

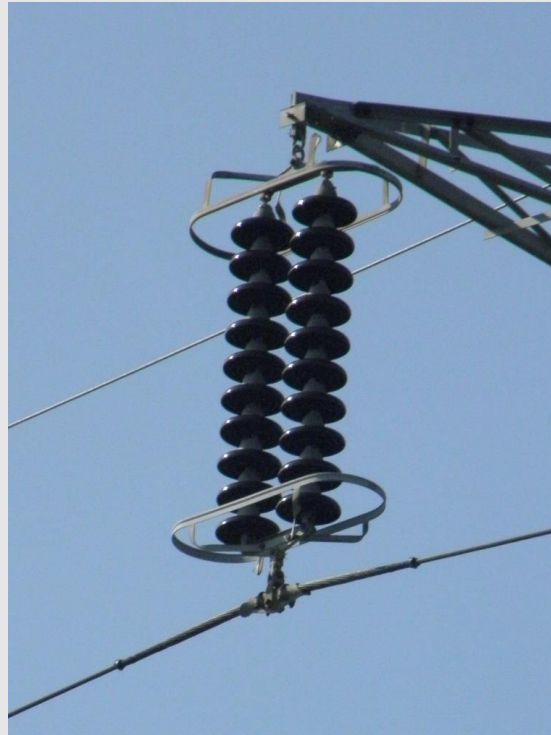


MODELOVANIE ZNEČISTENIA  
PROSTREDIA PRE POTREBY  
ENERGETICKÉHO PRIEMYSLU

Krammer, Kvassay, Forgáč, Očkay, Skovajsová,  
Hluchý, Skurčák, Pavlov

# Modelovanie znečistenia prostredia pre potreby energetického priemyslu

- Spolupráca s VUJE a.s., dáta SHMÚ
- Projekty: najskôr štúdia realizovateľnosti, neskôr ARIEN APVV-20-0548, VEGA Hluchý 2/0125/20.



# Riziká v energetike

Pôsobením prachu a spadových častíc na stožiaroch/izolátoroch sú oplyvnené elektrické vlastnosti (napr. merná el. vodivosť, odpor, ...).

Vznik s tým súvisiacich rizík – kolísania napätia, napät'ových preskokov, výbojov, až vzplanutia.

Možnosť poškodenia vodiča, stožiaru, vznik požiaru, materiálna škoda, výpadky prúdu, ...

## Fázy napät'ového preskoku cez izolátorový záves



(a) Safety Zone, (b) Forecast Zone, (c) Danger Zone, (d) Flashover

**Prevzaté z:** Zeng et al. (2020). Experimental Study on High Frequency Pulse Current Variation Characteristics of Pollution Discharge of Insulators. In IOP Conference Series: Earth and Environmental Science (Vol. 446, No. 4, p. 042006). IOP Publishing.

# Motivácia a problémy

- Jednotlivé vplyvy musia byť zohľadnené už pri dizajnovaní elektrických vysoko-napäťových stožiarov (vhodná voľba stupňa izolácie I. až IV. )
- Zrušená technická norma STN 330405
- 1 meranie znečistenia trvá 6 týždňov – vysoká časová a finančná náročnosť (nutnosť dlhej prevádzky vysokého počtu staníc).

## Motivácia:

- Snaha o modelovanie miery znečistenia na základe iných atribútov meraných SHMÚ (napr. PM10, PM2.5), ktoré sú ľahšie merateľné
- Potreba modelovania znečistenia pre environmentálne / ekologické účely.

## Problémy:

- Výrazne nízky počet záznamov
- Veľký počet vplyvov (počasie, doprava, stavebný priemysel, poľnohospodárska činnosť, .. )
- Výrazný stochastický charakter veličín

# Meranie stupňa znečistenia priamo v teréne a použitie rôznych typov izolátorov

Výňatok zo zrušenej normy STN 33 0405:

## 3 ZPRACOVÁNÍ VÝSLEDKU ANALÝZ

V každém měřeném místě je třeba ve zhruba dvouměsíčních intervalech provést po sobě 6 až 12 sběrů sedimentačních nádobek.

Z rozboru spadu jednotlivých sběrů se pro určité měřené místo stanoví odpovídající množství hodnot:

- a) celkové množství zachyceného spadu přepočteného na jednotku plochy za jednotku času  $S$  v  $\text{mg} \cdot \text{cm}^{-2}$  a den,
- b) množství rozpustných látek v sedimentu  $S_r$  v  $\text{mg} \cdot \text{cm}^{-2}$  a den,
- c) měrná el. vodivost 0,2 % roztoku rozpustných látek  $\gamma_{0,2}$  v  $\mu\text{S} \cdot \text{cm}^{-1}$ .

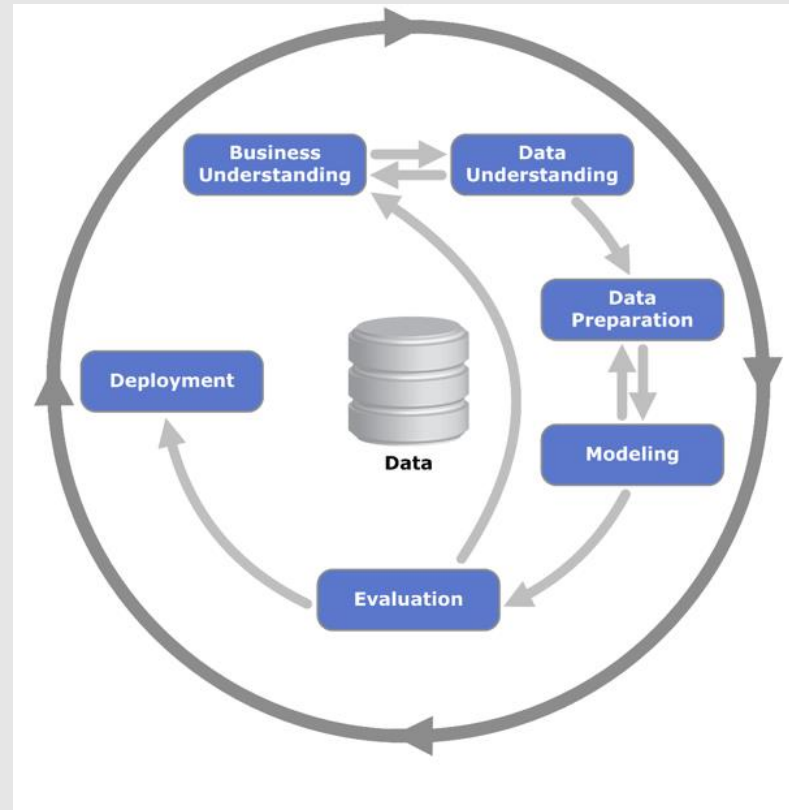
Z takto získaných souborů hodnot  $S$ ,  $S_r$  a  $\gamma_{0,2}$  se statisticky pomocí Weibullova rozložení stanoví maximální hodnoty  $S_m$ ,  $S_{r_m}$  a  $\gamma_{0,2_m}$ , které nebudou v daném místě překročeny s pravděpodobností 0,5 %. Všechny tři takto získané veličiny mají při hodnocení vlivu stupně znečištění na elektrickou pevnost izolace při znečištění stejnou váhu a je nutno je uvažovat současně. Kritériem je proto jejich součin.

$$p = S_m \cdot S_{r_m} \cdot \gamma_{0,2_m}$$

# Metodológia

## Metodológia CRISP – DM

Vychádza z priemyselného štandardu, s rozšírením pre dolovanie v údajoch.



