

# STUDIJNÍ MATERIÁL 1:

## Metoda rozložení citlivosti druhů (Species Sensitivity Distribution; SSD)

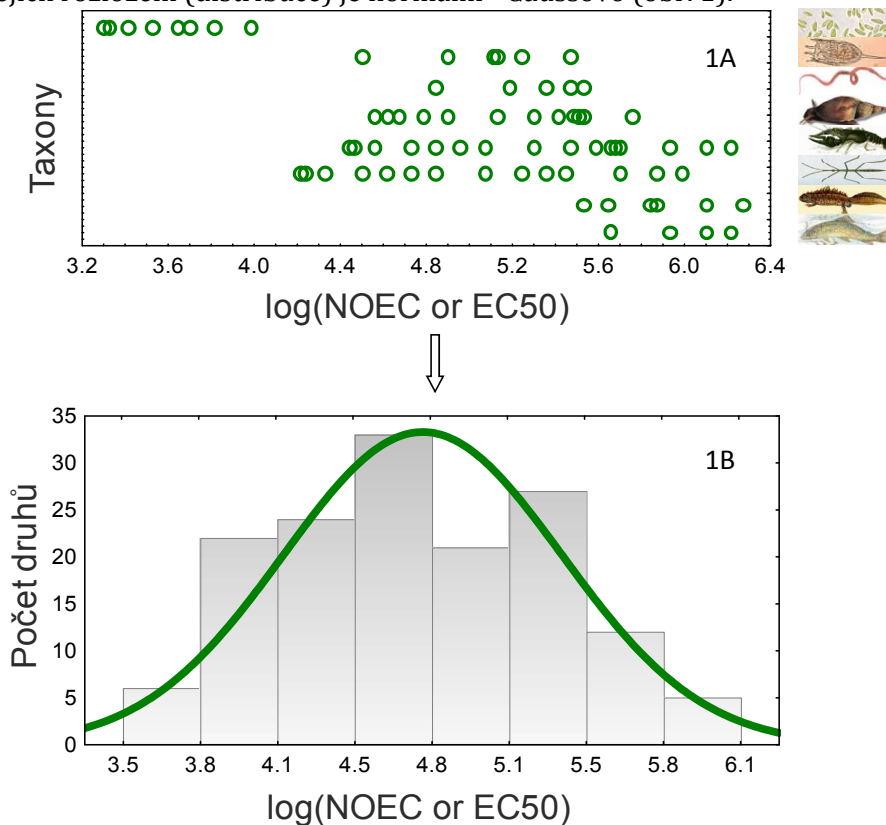
Mgr. Soňa Smetanová, RECETOX, 2013

### 1. Úvod do SSD

SSD model byl poprvé představen v roce 1989, ale všeobecné známosti se dočkala až s rozvojem výkonnějších počítačů (kolem roku 2000). Je hojně využíván pro hodnocení účinků a rizik toxických látek v ekosystémech a pro stanovení limitů toxických látek v prostředí a uplatňuje se ve směrnicích EU (např. v rámcové směrnici o vodách).

Model je založen na předpokladu, že různé druhy organismů žijící v určitém ekosystému (např. řeka, jezero či půda) jsou různě citlivé na určitou toxickou látku (např. koryš *Daphnia magna* je na rozdíl od řasy *Raphidocelis subcapitata* velmi citlivý na insekticidní látku chlorpyrifos). Tyto citlivosti jsou vyjádřeny nejčastěji jako NOEC či EC50 hodnoty.

Základním předpokladem SSD je, že pokud se vezmou citlivosti (**logaritmy** NOEC či EC50 hodnot) všech druhů organismů (žijících v daném ekosystému) na určitou toxickou látku, pak jejich rozložení (distribuce) je normální - Gaussovo (obr. 1).



Obr. 1. Princip SSD

1A: Různé druhy jsou různě citlivé na toxickou látku (tj. mají různé hodnoty NOEC či EC50). Citlivost se může výrazně lišit mezi druhy a také mezi vyššími taxonomickými skupinami (zde jsou například nejvíce citlivé řasy a naopak nejméně citliví jsou obratlovci – ryby a obojživelníci); zelená kolečka značí různé druhy organismů, osa y značí různé taxonomické skupiny.

