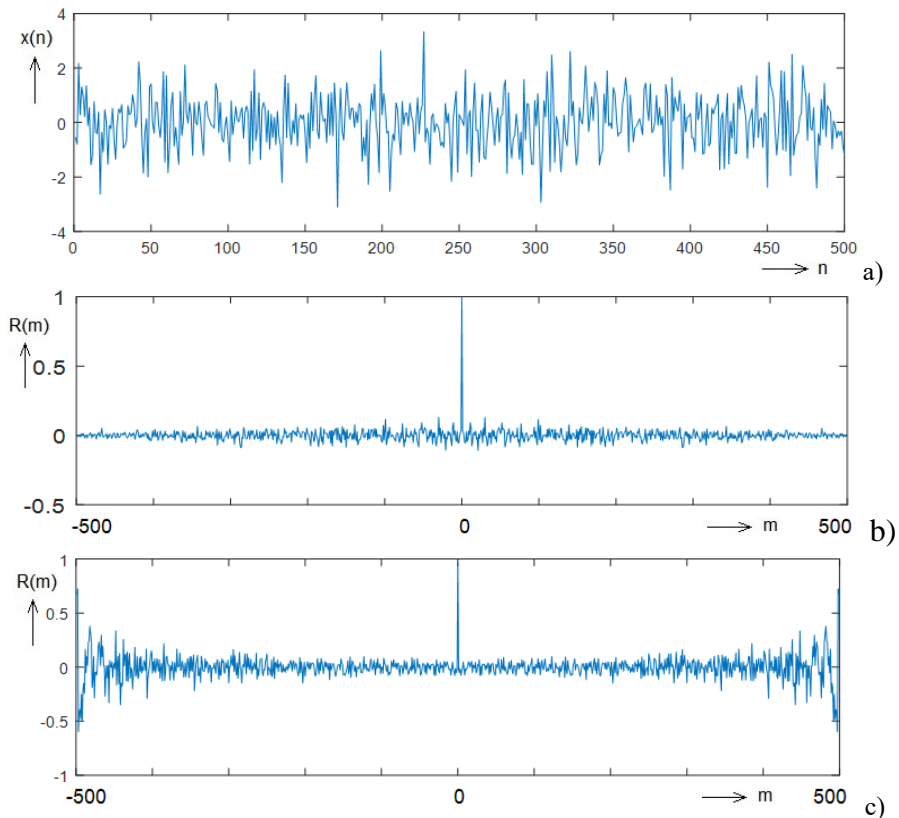
V tomto případě vskutku nelze výsledek demonstrovat pomocí nějaké zjednodušené verze. Můžeme jej určit pouze na základě rozboru vlastností náhodné posloupnosti, příp. si výsledek úvahy ověřit pomocí počítačové simulace.

Jsou-li každé dva bezprostředně sousední prvky náhodné posloupnosti nekorelované, při normálním rozložení pravděpodobnosti i na sobě nezávislé, pak i při posunu o jeden prvek musí být autokorelační posloupnost nulová. Jediná hodnota posunu, pro který je autokorelační posloupnost nenulová je nulový posun, tj. $m = 0$. Pro tuto hodnotu posunu je při nulové střední hodnotě hodnota autokorelační posloupnosti rovna rozptylu zadané náhodné posloupnosti.

Výsledek počítačové simulace v prostředí MATLAB[®] je na obr.11. Oba odhady jsou dle výše uvedené rozvahy pro $m \neq 0$ nulové. Drobné oscilace kolem nuly jsou dány mírou kvality generátoru náhodné posloupnosti. V případě vychýleného odhadu tyto oscilace s růstem hodnoty m klesají (viz princip výpočtu vychýleného odhadu), u nevychýleného odhadu naopak oscilace se zvětšováním vzájemného posunu posloupností (viz princip výpočtu nevychýleného odhadu). Hodnota autokorelační posloupnosti pro oba způsoby odhadu je rovna rozptylu vstupní posloupnosti, tedy dle zadání rovna jedné. □□□



Obr.10 a) Jednotková posloupnost s 500 prvky; b) vychýlený odhad, c) nevychýlený odhad autokorelační posloupnosti této posloupnosti



