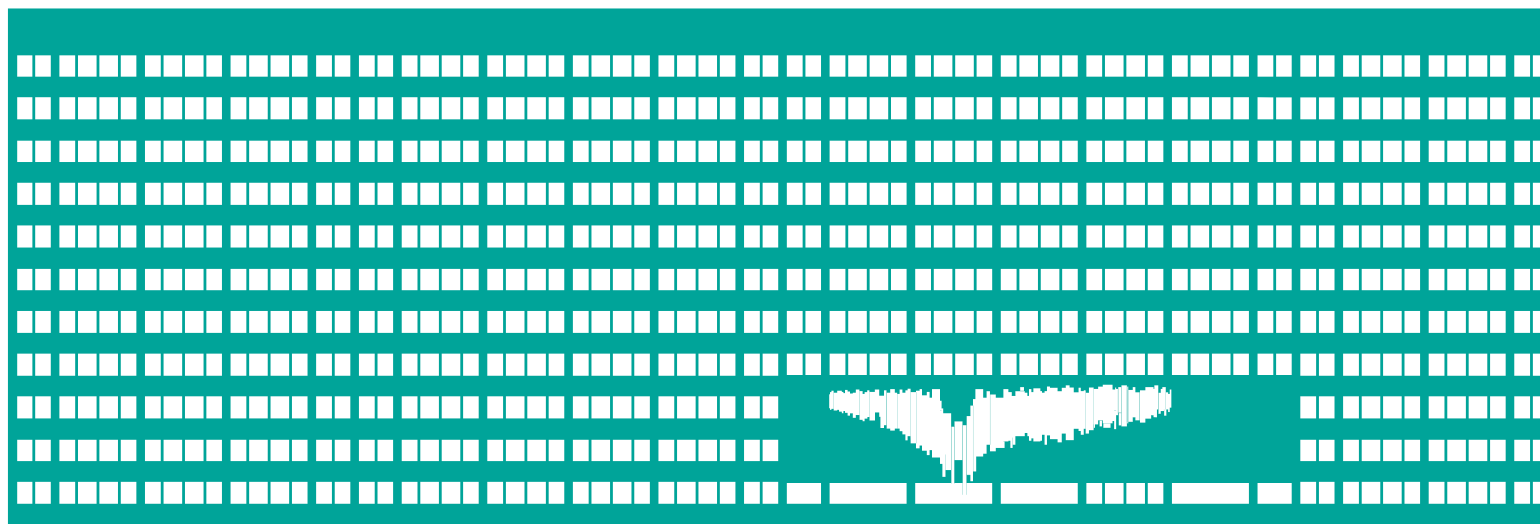


VŠB TECHNICKÁ
UNIVERZITA
OSTRAVA

VSB TECHNICAL
UNIVERSITY
OF OSTRAVA



www.vsb.cz

Základy geoinformatiky 10

Získávání geodat

Michal Kačmařík

A924, tel.: 5512

<http://gis.vsb.cz>

<https://gis.vsb.cz/pracoviste/lide/kacmarik/>

E-mail: michal.kacmarik@vsb.cz

Osnova přednášky

- Úvod
- Signály, informace
- Získávání geodat
- Primární geodata
 - Rozlišovací úroveň
 - Proces získávání hodnot primárních geodat fyzikálně-chemických veličin
 - Měření, vzorkování, kvantování, kódování
 - Proces získávání geodat ostatních veličin

Objekty, jevy, události a procesy reálného světa jsou popisovány geodaty.

Geodata můžeme získávat:

- přímým měřením vlastností objektů, jevů, událostí a procesů,
- nepřímým měřením těchto vlastností,
- průzkumem,
- následným zpracováním geodat získaných výše uvedenými postupy.

Z jiného pohledu můžeme geodata rozdělit na:

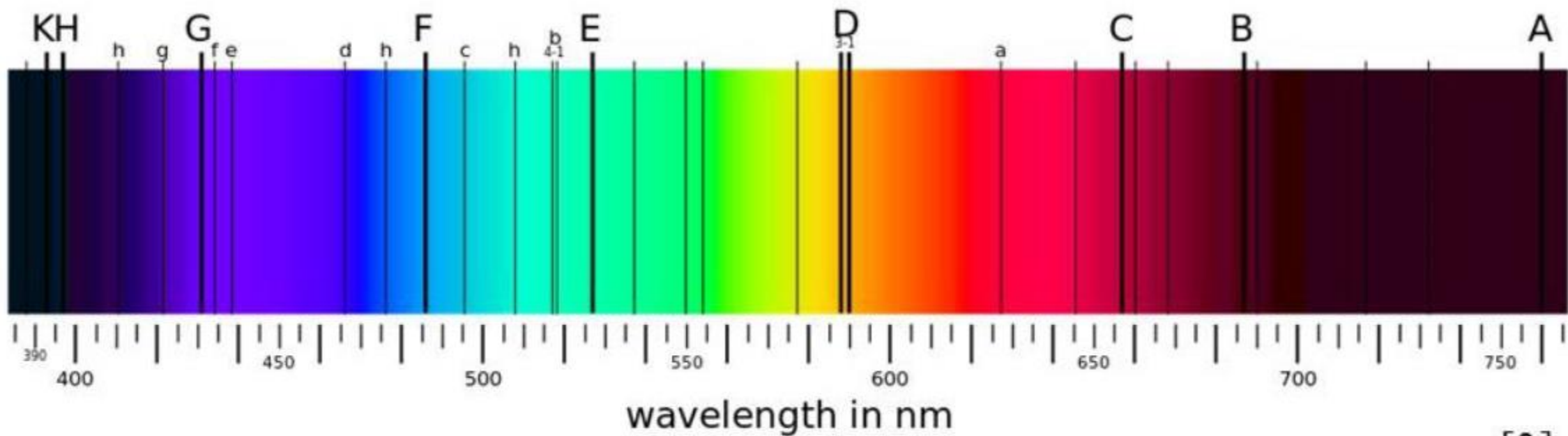
- **Geodata primární** – získávaná přímo v kontaktu s popisovanou realitou a získávaná pro daný účel. Požizovatel/uživatel je zpravidla nijak nemodifikuje. Tato geodata mohou být získávána:
 - **přímým měřením** – v tom případě měříme přímo hodnoty sledované vlastnosti,
 - **nepřímým měřením** – jedná se zpravidla o získávání hodnot vlastností, které nelze získat přímo, ale pouze odvozením z hodnot jiné vlastnosti, která je měřitelná přímo; příkladem může být hustota obyvatel, kterou nelze přímo měřit, ale lze ji odvodit ze znalosti rozložení obyvatelstva v daném prostoru.
- **Geodata sekundární** – data odvozená z primárních geodat různými analýzami. Jsou obvykle informačně mnohem bohatší (resp. informace jsou v nich obsaženy v mnohem koncentrovanější podobě).
- Základním rozdílem mezi primárními a sekundárními zdroji geodat tedy je, v jakém rozsahu máme pod kontrolou *co je sbíráno a jakým způsobem*.

