



# Umelá inteligencia pre predikovanie vplyvov životného prostredia na elektroenergetiku

*M. Kvassay, P. Krammer, R. Forgáč*

*Ústav informatiky Slovenskej akadémie vied  
Bratislava*



## 1. Úvod do problematiky

- Problém a doména: napäťové preskoky cez znečistené izolátory vedení vysokého napätia a ako im predchádzať
- Vstupné dáta poskytnuté SHMÚ a VUJE
- Zadanie úlohy od VUJE
- Proces CRISP-DM: teória a realita

## 2. Predbežné vyhodnotenie klasických klasifikačných modelov

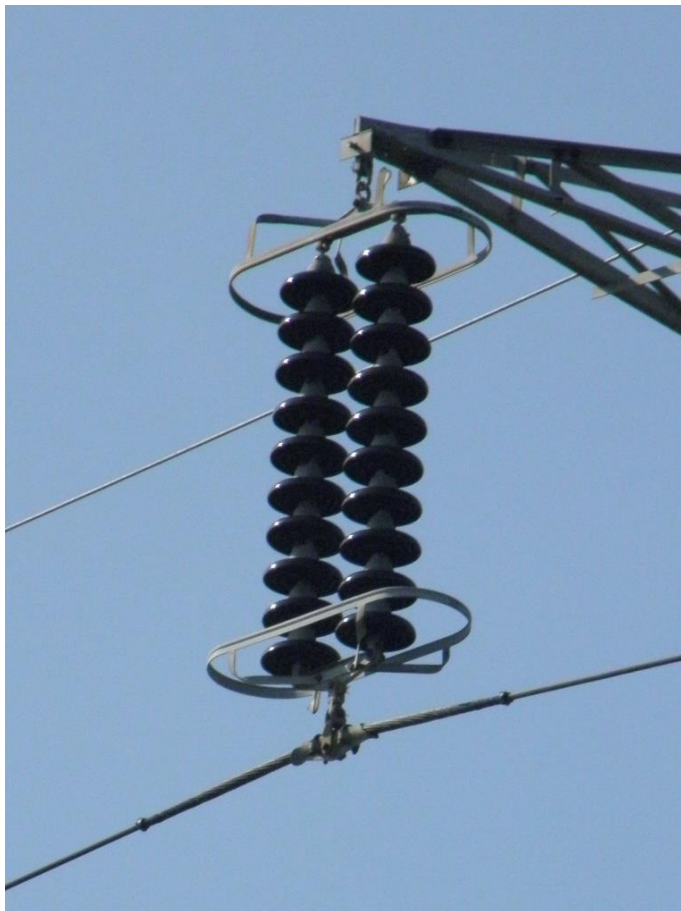
## 3. Pokus o konštrukciu nových atribútov z webových dát SHMÚ

## 4. Analýza predbežných výsledkov a návrh ďalšieho postupu



# 1. Úvod do problematiky

# Problém: znečistenie izolátorových závesov VVN a hrozba nežiadúcich napäťových preskokov

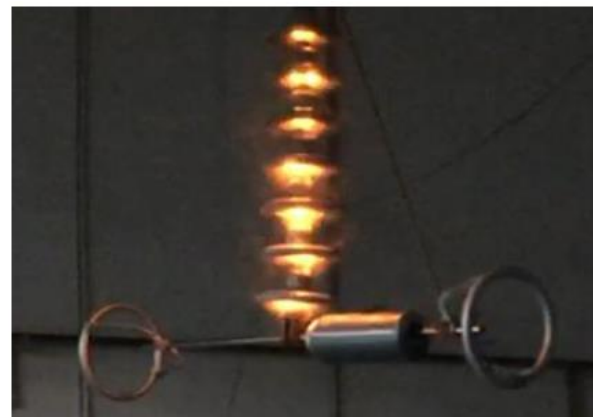


Nebezpečenstvo pri hmle, mrholení, daždi a vysokej vlhkosti v kombinácii s vodivými nečistotami na povrchu izolátorov

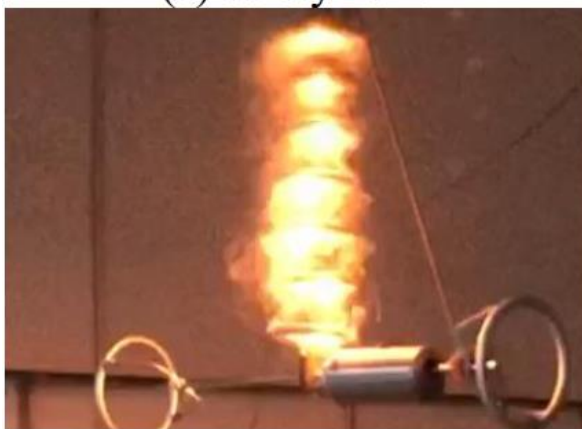
# Fázy napät'ového preskoku cez izolátorový záves