Obr.11 a) Náhodná posloupnost s nulovou střední hodnotou a jednotkovým rozptylem s 500 prvky; b) vychýlený odhad; c) nevychýlený odhad autokorelační posloupnosti této posloupnosti

Příklady ze života

Skupina látek běžně označovaná jako perzistentní organické polutanty (v angličtině se používá zkratka POPs a přeneseně se používá i v češtině) je tvořena látkami, jejichž základ je tvořen uhlíkovými řetězci nebo častěji v kombinaci s tzv. aromatickými/benzenovými jádry. Tyto látky vykazují v prostředí vysokou perzistenci, tj. odolnost proti přirozenému rozkladu např. působením slunečního záření, oxidačních činidel, případně vlivem živých organismů. Až na výjimky jde o látky umělého původu vytvářené člověkem záměrně pro jejich perzistentní vlastnosti nebo mimoděk jako vedlejší produkty jiných chemických procesů.

Právě vzhledem k jejich perzistenci se POPs mohou v prostředí jako je voda, vzduch, půda nebo sedimenty, ale také v živých tělech rostlin a živočichů včetně lidí, vyskytovat i poměrně dlouhou dobu po svém vzniku (řádově od roků až po stovky let). Tato skutečnost v některých případech vyvolává nebo může vyvolávat nežádoucí vliv na životního prostředí a lidské zdraví.

Klíčovým prostředím pro monitoring POPs je vnější ovzduší. V principu lze pro stanovení vzdušné koncentrace POPs rozlišit dva zásadní přístupy – aktivní a pasivní vzorkování ovzduší. Aktivní vzorkování spočívá v čerpání vzduchu skrze adsorpční filtry, které s vysokou účinností a při řízené teplotě zachycují na svém povrchu molekuly POPs. Jde o relativně dražší a komplikovanější metodu, vyžadující umístění zařízení na místě s přístupem k elektrické energii a možností náročné údržby. Získané údaje jsou nicméně vysoce přesné. Alternativou je pasivní vzorkování, kdy je adsorbční médium umístěno do polouzavřené schránky, kterou vzduch volně prochází. Tento levnější a jednodušší způsob vzorkování je nicméně zatížen určitou mírou nejistoty vyplývající z neznalosti skutečného množství vzduchu, které bylo za dobu expozice ve styku s adsorbčním médiem a teploty, při které adsorpce probíhala.

V České republice probíhají na observatoři ČHMÚ v Košetcích oba typy monitoringu ovzduší paralelně, díky čemuž jsou dlouhodobě získávána cenná data umožňující nejen posouzení samotné koncentrace POPs v ovzduší, ale také hodnocení vzájemného vztahu obou monitorovacích metod a odvození obecných principů a metod datové analýzy.