- vytvoření přesné a dobře zdokumentované databáze geoinformačního systému je kritickým místem jeho tvorby
- takovou databázi lze vybudovat pouze z kvalitních a dobře zdokumentovaných geodat, aby bylo možné kdykoliv posoudit vhodnost jejich použití pro různé účely

Signály, informace

- signály jsou funkce jedné nebo více nezávisle proměnných
- nesou informace:
 - o podstatě a vlastnostech svého zdroje
 - o podstatě a vlastnostech prostředí a
 - informace záměrně do signálu zakódované
- informace jsou v signálu reprezentovány změnami okamžité hodnoty fyzikální veličiny, kterou se signál představuje

Viditelné záření



[3]

https://images.slideplayer.cz/42/11335178/slides/slide_11.jpg

Signály, informace

Signálem můžeme obecně rozumět libovolný **fyzikální fenomén**, který může být **modelován funkcí prostoru nebo času** a je použitelný pro **přenos informací**.

Signály mohou vznikat v nejrůznějších zdrojích.



- **jednorozměrný (1D)** – je obvykle reprezentován časově závislou fyzikální veličinou (například průtok vody přehradou) nebo záznamem hodnot sledované veličiny podél nějaké linie (například záznam stavu železničního svršku pořízený diagnostickým vozem)
- **dvourozměrný (2D)** – obvykle je definován ve 2D prostoru (například reliéf terénu nebo rozložení znečištění těžkými kovy v půdách apod.)
- **třírozměrný (3D)** – obvykle charakterizuje prostorové (3D) objekty (například rozložení koncentrace užitečné suroviny v prostoru znázorněná v podobě 3D modelu ložiskového tělesa)
- **n-rozměrný** – bereme v úvahu prostor, čas i jiné parametry.

Z hlediska závislosti na čase může být signál:

- **statický** – signál není závislý na čase (viz. například vytištěná mapa, letecký snímek apod.)
- **dynamický** – takovýto signál je funkcí času (například záznam intenzity dopravy na křižovatce apod.)

Z hlediska spojitosti rozlišujeme signály:

- **spojité** – v tom případě jsou modelovány spojitou funkcí prostoru nebo času,
- **diskrétní** – modelované diskrétní funkcí prostoru nebo času.

- **Přenášená informace** reprezentovaná některým parametrem signálu může teoreticky nést jen **užitečný obsah** a představovat tak vlastní přenášenou zprávu, ale v praxi zpravidla obsahuje i **nežádoucí část**, kterou nazýváme **rušení** nebo častěji **šum** (angl. noise).

Vztah mezi informací a signálem můžeme popsat takto:

- jedinou informaci můžeme přenášet prostřednictvím různých signálů a současně
- jeden signál může přenášet i více informací

Pro popis signálů můžeme používat dva základní přístupy:

- **deterministický**
- **stochastický.**

Deterministickým nazýváme takový popis, v němž je chování signálu předem přesně dané, tj. pro každý časový okamžik (nebo pro každé místo) lze vypočítat jeho hodnotu.

Matematický model je v tomto případě často založen na funkcích definovaných na číselných množinách.

Deterministický popis signálu

Funkci f reprezentující deterministický popis signálu můžeme vyjádřit jako zobrazení prvků množiny A na prvky množiny B a schematicky vyjádřit:

$$(f : A \rightarrow B)$$

kde množina A reprezentuje **definiční obor** a množina B **obor hodnot**.

- signál, jehož definičním oborem i oborem hodnot jsou nekonečné a nespočítatelné číselné množiny, nazýváme *analogový* nebo též *spojitý*.
- signál, jehož definičním oborem i oborem hodnot jsou nekonečné, ale spočítatelné číselné množiny, se nazývá *číslicový* resp. *digitální*.

Zvláštní třídou deterministických signálů jsou signály periodické

$$a(x + kP) = a(x)$$

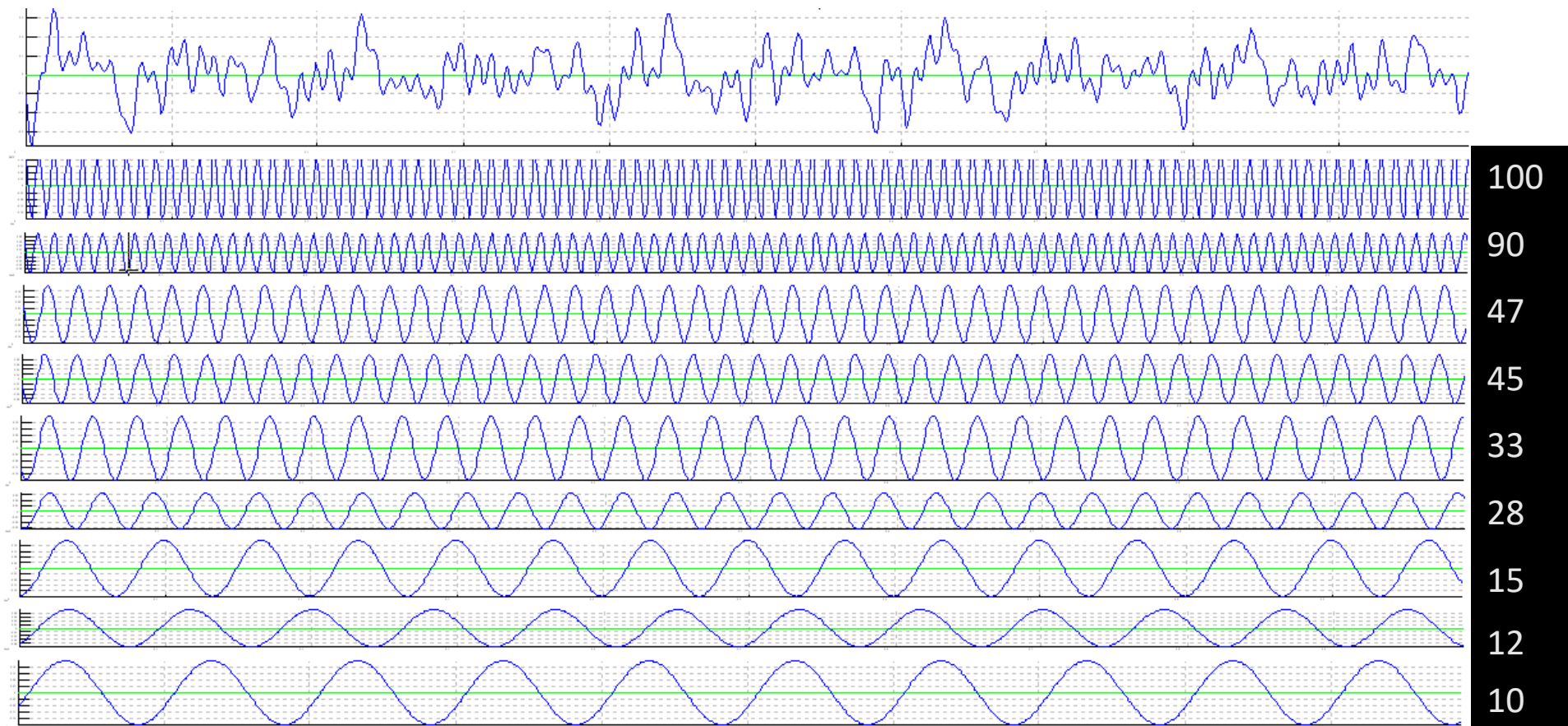
- Vedle deterministických signálů se setkáváme i se signály, u kterých neumíme přesně určit (vypočítat), jaké hodnoty v daném čase nebo místě nabudou.
- Jejich časový resp. prostorový průběh není možné popsat matematickým vztahem, jejich hodnoty jsou **náhodné** (stochastické).
- vznikají nějakým fyzikálním pochodem, který je ovlivňovaný náhodnými okolnostmi a podmínkami jejich vytváření
- případně může dojít i k situaci, kdy je ryze deterministický signál ovlivněn (rušen) náhodným šumem
- takové signály označujeme jako **stochastické** nebo **náhodné**