(a) safety zone



(b) forecast zone



(c) danger zone



(d) flashover

Figure 3. Discharge phenomena at different discharge stages

**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.

# Prevenca: 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.

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



# Ukážka nameraných dát od partnera VUJE, a.s.



	A	E	F	G	H	I	O	P	Q	W	X	Y	AE	AF	AG	AH	AI
1	Namerané hodnoty												Ročné výsledky				
2			Celkové množstvo spadu S (mg.cm <sup>-2</sup> .deň <sup>-1</sup> )			Množstvo rozpustných látok S <sub>r</sub> (mg.cm <sup>-2</sup> .deň <sup>-1</sup> )			Merná el. vodivosť 0,2% roztoku rozp. látok γ <sub>0,2</sub> (μS.cm <sup>-1</sup> )			Celkové množstvo spadu S <sub>m</sub> (mg.cm <sup>-2</sup> .deň <sup>-1</sup> )	Množstvo rozp. látok S <sub>rm</sub> (mg.cm <sup>-2</sup> .deň <sup>-1</sup> )	Merná vodivosť 0,2% roztoku pri 20°C (μS.cm <sup>-1</sup> )	p (súčín)	Stupeň znečistenia	
3	Doba merania		Číslo zberu			Číslo zberu			Číslo zberu								
4	Stanice	od	do	1	2	3	1	2	3	1	2	3					
5	Podunajské biskupice	13/05/2008	14/04/2009	0.00532007	0.01402077	0.01094552	0.00182158	0.00391103	0.00392174	1256.2	1411.6	1254.5	0.021485180	0.005984409	2192.68	0.28	I.
6	Senica 220 kV	13/05/2008	14/04/2009	0.00717915	0.01112768	0.0138279	0.00179479	0.00199302	0.00126439	1277.7	953.3	1069.6	0.024007898	0.002585061	2135.60	0.13	I.
7	Bošaca 400 kV	13/05/2008	14/04/2009	0.00589869	0.01066157	0.00720594	0.00214839	0.00288773	0.0012001	1475.9	1485.7	1211.6	0.017694859	0.003666455	2058.38	0.13	I.
8	Lemešany 400 kV	14/05/2008	15/04/2009	0.01019011	0.00708271	0.00557723	0.0024377	0.00124831	0.00051968	1402.1	1502.7	1261.6	0.013381296	0.004684182	2070.61	0.13	I. **)
9	Lemešany 220 kV	14/05/2008	15/04/2009	0.01070443	0.00688984	0.00468252	0.00252342	0.00156977	0.00087864	1333.9	1258.9	1178.4					
10	Rimavská Sobota 400 kV	14/05/2008	16/04/2009	0.00825066	0.00808458	0.00432356	0.00230911	0.00209481	0.00075542	1200.0	1220.5	1182.1	0.014782757	0.003673845	1949.54	0.11	I.
11	Levice 400 kV	15/05/2008	16/04/2009	0.00568439	0.00879714	0.00286094	0.00158584	0.00171978	0.00056255	1350.7	971.1	1007.1	0.016208572	0.004855633	2240.28	0.18	I.
12	Veľký Ďur 400 kV	15/05/2008	16/04/2009	0.00336456	0.01136877	0.0034985	0.00064291	0.00302167	0.00061612	1126.2	1033.9	1193.8	0.016471624	0.004889243	2484.27	0.20	I.
13	Bystričany 220 kV	15/05/2008	17/04/2009	0.00967578	0.01278853	0.0066166	0.00298441	0.00412533	0.00197159	1416.6	1297.3	1431.3	0.017324492	0.005810816	1992.46	0.20	I.
14	Križovany 400 kV	15/05/2008	17/04/2009	0.01148737	0.00796135	0.00451643	0.00342946	0.00235733	0.00092686	1310.7	1262.5	1311.6	0.016561182	0.004818413	2222.45	0.18	I. **)
15	Križovany 220 kV	15/05/2008	17/04/2009	0.00960248	0.00675054	0.00379852	0.0028797	0.00243234	0.00081435	1237.5	987.5	1078.6					
16	V427	05/05/2011	05/04/2012	0.0097	0.0190	0.0442	0.0031	0.0046	0.0180	1200.89	1268.75	821.43	0.106559	0.037356	1470.006	5.85	III.
17	V427	05/05/2011	05/04/2012	0.0153	0.0455	0.0153	0.0028	0.0104	0.0045	1291.07	984.82	929.46	0.113406	0.02115	1915.027	4.59	II.
18	V496	12/05/2010	20/04/2011	0.0177	0.0249	0.0359	0.0073	0.0081	0.0068	1041.1	1258.9	1192	0.078765523	0.054579695	2551.631659	10.97	IV.
19	Poznámky:																
20	**) - hodnoty boli vypočítané pre rozvodňu ako celok																

# Pokus o jednoduchšie určovanie stupňa znečistenia pomocou veličín pravidelne monitorovaných SHMÚ



## Hodinové koncentrácie znečisťujúcich látok - ale iba na vybraných miestach

Aktuálne zverejnené hodinové údaje majú len informatívny charakter, nie sú okamžite vyhodnocované.

Údaje sa kontrolujú a vyhodnocujú na základe technických parametrov prístroja nasledujúci pracovný deň v ranných hodinách poverenými pracovníkmi SHMÚ.