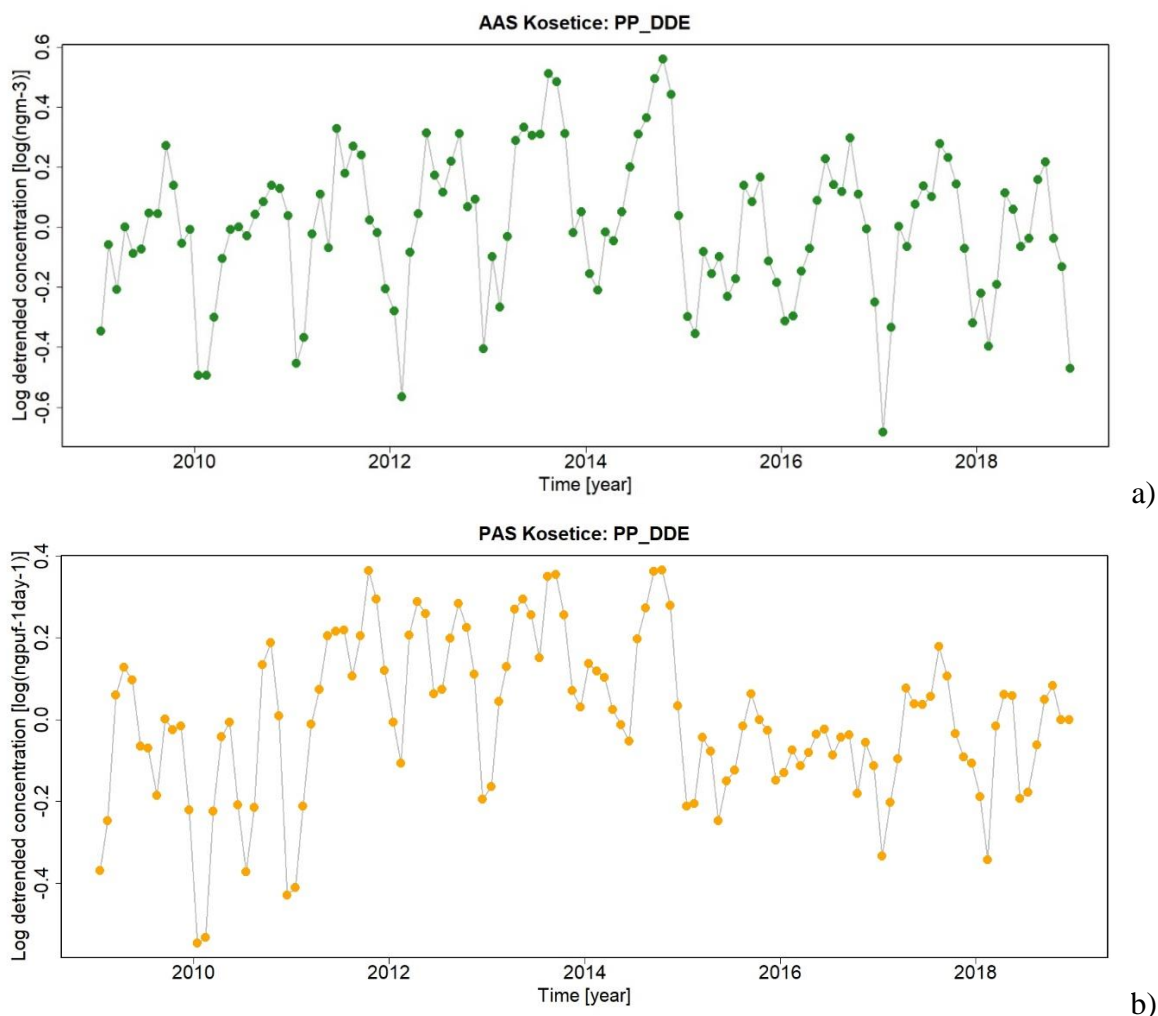
Ideální by bylo, kdyby oba typy měření znečištění obsahovaly stejnou podstatnou informaci o kvalitě měřeného prostředí. Pak by bylo možné a ekonomicky výhodnější používat pasivních vzorkovačů. Z dále uvedených dat je zřejmé, že je toto přání nereálné. Proto může analýza záznamů z obou typů vzorkovačů, případně s dalšími meteorologickými záznamy, napomoci k hledání příčin rozdílů v datových záznamech a zkoumání možností tyto rozdíly napravit.

Protože se v tomto místě zabýváme především použitím korelačních posloupností, použijme právě tohoto způsobu k vyjádření vlastností uvedených časových záznamů. Záznamy jsou z období let 2009 až 2019 s pravidelnou měsíční vzorkovací periodou (12 vzorků do roka), na kterou byly naměřené hodnoty přepočítány, protože některé hodnoty nebyly změřené, resp. byly změřené chybně.

Z pořízených záznamů byly rovněž odečteny lineární trendy, to aby byla odstraněna jejich nejvýznamnější nestacionární složka.

Základní vlastnosti korelačních posloupností těchto záznamů si předvedme na datech koncentrace látky p,p'-DDE, která je produktem rozkladu známého pesticidu 1,1,1-trichlor-2,2-bis(4-chlorfenyl)ethanu, označovaného jako DDT, ovšem s podstatně delším poločasem rozkladu (a tedy delším setrváním v prostředí).