Stochastický popis signálu

- popisujeme je buďto pravděpodobnostními nebo statistickými metodami
- statistické charakteristiky určujeme z jeho časového nebo prostorového průběhu
- nejpoužívanějšími statistickými charakteristikami jsou například střední hodnota, rozptyl, autokorelační funkce, spektrální hustota apod.

Některé pojmy z oblasti zpracování signálů

- Fourierova transformace** vyjadřuje daný signál (resp. jej modelující funkci) jako součet sinusových funkcí různé periody násobených koeficienty vyjadřujícími jejich amplitudy.



Ukázka vyjádření daného signálu (nahore) jako součtu sinusových signálů s různou periodou a amplitudou. Čísla vpravo udávají frekvenci jednotlivých sinusových signálů v [Hz]

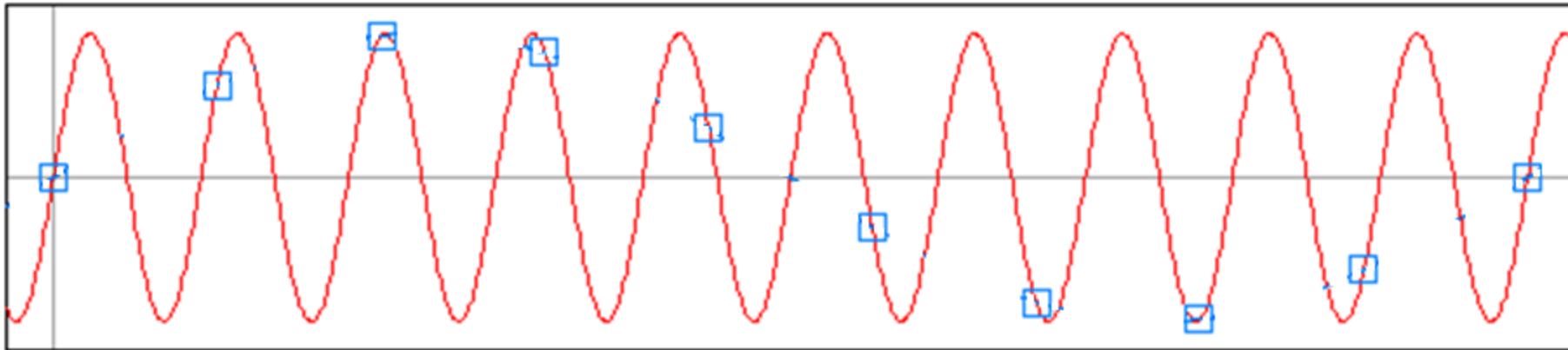
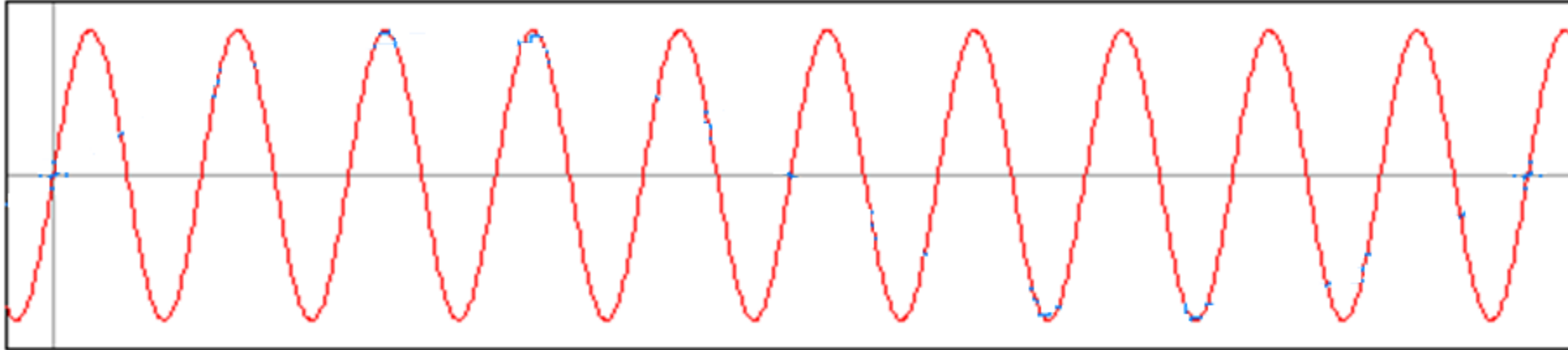
Některé pojmy z oblasti zpracování signálů

- **Fourierova analýza** se zabývá nalezením vhodné Fourierovy transformace pro analyzovaný signál.
- **Frekvenční spektrum** vychází z výsledků Fourierovy analýzy a vyjadřuje závislost amplitudy dané složky signálu (dané sinusové funkce) na její frekvenci (resp. periodě).

