

Identifikácia neštandardného správania odberateľov v energetickej sieti

Vedúci práce:
Ing. Marek Lóderer

Bc. Matúš Cuper

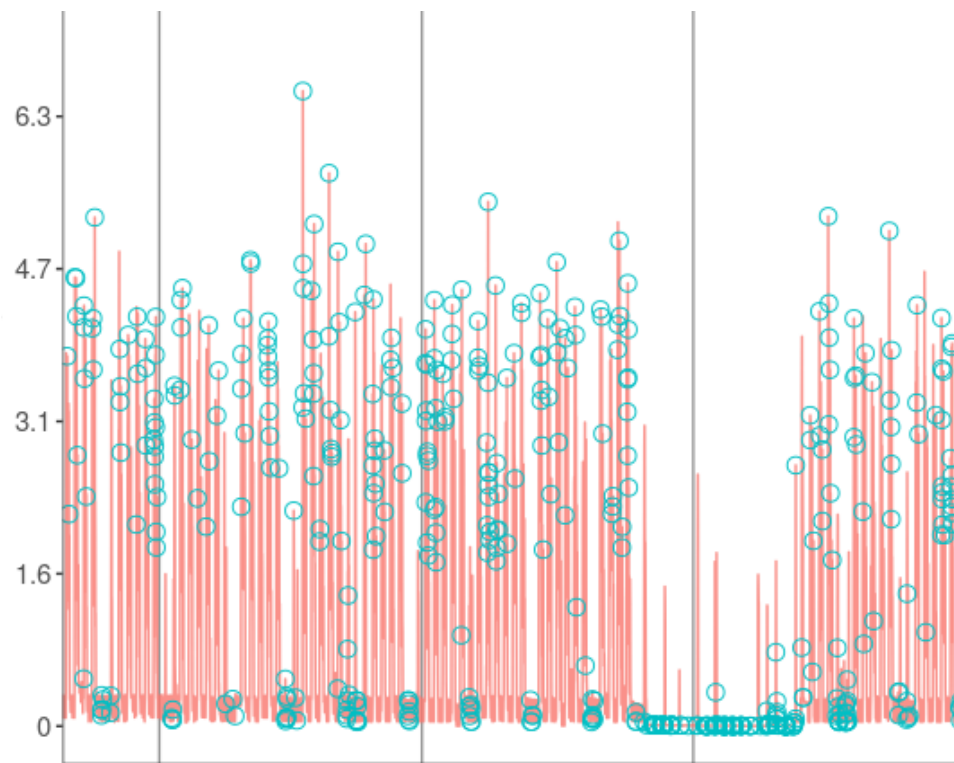
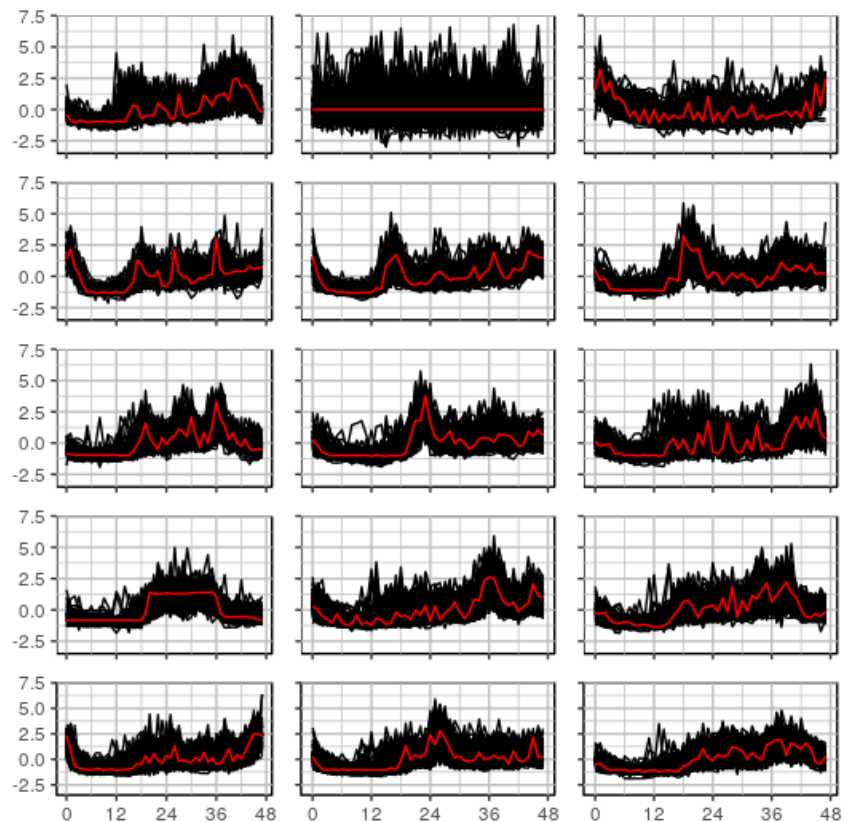
Motivácia