Druhý příklad bude pro jednoho zástupce z třídy polychlorovaných bifenyly (PCB). To jsou toxické perzistentní látky se dvěma benzenovými jádry, které se hojně využívaly zejména od 60. do 80. let 20. století jako lubrikanty výhodně pro svoji stabilitu a široký rozsah teplot, ve kterém zůstávají v kapalném stavu. Typickým využitím PCB bylo chlazení v elektrických zařízeních (zejména transformátorech) a v nátěrových hmotách a barvivech.



Obr.12 Časová řada koncentrace PP_DDE pomocí a) aktivního měřiče; b) pasivního měřiče ze stanice Košetice

Záznam z aktivního měřiče vykazuje pěkně viditelnou roční periodicitu, možná jen částečně ovlivněnou pozvolným nárůstem/trendem v období 2009 až 2014, poté pokles a relativně stálou základní úroveň dat v čase od roku 2015 do konce záznamu. Záznam z pasivního měřiče vykazuje frekvenční složky i s vyšší frekvencí (orientačně půlroční), kolísání základní linie lze hodnotit jako výraznější.

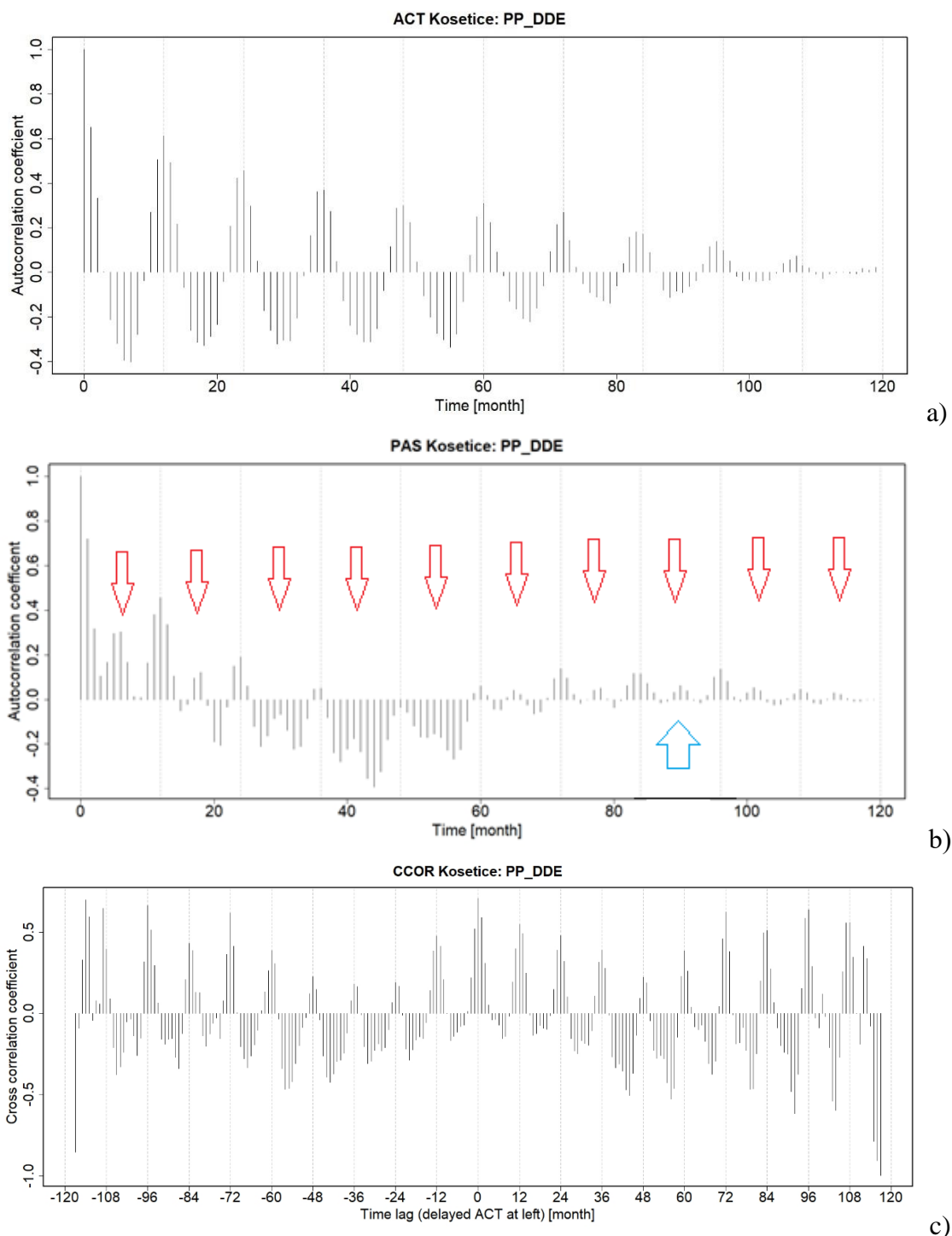
Jak se nyní tyto vlastnosti projeví v autokorelačních posloupnostech¹?

