

## **Posudek na diplomovou práci Bc. Roberta Pánika "Určení geologické provenience katodoluminiscenční spektroskopii apatitů a karbonátů "**

Diplomová práce se zabývá využitím katodoluminiscenční spektroskopie k rozpoznávání minerálů a hornin. Je to komplexní pojednání zahrnující popis laboratorní metody, její aplikaci, zpracování dat a diskuzi výsledků, přičemž jádro práce představuje zpracování dat. Úkolem je nalezení vhodné metody zpracování spektrálních křivek, jejichž tvar je ovlivněn kombinovanými vlivy příměsí, tak, aby bylo možné z naměřených dat rozhodnout, z jaké lokality pochází studovaný vzorek. Práce má 74 stran (při použití fontu 11 a řádkování asi 1,5) a je psána srozumitelnou angličtinou.

### **Obsah a členění práce**

První kapitola obsahuje výklad fyzikální podstaty katodoluminiscence se zaměřením na její využití ke studiu karbonátů a apatitů.

Ve druhé kapitole se autor zaměřuje na problém využití katodoluminiscence pro určování místa původu (geologické provenience). Je zde uveden na diplomovou práci dosti rozsáhlý přehled faktů a výsledků různých autorů. Jsou popsány též limity metody.

Třetí kapitola popisuje zkoumané materiály a laboratorní metody. Je popsán původ velkého počtu vzorků, přístroj použitý k získání katodoluminiscenčních spekter a měřící procedura.

Čtvrtá kapitola se zabývá zpracováním dat z katodoluminiscence a představuje jádro práce.

V rámci první části nazvané Analysis of carbonates obsahuje podrobný popis metod použitých ke zpracování dat a analýze. Normalizovaná a vyhlazená spektra byla proložena jednočlenným gaussovským modelem, jehož 3 parametry jsou označeny jako  $a$ -amplituda,  $b$ -lokalisace,  $c$ -šířka). Uvádí se, že autor ke stanovení těchto parametrů použil "nonlinear method of least-squares of curve fitting". Bližší popis metody jsem ale nenalezl. Za primární parametr pro porovnávání považuje autor lokalizaci ( $b$ ). Pro lepší rozlišení se používají reziduální spektra (Fig. 6), která jsou ještě normována kvůli potlačení vlivu divergence v okrajových částech. Používají se rovněž i dvoučlenné gaussovské modely. Autor uvádí, že v případě dolomitů narazil na problém asymetrie spekter (s kladnou šikmostí).

Dále se popisují další metody sloužící k popisu spektrální podobnosti a rozlišování spekter: korelace (Pearsonův a Spearmanův koeficient), korelační matice, popisující korelaci mezi vzorky (sample-sample), nazývaná correlation coefficient map, a její odvozenina disrelation matrix. Následuje analýza hlavních komponent (založená na korelační matici) a shluková analýza, jejíž varianta agglomerative hierarchical clustering je popsána podrobněji.

V odstavci 4.1.3 autor detailně popisuje algoritmus, který použil při analýze dat a který sestává z celkem 25 kroků. Algoritmus je zřejmě realizován pomocí excelovského spreadsheetu, který jsem při posuzování práce neměl k dispozici. Není jasné, zda ho sestavil autor práce. Jednotlivé kroky obsahují mj. fitování spekter, jejich normalizaci, korelační analýzu, aplikaci analýzy hlavních komponent a shlukovou analýzu, celý proces je, je zřejmě řízen zpracovatelem dat. Nemaje příslušný spreadsheet nedovedu posoudit míru automatizace, zdá se ale, že zásahy zpracovatele dat jsou podstatné a algoritmus je v podstatě prováděn zpracovatelem. Z popisu není jasné, zda označení "the whole dataset" v bodu 15 se vztahuje jen na reziduální spektra nebo na původní plus reziduální. V následující odstavci 4.1.4 se jmenuje "Meta-analysis of results". Zdá se, že se jedná o celkové vyhodnocení, které je sumací výsledků aplikace jednotlivých metod, které se nasčítají do stupňovité křivky a vypočte se směrnice, resp. tangens přímky

spojující první a poslední bod. Tím se kvantifikuje celkový výsledek, přičemž nejvyšší možná hodnota odpovídající porovnání vzorku se sebou samým je 1, jak je znázorněno na Fig. 9. Kapitola dále uvádí výsledky analýzy apatitů.