Dátum a čas merania: 20.05.2021 - 10:00

Zobraziť

|<<

<

>

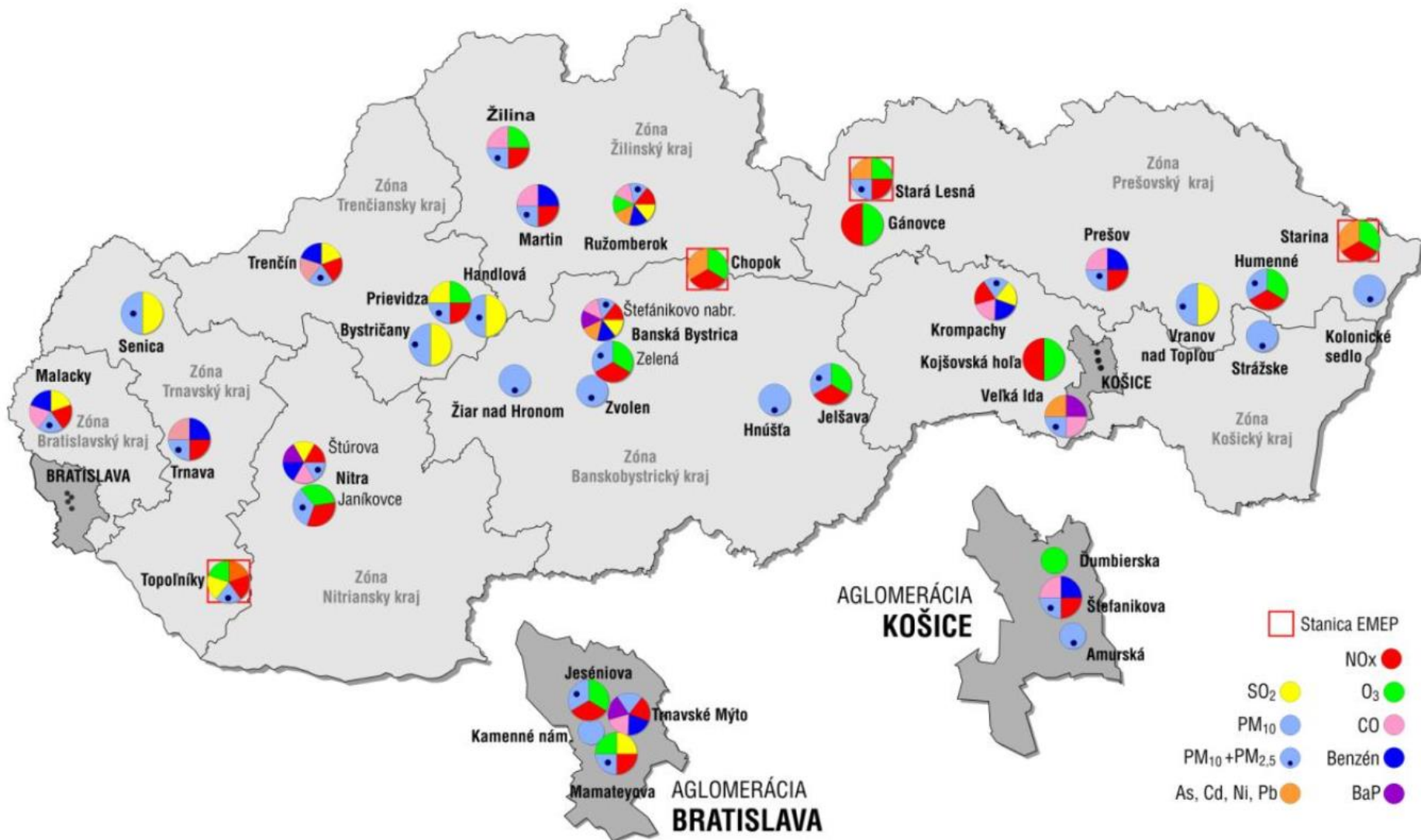
>>|

### Stanice SHMÚ

Stanica	PM <sub>10</sub>	PM <sub>2.5</sub>	O <sub>3</sub>	SO <sub>2</sub>	CO	NO <sub>2</sub>	NO <sub>x</sub>	Benzén	Hg
Bratislava, Kamenné nám.	5	4							
Bratislava, Trnavské Mýto	12	5			260	32	79	0.3	
Bratislava, Jeséniova	9	6	64	7		4	6		
Bratislava, Mamateyova	5	3	61	3		9	14		
Rovinka, mobilná AMS	11			6	132	12	14	0.2	
Pezinok	16	5			142	12	12		
Malacky, Mierové námestie	9	6		5	571	26	59	0.2	
Nitra, Janíkovce	12	6	79			4	5		
Nitra, Štúrova	13	6		29	302	18	33	0.2	
Topoľníky, Aszód, EMEP	7	5	63	4		4	5		1.1
Senica, Hviezdoslavova	9	4		15					
Trnava, Kollárova	15	5			660	35	85	0.3	
Sereď, Vinárska	14	6				11	17		
Trenčín, Hasičská	8	4		1	174	8	11	0.2	



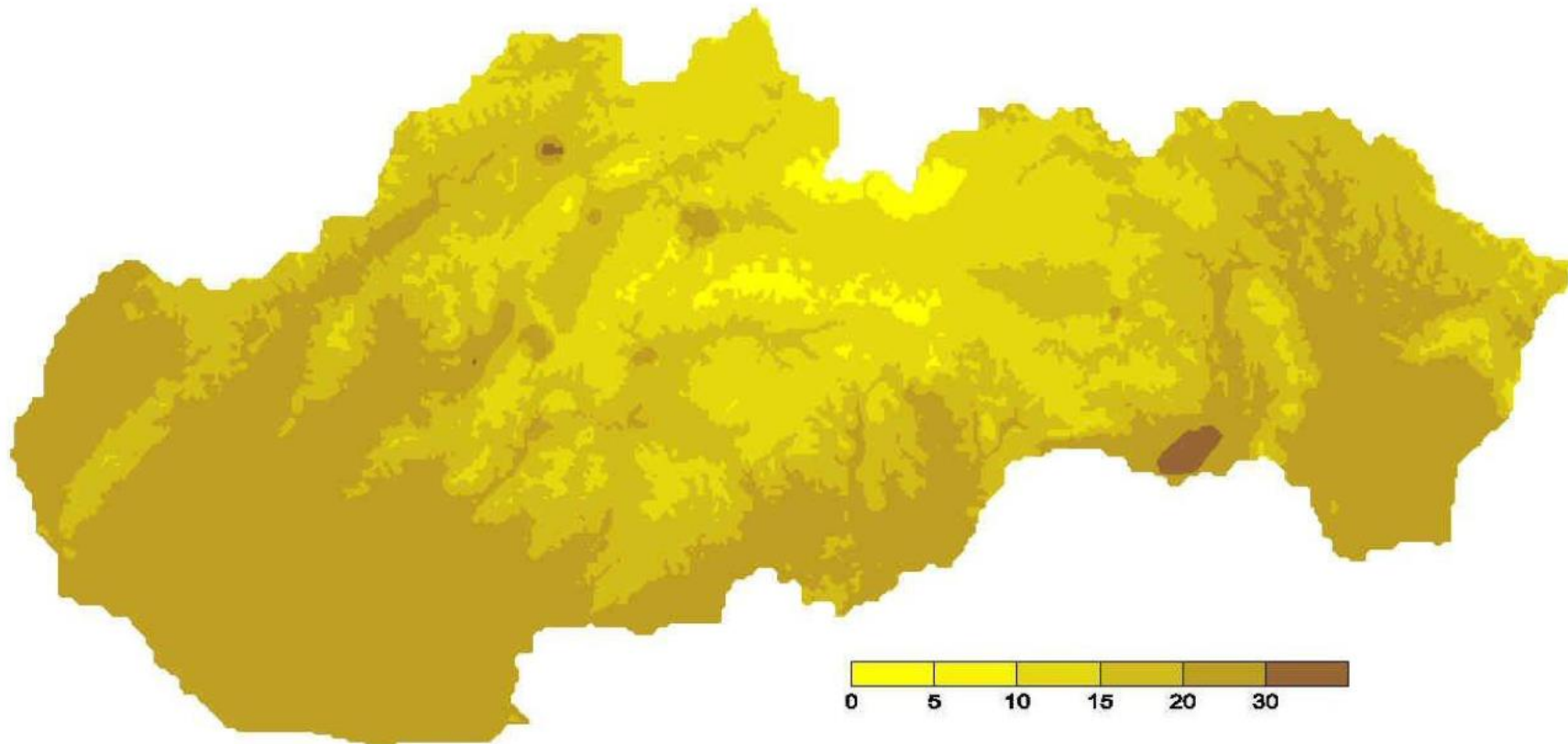
# Rozmiestnenie meracích staníc SHMÚ a ich program