- Identifikácia neštandardného správania
 - evolúcia dát, neoznačené datasety, šum, definícia anomálie...
 - klasifikácia, zhukovanie, štatistické metódy...
 - lokálne a globálne anomálie, zlomy...
- Zamedzenie nelegálnym odberom, identifikácia chybných zariadení, optimalizácia distribúcie elektriny
- Správanie rozdielne od správania okolitých inštancií

Existujúce riešenia

- Zhlukovanie pomocou k-medoidov
 - rýchle spracovanie veľkých datasetov
 - definícia normálnych dát
 - evolúcia dát
- Extrémna Studentova odchýlka (ESD)
 - Grubbov test
 - dekompozícia časových radov
 - identifikácia až do 50% anomálií

Existujúce riešenia



Navrhnutý proces identifikácie anomálií

1. Zhlukovanie časových radov
2. Skórovanie inštancií a zhlukov na základe zhlukovania
3. Výber inštancií s vysokým skóre
4. Vyhľadovanie časových radov
5. Skórovanie inštancií pomocou metódy S-H-ESD

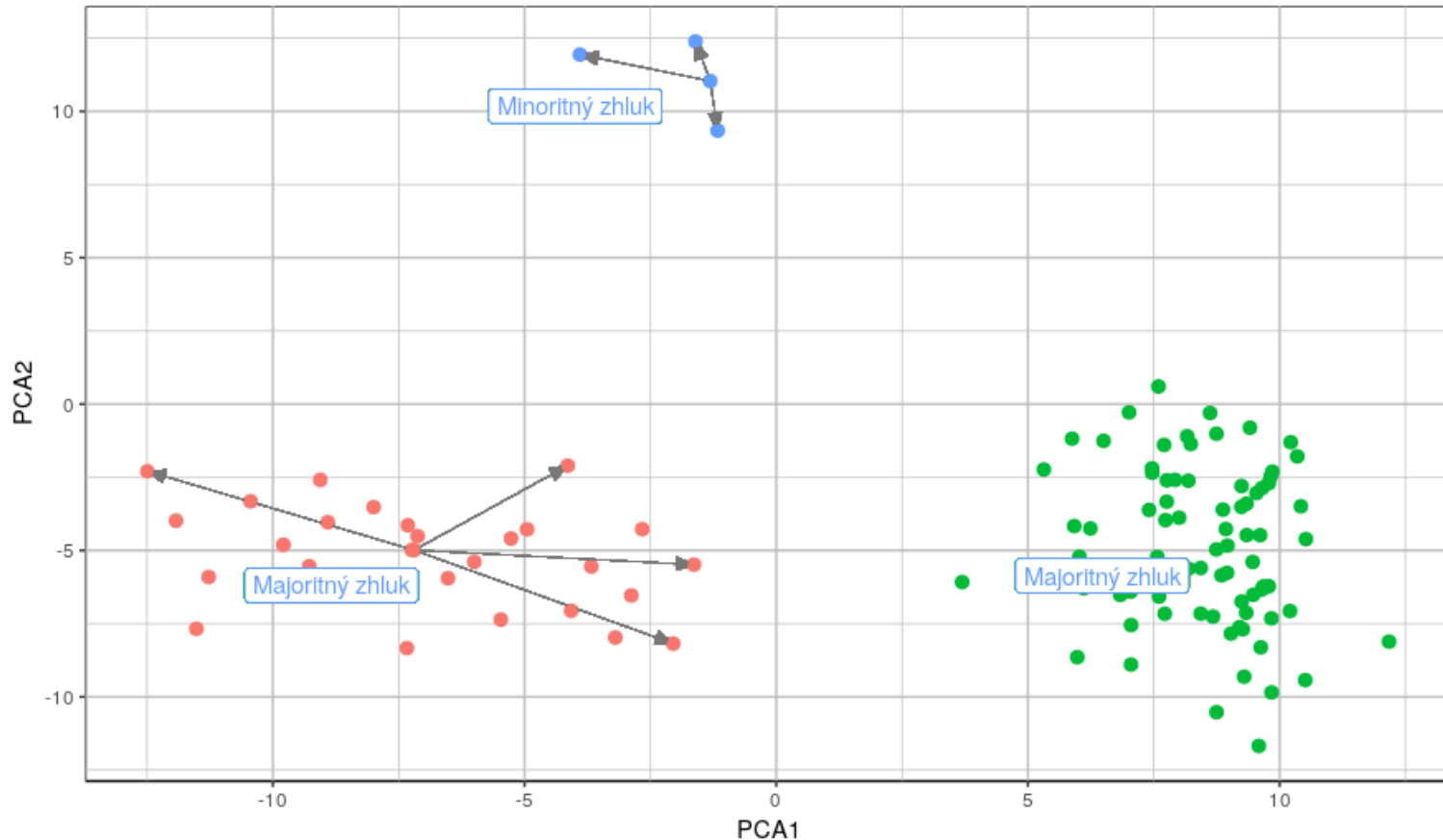
1. Zhlukovanie časových radov

- Predspracovanie dát
 - normalizácia (z-score)
 - orezanie (dataset nezačína a nekončí v strede týždňa)
 - rozdelenie (pracovné dni a dni voľna)
- Agregovanie datasetu po posuvných oknách
 - veľkosť okna (2 týždne)
 - posun okna (1 týždeň)
- Zhlukovanie pomocou k-medoids
 - dištančná metrika (GAK)