## Dátový problém: Opakujúce sa stĺpce výsledkov 6T meraní v teréne

Dáta za rok 2008:

| meranie | S (2008) |         |         |         |         |
|---------|----------|---------|---------|---------|---------|
|         | 1        | 2       | 3       | 4       | 5       |
| 1       | 0.00532  | 0.01402 | 0.01095 | 0.00768 | 0.00282 |
| 2       | 0.01162  | 0.01347 | 0.00447 | 0.00495 | 0.00224 |
| 3       | 0.00718  | 0.01113 | 0.01383 | 0.00459 | 0.00340 |
| 4       | 0.00590  | 0.01066 | 0.00721 | 0.00301 | 0.00135 |
| 5       | 0.00700  | 0.01338 | 0.00877 | 0.00481 | 0.00376 |
| 6       | 0.00620  | 0.00570 | 0.00557 | 0.00396 | 0.00294 |
| 7       | 0.01002  | 0.01325 | 0.00357 | 0.00606 | 0.00161 |
| 8       | 0.00751  | 0.00536 | 0.00588 | 0.00174 | 0.00114 |
| 9       | 0.00845  | 0.00582 | 0.00258 | 0.00235 | 0.00087 |
| 10      | 0.01003  | 0.00223 | 0.00318 | 0.00229 | 0.00102 |
| 11      | 0.01019  | 0.00708 | 0.00558 | 0.00525 | 0.00463 |
| 12      | 0.00697  | 0.01101 | 0.00463 | 0.00385 | 0.00130 |
| 13      | 0.00780  | 0.01892 | 0.00682 | 0.00418 | 0.00202 |
| 14      | 0.01052  | 0.00981 | 0.00558 | 0.00636 | 0.00224 |
| 15      | 0.00825  | 0.00808 | 0.00432 | 0.00221 | 0.00111 |
| 16      | 0.00568  | 0.00880 | 0.00286 | 0.00766 | 0.00069 |
| 17      | 0.00336  | 0.01137 | 0.00350 | 0.00364 | 0.00096 |
| 19      | 0.00961  | 0.00962 | 0.00297 | 0.00422 | 0.00099 |
| 20      | 0.00968  | 0.01279 | 0.00662 | 0.00788 | 0.00825 |
| 21      | 0.01149  | 0.00796 | 0.00452 | 0.00751 | 0.00595 |

Nadväzujúce dáta za rok 2009:

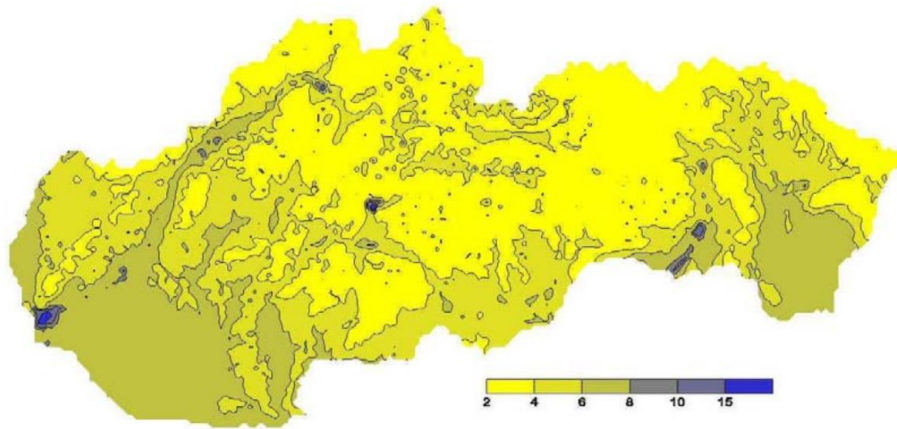
| meranie | S (2009) |         |         |
|---------|----------|---------|---------|
|         | 1        | 2       | 3       |
| 1       | 0.00224  | 0.01095 | 0.00768 |
| 2       | 0.00151  | 0.00447 | 0.00495 |
| 3       | 0.00137  | 0.01383 | 0.00459 |
| 4       | 0.00130  | 0.00721 | 0.00301 |
| 5       | 0.00183  | 0.00877 | 0.00481 |
| 6       | 0.00229  | 0.00557 | 0.00396 |
| 7       | 0.00230  | 0.00357 | 0.00606 |
| 8       | 0.00157  | 0.00588 | 0.00174 |
| 9       | 0.00194  | 0.00258 | 0.00235 |
| 10      | 0.00079  | 0.00318 | 0.00229 |
| 11      | 0.00255  | 0.00558 | 0.00525 |
| 12      | 0.00238  | 0.00463 | 0.00385 |
| 13      | 0.00256  | 0.00682 | 0.00418 |
| 14      | 0.00393  | 0.00558 | 0.00636 |
| 15      | 0.00189  | 0.00432 | 0.00221 |
| 16      | 0.00178  | 0.00286 | 0.00766 |
| 17      | 0.00186  | 0.00350 | 0.00364 |
| 19      | 0.00171  | 0.00297 | 0.00422 |
| 20      | 0.00244  | 0.00662 | 0.00788 |
| 21      | 0.00261  | 0.00452 | 0.00751 |

Identické výsledky meraní v teréne v dvoch po sebe nasledujúcich rokoch  
(chybný export z databázy pri integrácii dát)

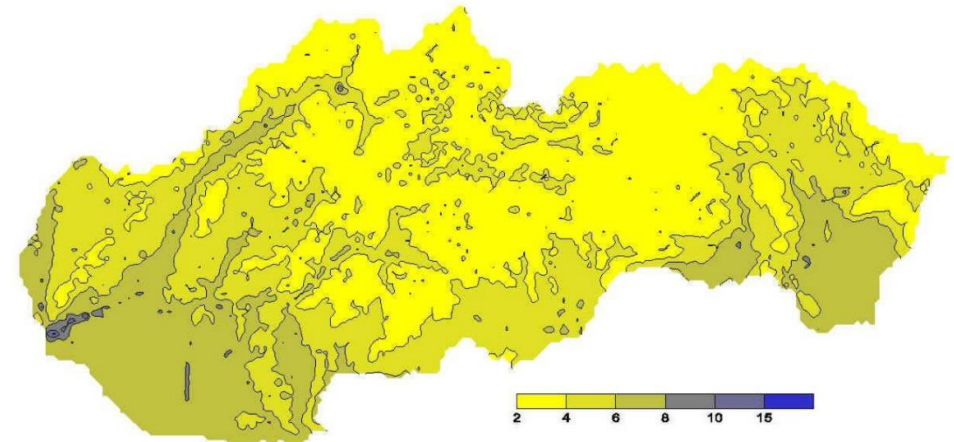
## Problémy v dátach z rokov 2014-15

1. Veličina NO<sub>2</sub> v roku 2015 dosahovala 2 až 3 násobné hodnoty oproti roku 2014, pričom simulácie SHMÚ pre tieto roky udávali zhruba rovnaké hodnoty NO<sub>2</sub>:

Obr. 5.5 Priemerná ročná koncentrácia NO<sub>2</sub> ( $\mu\text{g}\cdot\text{m}^{-3}$ ), rok 2014 na území Slovenskej republiky.



Obr. 5.5 Priemerná ročná koncentrácia NO<sub>2</sub> ( $\mu\text{g}\cdot\text{m}^{-3}$ ), rok 2015.



