Doplňující návod na zpracování experimentálních dat v předmětu

Metody chemického výzkumu – praktikum

Ve všech úlohách, kde se stanovuje koncentrace analytu ve vzorku, je třeba prezentovat výsledek odpovídajícím způsobem. Uvádíme tedy interval spolehlivosti, který vypočítáme v závislosti na tom, zda byla nebo nebyla ke kvantifikaci použita kalibrační závislost. Změříme-li několikrát signál vzorku a koncentraci analytu v něm vypočítáme z rovnice kalibrační závislosti, nestačí výpočet nejistoty dle Dean-Dixona, Horna, Studentova rozdělení nebo jiného přímého postupu pro zpracování souboru opakovaných měření. Tím bychom zanedbali velmi významný příspěvek rozptylu bodů v kalibrační závislosti. Proto je třeba použít vztahy na str. 163 skript Metody chemického výzkumu – praktikum, které platí pro kalibrační přímku. Má-li kalibrační přímka kladný úsek, použijeme horní vztah, prochází-li počátkem soustavy souřadnic (úsek je nulový – je třeba otestovat), použijeme vztah dolní. Nezapomeneme vynést do grafu všechny body kalibrační závislosti. To znamená, že měřil-li se každý bod vícekrát (např. 3×), nevynášíme průměry těchto bodů ani chybové úsečky (pokud nejsou přímo požadovány), ale vyneseme pro jednu koncentraci všechny hodnoty signálu nad sebou. Počet bodů v grafu nebude roven počtu proměřených koncentrací, ale počtu opakování měření každé koncentrace násobený počtem koncentrací (např. 3 opakování krát 5 koncentrací, tj. n = 15 bodů). Z technických důvodů se může stát, že různé koncentrace nejsou proměřeny se stejným počtem opakování. Některé např. 5×, některé 3×, jiné zase 2×. To však nijak nemění postup zpracování dat. Jen celkový počet bodů v grafu nebude násobkem počtu proměřených koncentrací, ale součtem zahrnujícím opakovaná měření všech koncentrací. Můžeme tudíž statisticky vyhodnotit i kalibrační přímku ze dvou koncentrací, kdy jedna byla proměřena 2× a druhá jen jednou, protože již máme k dispozici 3 body, které mají vzhledem k přímce nenulový rozptyl. V Excelu nebo Originu můžeme vícekrát zapsat do levého sloupce stejnou hodnotu koncentrace a do pravého sloupce příslušné signály, např.:

c [mg l^{-1}]	Signál
0	2487
0	2451
0	2508
10	6602
10	6584
10	6633
10	6592
20	10538
20	10499
30	14685
30	14546
30	14472
30	14710
40	18544
40	18613

Označíme data v obou sloupcích a proložíme je (podle příslušného programu: přidat spojnici trendu v Excelu nebo lineární fit v Originu) přímkou. Z m opakovaných měření signálu vzorku vypočítáme jeho průměr, který dosadíme do rovnice kalibrační přímky, z níž vypočítáme příslušnou koncentraci. Ze vztahů na str. 163 skript plyne, že převažujícím příspěvkem k nejistotě je rozptyl bodů kalibrační

přímky. Je možné využít i složitějších vztahů, které zahrnují též rozptyl opakovaného měření signálu vzorku. Většinou však nebývá nárůst nejistoty oproti vztahům ve skriptech významný.

Stejná pravidla platí i pro sestrojení a vyhodnocení grafu kalibrační závislosti z přídavků standardu. Nejistotu je však třeba vypočítat podle příslušného vztahu v souboru Doplnění teorie ve Studijních materiálech předmětu. Tento vztah je uveden pro přídavky standardu v jednotkách objemu. Jestliže máme kalibrační závislost přídavků standardu v jednotkách koncentrace, pak již nebudeme přepočítávat nejistotu z objemu na koncentraci.

Nezapomeneme započítat ředění vzorku, a to i při stanovení nejistoty podle vztahů na str. 163. Byl-li vzorek zředěn např. 100×, musíme 100× vynásobit nejen koncentraci, ale i její nejistotu. Pokud se měla zjistit hmotnost analytu ve vzorku, musíme přepočítat zjištěnou koncentraci na mg nebo g. Opět je třeba stejně přepočítat i nejistotu z koncentrace na hmotnost. I když někdy pracujeme při analýze s látkovým množstvím, obsah ve vzorku převedeme i s nejistotou na hmotnost v mg nebo g.