2. Skórovanie inštancií a zhlukov na základe zhlukovania

- Skórovanie inštancií
 - rôzne hodnoty pre rôzne inštancie
 - relatívna vzdialenosť inštancie v rámci zhuku
- Skórovanie zhlukov
 - rôzne hodnoty pre rôzne zhluky
 - penalizácia pre najmenšie zhluky

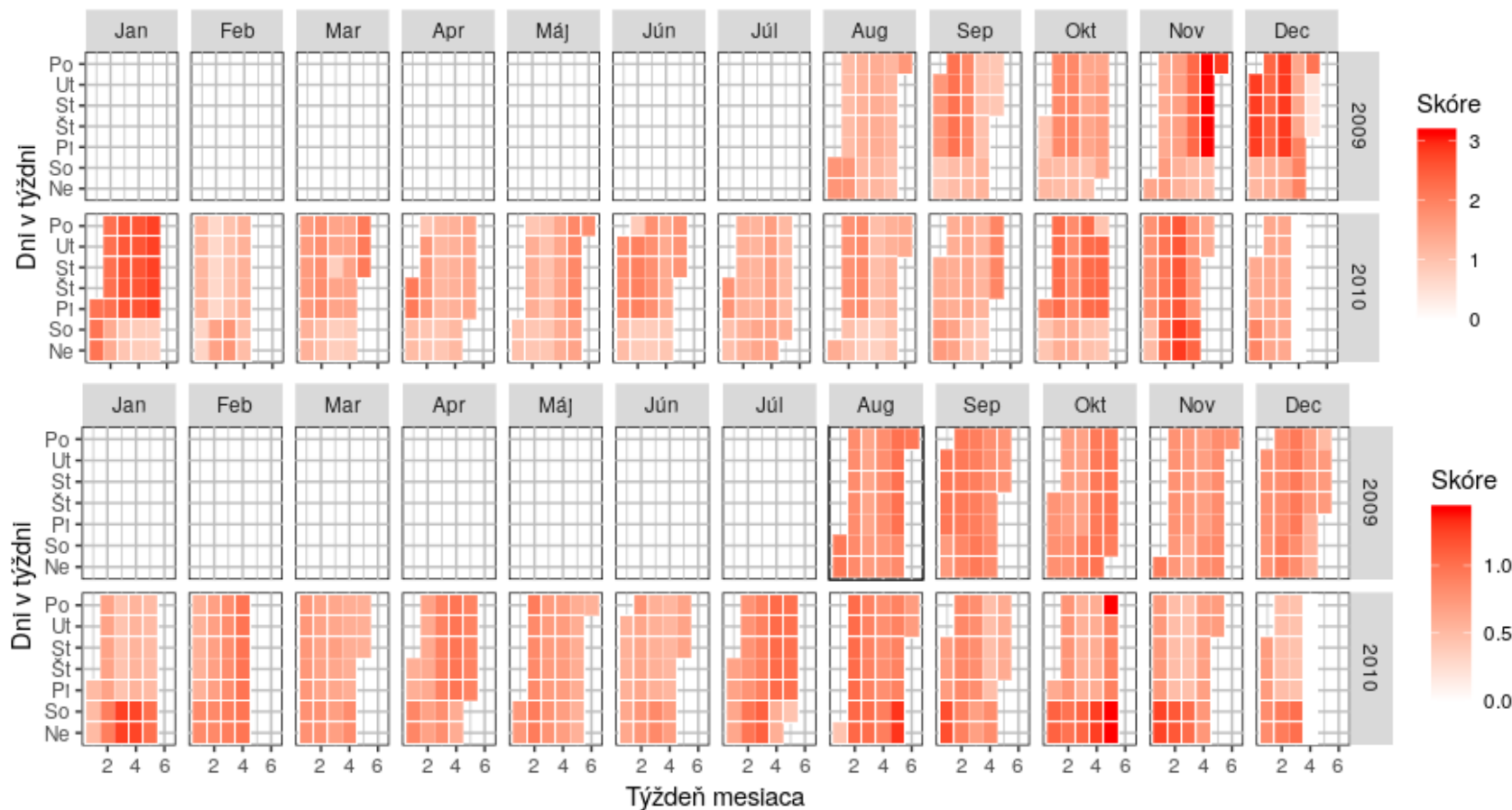
2. Skórovanie inštancií a zhlukov na základe zhlučovania



3. Výber inštancií s vysokým skóre

- Intervalom medzikvartilového pravidla pre anomálie
 $\langle Q1 - 1.5 * IQR, Q3 + 1.5 * IQR \rangle$
- Metódou FeaClip
- Na základe podozrenia
- Vizualizáciou