2. Veličina g<sub>0.2</sub> v rokoch 2014-15 dosahovala iba 1/9 až 1/10 hodnoty oproti predchádzajúcim rokom:
3. Podobné chyby v dátach v roku 2008

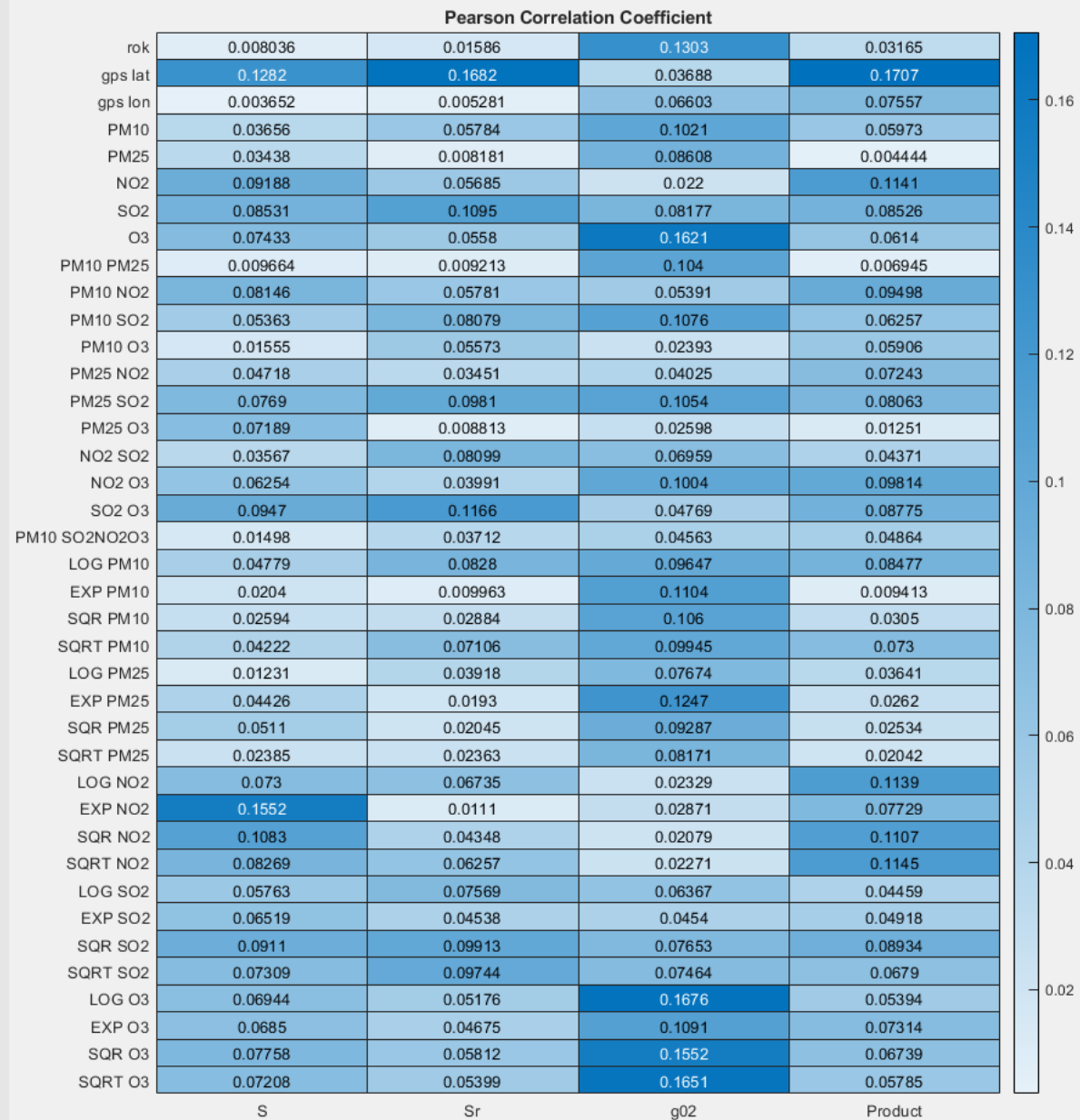
Celkovo : výrazne vysoký počet chybných dát, ktoré museli byť odstránené;  
Zostávajúcich počet záznamov cca 450.



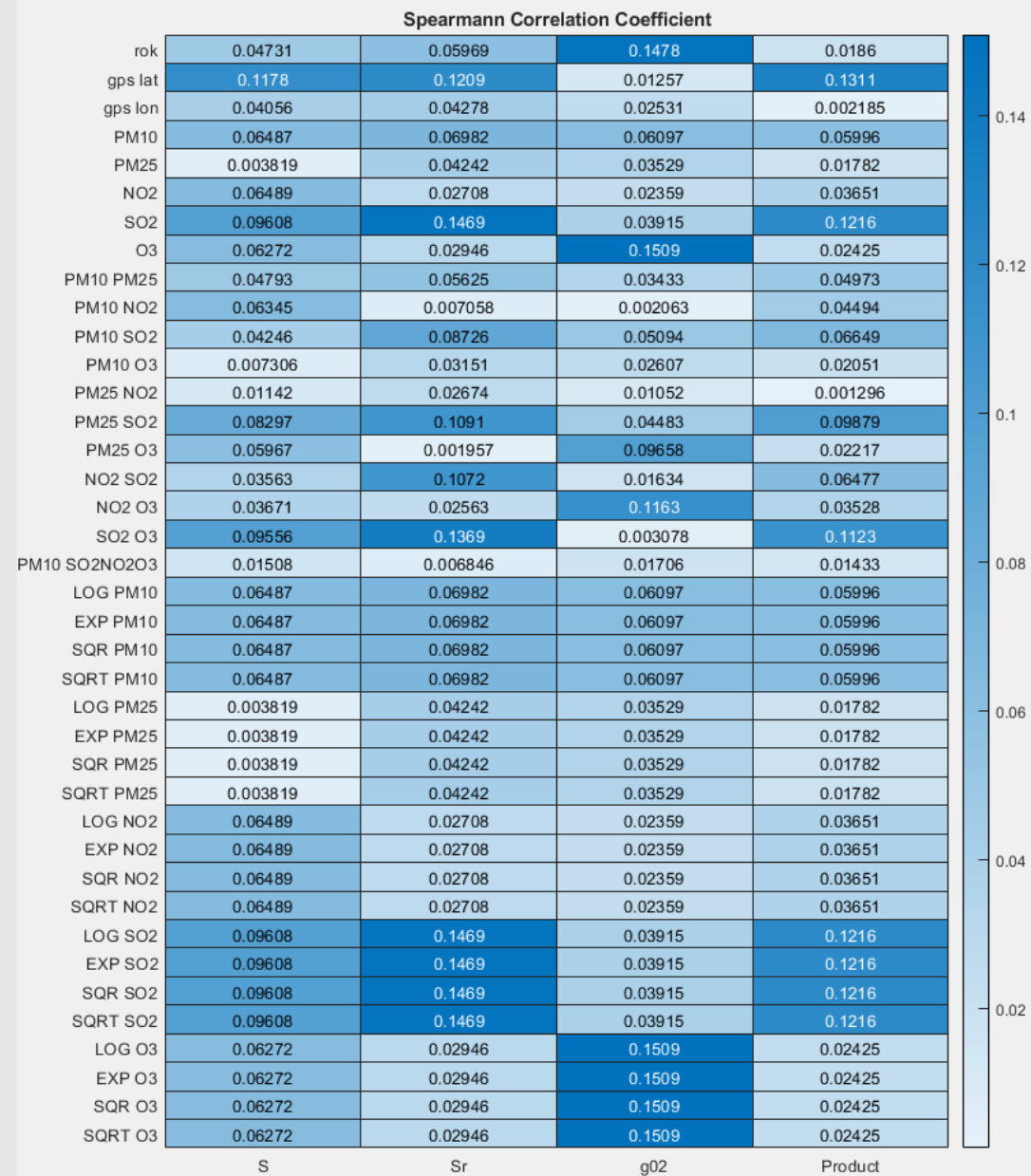
# Prvotné experimenty pri modelovaní

- Korelačná analýza
- doplnenie vstupných atribútov: Nelineárne transformácie vstupov, súčiny vstupov (kombinácie)
- Modelovanie za použitia klasifikačného (problém extrémne nevyváženého počtu vzoriek pri vyšších mierach znečistenia) aj regresného prístupu.
- Modelovanie použitím rozličných modelov (Gaussian Process, SMO Regression, Random Forest, Linear Regression, Regression Tree M5P, Multilayer Perceptron, Radial Basis Function Network).
- Použitie 40-násobnej krížovej validácie (pre dosiahnutie stabilných odhadov presnosti modelu).