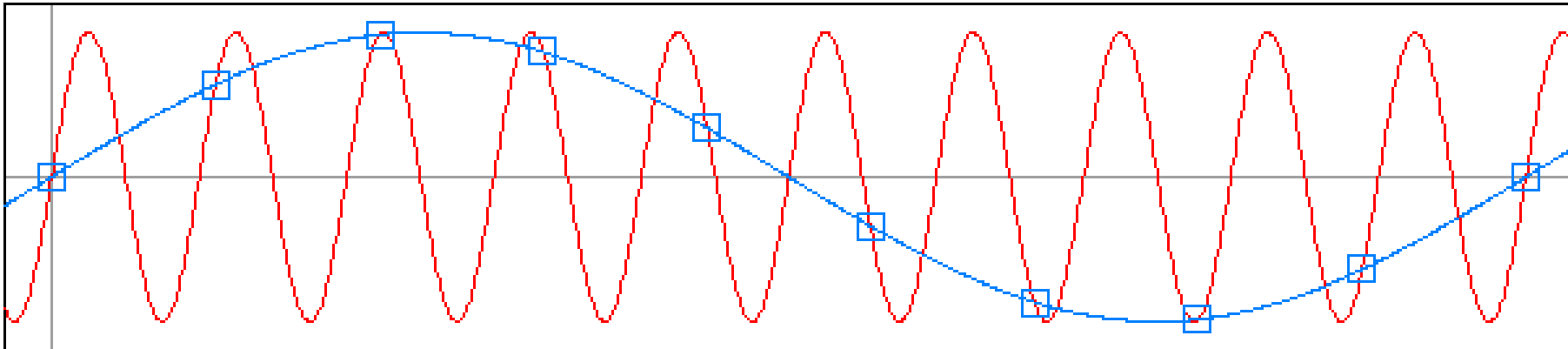
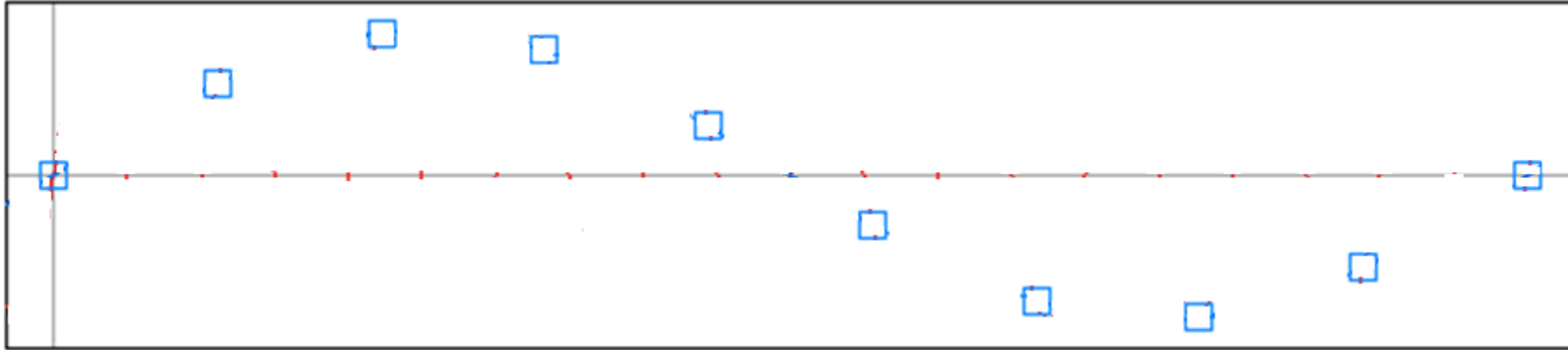
Některé pojmy z oblasti zpracování signálů

- **Aliasing** je jev spojený se vzorkováním signálu a je důsledkem jeho **podvzorkování**. Při rekonstrukci signálu ze vzorků nejsme schopni jednoznačně rozhodnout o skutečném průběhu původního signálu.

Některé pojmy z oblasti zpracování signálů



Některé pojmy z oblasti zpracování signálů



Některé pojmy z oblasti zpracování signálů

- **Interference** vzniká superpozicí dvou nebo více signálů se stejnou nebo velice blízkou frekvencí.

Získávání geodat

Získávání geodat

V první řadě můžeme vlastnosti, popisované geodaty (a tím i geodata jako taková) rozdělit na:

- kvalitativní
- kvantitativní

Získávání geodat

- **Kvalitativní geodata** je možné definovat jako kvalitu odlišující jeden objekt reálného světa od druhého, s nízkou až žádnou variabilitou v rámci objektu reálného světa. Příkladem může být využití půdy, druh průmyslu, půdní typ, klimatická zóna, vegetační pokryv.
- Získávají se zpravidla průzkumem.

Získávání geodat

- **Kvantitativní geodata** jsou na druhé straně měřená (ať už přímo nebo nepřímo) a jsou jim přiřazované numerické hodnoty vztažené k určitému standardu – **měrné jednotce**.
- Umožňují poměřovat individuální jevy mezi sebou. I v rámci jednoho geopravku mohou vykazovat značnou variabilitu.
- Příkladem proměnné může být nadmořská výška, koncentrace znečišťujících látek v půdě a ve vodě, vzdálenost nebo velikost částic.

Primární kvantitativní geodata můžeme podle způsobu jejich získávání rozdělit na:

- **geodata měřitelná různými fyzikálními nebo chemickými postupy** a tedy získatelná různými snímači či laboratorními měřeními a analýzami (obvykle měření v terénu či laboratoři)
- **geodata získávaná pozorováním resp. průzkumem**, tedy bez využití jakýchkoliv fyzikálních či chemických měření (např. dotazníkové šetření).