# SHMÚ numerickými algoritmami rozpočíta znečistenie na celé územie Slovenska (ale iba ročné priemery)



Obr. 5.12 Priemerná ročná koncentrácia  $PM_{10}$  [ $\mu\text{g}\cdot\text{m}^{-3}$ ], rok 2016.



# Ukážka našich vstupných dát od VUJE+SHMÚ



meranie	Vstupné atribúty							Výstupné atribúty		
	lat	lon	PM10	PM2.5	NO2	SO2	O3	S	Sr	g0.2
1	48.11267	17.21781	21.00	14.70	7.40	10.12	47.86	0.02150	0.00600	2192.70
2	48.30581	17.03103	21.06	14.74	6.90	4.61	57.75	0.02440	0.00400	2310.90
3	48.66647	17.39286	21.82	15.27	6.30	4.09	60.56	0.02400	0.00260	2135.60
4	48.79436	17.83872	21.00	14.70	6.30	4.27	48.70	0.01770	0.00370	2058.40
5	49.11592	18.42628	17.68	15.03	7.00	4.02	45.47	0.01960	0.00440	2002.50
6	49.21431	18.86353	29.33	24.93	5.40	3.65	48.50	0.01100	0.00610	1697.20
7	49.09164	18.97614	29.50	25.08	5.60	4.00	48.88	0.02130	0.00530	2598.90
8	49.08581	19.49019	25.18	21.40	6.60	4.41	61.13	0.01310	0.00340	2251.80
9	48.79694	19.36706	21.95	18.65	5.60	3.60	44.86	0.01290	0.00440	2170.30
10	48.96456	20.53486	17.68	15.03	6.70	4.46	61.12	0.01590	0.00500	1924.20
11	48.85797	21.25386	27.45	19.21	6.90	4.66	55.35	0.01340	0.00470	2070.60
12	48.56881	22.04022	21.00	14.70	7.20	4.68	51.77	0.01660	0.00420	2459.40
13	48.85622	21.84814	24.70	17.29	7.50	4.57	53.96	0.02750	0.00440	2264.30
14	48.60181	21.04483	31.73	21.97	6.60	4.39	43.46	0.01850	0.00820	1652.90
15	48.38006	20.04481	22.55	15.78	6.80	4.22	47.70	0.01480	0.00370	1949.50
16	48.23336	18.60417	21.00	14.70	8.00	4.50	56.36	0.01620	0.00490	2240.30
17	48.20564	18.42531	21.00	14.70	6.10	4.43	59.00	0.01650	0.00490	2484.30
19	48.55900	18.74136	22.56	19.18	5.60	4.62	37.68	0.02000	0.00500	1990.00
20	48.66664	18.51206	28.54	19.01	6.70	19.40	48.96	0.01730	0.00580	1992.50
21	48.33717	17.65869	21.42	15.00	7.00	4.68	59.82	0.01660	0.00480	2222.50