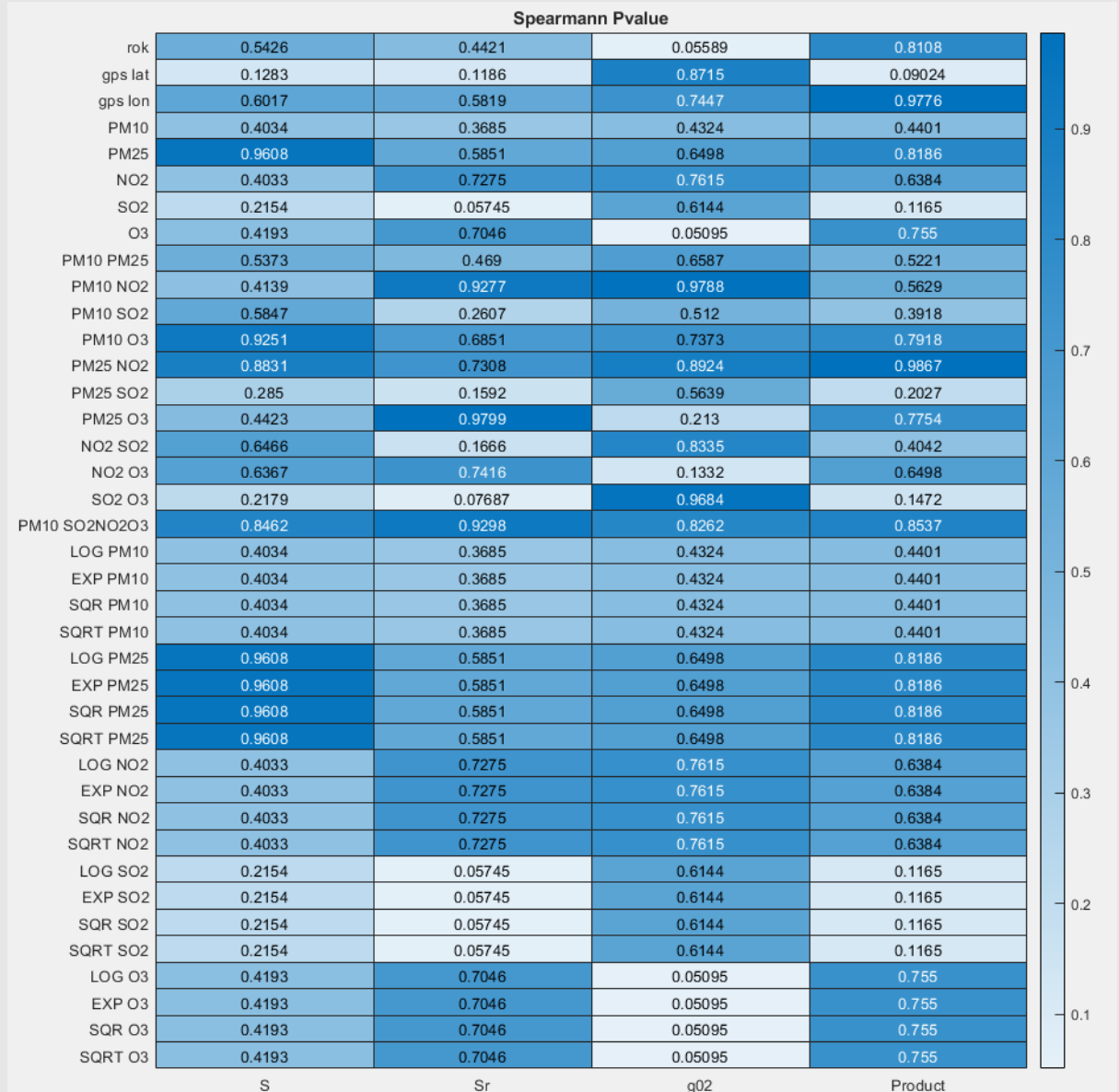
- Významnosť jednotlivých vstupných atribútov (vyjadrená absolútnou hodnotou Pearsonovho koeficientu) pre jednotlivé cieľové atribúty.



- Významnosť jednotlivých vstupných atribútov (vyjadrená absolútnou hodnotou Spearmanovho koeficientu) pre jednotlivé cieľové atribúty stanovené VUJE.



- o Matica p-hodnôt Spearmanovho korelačného koeficientu pre jednotlivé cieľové atribúty stanovené VUJE. Konvenčne sa za štatisticky významné považujú tie hodnoty Spearmanovho korelačného koeficientu, pre ktoré je p-hodnota menšia ako 0.05 -- tie vierohodne potvrdzujú existenciu nešpecifikovanej funkčnej závislosti medzi príslušným vstupným a cieľovým atribútom.



# Zistenia

- Použitie normalizácie / štandardizácie dát má minimálny resp. zanedbateľný vplyv na dosiahnutú presnosť
- Nevhodnosť modelovať priamo súčin cieľových veličín (v príp. Počítaných kvantilov najlepší model - korelácia iba 0.07)
- Model Radial Basis Function
  - CGD vs bez CGD;
  - s použitím CGD sú takmer vo všetkých prípadoch lepšie výsledky
- Výrazná variabilita v dosiahnutých presnostiach; zdá sa že modely odolnejšie voči chybám a menej citlivejšie na počet záznamov dávajú lepšie výsledky.
- Výrazný stochastický charakter / šum obsiahnutý v dostupných dátach
- Potreba vyššieho počtu záznamov pre lepšie generalizovanie závislostí

- V niektorých prípadoch závažný vplyv efektu preučenia
  - Príliš vysoký počet parametrov modelu
  - resp. nízky počet dátových záznamov / vzoriek
- Napr. Pri modeli MultiLayer Perceptron, pri odhade 0.995-kvantilu g02

**Model: Multilayer Perceptron, 2 hidden layers, max. epochs = 3500**

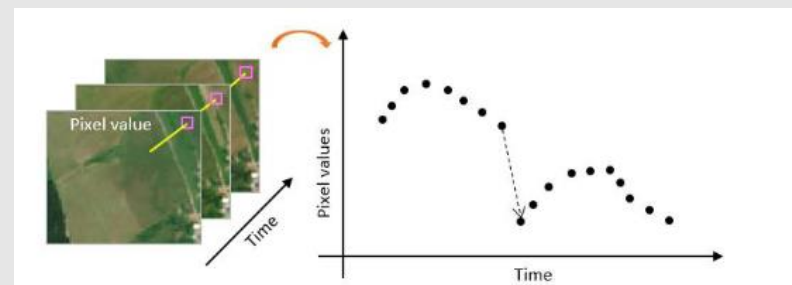
|                             |            |
|-----------------------------|------------|
| Correlation coefficient     | 0.3776     |
| Mean absolute error         | 496.7104   |
| Root mean squared error     | 639.6246   |
| Relative absolute error     | 151.7891 % |
| Root relative squared error | 153.4669 % |

**Model: Multilayer Perceptron, 2 hidden layers, max. epochs = 8500**

|                             |            |
|-----------------------------|------------|
| Correlation coefficient     | 0.2123     |
| Mean absolute error         | 574.903    |
| Root mean squared error     | 802.0972   |
| Relative absolute error     | 175.6839 % |
| Root relative squared error | 192.4494 % |