1B: Distribuce (rozložení) těchto citlivostí (resp. logaritmů EC50 nebo logaritmů NOEC hodnot) má vlastnosti rozložení normálního (Gaussova); sloupce značí počet druhů s daným rozmezím logEC50 (či logNOEC) hodnot, zelená čára označuje proložení Gaussovou funkcí.

Toto rozložení je symetrické a definované pouze dvěma parametry – aritmetickým průměrem a směrodatnou odchylkou a proto je velmi jednoduché s ním pracovat (rovn. 1 a 2, obr. 2):

Rovnice pro výpočet aritmetického průměru  $\mu$  a směrodatné odchylky  $\sigma$  pro SSD:

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i$$

(1)

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2}$$

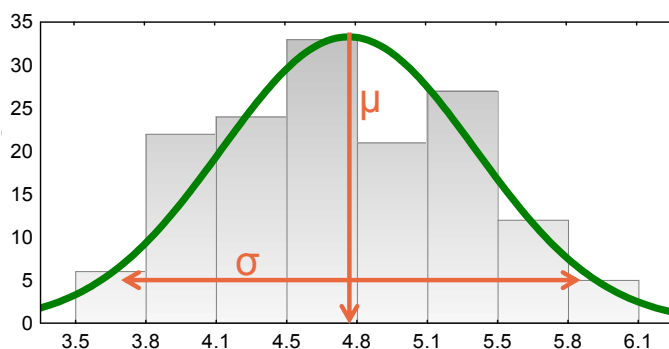
(2)

$\mu$  ... aritmetický průměr

$\sigma$  ... směrodatná odchylka

$n$  ... celkový počet druhů

$x_i$  ... citlivita i-tého druhu (logEC50 či logNOEC hodnota) na konkrétní toxickou látku



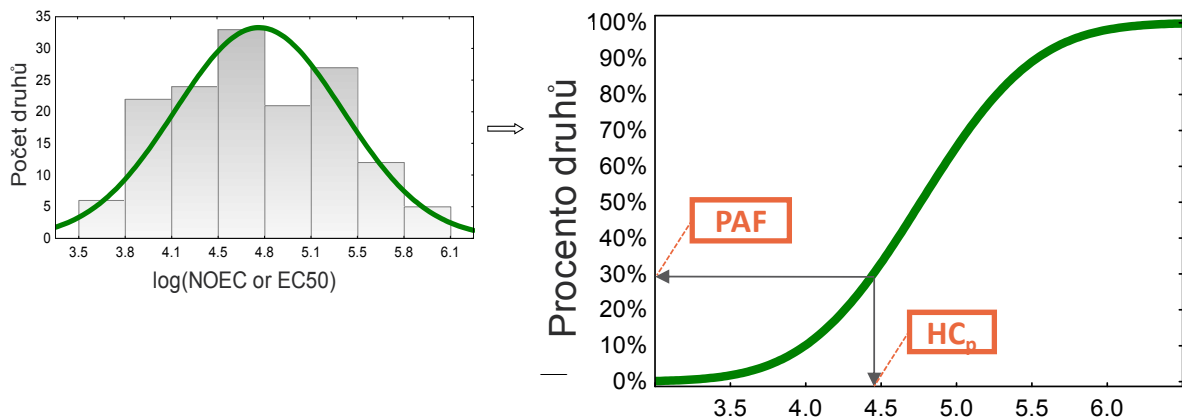
**Obr. 2. Parametry normálního (Gaussova) rozložení**

Směrodatná odchylka ( $\sigma$ ) popisuje míru variability dat a aritmetický průměr ( $\mu$ ) jejich střední hodnotu

Výše popsané Gaussovo rozložení se dá vyjádřit také jako kumulativní distribuční funkce (ukazuje totéž s tím rozdílem, že na ose y nejsou četnosti druhů s danou kategorií citlivosti, ale procento organismů s danou nebo vyšší citlivostí (např. danou nebo nižší hodnotou EC50). A z této kumulativní distribuční funkce se mohou lehce vypočítat hodnoty  $HC_p$  či PAF (obr. 3).