Výsledky jsou znovu shrnuty v kapitole páté a diskutovány v kapitole šesté. Na závěr přichází jednostránkový závěr práce.

## **Připomínky**

Práce je celkově velmi kvalitní, takže místo pochval se soustředím na některé detaily ze zpracování dat. Autor se opírá o představu, že studovaná spektra lze dobře popsat gaussovským modelem. Porovnání spekter je pak založeno na akumulaci výsledků dílčích metod, zjednodušeně lze říci, že jsou tu 3 hlediska – shoda gaussovských křivek, korelace a analýza hlavních komponent a shluková analýza. Přístup autora bych označil jako heuristickou kombinaci nástrojů známých ze statistiky, opírající se jeho hlubokou znalostí reálií (geologických, chemických, fyzikálních). Vychází z představy, že akumulace různých hledisek dá v součtu kvalitnější možnost rozpoznání proveniencí vzorku. Neřeší se, jak metoda (celkově nebo v dílčí části) bude fungovat, když bude předpoklad splněn jen do určité míry nebo dokonce není vůbec vhodný. Komplexnost algoritmu, která má být jeho předností, se pak může stát jeho největší slabostí, když není dobře vidět, kde se do akumulovaného skóre (tj. celkové korelace, Fig. 9) infiltrují artefakty. Algoritmus pak může provádět jenom opatrně osoba velmi znalá a to, co, předpokládám, chtěl autor dosáhnout, tj. poskytnout nástroj pro analýzu ostatním, je ohroženo.

Myslím tedy, že autor by měl při obhajobě vysvětlit, jak je to se splněním předpokladu gaussovského průběhu spekter, jaký charakter může mít například korelace (korelační koeficienty) a na ní založená analýza hlavních komponent v případě předpokládaného typu odchylek od modelu.

## **Drobnosti**

Není jasné, proč se spektra, standardizovaná podle vzorce (7) nazývají dynamická (není uvedena ani citace). Autor by mohl název vysvětlit.

Uvádí se, že "...first principal component can be thought of as the main axis of an ellipsoid enclosing all analysed data". Autor by se měl zamyslet, jak bude uvedený elipsoid, resp. shluk bodů vypadat.

Práci by prospělo členění, kde se jasně oddělí metody a výsledky. Nebo alespoň upozornit čtenáře, že text je vystavěn jinak.

Apendix 1. Není uvedeno, co znamenají čísla v tabulce A1. Vyhlášení první či druhé komponenty za lepší je (s ohledem na nejasné rozdělení těchto charakteristik a možnosti přiřazení jejich statistické významnosti) dosti odvážné a subjektivní (obzvlášť při poměru 3/5).

Tabulka A2 má dokumentovat úspěšnost diskriminace. Autor vybral případ 227, který v předchozí tabulce A2 jako jediný koreluje s oběma komponentami vybranými pro analýzu původních dat (PC2) a reziduálních dat (PC1). Jak by vypadal podobný graf pro ostatní případy?

Tabulka A3 ukazuje > 80% úspěšnost zpětného zařazení. Lze zdůvodnit použití Minkovského vzdálenosti ( $p=3$ ) nebo je to jen konkrétní případ po doplnění databáze to bude jinak?

Autor děkuje Kriminalistickému ústavu, což je při zaměření práce nezvyklé. Měl by tedy alespoň při obhajobě specifikovat zač.

### **Celkové hodnocení**

Přes výše uvedené výhrady, pokládám předloženou diplomovou práci za velmi kvalitní. Dokumentuje autorovy hluboké znalosti o chemickém složení a charakteru minerálů a hornin důležité pro použitou laboratorní metodu a rovněž znalost její fyzikální podstaty. Cenný je i autorův pokus o vytvoření algoritmu pro rozpoznávání geologické proveniencí. Osobně bych práci hodnotil známkou výborně.

Josef Ježek, 29.5.2015