3. Výber inštancií s vysokým skóre



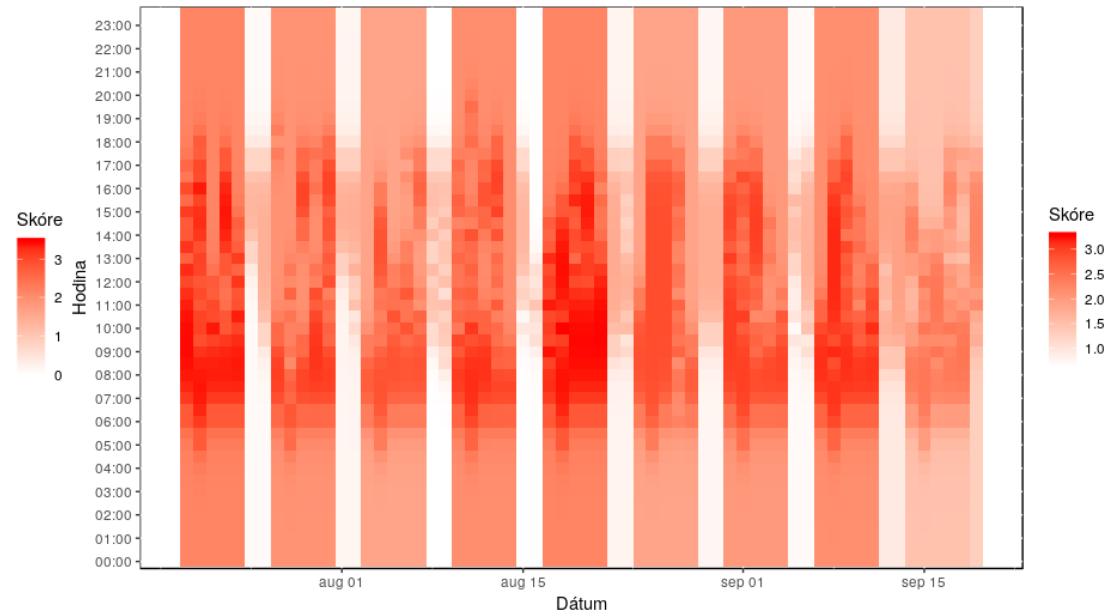
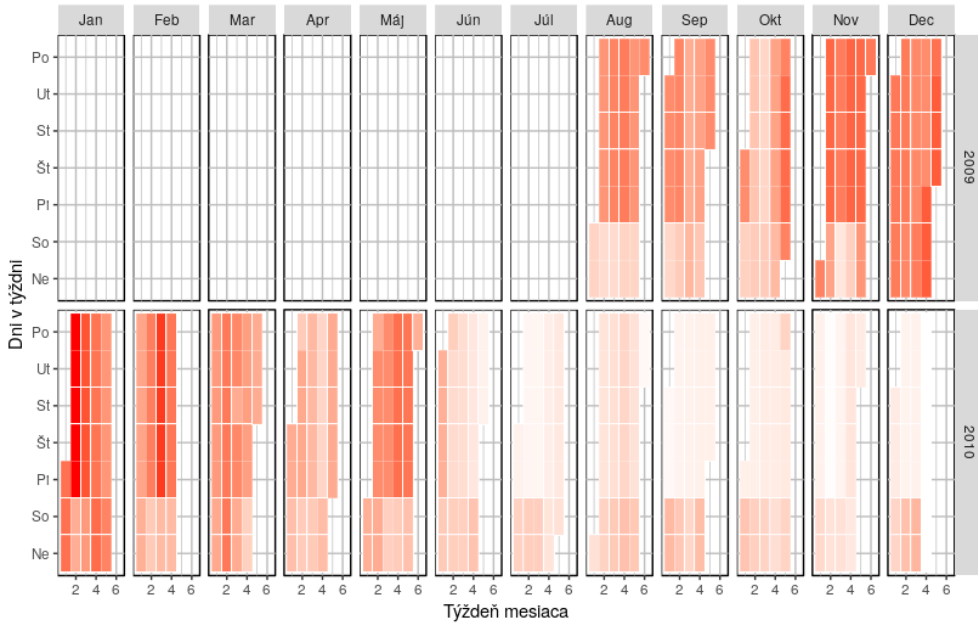
4. Vyhladzovanie časových radov

- Pred analýzou S-H-ESD
 - eliminácia množstva menších lokálnych anomálií
- Po analýzou S-H-ESD
 - výstupom analýzy je iba príznak
 - intervaly s hustým výskytom anomálií sú zlúčené
 - intervaly s nízkym výskytom sú vyhladené