**$HC_p$  (Hazard Concentration for p-percent of species)** je taková koncentrace dané toxické látky, která podle SSD modelu negativně ovlivní p-procent druhů organismů (například  $HC_{15}$  metanolu je taková koncentrace metanolu, která má negativní efekt na 15 % všech druhů organismů žijících v jednom ekosystému).

**Oproti tomu PAF (Potentially Affected Fraction)** označuje frakci (nebo procento) organismů, která bude negativně ovlivněna působením určité koncentrace dané toxické látky (tudíž při působení metanolu v koncentraci rovné  $HC_{15}$  bude PAF rovno 15 %).



**Obr. 3.** Zobrazení SSD jako kumulativní distribuční funkce a odvození PAF (Potentially Affected Fraction) a HC<sub>p</sub> (Hazard Concentration for p percent of species). Oba tyto údaje se tedy týkají stejné věci (negativní efekt na různě velkou část druhů), ale popisují ji z opačných konců.

HC<sub>p</sub> se používá pro prospektivní analýzu rizik, a sice pro stanovení PNEC (Predicted no-effect concentration) limitů – maximálních koncentrací toxických látek, které ještě nebudou mít významné negativní účinky na ekosystémy. V Evropě byla dohodnuta tato hranice na 5 %. Limitní „bezpečná“ koncentrace určité toxické látky v životním prostředí je tedy taková, která negativně ovlivní 5 % druhů a je označována jako HC<sub>5</sub>.

Stanovení PNEC hodnot SSD metodou má oproti ostatním metodám výhodu v menších nejistotách (limit odvozujeme za použití velkého množství taxonů a druhů, nejen z otestování tří standardních organismů, jak tomu je v případě klasických postupů). Proto i faktory nejistoty (Assessment factors) aplikované na hodnoty HC<sub>5</sub> pro získání PNEC hodnot jsou řádově nižší (tab. 1).

**Tab. 1.** Faktory nejistoty, kterými musí být poděleny maximální „bezpečné“ zjištěné koncentrace látek pro získání PNEC hodnot; tyto faktory se liší podle zvolené metody. (tabulka převzata z guidance dokumentu EU CIS-WFD No. 27 pro stanovení norem environmentální kvality).

Available data	Assessment factor
At least one short-term L(E)C50 from each of three trophic levels (fish, invertebrates (preferred <i>Daphnia</i> ) and algae) (i.e. base set)	1000 <sup>a)</sup>
One long-term EC10 or NOEC (either fish or <i>Daphnia</i> )	100 <sup>b)</sup>
Two long-term results (e.g. EC10 or NOECs) from species representing two trophic levels (fish and/or <i>Daphnia</i> and/or algae)	50 <sup>c)</sup>
Long-term results (e.g. EC10 or NOECs) from at least three species (normally fish, <i>Daphnia</i> and algae) representing three trophic levels	10 <sup>d)</sup>
Species sensitivity distribution (SSD) method	5-1 (to be fully justified case by case) <sup>e)</sup>
Field data or model ecosystems	Reviewed on a case by case basis <sup>f)</sup>

Hodnoty PAF se naopak používají pro kvantitativní zhodnocení ekotoxikologického stavu ekosystému (tj. pro retrospektivní analýzu rizik). Po změření koncentrace určité toxické látky (např. v řece) se aplikuje model SSD a vypočítá se hodnota PAF. Výstupem je nejen informace, jestli tato hodnota přesáhla onen stanovený limit 5 %, ale také informace o míře poškození (např. látka A mající PAF = 6 % sice přesahuje limit, ale rozhodně páchá méně škod než látka B

s PAF = 62 %), což je opět velká výhoda oproti tradičnímu hodnocení, které je pouze kvalitativní („koncentrace látky v prostředí je pod/nad PNEC hodnotou“).

SSD je tedy statistický model velice vhodný pro hodnocení ekotoxikologických rizik a účinků toxických látek v prostředí. V následující kapitole se dozvíme, jaká vstupní data musíme mít, abychom takový model mohli vytvořit.