Popis objektů reálného světa probíhá současně v několika doménách:

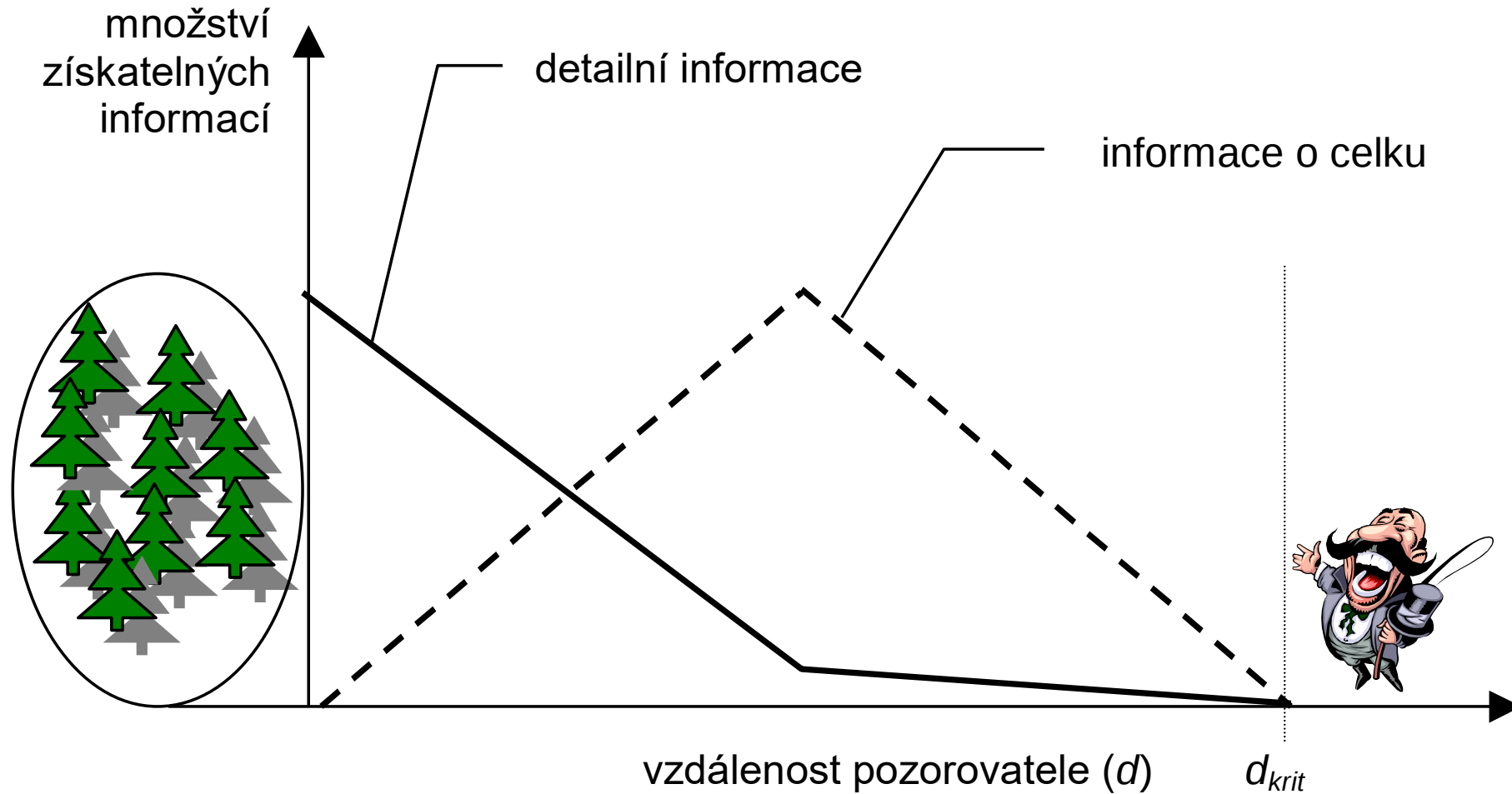
- **prostorová** – prostorové vymezení místa, které popisujeme,
- **časová** – časové vymezení okamžiku (resp. doby), kdy máme popis provádět, případně jak často,
- **tematická** – vymezení, které atributy jsou předmětem popisu.

Pro každou z těchto domén můžeme definovat i tzv. **rozlišovací úroveň**.

Rozlišovací úrovně

- prostorové rozlišení
- časové rozlišení
- tematické rozlišení

Prostorové rozlišení



Prostorové rozlišení, které si zvolíme pro získávání geodat, by mělo odpovídat:

- velikosti popisovaných geoprvků,
- prostorové tvářnosti modelovaného objektu reálného světa (tj. proměnlivosti jeho geometrických i tematických vlastností v prostoru),
- účelu, pro který jsou geodata pořizována,
- ekonomickým možnostem organizace zadávající sběr geodat.

Časové rozlišení, které si zvolíme pro získávání geodat, by mělo vycházet z:

- proměnlivosti geometrických a tematických vlastností popisovaného objektu reálného světa v čase
- účelu, pro který jsou geodata pořizována (a z něj vyplývajících nároků na aktuálnost/aktualizaci geodat)
- ekonomických možností organizace zadávající sběr geodat.

Tematickou rozlišovací úrovní rozumíme:

- výběr tematických vlastností objektů reálného světa, pro které budeme získávat geodata,
- výběr hodnotových domén pro jednotlivé vlastnosti.

Výběr by měl odpovídat:

- účelu, pro který jsou geodata pořizována,
- ekonomickým možnostem organizace zadávající sběr geodat.

Proces získávání hodnot primárních geodat fyzikálně-chemických veličin

Proces získávání hodnot primárních geodat fyzikálně-chemických veličin

Hodnoty jednotlivých sledovaných atributů získáváme v procesu, skládajícím se ze čtyř podprocesů:

- měření
- vzorkování
- kvantování,
- kódování.

Proces získávání hodnot primárních geodat fyzikálně-chemických veličin

Měření je podproces, který sleduje měnící se hodnoty daného atributu a umožňuje určovat jejich velikost.

Vzorkování je podproces, který rozhoduje o tom, kdy bude změřená hodnota daného atributu odebrána (kdy bude odebrán vzorek) pro další zpracování v počítači.

Kvantování je podproces, který přiřadí změřené hodnotě číselnou úroveň.

Kódování je podproces, který tuto číselnou úroveň převede do podoby vhodné pro další zpracovávání (například převedení do fyzikálních jednotek apod.).

Výsledkem je získání digitálních (geo)dat, vhodných pro další zpracovávání (v geoinformačních systémech)

Voltmetr



Měření identifikované vlastnosti se zpravidla provádí pomocí tzv. **čidla** snímajícího fyzikální veličinu tuto vlastnost reprezentující.

Čidlo je definováno jako **zařízení na zpracování informace vázané na nějaký druh energie**.