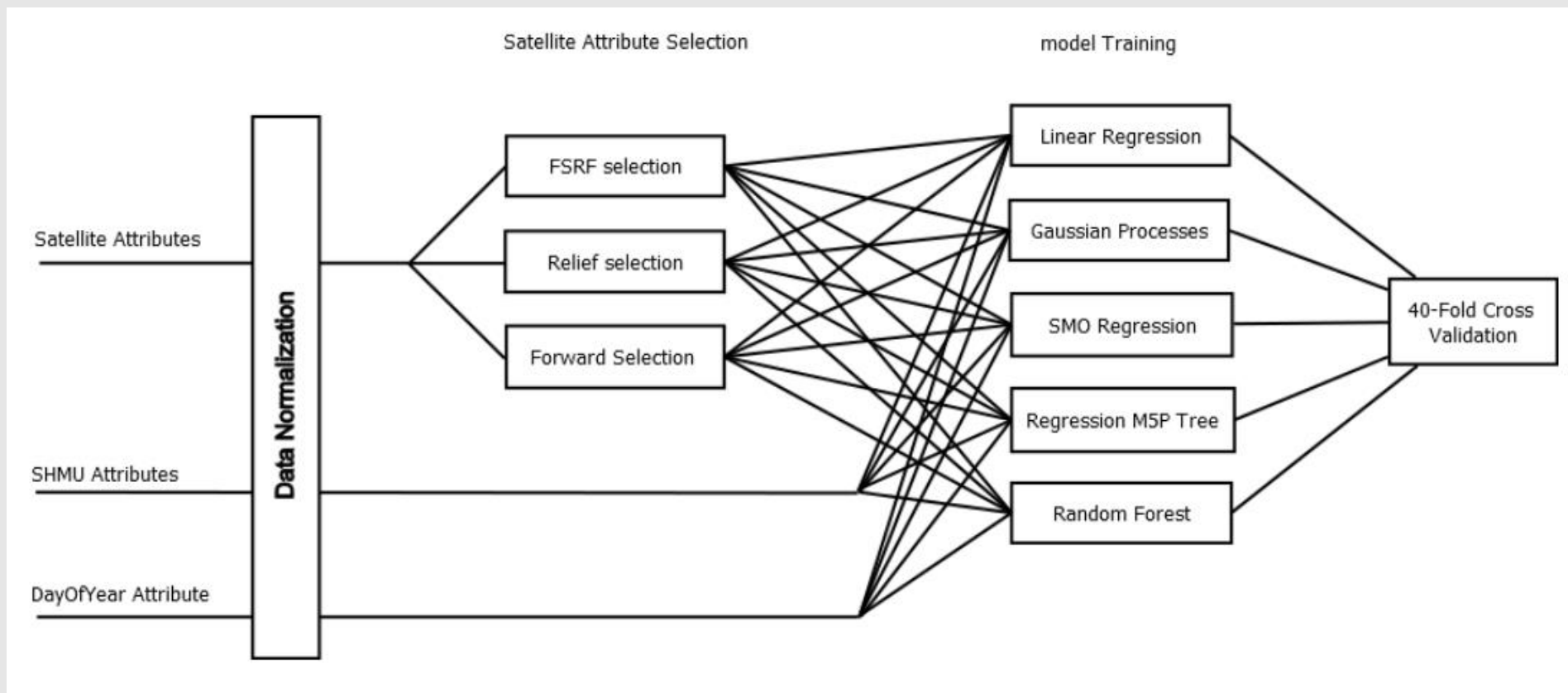
# Doplnenie vstupných atribútov o satelitné atribúty

- Doplnenie vstupných atribútov definovaných zo satelitných snímok
- Snaha o zohľadnenie vplyvov z okolia ako napr. stavebná či poľnohospodárska činnosť, na úroveň spadu v danej lokalite.
- Počet záznamov 450 -> 400 (dostupné satelity)
- Zadefinovaných rozličných 180 atribútov (funkcie min, max, mean, q25, q50, q75, aplikované pre časový aj priestorový aspekt, každá v 5 variantoch (abs, abs\_dif, dif, norm, ...)).
- Pre rôznych 13 frekvenčných kanálov
- $36 \times 5 \times 13 = 2340$  atribútov
- Metódy pre výber atribútov : FSRF, Relief, Forward selection



**Table 1.** List of used satellite spectral bands including normalized difference indices.

| Name  | Pixel Size | Wavelength | Description                               |
|---|------------|------------|---|
| AOT   | 10 m       |            | Aerosol optical thickness                 |
| B1  | 10 m       | 443.9 nm   | Aerosols                                  |
| B2  | 10 m       | 496.6 nm   | Blue                                      |
| B3  | 10 m       | 560.0 nm   | Green                                     |
| B4  | 10 m       | 664.5 nm   | Red                                       |
| B6  | 20 m       | 740.2 nm   | Red Edge 2                                |
| B8  | 10 m       | 835.1 nm   | NIR                                       |
| L1 B10 cir                                    | 60 m       | 1373.5 nm  | Cirrus                                    |
| B11   | 20 m       | 1613.7 nm  | SWIR 1                                    |
| NDVI (normalized difference vegetation index) | 10 m       |            | $NDVI = (B8 - B4)/(B8 + B4)$              |
| NDWI (normalized difference water index)      | 10 m       |            | $NDWI = (B3 - B8)/(B3 + B8)$              |
| NDSI (normalized difference soil index)       | 20 m       |            | $NDSI = (B3 - B11)/(B3 + B11)$            |
| Moisture index                                | 20 m       |            | $moisture\ index = (B8 - B11)/(B8 + B11)$ |