Odhady průběhů autokorelačních posloupností jsou počítány pomocí vzorce pro vychýlený odhad. Proto se hodnoty maxim v obou zobrazených autokorelačních posloupnostech se zvětšováním posunu zmenšují.

Autokorelační posloupnost dat z aktivního měřiče vykazuje téměř až dokonalou roční periodicitu. Vzhledem k tomu, že záznam začíná a končí na přelomu roku v zimním období a minima jsou v časové řadě užší než maxima v letních měsících, jsou maxima autokorelační posloupnosti strmější než minima.

Autokorelační posloupnost pro data z pasivního měřiče sice také vykazuje maxima ukazující na roční periodicitu originálních dat. Maxima jsou ale méně výrazná než v případě aktivního měřiče. Kromě toho můžeme v grafu vidět i maxima (i lokální, a ne příliš výrazná) pro poloviční časové intervaly, což signalizuje výše zmíněnou půlroční periodicitu (červené šipky). V grafu

¹ V příkladech zde uvedených jsou hodnoty korelačních posloupností standardizovány, tedy odpovídají hodnotám Pearsonova korelačního koeficientu pro daný posun. To znamená, že hodnoty autokorelačních posloupností pro nulový posun jsou vždy rovny jedné – není podobnější posloupnost než ta samá neposunutá posloupnost.



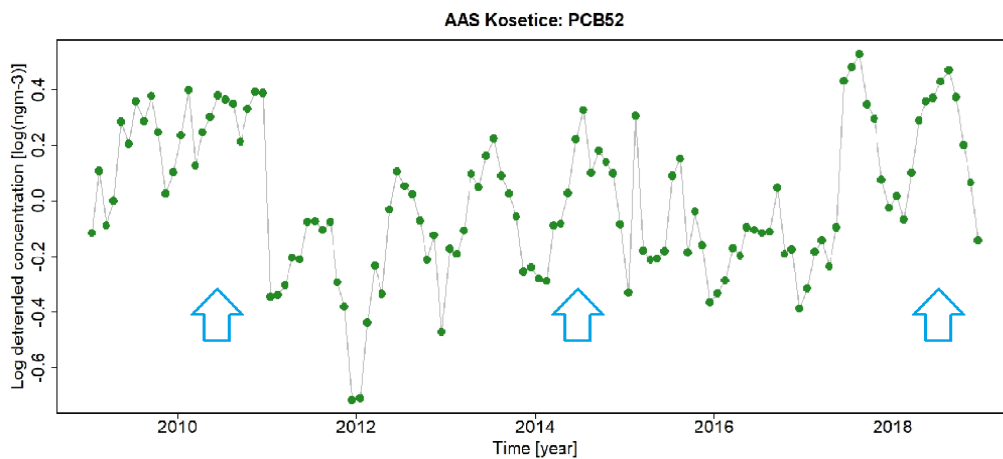
Obr.13 Autokorelační posloupnosti časových řad koncentrace PP_DDE ze stanice Košetice z a) aktivního měřiče; b) pasivního měřiče (šedé svislé linie v grafech vyznačují roční periodicitu – celočíselné násobky 12) a jejich vzájemná korelační posloupnost

autokorelační posloupnosti je též velice zřejmý vliv pomalého kolísání základní linie, s druhým nevýrazným maximem pro posun zhruba 90 vzorků (modrá šipka). To odpovídá orientačně periodě 7,5 roku. Pokud bychom naznali, že první minimum pomalého kolísání je právě na začátku záznamu, pak druhé minimum by mělo být přibližně v polovině roku 2016. Z grafu originálních dat na obr.12b) by tak opravdu mohlo být. V záznamu z aktivního měřiče bychom toto pomalé kolísání mohli také identifikovat, je však o něco méně významné, v autokorelační posloupnosti nezřetelné, částečně i vlivem výpočtu pomocí vychýleného odhadu.