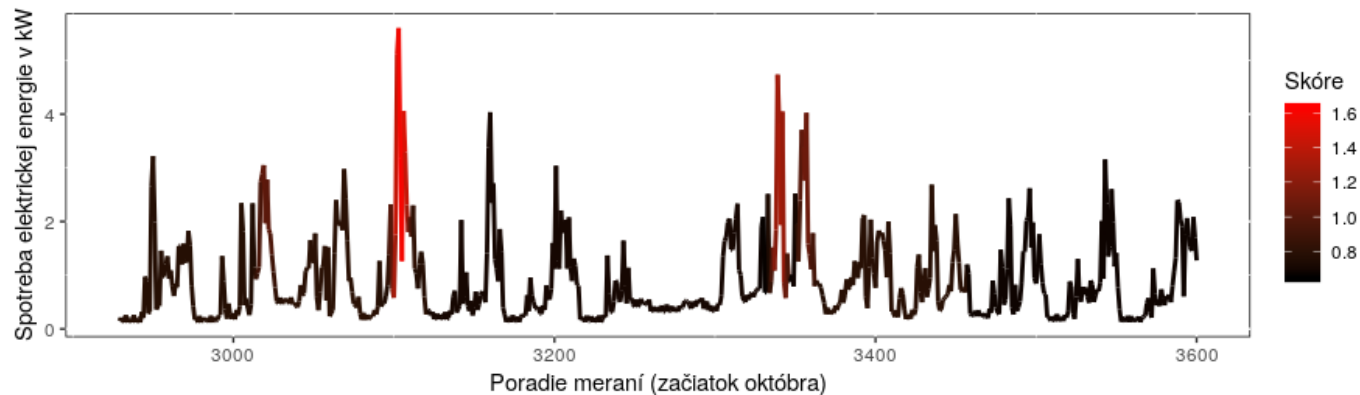
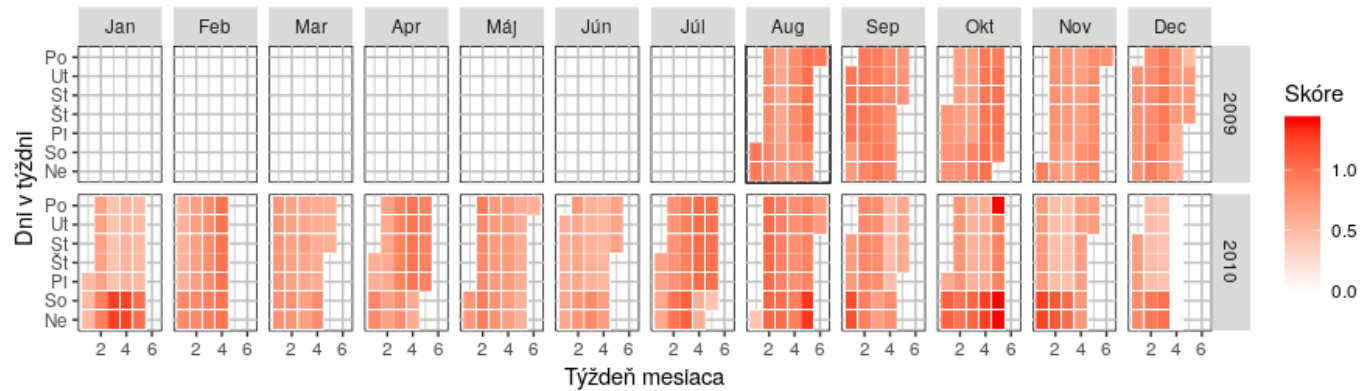
5. Skórovanie inštancií pomocou metódy S-H-ESD

- Kombinovanie vyhladených príznakov so skóre zo zhlučovania
- Nižšia granularita
- Výsledkom nie je príznak anomálnosti, ale číslo označujúce mieru podozrenia daného merania