Pro každé čidlo je nezbytné znát jeho charakteristiky, které popisují mimo jiné:

- vztah mezi ustálenou hodnotou snímané veličiny a výstupním signálem z čidla (tzv. **převodní charakteristiku**),
- **citlivost čidla** (tj. minimální změnu snímané veličiny, kterou je čidlo schopné zaznamenat),
- **linearitu čidla** (linearitu závislosti výstupního signálu na hodnotě snímané veličiny),
- **časovou stálost**,
- **závislost na podmínkách měření** (teplota, tlak apod.),

- Abychom mohli měřené veličiny zpracovávat prostřednictvím digitální techniky, je nezbytné jejich hodnoty převést na posloupnost číselných hodnot.
- Využívá se přitom skutečnosti, že za určitých okolností může být spojitý signál plně reprezentován posloupností vzorků, které jsou určeny okamžitými resp. lokálními hodnotami spojitého signálu získanými v přesně definovaných časových okamžicích nebo místech v prostoru.
- **Vzorkování** spočívá v odebrání určitého vzorku spojitého vstupního signálu. Přitom se provádí odečítání hodnot měřené veličiny buďto v pravidelných intervalech Δt (a pak mluvíme o **pravidelném vzorkování**) nebo v nepravidelně (náhodně) zvolených časových okamžicích (a pak mluvíme o **nepravidelném** resp. o **náhodném vzorkování**).

Základním problémem vzorkování je volba vhodného časového intervalu Δt . Ten musí být v relaci s maximální frekvencí $f_{\max}$, obsaženou ve vstupním signálu snímače měřené veličiny (reprezentuje nejrychleji se měnící složku signálu).

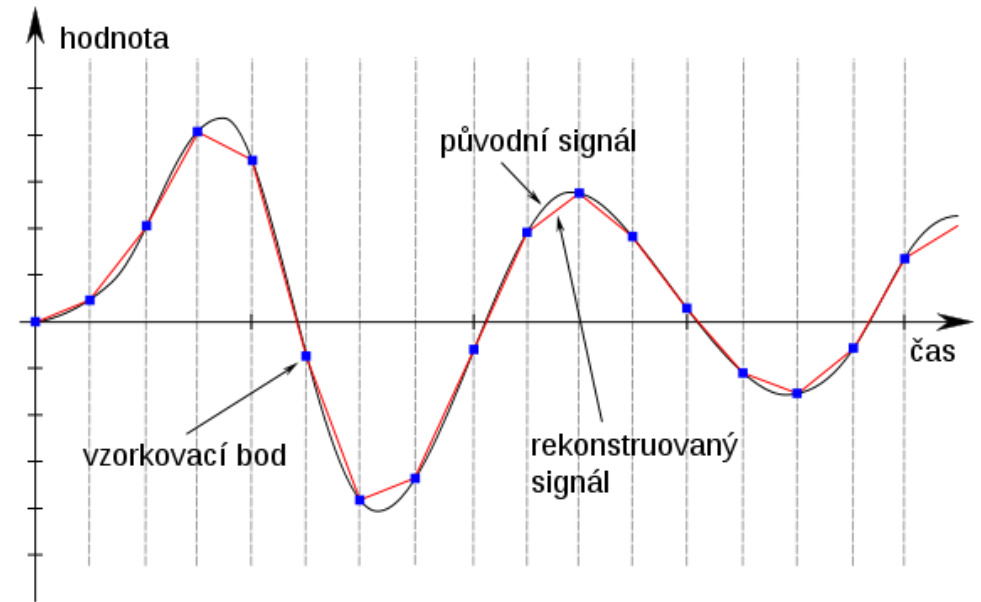
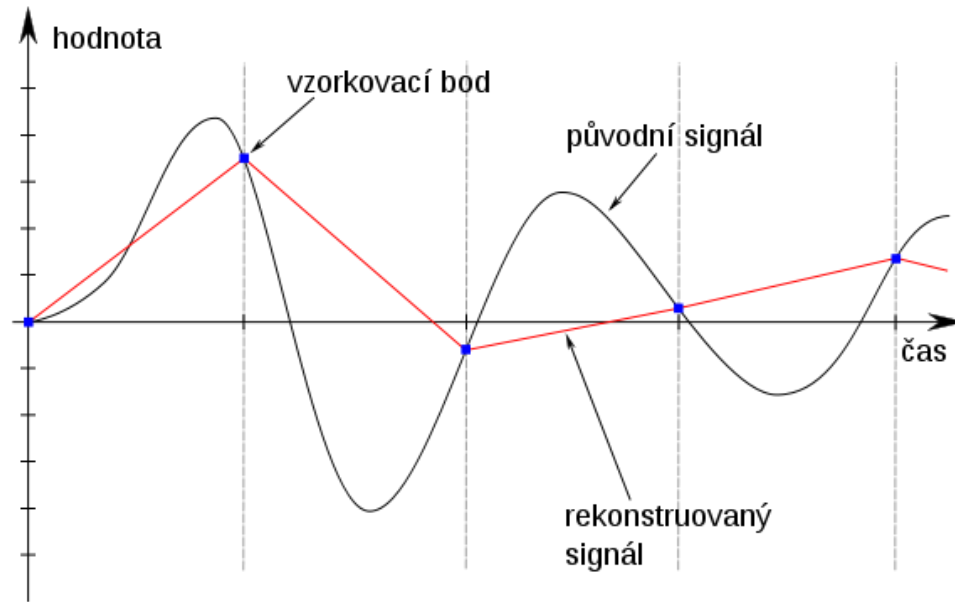
Shannonovo kritérium:

$$\frac{1}{\Delta t} = f_{vz} \geq 2 f_{\max}$$

$$f_{vz} = (2.5 \dots 3.0) f_{\max}$$

Vzorkování

- Převzorkování,
- Podvzorkování,
- Zkreslení.