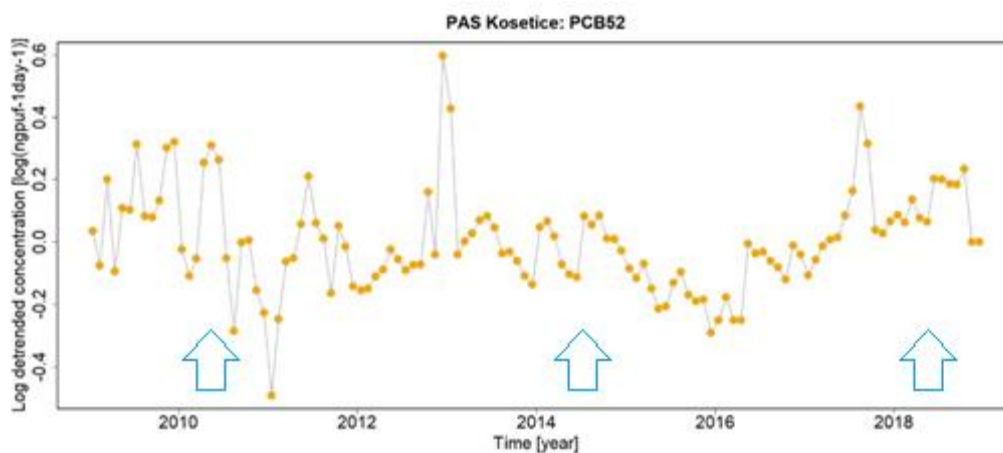
Odhad vzájemné korelační posloupnosti je počítán pomocí vzorce pro nevychýlený odhad, proto se velikost výchytky se zvětšující se hodnotou posunu zvětšuje. Ve zmíněném vzorci je za posloupnost x_1 považován záznam z aktivního měřiče, posunovaná posloupnost x_2 je dána daty z pasivního měřiče. Pro výpočet je použita váhovaná varianta vztahu (4), tj. se záporným znaménkem pro posun x_2 . Protože vzájemná korelační posloupnost není obecně symetrická podle osy y , zobrazuje graf posun oběma směry.

Absolutní maximum se vyskytuje pro nulový posun, i když teoreticky to u vzájemné korelace být nemusí. To, že to tak je, signalizuje podobnost obou posloupností. Vlastně zde nastává to, co platí pro autokorelační posloupnosti, kde absolutní maximum v počátku časové osy být musí. Je-li tento extrém překonán velikostmi korelační posloupnosti pro velké hodnoty posunu, je to dáno chybami vyplývajícími z malého počtu vzorků zahrnutých do výpočtu nevychýleného odhadu. O podobnosti obou originálních záznamů svědčí i skutečnost, že obě strany vzájemné korelační posloupnosti, pro kladný i záporný posun, se v zásadě podobají, o korelační posloupnosti lze zhruba hovořit jako o symetrické.

V průběhu vzájemné korelační posloupnosti jsou jasně viditelné dvě zásadní tendence – s roční periodou (12 vzorků) a dříve zmíněná pomalá, cca 7,5roční, s periodou cirká 90 vzorků, více patrná u záznamu z pasivního měřiče. To, že se tato pomalá oscilace projevila i ve vzájemné korelaci potvrzuje, že je vskutku přítomna i v datech z aktivního měřiče, ač se to v jejich autokorelační posloupnosti výrazně neprojevuje. Zvýšení frekvence kmitů na okrajích vzájemné korelační posloupnosti (pro absolutní hodnotu posunu přibližně větší než 100) můžeme



a)