5. Skórovanie inštancií pomocou metódy S-H-ESD



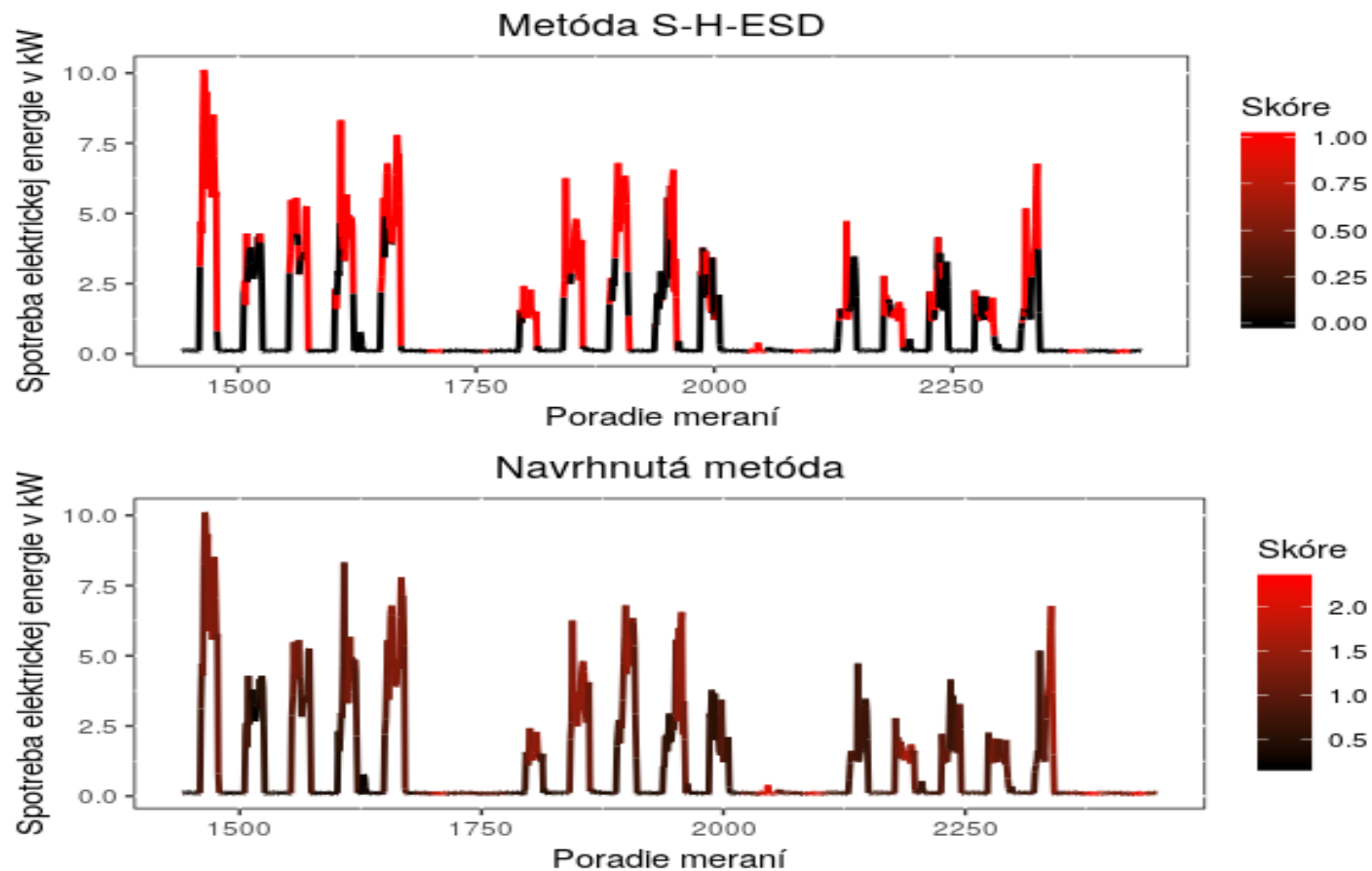
5. Skórovanie inštancií pomocou metódy S-H-ESD