## 2. Ekotoxikologická data vhodná pro SSD

Jak je již zmíněno výše, vstupními daty pro vytvoření SSD modelu jsou nejčastěji hodnoty NOEC či EC50. Původně byly využívány spíše hodnoty NOEC, které označují takovou koncentraci látky, která nemá statisticky významný chronický účinek na organismy. Pokud se vytvoří SSD model z NOEC hodnot, pak např. PAF = 5 % tedy znamená, že 5 % druhů je chronicky ovlivněno „více než nevýznamně“. V poslední době se ovšem od používání NOEC hodnot všeobecně upouští a proto se stále více i v SSD modelech využívají akutní EC50 hodnoty. EC50 je taková koncentrace látky, která akutně způsobuje 50% efekt na organismech. Pokud se vytvoří SSD model z EC50 hodnot, pak např. PAF = 5 % znamená, že 5 % druhů je akutně ovlivněno „minimálně z padesáti procent“. Tyto rozdíly si je nutné vždy uvědomit. V následujícím textu již bude vše vysvětlováno na příkladu použití akutních EC50 hodnot.

Data ekotoxicity se nejčastěji získávají z online databází (např. US EPA ECOTOX: <http://cfpub.epa.gov/ecotox> či IUCLID Chemical Data Sheets: <http://esis.jrc.ec.europa.eu>) a publikovaných vědeckých článků. Není tedy třeba další testování organismů, což je výhodné jak z hlediska ochrany organismů, tak finančně a časově.

Pouhé použití všech (nezkontrolovaných) EC50 hodnot z online databází pro tvorbu SSD modelu však může vést k velkému zkreslení a proto je nutná jejich kontrola a přebrání. Databáze vhodná pro tvorbu SSD modelu akutních toxických účinků určité látky na sladkovodní druhy může vzniknout například následujícím postupem:

- Stažení základní databáze EC50 (včetně LC50 a IC50) hodnot pro všechny sladkovodní druhy z US EPA ECOTOX online databáze.<sup>1</sup>
- Výběr pouze akutních EC50 hodnot (obvykle se používají EC50 hodnoty získané z testů s délkou expozice 1-4 dny či 1-7 dní).<sup>2</sup>
- Výběr pouze vhodných akutních efektů, které mají přímý vliv na změnu abundance a složení společenství organismů (např. efekty na růst a biomasu rostlin a mortalita a imobilizace u živočichů).<sup>3</sup>
- Výběr pouze takových EC50 hodnot, které byly odvozeny z testování čistých látek, ne směsných produktů (důležité zejména v případě pesticidů). Rozumná hranice čistoty je 90%.

Důležitá je také kontrola a odstranění opakujících se stejných záznamů v databázi (replikací) a významně podezřelých EC50 hodnot. Například pokud se v databázi vyskytuje deset různých EC50 pro druh *Daphnia magna* (od různých autorů), které mají hodnotu kolem 5 µg/l a jednu EC50 o hodnotě 1000 µg/l, pak je nutné se nad touto podezřele vysokou hodnotou zamyslet. Doporučuje se nahlédnout do vědeckého článku, ze kterého byla hodnota do databáze vložena a po úvaze ji buď z databáze vypustit, opravit nebo ji tam ponechat.

V dalším kroku je nutné v databázi přepočítat všechny hodnoty EC50 na stejné jednotky (obvykle µg/l).

<sup>1</sup> V případě tvorby SSD modelu chronických toxických účinků se používají NOEC, LOEC a EC10 hodnoty.

<sup>2</sup> V případě tvorby SSD modelu chronických toxických účinků se používají naopak chronické NOEC, LOEC a EC10 hodnoty.

<sup>3</sup> V případě tvorby SSD modelu chronických toxických účinků se může použít i změna reprodukce.

Protože do SSD modelu vstupují druhové hodnoty EC50 (vždy pouze jedna EC50 pro jeden druh), je také potřeba rozhodnout, co udělat v případě přítomnosti více hodnot pro jeden druh. Obecně jsou používány dva postupy – buď se pro daný druh použije ta EC50, která má nejnižší hodnotu nebo se použije geometrický průměr všech EC50 hodnot pro tento druh. Tento krok významně ovlivní výsledek i smysl celého SSD modelu a proto je vždy důležité uvážit výběr správného postupu a uvést ho spolu s výsledky.

### 3. Tvorba SSD modelu

Když máme vytvořenou databázi vhodných EC50 hodnot pro určitou toxickou látku, pak můžeme vytvořit SSD model a použít ho pro predikci efektů látky na ekosystém a pro odvození limitních koncentrací (PNEC hodnot).