# Porovnanie presnosti natrénovaných modelov pre cieľovú veličinu S – celkový spad.

|   | Linear Model | Gaussian Process | SMO Regression | M5P Tree | Random Forest |
|---|--------------|------------------|----------------|----------|---------------|
| Models using attributes SHMU, without Day of Year and without Satellites                  |              |                  |                |          |               |
| Correlation coefficient   | 0.2195       | 0.2289           | 0.2377         | 0.1921   | 0.2243        |
| Relative absolute error   | 0.960381     | 0.958122         | 0.838752       | 0.963788 | 0.996284      |
| Root relative sq. error   | 0.975826     | 0.971225         | 1.013767       | 0.983825 | 1.007389      |
| Models using attributes SHMU and Day of Year, without Satellites                          |              |                  |                |          |               |
| Correlation coefficient   | 0.3044       | 0.3228           | 0.3221         | 0.3655   | 0.3810        |
| Relative absolute error   | 0.953682     | 0.931039         | 0.81304        | 0.873916 | 0.867675      |
| Root relative sq. error   | 0.952256     | 0.944247         | 0.990007       | 0.933188 | 0.960127      |
| Models using attributes SHMU, Satellites (FSRF-TEST 20 attr), without Day of Year         |              |                  |                |          |               |
| Correlation coefficient   | 0.3962       | 0.3962           | 0.3936         | 0.4011   | 0.5015        |
| Relative absolute error   | 0.929776     | 0.91984          | 0.776143       | 0.913888 | 0.834525      |
| Root relative sq. error   | 0.918874     | 0.916169         | 0.971942       | 0.928474 | 0.863463      |
| Model using attributes SHMU, Satellites (Relief 20 attr), without Day of Year             |              |                  |                |          |               |
| Correlation coefficient   | 0.2934       | 0.3322           | 0.2697         | 0.3291   | 0.3574        |
| Relative absolute error   | 0.998795     | 0.958923         | 0.827651       | 0.952494 | 0.930419      |
| Root relative sq. error   | 0.969389     | 0.945013         | 1.009543       | 0.972802 | 0.936803      |
| Models using attributes SHMU, Satellites (Forward selection 20 attr), without Day of Year |              |                  |                |          |               |
| Correlation coefficient   | 0.3754       | 0.3691           | 0.2771         | 0.2891   | 0.4214        |
| Relative absolute error   | 0.977038     | 0.942754         | 0.808266       | 1.005264 | 0.851436      |
| Root relative sq. error   | 0.934184     | 0.928854         | 0.999298       | 1.010215 | 0.905755      |
| Models using attributes SHMU, Satellites (Forward selection 10 attr), Day of Year         |              |                  |                |          |               |
| Correlation coefficient   | 0.4288       | 0.4475           | 0.4104         | 0.4263   | 0.3976        |
| Relative absolute error   | 0.948856     | 0.89844          | 0.778747       | 0.922759 | 0.861133      |
| Root relative sq. error   | 0.911549     | 0.892596         | 0.962801       | 0.916577 | 0.921097      |
| Models using attributes SHMU, Satellites (Forward selection 5 attr), Day of Year          |              |                  |                |          |               |
| Correlation coefficient   | 0.4635       | 0.4508           | 0.4506         | 0.3769   | 0.4381        |
| Relative absolute error   | 0.923401     | 0.896762         | 0.762388       | 0.924544 | 0.843924      |
| Root relative sq. error   | 0.886452     | 0.890677         | 0.952388       | 0.948177 | 0.898847      |
| Models using attributes SHMU, Satellites (FSRF-TEST 10 attr), Day of Year                 |              |                  |                |          |               |
| Correlation coefficient   | 0.4363       | 0.4339           | 0.4242         | 0.4686   | 0.5400        |
| Relative absolute error   | 0.900255     | 0.892416         | 0.761173       | 0.868294 | 0.806506      |
| Root relative sq. error   | 0.899348     | 0.899034         | 0.956712       | 0.885832 | 0.84258       |
| Models using attributes SHMU, Satellites (FSRF-TEST 5 attr), Day of Year                  |              |                  |                |          |               |
| Correlation coefficient   | 0.3956       | 0.3884           | 0.3983         | 0.4255   | 0.5183        |
| Relative absolute error   | 0.925902     | 0.911766         | 0.768899       | 0.885788 | 0.805674      |
| Root relative sq. error   | 0.918372     | 0.919311         | 0.969704       | 0.905563 | 0.853435      |
| Models using attributes SHMU, Satellites (Relief 10 attr), Day of Year                    |              |                  |                |          |               |
| Correlation coefficient   | 0.4016       | 0.4063           | 0.3959         | 0.4021   | 0.3987        |
| Relative absolute error   | 0.934508     | 0.910194         | 0.781673       | 0.908593 | 0.863821      |
| Root relative sq. error   | 0.921802     | 0.913254         | 0.960806       | 0.921916 | 0.916560      |
| Models using attributes SHMU, Satellites (Relief 5 attr), Day of Year                     |              |                  |                |          |               |
| Correlation coefficient   | 0.3533       | 0.3364           | 0.3490         | 0.3028   | 0.3347        |
| Relative absolute error   | 0.922774     | 0.915231         | 0.784270       | 0.912179 | 0.904339      |
| Root relative sq. error   | 0.938938     | 0.941382         | 0.975032       | 0.972598 | 0.956348      |
|   | Linear model | Gaussian process | SMO Regression | M5P Tree | Random forest |

Porovnanie presnosti  
natrénovaných modelov  
pre cieľovú veličinu  $S_r$  –  
množstvo rozpustných  
látok.

|  | Linear Model | Gaussian Process | SMO Regression | M5P Tree | Random Forest |
|--|--------------|------------------|----------------|----------|---------------|
| Models using attributes SHMU, without Day of Year, and without Satellites                |              |                  |                |          |               |
| Correlation coefficient  | 0.2302       | 0.1822           | 0.1934         | 0.2314   | 0.1898        |
| Relative absolute error  | 0.961795     | 0.974067         | 0.855066       | 0.960243 | 1.006067      |
| Root relative sq. error  | 0.971534     | 0.981626         | 1.014425       | 0.971155 | 1.016997      |
| Models using attributes SHMU, Day of Year without Satellites                             |              |                  |                |          |               |
| Correlation coefficient  | 0.2582       | 0.2236           | 0.2305         | 0.2053   | 0.2643        |
| Relative absolute error  | 0.952205     | 0.962065         | 0.857284       | 0.962442 | 0.913962      |
| Root relative sq. error  | 0.964646     | 0.973101         | 1.006680       | 0.988054 | 1.005748      |
| Models using attributes SHMU, Satellites (FSRF-TEST 20 attr) without Day of Year         |              |                  |                |          |               |
| Correlation coefficient  | 0.1428       | 0.1786           | 0.1742         | 0.065    | 0.3264        |
| Relative absolute error  | 0.991643     | 0.970063         | 0.859727       | 1.034066 | 0.920992      |
| Root relative sq. error  | 0.998483     | 0.984765         | 1.012168       | 1.097823 | 0.945923      |
| Models using attributes SHMU, Satellites (Relief 20 attr) without Day of Year            |              |                  |                |          |               |
| Correlation coefficient  | 0.0927       | 0.0648           | 0.1264         | 0.0789   | 0.1826        |
| Relative absolute error  | 1.016839     | 1.016156         | 0.877836       | 1.064591 | 0.967896      |
| Root relative sq. error  | 1.014047     | 1.007755         | 1.023097       | 1.049356 | 1.005119      |
| Models using attributes SHMU, Satellites (Forward selection 20 attr) without Day of Year |              |                  |                |          |               |
| Correlation coefficient  | 0.1883       | 0.1210           | 0.233          | 0.1469   | 0.1377        |
| Relative absolute error  | 1.005077     | 0.986239         | 0.835902       | 1.031335 | 0.953701      |
| Root relative sq. error  | 0.995576     | 1.000356         | 0.992946       | 1.015140 | 1.007062      |
| Models using attributes SHMU, Satellites (Forward selection 10 attr), Day of Year        |              |                  |                |          |               |
| Correlation coefficient  | 0.2398       | 0.3101           | 0.3535         | 0.1445   | 0.2839        |
| Relative absolute error  | 1.022736     | 0.948921         | 0.815373       | 1.054828 | 0.870082      |
| Root relative sq. error  | 0.995421     | 0.951980         | 0.966335       | 1.135082 | 0.961922      |
| Models using attributes SHMU, Satellites (Forward selection 15 attr), Day of Year        |              |                  |                |          |               |
| Correlation coefficient  | 0.2611       | 0.3029           | 0.3283         | 0.2056   | 0.2714        |
| Relative absolute error  | 1.007581     | 0.959974         | 0.805193       | 0.998895 | 0.878738      |
| Root relative sq. error  | 0.982944     | 0.955482         | 0.972581       | 1.021172 | 0.962601      |
| Models using attributes SHMU, Satellites (FSRF-TEST 10 attr), Day of Year                |              |                  |                |          |               |
| Correlation coefficient  | 0.2888       | 0.323            | 0.3305         | 0.3005   | 0.3698        |
| Relative absolute error  | 0.947149     | 0.919711         | 0.804691       | 0.937511 | 0.866946      |
| Root relative sq. error  | 0.964184     | 0.945773         | 0.969808       | 0.960605 | 0.928711      |
| Models using attributes SHMU, Satellites (FSRF-TEST 5 attr), Day of Year                 |              |                  |                |          |               |
| Correlation coefficient  | 0.3299       | 0.3438           | 0.36           | 0.3284   | 0.3999        |
| Relative absolute error  | 0.936913     | 0.913589         | 0.793238       | 0.938661 | 0.866407      |
| Root relative sq. error  | 0.945296     | 0.937339         | 0.965524       | 0.945354 | 0.921331      |
| Models using attributes SHMU, Satellites (Relief 10 attr), Day of Year                   |              |                  |                |          |               |
| Correlation coefficient  | 0.2194       | 0.2357           | 0.2848         | 0.2272   | 0.3102        |
| Relative absolute error  | 0.964405     | 0.956218         | 0.846194       | 0.963997 | 0.869070      |
| Root relative sq. error  | 0.986047     | 0.975347         | 0.987452       | 0.988660 | 0.949752      |