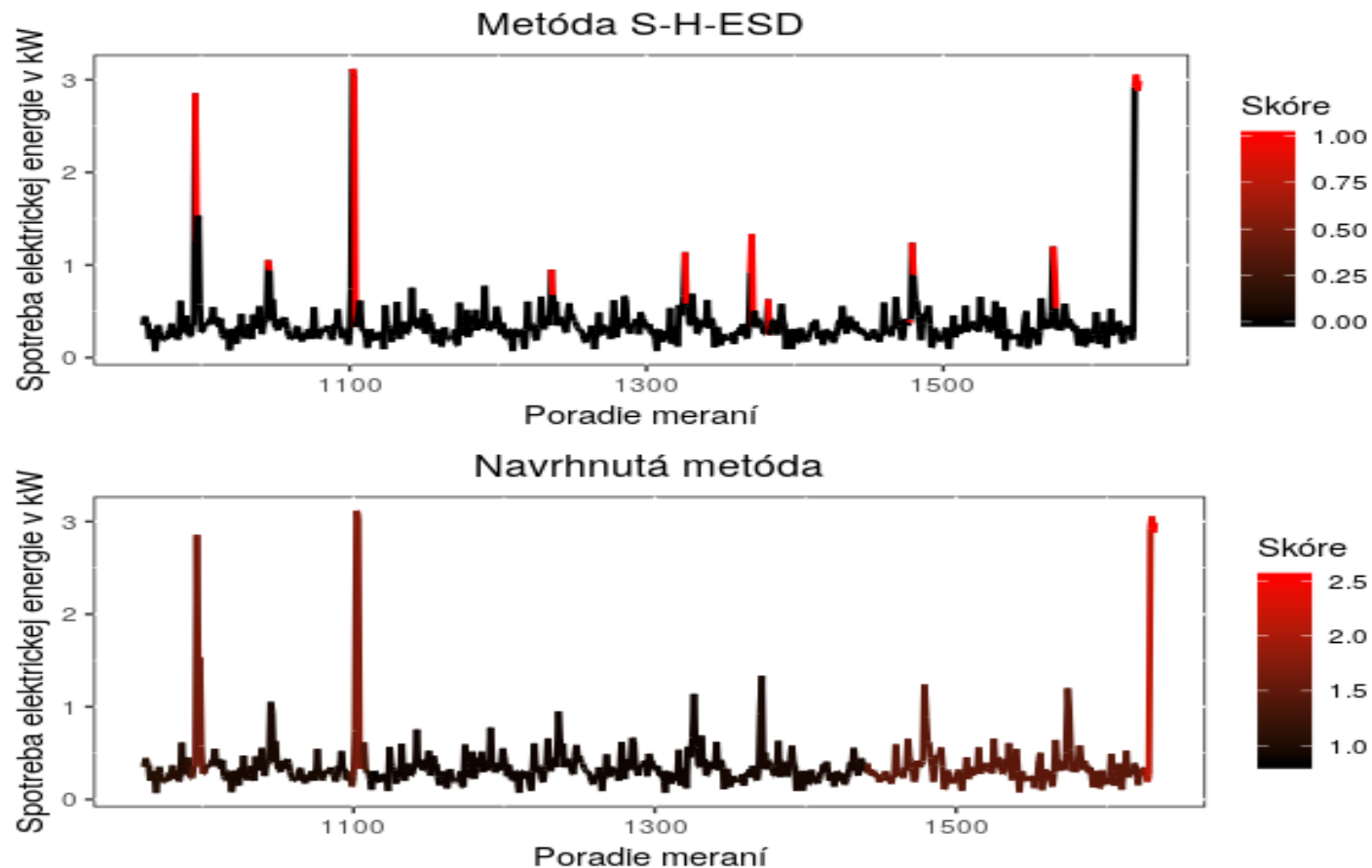
Overenie

- Zhlukovanie
 - experimenty
 - validačné zhlukovanie indexy
- Navrhnutá metóda
 - porovnanie so samotnou metódou S-H-ESD
 - porovnanie so samotným zhlukovaním
 - porovnanie s metódou FeaClip

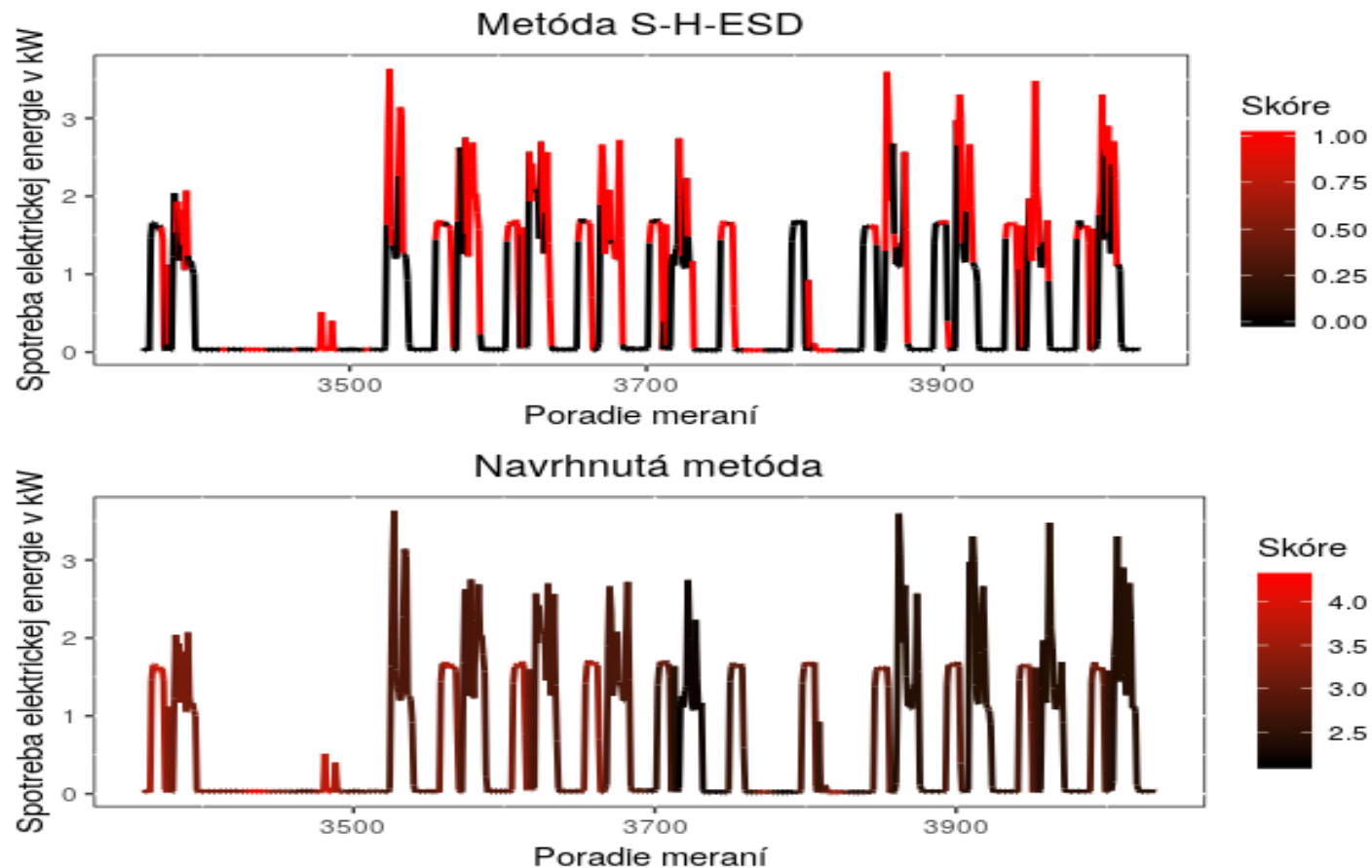
Overenie



Overenie



Overenie



Prínos

- Zhlukovanie
 - nižšia granularita
 - možnosť online spracovania
- S-H-ESD
 - rýchlejšie (analyzovaní sú iba podozriví odberatelia)
 - príznak anomálnosti nahradený mierou podozrenia