Za tvorbou modelu se skrývá (jak již vyplývá z předchozího textu) statistika. Naštěstí není důvod k velkým obavám těch, kteří s ní velkou zkušenost nemají, protože existuje software vytvořený přímo za tímto účelem. Jmenuje se ETX 2.0 a je zdarma ke stažení. Více o práci s tímto softwarem bude vysvětleno ve cvičení. Zde bude následovat především stručné vysvětlení teorie.

#### 3A. Základní princip tvorby SSD modelu:

- Data z připravené databáze se zobrazí jako histogram (tj. zobrazení jejich četností) a proloží se Gaussovou (normální) distribuční funkcí. Rovnice hustoty Gaussovy distribuční funkce je:

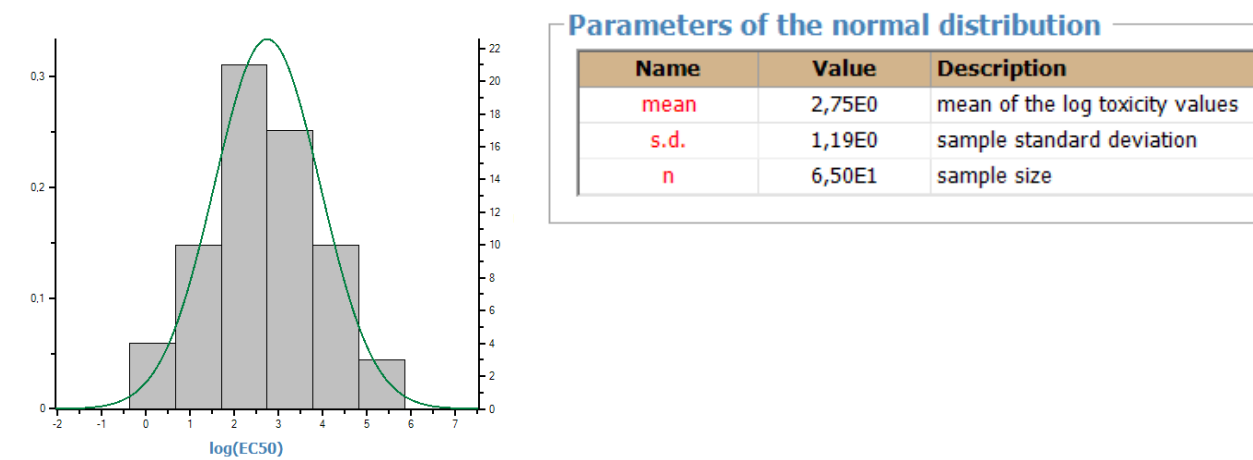
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (3)$$

$\mu$  ... aritmetický průměr Gaussovy distribuční funkce

$\sigma$  ... směrodatná odchylka Gaussovy distribuční funkce

$x$  ... hodnota na ose x (v našem případě  $\log(\text{EC50})$ ), pro kterou funkce vypočítá hustotu pravděpodobnosti  $f(x)$

Toto proložení je vlastně základní SSD model s potřebnými parametry  $\mu$  (průměr) a  $\sigma$  (směrodatná odchylka)