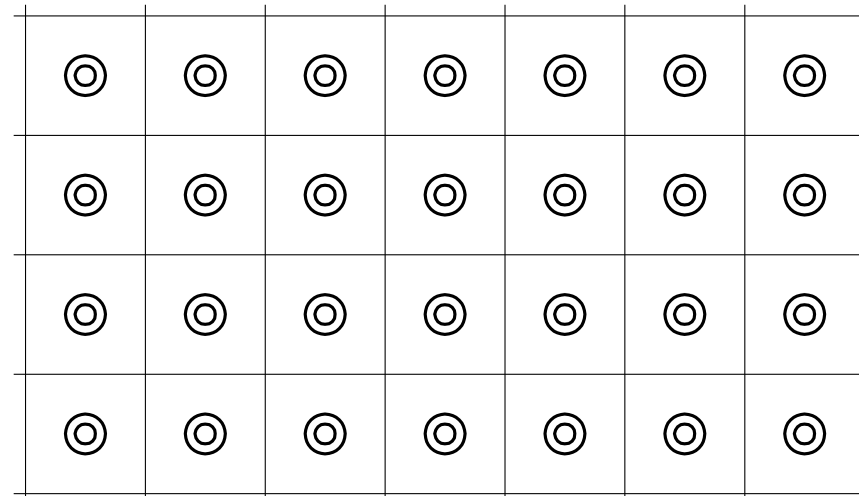
Vzorkování – prostorová doména

Na vzorkovací algoritmus máme následující požadavky:

- pokrytí vzorkované oblasti musí být rovnoměrné (na globální úrovni)
- je nežádoucí absolutní pravidelnost z důvodu možného vzniku **interference**
- výpočet rozmístění vzorků musí být efektivní.

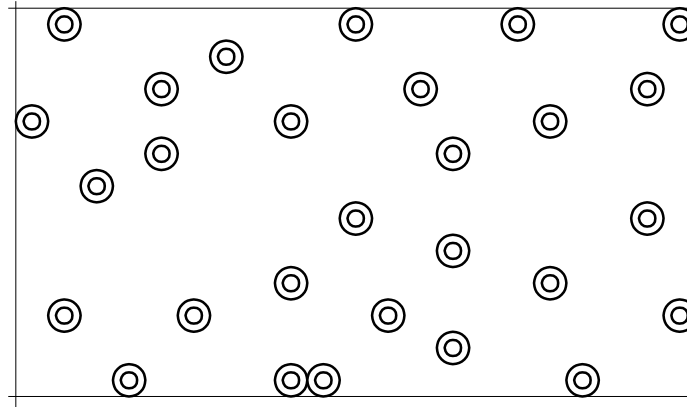
Vzorkování – prostorová doména

- **Pravidelné** – vzorky jsou rozmístěny v pravidelné síti s nejčastěji čtvercovou základní buňkou. Toto vzorkování neodstraňuje rušivé interference.



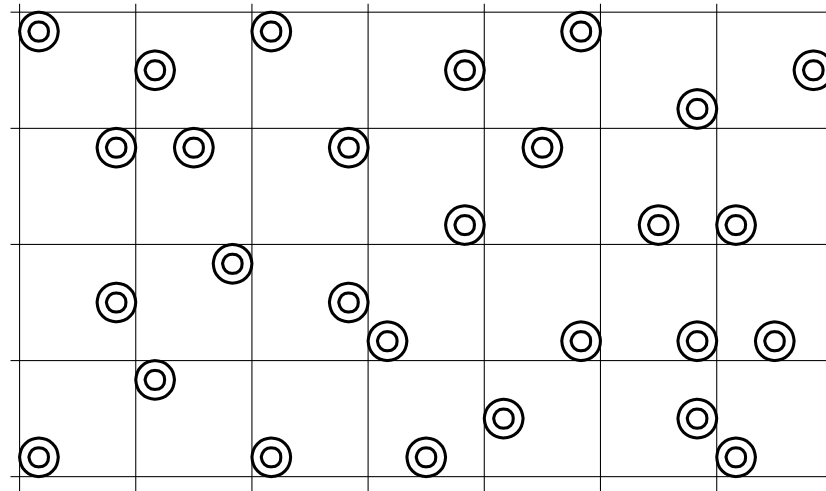
Vzorkování – prostorová doména

- **Náhodné vzorkování** – ve vzorkované oblasti je rozmístěno N nezávislých náhodných vzorků s rovnoměrným rozložením pravděpodobnosti. Vzorky mohou vytvářet větší shluky. Výsledek vzorkování je zatížen větším šumem.



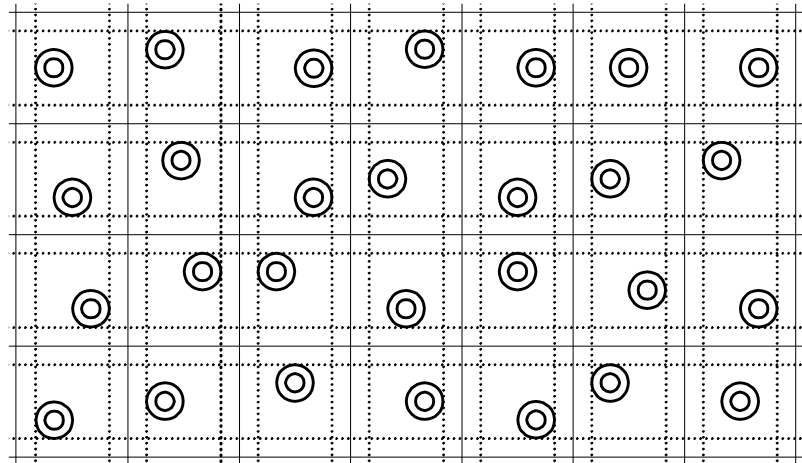
Vzorkování – prostorová doména

- **Roztřesení** – v $K \times K$ shodných podintervalech je rozmístěno $K \times K$ nezávislých náhodných vzorků. Tím se sníží pravděpodobnost větších shluků a dosáhne se rovnoměrnějšího pokrytí vzorkované oblasti.



Vzorkování – prostorová doména

- **Částečné roztřesení** – v $K \times K$ shodných podintervalech („zmenšených“ buňkách, nepokrývajících původní oblast beze zbytku) je rozmístěno $K \times K$ nezávislých náhodných vzorků. Tento postup zamezuje vzniku shluků.



Vzorkování – prostorová doména

- Někdy může být výhodné provést adaptivní zahušťování sítě vzorků. To se provádí zpravidla lokálně na základě zvýšené variability sledované veličiny.
- Tato lokální variabilita nemusí být známa předem, proto se musí vzorkovací algoritmus automaticky přizpůsobovat dosaženým výsledkům (adaptovat se).
- Příkladem může být **progresivní vzorkování** při sběru dat o reliéfu pomocí fotogrammetrie.