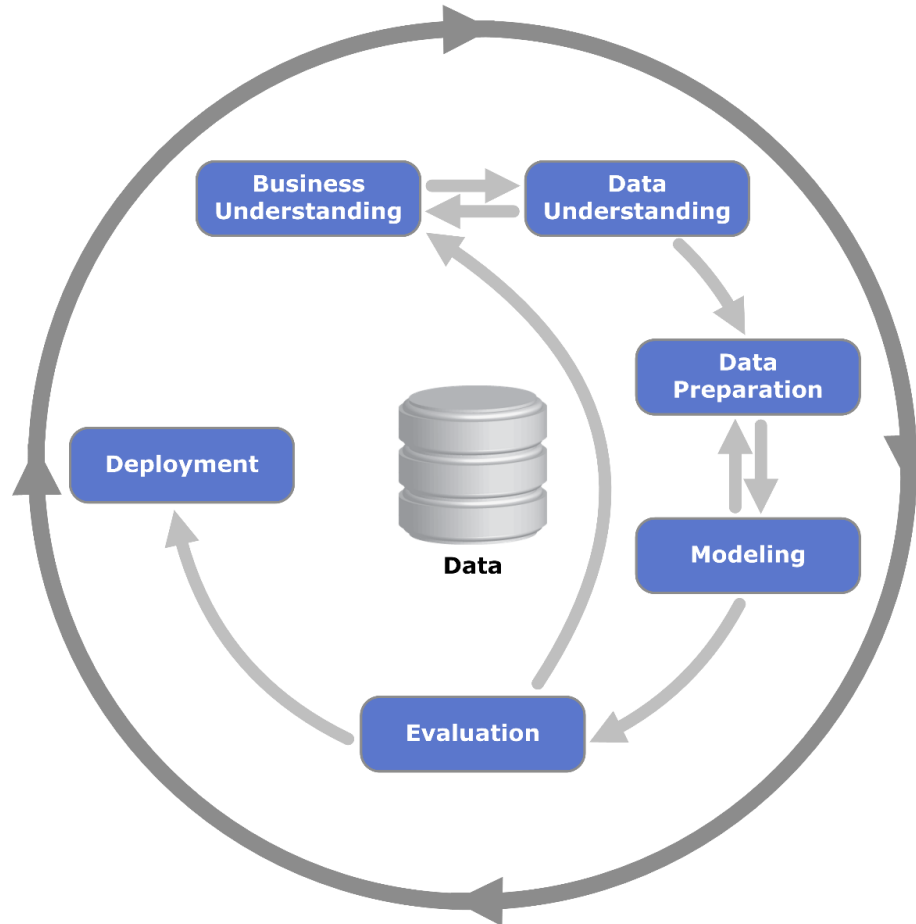
- Na prvý pohľad všetko sedí a dáva zmysel ... ale zdanie často klame!



## 1. Business Understanding

Pochopiť potreby a ciele zákazníka

**VUJE**: náhrada náročných meraní v teréne teoretickým výpočtom z verejne dostupných dát pomocou techník umelej inteligencie



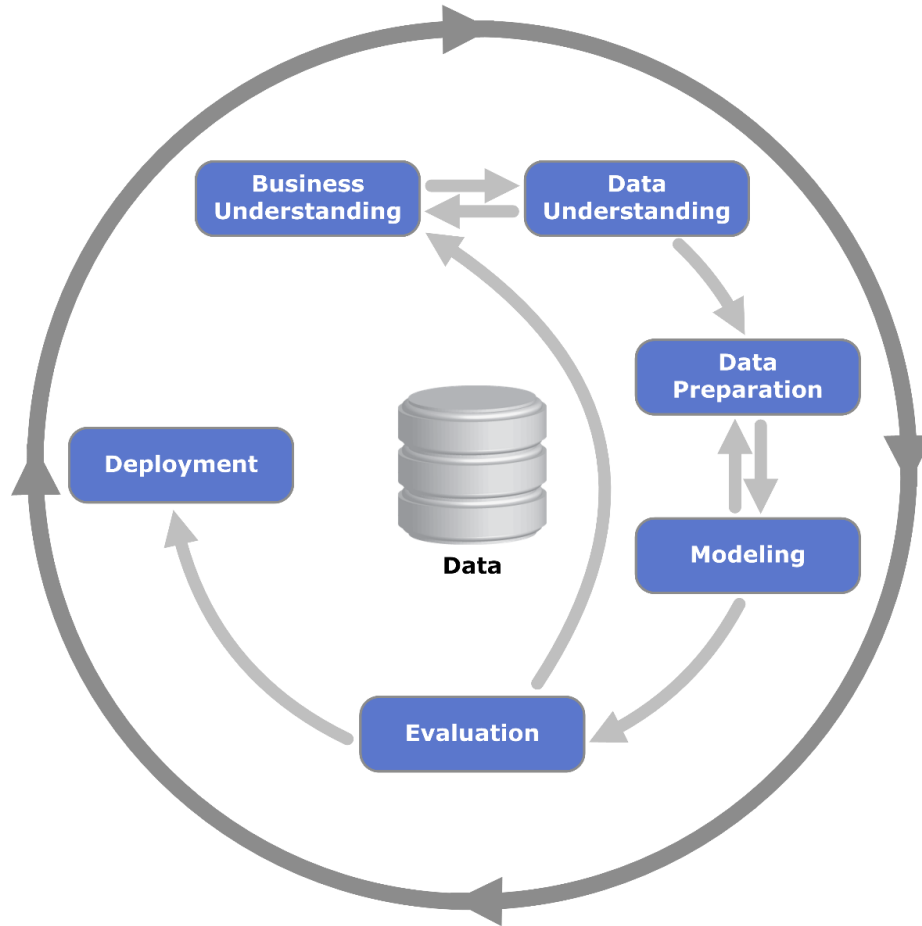
## 2. Data Understanding

prvotný zber dát a prvý odhad ich kvality, vnútornej štruktúry a možných latentných vzorov v nich prítomných (explorácia dát)

## 3. Data Preparation

Príprava finálneho datasetu pre modelovanie: integráciu dát, čistenie dát, tvorbu odvodených atribútov, transformáciu dát (normalizácia, štandardizácia), selekcia atribútov...

- Obrázok: By Kenneth Jensen - Own work based CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=24930610>



## 4. Modeling:

Výber a aplikácia viacerých prípustných modelovacích techník: klasifikácia, klastrovanie, regresia...