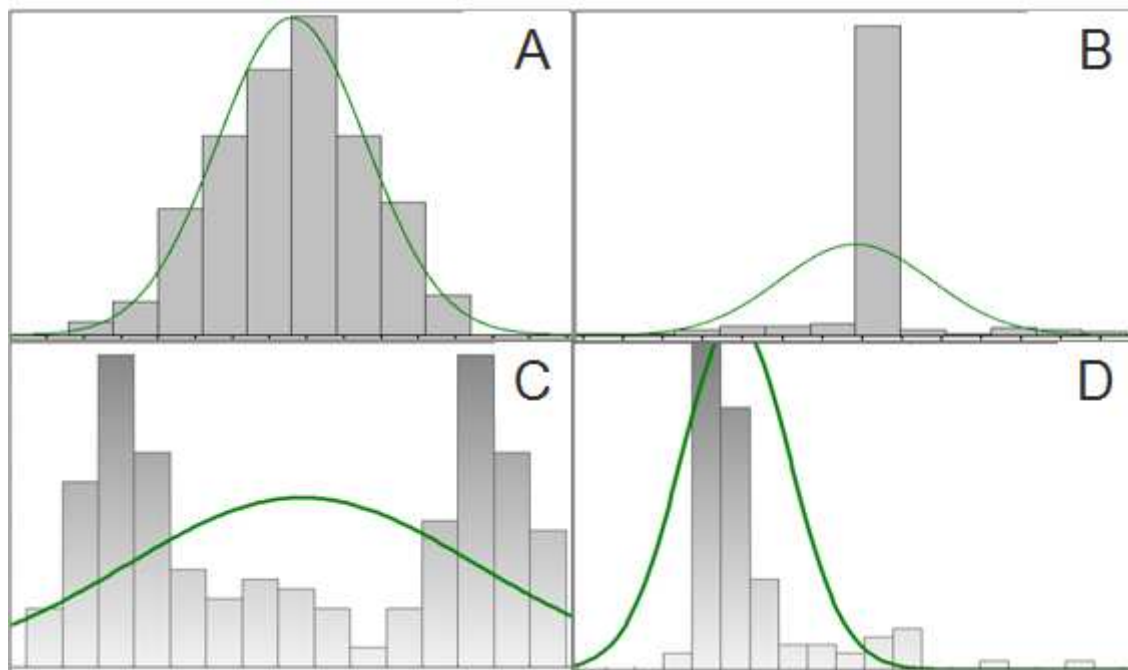
**Obr. 4.** Proložení histogramu  $\log(\text{EC50})$  hodnot pro insekticid karbofuran Gaussovou hustotní distribuční funkcí v softwaru ETX 2.0, parametry této funkce a celkový počet hodnot v databázi.

Graf: na pravé ose y jsou četnosti jednotlivých kategorií hodnot  $\log(EC50)$ , na levé pak jejich hustoty pravděpodobnosti -  $f(\log(EC50))$ .

- V dalším kroku bychom si měli ověřit, zdali proložení četností  $\log(EC50)$  hodnot touto Gaussovou funkcí je smysluplné. K tomu nám slouží tzv. testy normality. Pro SSD se nejčastěji využívá Anderson-Darling nebo Kolmogorov-Smirnov test. Zjednodušeně řečeno tyto testy porovnávají, jak moc se rozložení četností prokládaných hodnot (sloupce v histogramu na obr. 4) liší od rozložení „modelově normálního“ (zelená křivka na obr. 4). Pro jejich bližší pochopení je možné prostudovat některou z početných statistických učebnic a jiných zdrojů. Nicméně jejich výpočty jsou zahrnuty v téměř všech statistických softwarech včetně ETX 2.0. Důležité je však dodat, že data se od normálního rozložení statisticky významně neliší, pokud je výsledkem takovýchto testů hodnota  $p \geq 0,05$ .

Pokud je tato hodnota nižší, je vhodné se podívat na histogram a opticky a selským rozumem posoudit, jestli jsou rozdíly opravdu tak markantní nebo zda-li se jedná pouze o přílišnou „přísnost“ testů normality (ke kterému může dojít, pokud je v databázi velké množství dat). Pro takové optické posouzení mohou posloužit i tzv. Q-Q ploty či P-P ploty. Důležité je všimnout si nesymetričnosti v histogramu a přítomnosti „dvojitých“ vrcholů (obr. 5).

Pokud i optické posouzení potvrdí, že data opravdu normální nejsou, je nutné se zamyslet nad možným důvodem a případné výsledky vycházející z takového SSD modelu brát s rezervou.



**Obr. 5. Normální rozložení (zelená křivka) versus rozložení dat (šedé sloupce):** Rozložení dat je normální (A), nenormální s příliš vysokou četností středních hodnot (B), nenormální se dvěma vrcholy (C), nenormální nesymetrické (D)

Nyní, když je sestaven SSD model, může být použit pro výpočet HCp a PAF hodnot.