Vzorkování – prostorová doména

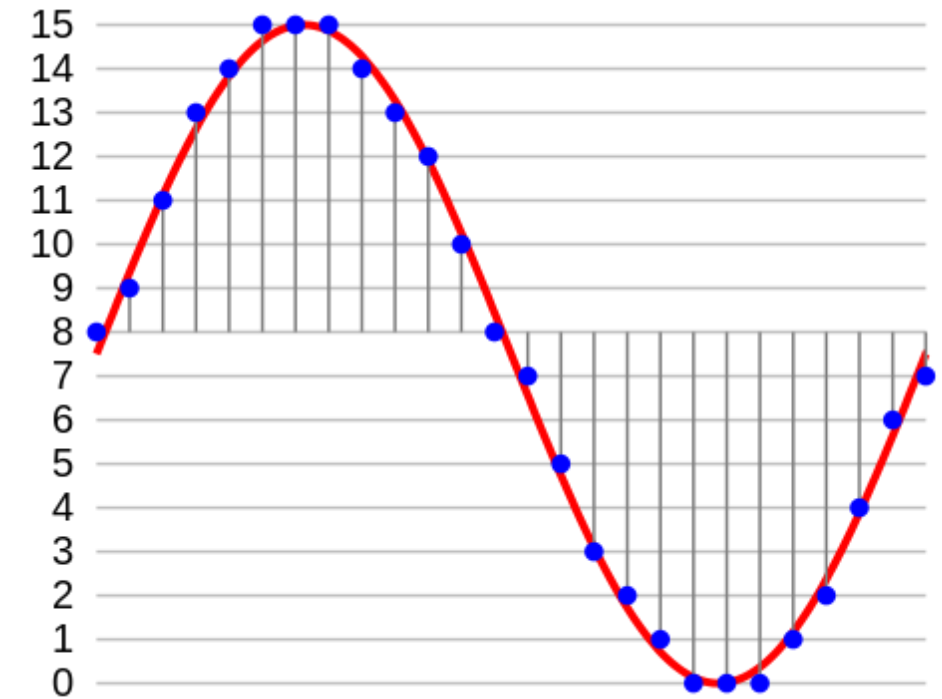
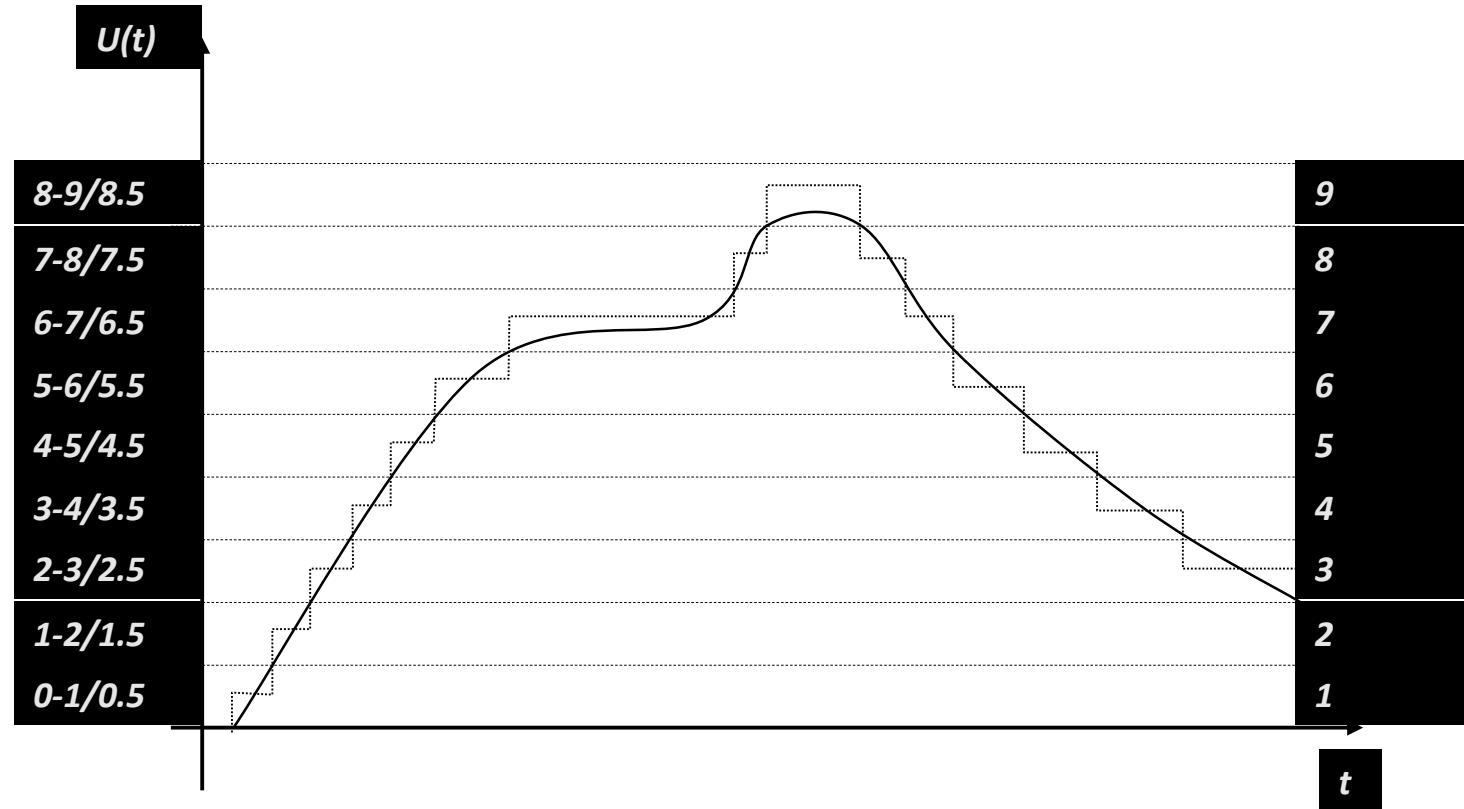
- **Selektivní vzorkování** nevychází z žádného apriorního vzoru rozmístění vzorků v prostoru, jednotlivé vzorky se rozmisťují tak, aby co nejlépe postihly variabilitu sledované vlastnosti.

Můžeme rozlišit několik strategií vzorkování:

- pravidelné opakování měření vždy po uplynutí zvoleného časového intervalu – jedná se o **pravidelné vzorkování**,
- nepravidelné opakování měření vždy po uplynutí náhodně zvoleného časového intervalu – **náhodné vzorkování**,
- průběžné sledování změn a zaznamenávání jen významných změn – obdoba **selektivního vzorkování**.

- Kvantování znamená, že je vytvořen omezený počet **kvantovacích intervalů**, které plně pokrývají rozsah měření a jsou navzájem disjunktní. Pro každý interval je dána minimální a maximální hodnota a dále střední hodnota, kterou daný interval reprezentuje.

Kvantování



Děkuji za pozornost

Michal Kačmařík

michal.kacmarik@vsb.cz

www.vsb.cz