Jestliže provádíme opakované měření pro stanovení veličin bez použití kalibrační závislosti (např. retenční faktor, kapacitní faktor, počet teoretických pater a jejich výškový ekvivalent), použijeme pro odhad nejistoty Dean-Dixonův (pro malý počet opakování) nebo jiný odpovídající přímý postup. Platí to i pro výpočet koncentrace bílkovin ve vzorku z empirických vzorců. Pro každou vlnovou délku vypočítáme průměrnou absorbanci z jejího opakovaného měření. Do vzorce pak dosazujeme průměry příslušných absorbancí. Dean-Dixonovým postupem vypočítáme nejistotu ke každé absorbanci. Nejistotu koncentrace vypočítáme ze zákona šíření chyb. Zde postačí pro stanovení relativní nejistoty koncentrace odmocnina ze součtu kvadrátů relativních nejistot dosazených absorbancí. Kdybychom počítali koncentraci z příslušného empirického vzorce dosazováním různých dílčích absorbancí, pak nevíme, jaké kombinace vlastně zvolit, případně jestli dosadit všechny možné kombinace absorbancí do vzorce. To by ovšem počet vypočtených dílčích koncentrací převýšil počet opakovaného měření absorbancí.

Zaokrouhlování IS

Nejistotu zaokrouhlíme na 2 platné číslice. Na stejný řád, v jakém je druhá platná číslice (zleva) nejistoty, zaokrouhlíme i výslednou hodnotu, kterou je nejčastěji aritmetický průměr, někdy medián. Uvádění více číslic nejistoty by mohlo vést k paradoxu, že snad známe hodnotu nejistoty přesněji než samotného výsledku. Je to vlastně odhad, i když se musí vypočítat. Nejde o počet desetinných míst. Např. u čísla (257,3 ± 1,2) V s je nejistota 1,2 V s menší než půl procenta, u výsledku (0,0026 ± 0,0012) mg l^{-1} činí nejistota již 46 %. Je tudíž dosti nepřesný. Nastavení zobrazení určitého počtu desetinných míst v buňce Excelu nemá se správným zápisem nic společného. Někdy to může být náhodou správně, jindy nikoliv. Je to jen estetická záležitost. Nicméně u dílčích výpočtů pracujeme s mnohem větším počtem číslic a zaokrouhlujeme až výslednou hodnotu, abychom do výsledku nevnášeli chyby dílčích zaokrouhlování. Je také třeba si uvědomit, že standardizace odměrného roztoku by měla zpřesnit znalost jeho aktuální koncentrace, proto nedává smysl zaokrouhlit faktor odměrného roztoku jen na 1 nebo 2 číslice, i když jeho stanovení je také zatíženo experimentální chybou. Uvádíme jej na 4 platné číslice.

Testy shodnosti výsledků

Ve skriptech je uvedeno několik testů shody výsledků. Jejich použití je však omezeno na dvojici hodnot (párové testy) a neměli bychom testovat více výsledků po dvojicích systémem „každý s každým“. Dopustili bychom se určité chyby ve vyhodnocení. Měli bychom správně testovat všechny hodnoty navzájem najednou. Tento postup se nazývá analýza rozptylu, neboli ANOVA (Analysis of Variance). V tomto cvičení si ale dovolíme tuto skutečnost zanedbat. Je tu ještě další problém s využitelností „běžných“ párových testů ve skriptech. Nejsou vhodné tam, kde jsme dospěli k výsledku využitím kalibrační závislosti a nejistota je dána převážně jejím příspěvkem. Obecnějším kritériem shody může být vztah:

$$|x_1 - x_2| < \sqrt{(\delta x_1)^2 + (\delta x_2)^2}$$

kde x_1, x_2 jsou testované hodnoty a δx_1, δx_2 jsou jejich nejistoty. Když zjišťujeme pomocí statistických testů shodu dvou výsledků, měly by se jejich nejistoty vztahovat ke stejné pravděpodobnosti, obvykle 95 %. Pokud víme, že tomu tak není, musíme některou z nejistot přepočítat na stejnou pravděpodobnost. To nemusí být snadné, jestliže neznáme způsob jejího výpočtu a počet opakování. I když v případě výpočtu dle Dean-Dixona nebo Studentova rozdělení vyměníme koeficient *K* nebo *t* za jiný, odpovídající požadované pravděpodobnosti, nezískáme vzhledem k zaokrouhlování bez znalosti původních dat také původní rozpětí nebo směrodatnou odchylku. Takto přepočítaný interval spolehlivosti bude zatížen určitou početní chybou.