Porovnanie presnosti  
 natrénovaných  
 modelov pre cieľovú  
 veličinu g02 – merná el.  
 vodivosť.

|   | Linear Model | Gaussian Process | SMO Regression | M5P Tree | Random Forest |
|---|--------------|------------------|----------------|----------|---------------|
| Models using attributes SHMU, without Day of Year, and without Satellites                 |              |                  |                |          |               |
| Correlation coefficient   | -0.0888      | 0.006            | -0.0239        | 0.076    | 0.1191        |
| Relative absolute error   | 1.013206     | 1.011356         | 0.996436       | 1.000262 | 1.041810      |
| Root relative sq. error   | 1.014285     | 1.005392         | 1.018875       | 1.000869 | 1.044912      |
| Models using attributes SHMU, Day of Year without Satellites                              |              |                  |                |          |               |
| Correlation coefficient   | 0.4113       | 0.4373           | 0.4342         | 0.6206   | 0.6339        |
| Relative absolute error   | 0.903762     | 0.894292         | 0.886614       | 0.755594 | 0.756144      |
| Root relative sq. error   | 0.911274     | 0.897203         | 0.905929       | 0.782397 | 0.778181      |
| Modeling using attributes SHMU, Satellites (FSRF-TEST 20 attr), without Day of Year       |              |                  |                |          |               |
| Correlation coefficient   | 0.4195       | 0.4314           | 0.405          | 0.3931   | 0.4062        |
| Relative absolute error   | 0.904127     | 0.892004         | 0.914068       | 0.91234  | 0.910343      |
| Root relative sq. error   | 0.908831     | 0.899933         | 0.919583       | 0.925257 | 0.924097      |
| Models using attributes SHMU, Satellites (Relief 20 attr), without Day of Year            |              |                  |                |          |               |
| Correlation coefficient   | 0.4135       | 0.3942           | 0.3798         | 0.3853   | 0.4601        |
| Relative absolute error   | 0.903282     | 0.905045         | 0.910703       | 0.912017 | 0.881088      |
| Root relative sq. error   | 0.912632     | 0.918536         | 0.935326       | 0.929368 | 0.889698      |
| Models using attributes SHMU, Satellites (Forward selection 20 attr), without Day of Year |              |                  |                |          |               |
| Correlation coefficient   | 0.3155       | 0.3276           | 0.3257         | 0.3426   | 0.4708        |
| Relative absolute error   | 0.942383     | 0.931576         | 0.946136       | 0.910879 | 0.873131      |
| Root relative sq. error   | 0.953646     | 0.946278         | 0.955216       | 0.951316 | 0.883867      |
| Models using attributes SHMU, Satellites (Manual selection 3 attr), Day of Year           |              |                  |                |          |               |
| Correlation coefficient   | 0.5163       | 0.5245           | 0.5092         | 0.6275   | 0.6004        |
| Relative absolute error   | 0.855031     | 0.848659         | 0.849584       | 0.76101  | 0.78007       |
| Root relative sq. error   | 0.855314     | 0.849386         | 0.863360       | 0.77716  | 0.800759      |
| Models using attributes SHMU, Satellites (Forward selection 5 attr), Day of Year          |              |                  |                |          |               |
| Correlation coefficient   | 0.5562       | 0.5565           | 0.5305         | 0.6058   | 0.6434        |
| Relative absolute error   | 0.827117     | 0.82688          | 0.835880       | 0.788339 | 0.761934      |
| Root relative sq. error   | 0.830543     | 0.829219         | 0.852279       | 0.795252 | 0.764349      |
| Models using attributes SHMU, Satellites (FSRF-TEST 5 attr), Day of Year                  |              |                  |                |          |               |
| Correlation coefficient   | 0.5635       | 0.5661           | 0.5354         | 0.5608   | 0.626         |
| Relative absolute error   | 0.822849     | 0.82399          | 0.850483       | 0.825195 | 0.762599      |
| Root relative sq. error   | 0.825459     | 0.822494         | 0.849352       | 0.830468 | 0.779014      |
| Models using attributes SHMU, Satellites (Relief 5 attr), Day of Year                     |              |                  |                |          |               |
| Correlation coefficient   | 0.4891       | 0.5099           | 0.4858         | 0.5494   | 0.6368        |
| Relative absolute error   | 0.87403      | 0.861349         | 0.869334       | 0.822957 | 0.759978      |
| Root relative sq. error   | 0.872218     | 0.858429         | 0.881012       | 0.837942 | 0.769692      |
|   | Linear model | Gaussian process | SMO Regression | M5P Tree | Random forest |