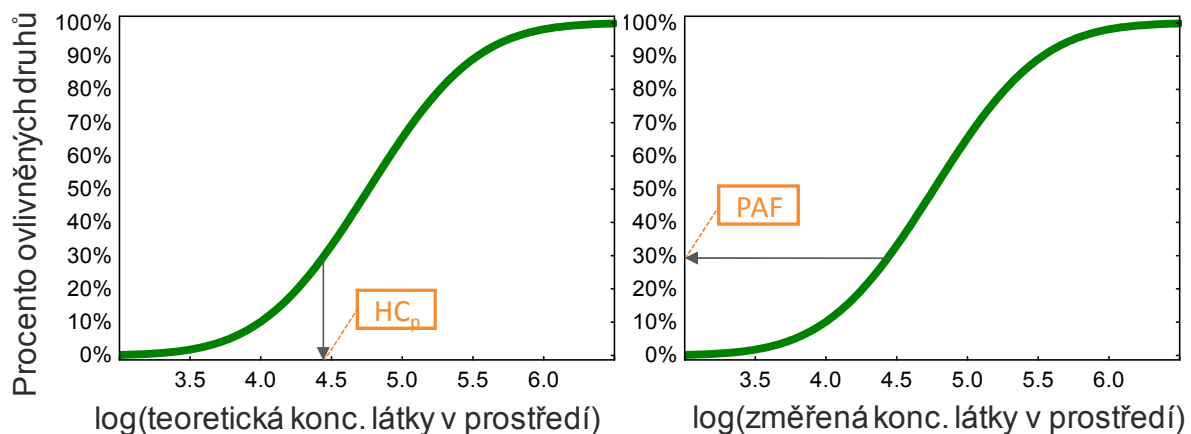
### 3B. Výpočet HCp a PAF:

Pro výpočet hodnot HCp a PAF nám tedy stačí znát pouze průměr a směrodatnou odchylku SSD modelu. Samotný model však musí být vyjádřený (jak je již uvedeno výše) jako kumulativní Gaussova distribuční funkce:

$$F(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

(4)

Je důležité si uvědomit, že pro účel následujících výpočtů hodnoty na ose x přestávají mít význam  $\log(\text{EC}_{50})$ , ale jsou chápány buď jako logaritmy koncentrace studované toxické látky v prostředí (v případě odvozování hodnot PAF) nebo jako „koncentrace způsobující daný efekt na společenství organismů“ (v případě odvozování hodnot HCp).



Obr. 6. Různý význam osy x při odvození HCp a PAF hodnot

Výpočet HCp a PAF hodnot z uvedeného vzorečku není úplně triviální, naštěstí lze opět využít softwaru ETX 2.0 (jedním z jeho základních výstupů je hodnota HC<sub>5</sub> i s intervaly spolehlivosti) a také Excelu. Tam existuje funkce „NORMDIST“ pro výpočet PAF a funkce „NORMINV“ pro výpočet libovolné hodnoty HCp. Jediné, co stačí znát je zmiňovaný průměr a směrodatná odchylka SSD modelu pro danou toxickou látku.

## 4. Limitace a modifikace SSD metody

Uvedená stručná teorie a postup tvorby základního SSD modelu jsou nutné k pochopení principů této metody. Praxe je však ještě o něco složitější.

Pro tvorbu modelu z dat sensitivity organismů se nepoužívá pouze vysvětlená Gaussova distribuce, ale častá je například distribuce logistická, triangulární a jiné. V poslední době se navíc začínají uplatňovat i neparametrické techniky či Bayesovská statistika. V poslední době se také čím dál víc začíná diskutovat o různých limitacích či omezeních této metody. Např. není zcela prokázáno, že limit HC<sub>5</sub> je opravdu dostatečně bezpečný, aby nebyl poškozen ekosystém. Dále se neustále hledá minimální počet hodnot druhových sensitivit pro danou toxickou látku (např. EC<sub>50</sub>), který je dostatečný pro tvorbu nezkreslujícího SSD modelu (EU doporučuje alespoň 10 druhových hodnot z různých taxonomických skupin). V neposlední řadě se vede i diskuze nad tím, že specificky působící látky (např. herbicidy či insekticidy) jsou výrazně toxické pouze pro určité taxony a jiné taxony

by měly být víceméně necitlivé. Proto je často narušena normalita takového rozložení a kvalitní SSD model je těžko vytvořitelný. Často se proto tvoří