Voľba modelov a ich štruktúry; trénovanie

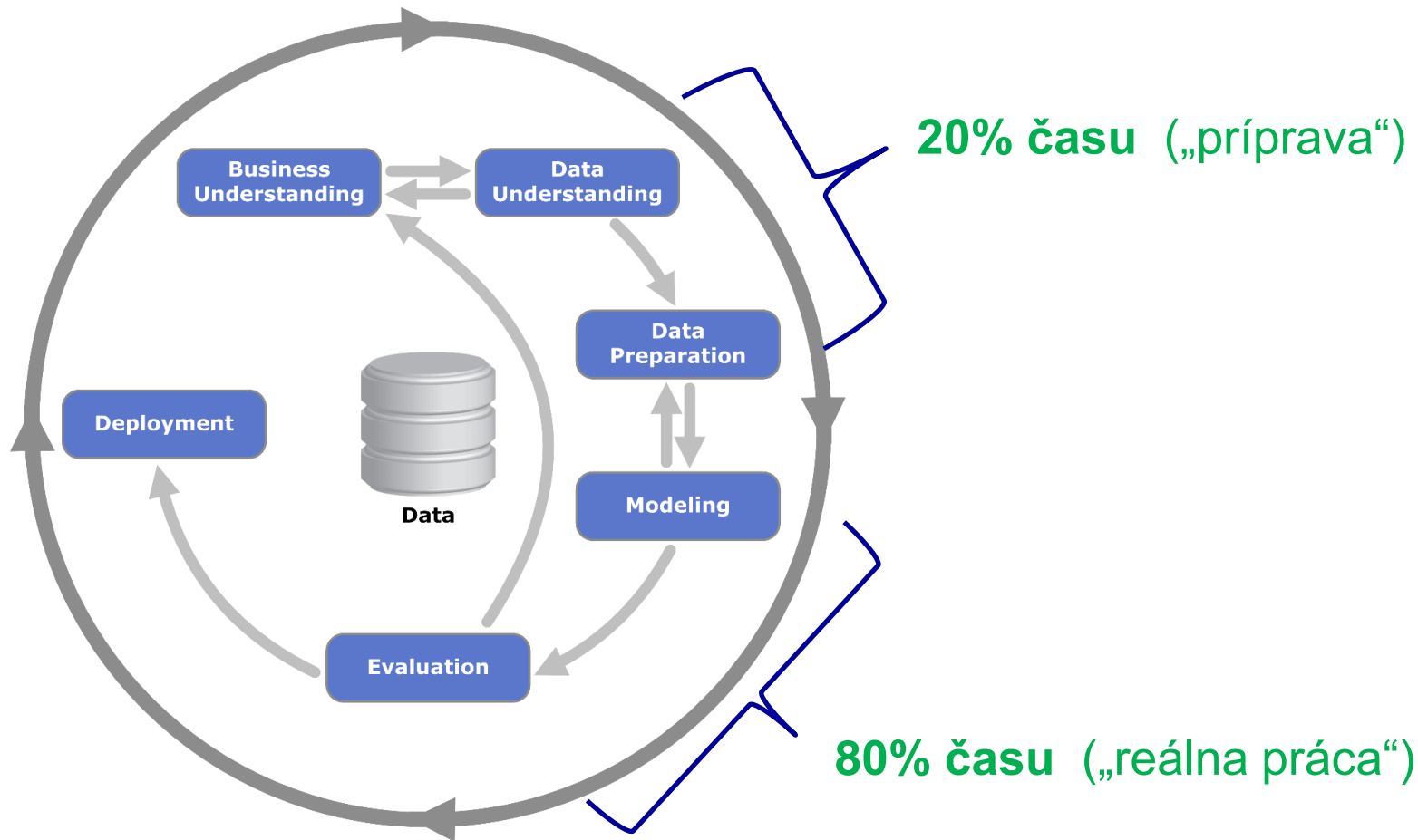
**Cieľ** = nájsť optimálne parametre pre každú použitú techniku a porovnať dosiahnuté výsledky.

## 5. Evaluation

Najlepšie modely sa ešte raz dôkladne prekontrolujú, či všetky metodické kroky pri ich tvorbe boli vykonané správne, či využívajú iba povolené typy vstupných dát a či naozaj dokážu uspokojiť potreby používateľa („*business requirements*“).

- Obrázok: By Kenneth Jensen - Own work based CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=24930610>

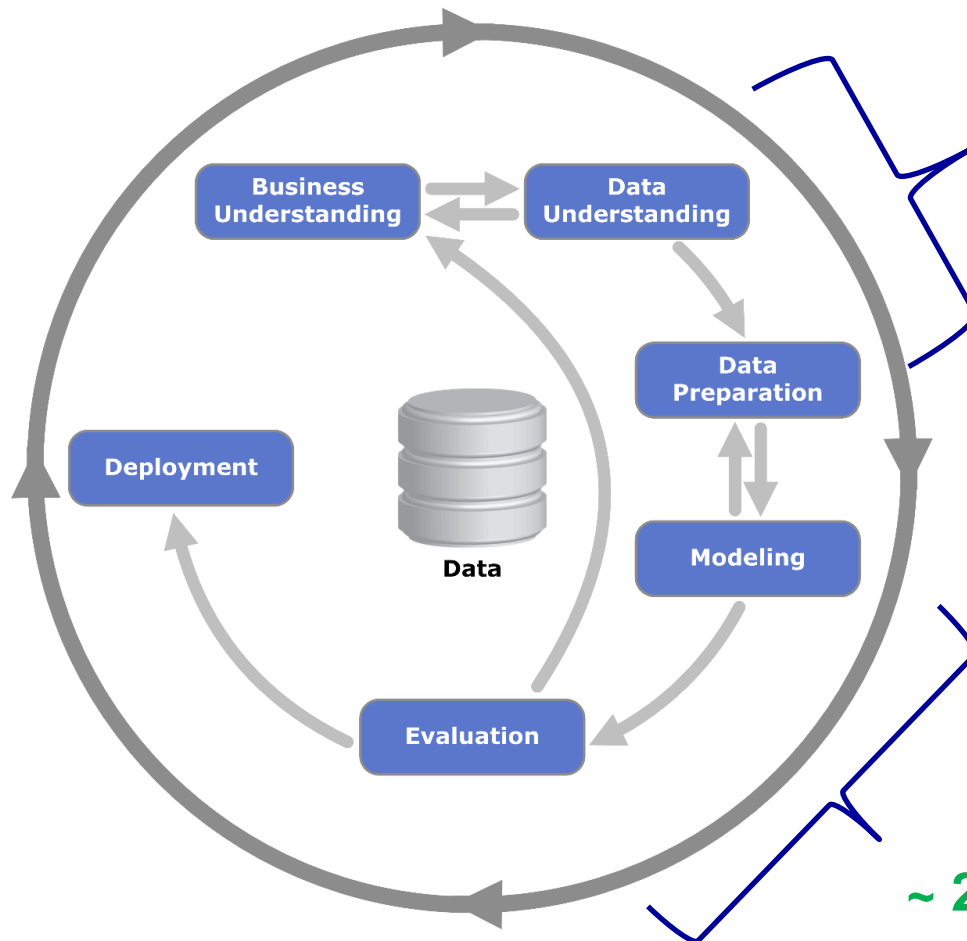
# CRISP-DM Proces (3) – časové očakávania:



- Obrázok: By Kenneth Jensen - Own work based CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=24930610>



# CRISP-DM Proces (3) – časová realita:



**~ 80% času a energie:**

## Časová os spolupráce:

- 4/2020 – prvé vzorky dát
- ... – oboznamovanie sa s dátami
- 1.10.2020 – začiatok zmluvnej spolupráce
- ... – explorácia dát, ich podrobné kontroly a opravy
- 8.3.2021 – dodanie dát v konečnej podobe
- 26.3.2021 – záverečná prezentácia
- 31.3.2021 – koniec zmluvnej spolupráce, odovzdanie záverečnej správy

**~ 20% času a energie**

- Obrázok: By Kenneth Jensen - Own work based CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=24930610>

# Úvodné problémy vo fáze porozumenia a prípravy dát:



- **Málo dát:** spočiatku ~210, ale nakoniec kvôli chybám iba 168 záznamov (každý záznam reprezentuje 5-11 šesťtyždňových terénnych meraní)
- Veľká **chybovosť dát** – viacnásobné **zdĺhavé kontroly a opravy:**
  - **Po každej oprave sme museli odznova integrovať všetky dáta a „postihnuté“ modely odznova pretrénovať...**
  - Konkrétne príklady závažných chýb v dátach na záver, ak bude čas a záujem...
- **Chýbali doménové znalosti** a konzultácie s expertami
  - **VUJE nedokázalo nájsť experta na šírenie znečistenia v atmosfére a jeho spad**
    - Ani univerzity, ani SHMÚ nevedeli pomôcť; **museli sme si pomôcť sami...**
- **Tlak zadávateľa**
  - na riešenie pomocou neurónových sietí;
  - na dvojročný projekt so zárukou úspešného riešenia.
- **Konečná dohoda: predbežná štúdia uskutočniteľnosti 10/2020 – 3/2021 a následný dlhodobejší (APVV) projekt.**

# Zadanie: Štúdia realizovateľnosti...



- ... vhodných prístupov pre vytvorenie modelu pre hodnotenie znečistenia prostredia v oblasti elektroenergetiky a pre určenie
  - celkového množstva spadu ( $S$ ),
  - množstva rozpustných látok ( $S_r$ ),
  - mernej elektrickej vodivosti 0,2% roztoku rozpustných látok ( $g_{0.2}$ )
- Uprednostniť prístupy na báze neurónových sietí (NS)
- Primárne vstupné údaje:
  - ročné priemerné koncentrácie  $PM_{10}$ ,  $PM_{2.5}$ ,  $NO_2$ ,  $SO_2$ ,  $O_3$
  - Výsledky meraní znečistenia prostredia ( $S$ ,  $S_r$ ,  $g_{0.2}$ , **stupeň znečistenia**)
- Štúdia má obsahovať
  - analyzované prístupy
  - popis vhodných prístupov na riešenie predmetu objednávky
  - návrh konceptuálneho modelu (resp. modelov) a tiež zhodnotenie ich výhod, nevýhod a rizík (s uvedením možností ich minimalizácie)



1. Analyzovať existujúce dáta a vyťažiť z nich čo najviac informácií
  - Regresia klasickým strojovým učením
  - Klasifikácia klasickým strojovým učením
  - Klasifikácia neurónovými sieťami
    - Pokúsiť sa rozmnožiť existujúce dáta simuláciou/zašumením, aby sa dali aplikovať vhodné neurónové siete
  
2. Teoreticky pochopiť doménu, aby sme vedeli zhodnotiť predbežné výsledky a navrhnúť ďalšie zmysluplné kroky
  - Štúdium vysokoškolských učebníc o atmosférickom znečistení
  - Rýchly prehľad odbornej literatúry: užitočné grafy a obrázky



# Predbežné vyhodnotenie klasických klasifikačných modelov

# Klasifikácia stupňa znečistenia klasickými metódami strojového učenia:



Veľmi málo vzoriek vo vyšších stupňoch znečistenia II-IV =>  
=> nutnosť zlúčiť tieto vzorky do jednej triedy.

Aj napriek tomu boli dve výsledné triedy silne nevyvážené:

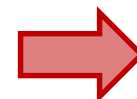
- Trieda „a“ = I. stupeň znečistenia (159 vzoriek)
- Trieda „B“ = II-IV. stupeň znečistenia (9 vzoriek)

Nedá sa teda použiť viac ako 9-násobná krížová validácia,  
ak má byť v každej testovacej podmnožine aspoň jedna vzorka triedy B

Použité klasické algoritmy:

- Naive Bayes
- Logistická regresia (Maximum Entropy)
- SMO SVM (Lineárny kernel)
- K-Nearest Neighbours (lazy/lbk vo Weke)
- Strom J48
- Random Forest

**Prezentované  
výsledky sú  
iba predbežné**





# Klasifikácia pre dve nevyvážené triedy

(trieda „a“ = I.st. = 159 ks, trieda B = II-IV.st. = 9 ks):



Väčšina klasifikátorov pri takto silne nevyvážených triedach ignoruje menšiu triedu.