Zhrnutie

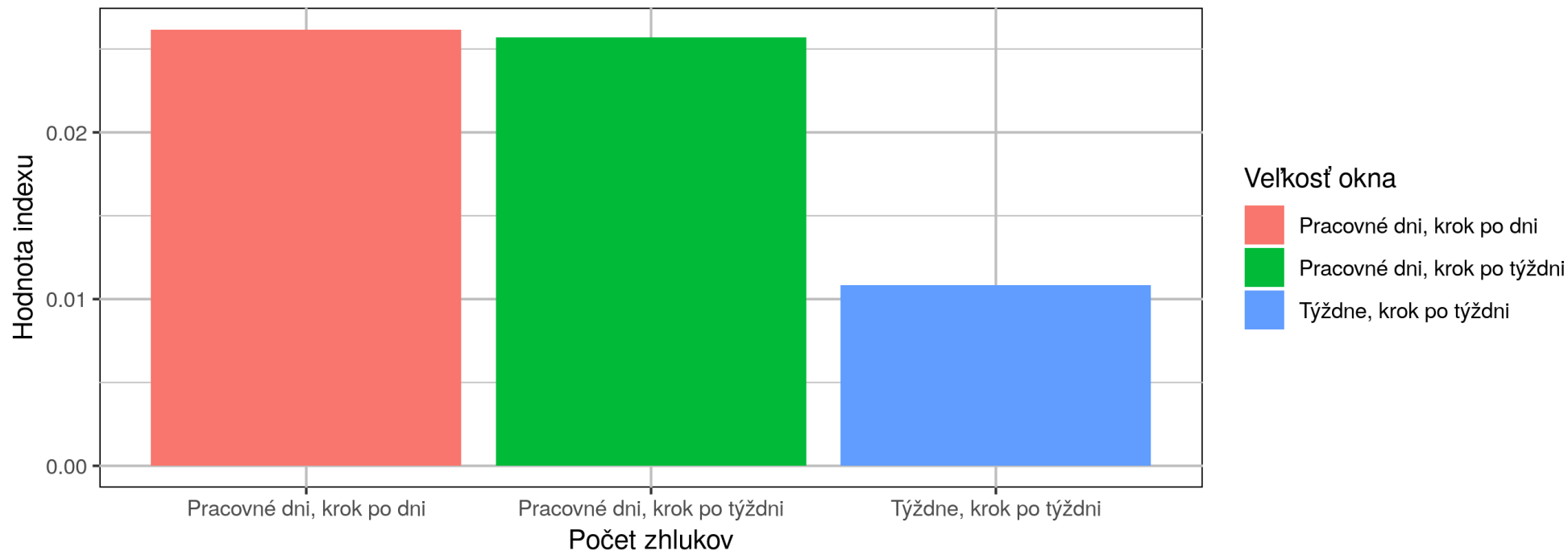
- Riešenie
 - kombinácia zhukovania a štatistickej metódy S-H-ESD
 - vhodné najmä na globálne anomálie
 - možnosť online spracovania
- S-H-ESD
 - rýchlejšie (analyzovaní sú iba podozriví odberatelia)
 - príznak anomálnosti nahradený mierou podozrenia

Otázka č. 1

- Ako by bolo potrebné upraviť metódu, aby sme ju mohli použiť v on-line režime?
- Krok posuvného okna by sa znížil z týždňov na dni, prípadne hodiny
- Experimenty dokazujú porovnateľné výsledky pri rôznych veľkostiach kroku

Online spracovanie

Silhouettov index (max.)



Otázka č. 2

- Pri identifikácii anomálie v rámci určitého zhluku by stálo za úvahu zistiť či je anomáliou aj voči iným zhlukom. Aký by bol postup v tomto prípade?
- Súčasne určujeme anomálnosť v rámci zhluku
- Je možné rozšíriť výpočet skóre o ďalšie zhlukovanie, napr. zhlukovanie založené na hustote

$$skore_i = skore_{instancia_i} * skore_{zhluk_j}, \text{ pre } instancia_i \in zhluk_j$$

Otázka č. 3

- V čom je prínos navrhnutej metódy?
- Väčšina metód sa zameriava na lokálne anomálie a zlomy
- Metóda je použiteľná aj na iné domény
- Miera podozrenia nie iba príznak anomálnosti