Testy významnosti trendů, úseků, směrnic

V tomto praktiku se použijí zejména tam, kde je třeba zjistit, zda je úsek lineární kalibrační závislosti významný. Jednak v případě, že je po provedení lineární regrese záporný, a také tam, kde je i po odečtení pozadí či blanku kladný, ale teoreticky by měl být nulový. Záporný úsek smysl nedává, a tudíž by měl po otestování vyjít jako nevýznamný. Regrese by se měla poté provést znovu se zafixovaným nulovým úsekem. Významný záporný úsek může vyjít z důvodu malé citlivosti pro nízké koncentrace. Směrnice je nižší než pro vyšší koncentrace a proložení celé závislosti přímkou nedává smysl. Je třeba koncentrační rozmezí rozdělit a proložit jen část, kterou je možno považovat za lineární a do níž spadají signály analytu, pokud to z hlediska limitu kvantifikace a rozptylu bodů dává smysl. Testování významnosti směrnice se využije např. při zjišťování míry interference osnovy vzorku. To je případ např. stanovení mědi ve víně metodou AAS. Zkoumáme vliv přítomnosti etanolu, draselné a vápenaté soli na signál analytu. Závislost signálu analytu na koncentraci interferentu nemusí být lineární, dokonce ani monotónní ve vyšetřovaném intervalu koncentrací. Přesto můžeme celou nebo postupně části této závislosti proložit přímkou a testovat, zda se její směrnice významně liší od nuly. Pokud tedy t-test ukáže, že je směrnice statisticky významná, pak lze říci, že je významná i matriční interference. Může být kladná i záporná, zvyšuje nebo snižuje signál analytu.

Limit detekce (LOD)

Tento pojem by měl vyjadřovat skutečnost, že signál analytu už je spolehlivě rozlišitelný od kolísavého signálu pozadí. Existuje mnoho vztahů pro jeho výpočet. Pokud jsou založeny na rozptylu bodů kalibrační přímky, nebývají u námi využívaných metod prakticky použitelné. Vystupuje v nich např. úsek kalibrační přímky nebo predikční pásy. Body kolem kalibrační přímky nemusejí být na pohled příliš rozptýleny, přesto však *LOD* vyjde nesmyslně vysoký, např. jednotky až desítky mg l^{-1}, což bývá i uprostřed kalibrační závislosti. Rozptyl těchto bodů nemusí vůbec souviset s citlivostí kolem úrovně pozadí. Je mnohem lepší využít definici 3-sigma:

$$LOD = \frac{3\sigma}{S}$$

kde σ je směrodatná odchylka z 10 opakovaných měření signálu pozadí a *S* je směrnice kalibrační závislosti pro nízké koncentrace. Pro koncentrační rozsahy kalibračních závislostí pro roztoky zhruba od 0,1 mg l^{-1} u metod v našem praktiku je vhodné vypočítat *S* přímky tvořené jen nulovou a první nenulovou koncentrací (např. 0 a 0,1 mg l^{-1}). Není vhodné použít směrnici celé kalibrační přímky, která zahrnuje koncentrace i o 2-3 řády vyšší než *LOD*, a tedy nevypovídá nic o citlivosti kolem *LOD*. Použitím jen 1 koncentračního bodu je sice *S* formálně nepřesnější, ale s velkou nejistotou i půl řádu se u *LOD* počítá. Vzhledem k zachování matrice vzorku je někdy vhodné zjišťovat *LOD* z dostatečně zředěného vzorku, u něhož již známe koncentraci analytu. V chromatografii se vychází z maximálního rozkmitu *h* základní linie v okolí píku, přičemž šířka tohoto okolí je desetinásobek šířky píku na základně. Tento rozkmit se vynásobí příslušným koeficientem *k*, jehož velikost závisí na dohodě (někdy 2, 3). Koncentrace odpovídající této výšce *k.h* je *LOD*. Pracuje se s výškou píku, nikoliv plochou.

Vyjde-li množství analytu ve vzorku nulové, když nenaměříme žádný nebo téměř žádný signál, měli bychom uvést hodnotu *LOD* a napsat, že množství analytu je < *LOD.* Vzorek by totiž mohl obsahovat naší metodou již nezjistitelné množství analytu, čímž bychom se dopustili chyby. Také hodnotu pod limitem kvantifikace (*LOQ*) bychom měli náležitě okomentovat, neboť je zatížena poměrně velkou chybou, jak již význam *LOQ* napovídá.