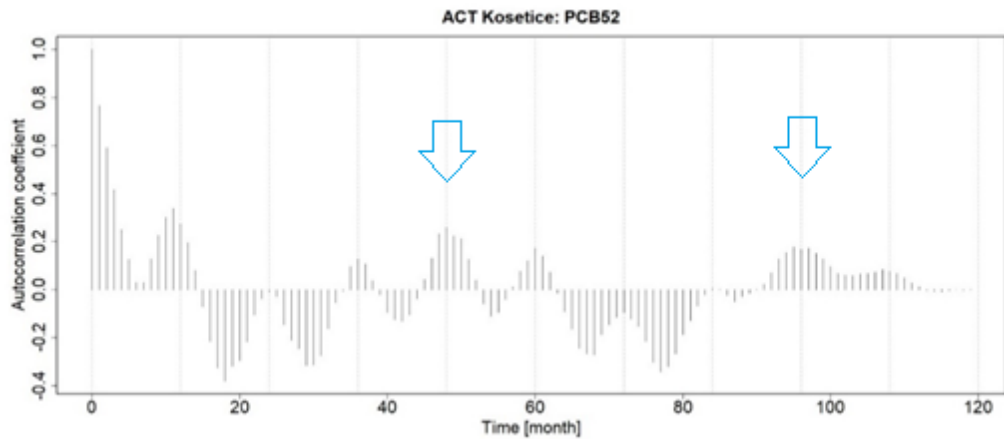
b)

Obr.14 Záznam koncentrace PCB52 z a) aktivního měřiče a b) jeho autokorelační posloupnosti; c) záznam koncentrace PCB62 z pasivního měřiče

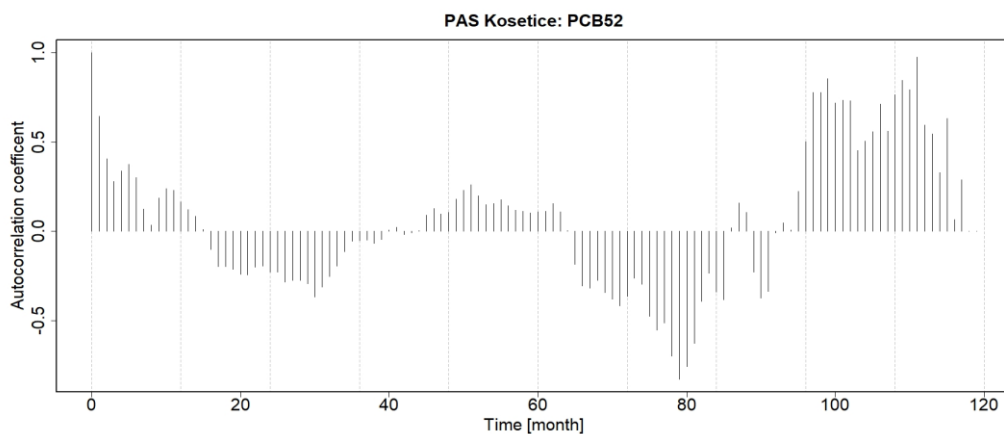
přisoudit, kromě nejistoty způsobené nevychýleným odhadem, výskytu rychlejších oscilací na začátku i konci obou originálních posloupností.

Trochu jiný pohled nabízí korelační analýza záznamů koncentrace PCB52.