# Štatistické testovanie spresnenia modelov

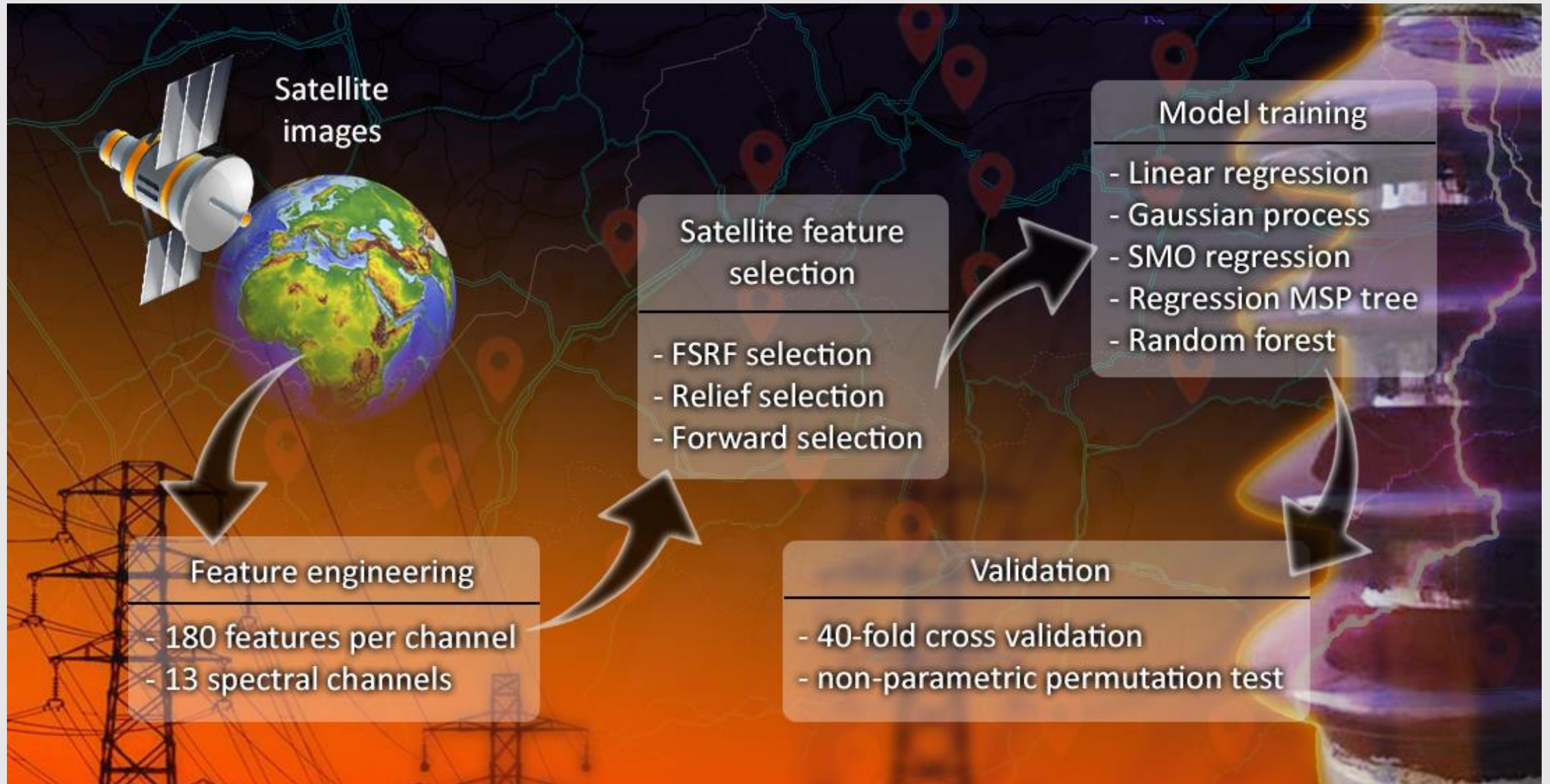
Neparametrický permutačný test

Trénovanie a validácia opakované 30-krát,  
pre rozličné hodnoty Seedu modelu RF

$X_A$  a  $X_B$  reprezentujú presnosť modelov  
(Pearson Correlation)

Test statistic: absolute difference of mean values (abs (mean ( $X_A$ ) – mean ( $X_B$ )).  
Number of permutations in the statistical test = 1,000,000;  
Significance level alpha = 0.05;  
Null hypothesis:  $X_A$  and  $X_B$  have the same mean value;  
Alternative hypothesis:  $X_A$  and  $X_B$  do not have the same mean value.

| Target attribute S:       |                                 |                 |
|---------------------------|---------------------------------|-----------------|
| Used attributes for $X_A$ | Used attributes for $X_B$       | <i>p</i> -value |
| SHMU                      | SHMU + Day of Year              | 0.000238        |
| SHMU                      | SHMU + Satellites               | 0.000179        |
| SHMU + Day of Year        | SHMU + Day of Year + Satellites | 0.000617        |
| SHMU + Satellites         | SHMU + Day of Year + Satellites | 0.000698        |
| Target Attribute Sr:      |                                 |                 |
| Used attributes for $X_A$ | Used attributes for $X_B$       | <i>p</i> -value |
| SHMU                      | SHMU + Day of Year              | 0.000297        |
| SHMU                      | SHMU + Satellites               | 0.000119        |
| SHMU + Day of Year        | SHMU + Day of Year + Satellites | 0.000917        |
| SHMU + Satellites         | SHMU + Day of Year + Satellites | 0.002740        |
| Target Attribute g02:     |                                 |                 |
| Used attributes for $X_A$ | Used attributes for $X_B$       | <i>p</i> -value |
| SHMU                      | SHMU + Day of Year              | 0.000249        |
| SHMU                      | SHMU + Satellites               | 0.000314        |
| SHMU + Day of Year        | SHMU + Day of Year + Satellites | 0.012501        |
| SHMU + Satellites         | SHMU + Day of Year + Satellites | 0.006714        |



# Záver

- Úloha sa ukazuje aj naďalej ako veľmi problematická, vzhľadom na nízky počet dostupných meraní, výrazný stochastický charakter a veľa externých vplyvov.
- Pridaním atribútov vypočítaných zo satelitných snímok sa podarilo zvýšiť presnosť regresných modelov a to pre všetky 3 cieľové veličiny.
- Miera zvýšenia presnosti (po pridaní satelitných atribútov) sa v prípade všetkých 3 cieľových veličín sa ukázala ako štatisticky významná pri použití permutačného testu, testovacia štatistika : absolútny rozdiel stredných hodnôt [2].
- Podarilo sa identifikovať štatistické rozdelenie intenzity znečistenia PM10 a PM2.5, potvrdiť, že sa jedná o Weibullovo rozdelenie pravdepodobnosti a určiť jeho parametre.
- Odpublikované 3 články [1-3], práca na ďalšom aktuálne prebieha, 1 prerušený (špecifická téma, len okrajovo súvisí s informatikou), pripravované ďalšie experimenty..
- Aj napriek zvýšeniu presnosti modelov, nie je ich presnosť dostatočná na ich aplikovanie do praxe (celková miera určujúca stupeň znečistenia je daná 3 cieľovými veličinami).

# Možnosti ďalšieho smerovania výskumu

Aj napriek doterajšej relatívne nízkej presnosti modelov, veríme že dodatočné vysoko-relevantné vstupy umožnia výrazné zvýšenie presnosti modelov.

- Zdroje znečisťovateľov (podniky – ich pozícia, intenzita a typ znečistenia) **(nedostatočne podrobné, krátke obdobie)**
- Údaje o doprave **(nedostatočne podrobné)**
- Merané zvukové/ultrazvukové záznamy a pridružené merania (relatívna vlhkosť, PM100, ...); **(dostupné len od 2022 – krátke obdobie)**
- Údaje z tzv. mokrej depozície (merané údaje o znečistení len počas zrážok samostatne) **(očakáva sa dodanie)**
- Úrovně znečistenia NO<sub>2</sub>, O<sub>3</sub>, SO<sub>2</sub>, PM<sub>10</sub>, PM<sub>2.5</sub>, CO na dennej / hodinovej báze, od SHMÚ **(aktuálne už dostupné za cca 10 rokov)**
- Radarové snímky z okolia daných lokalít **(dostupné, publikácia sa pripravuje)**

# Hlavné scientometrické výstupy

- [1] KRAMMER, Peter - KVASSAY, Marcel - FORGÁČ, Radoslav - OČKAY, Miloš - SKOVAJSOVÁ, Lenka - HLUCHÝ, Ladislav - SKURČÁK, Ľuboš - PAVLOV, Ľuboš: Regression analysis and modeling of local environmental pollution levels for the electric power industry needs. In Computing and informatics, 2022, vol. 41, no. 3, p. 861-884. (2021: 0.455 - IF, Q4 - JCR, 0.226 - SJR, Q4 - SJR). ISSN 1335-9150. [https://doi.org/10.31577/cai\\_2022\\_3\\_861](https://doi.org/10.31577/cai_2022_3_861) Typ: ADDA
- [2] KRAMMER, Peter - KVASSAY, Marcel - MOJŽIŠ, Ján - KENYERES, Martin - OČKAY, Miloš - HLUCHÝ, Ladislav - PAVLOV, Ľuboš - SKURČÁK, Ľuboš: Using satellite imagery to improve local pollution models for high-voltage transmission lines and insulators. In Future internet, 2022, vol. 14, no. 4, art. no. 99. (2021: 0.793 - SJR, Q2 - SJR). ISSN 1999-5903. <https://doi.org/10.3390/fi14040099> Typ: ADMB
- [3] PAVLOV, Ľuboš - SKURČÁK, Ľuboš - HLUCHÝ, Ladislav - KVASSAY, Marcel - KRAMMER, Peter - KREMLER, Martin. Use of artificial intelligence to determine the degree of environmental pollution. In Proceedings of the 11th International Scientific Symposium on Electrical Power Engineering : ELEKTROENERGETIKA 2022. - Košice: Technical University of Košice, 2022, p. 88-92. ISBN 978-805534104-0. (ELEKTROENERGETIKA 2022 : International scientific symposium) Typ: ADNB