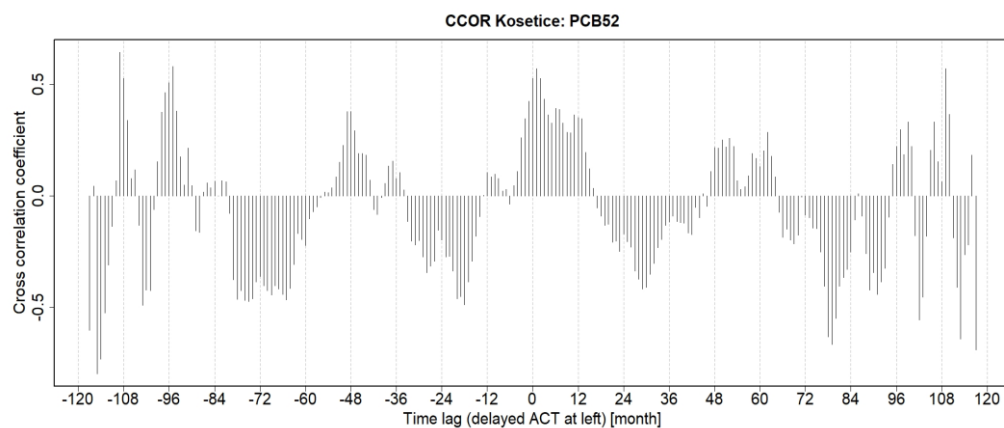
Autokorelační posloupnost záznamu z aktivního měřiče (obr.14a) vykazuje kromě standardní roční periodicity i jasné pomalé kolísání s čtyřletou periodou (posun o 48 vzorků, resp. o 96 vzorků) (obr.15a). Maximum autokorelační posloupnosti pro tuto hodnotu posunu je značně větší než okolní extrémů pro posun o 36 a 60 vzorků, byť by se při výpočtu algoritmem pro vychýlený odhad měla maxima spíše zmenšovat. Podporu pro takovou periodicitu nabízí i druhý pozitivní extrém pro posun o 96 vzorků, tedy právě dvojnásobek 48. Že by v originálních datech mohla tato periodicitu být, naznačují v obr.14 modré šipky v místech odhadu lokálních



a)



b)



c)

Obr.15 Autokorelační posloupnosti koncentrace PCB52 z a) aktivního měřiče a b) jeho autokorelační posloupnost a c) jejich vzájemná korelační posloupnost

maxim v obr.14a. A tak je otázkou další, tentokrát již ani ne matematické analýzy, zda je toto zjištění jen artefakt, či zda existují jiné důvody, proč by tak mohlo objektivně být. Kupodivu se podobná periodičita, ovšem s lehce prodlouženou periodou vyskytuje i v datech z pasivního měřiče (obr.14b). To prodloužení může být způsobeno relativně vysokou koncentrací PCB52 v letech 2009 až 2011. Pomoc pro zkoumání tohoto jevu by mohla nabídnout i frekvenční analýza, ale to už je teď mimo tuto kapitulu s příklady výpočtu korelačních posloupností, ale můžeme ji použít v další kapitole.

V autokorelační funkci dat z pasivního měřiče jasně převládá uvedené pomalé kolísání úrovně, byť i roční periodičita a jí odpovídající lokální maxima jsou v jejím průběhu patrné, i když ne příliš výrazně. Navíc se ta lokální maxima nevyskytují přesně v násobcích 12, což lze přičíst i jistému kolísání délky „ročních“ period v originálních datech.

Vzájemná korelační posloupnost už vůči hodnotě nulového posunu symetrická, ani zhruba, není. Dokonce ani extrém v okolí nulového posunu. Proč tomu tak je, je třeba posoudit globálně z obou celých záznamů, což je na základě pouze vizuálního zkoumání značně obtížné. (Pomůckou v tom případě může být připomínka, že vzhledem k použitému vzorci, kladné hodnoty posunu ve vzájemné korelační posloupnosti odpovídají posunu záznamu z pasivního měřiče vpravo.) Zhruba lze snad jen konstatovat, že podobnost obou posuzovaných originálních záznamů je pro malé hodnoty kladného posunu větší než pro jeho záporný směr.

V průběhu korelační posloupnosti je dominantní pomalé kolísání (s výše zmíněnou periodičitou cirká 48 vzorků). To je známkou významné přítomnosti této frekvenční složky v obou originálních posloupnostech.

Lokální maxima odpovídající roční oscilaci (násobky 12) jsou v popisované korelační posloupnosti rovněž přítomná. Jsou však superponována na křivce reprezentující výše zmíněnou nízkou frekvenci, a tudíž ne tolik výrazná.

Zvýšená frekvence oscilací na krajích vzájemné korelační posloupnosti je dána zvýšenou frekvencí na začátcích obou originálních záznamů. Zvětšení výchylek v těchto úsecích posloupnosti vyplývá ze způsobu výpočtu, tj. pomocí nevychýleného odhadu korelační posloupnosti.

□□□