1. Typická chybová matica (confusion matrix):

<u>a</u>	<u>b</u>	<u>&lt;-- classified as</u>
159	0	a = a
9	0	b = B

Celková presnosť (accuracy) =  
94.64% = 159 / 168

2. Ojedinelé výnimky, ako napr. **Naive Bayes:**

<u>a</u>	<u>b</u>	<u>&lt;-- classified as</u>
86	73	a = a
3	6	b = B

Celková presnosť (accuracy) =  
54.76% = 92 / 168

- Treba zvoliť vhodnú metriku kvality modelu:  
makro F-skóre (macro F-measure) (?)

# Klasifikácia pre dve vyvážené triedy

( $a = I = 159$  ks,  $B = II-IV = 9$  ks):



Po vyvážení tried sú klasifikátory nútené brať obe triedy rovnako vážne.  
(Za istú cenu: 1 vzorka triedy B má rovnakú váhu ako 17+ vzoriek triedy „a“)

Zatiaľ najlepšie modely:

## 1. SMO SVM

(confusion matrix):

a	b	<-- classified as
54.42	29.58	a = a
28	56	b = B

a      b   <-- classified as (reálne)

103	56	a = a
3	6	b = B

## 2. Naive Bayes

(confusion matrix):

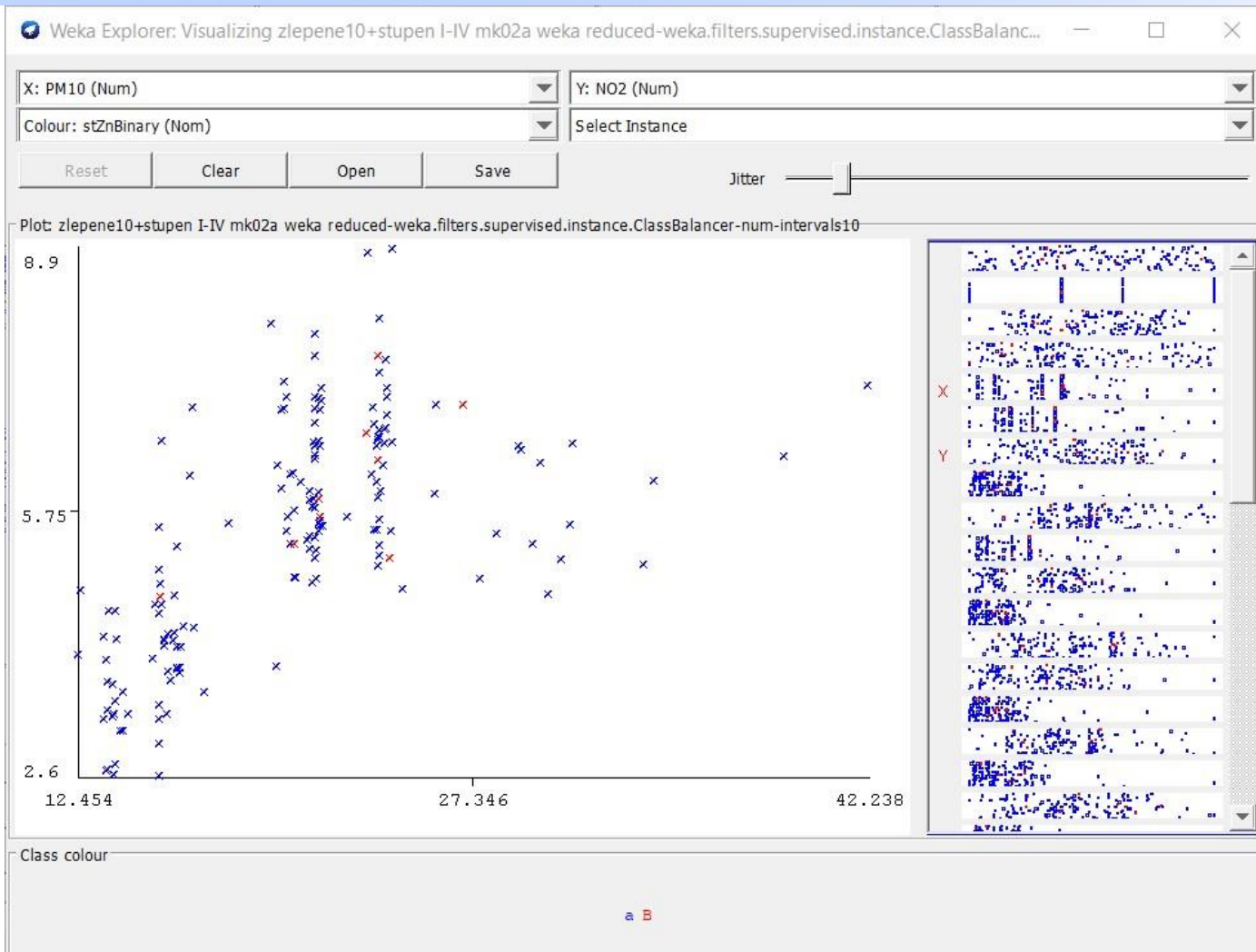
a	b	<-- classified as
44.38	39.62	a = a
18.67	65.33	b = B

a      b   <-- classified as (reálne)

84	75	a = a
2	7	b = B

- **Ktorý model je lepší? => Dôležitosť voľby vhodnej metriky**
- Výsledky by sa mohli zlepšiť po dodaní nových typov vstupných dát

# Hlavná príčina neuspokojivej presnosti klasifikátorov:



- Vzorky oboch tried sú silne premiešané a nedajú sa čisto oddeliť



# Pokus o konštrukciu nových atribútov pre klasické strojové učenie z webových dát SHMÚ

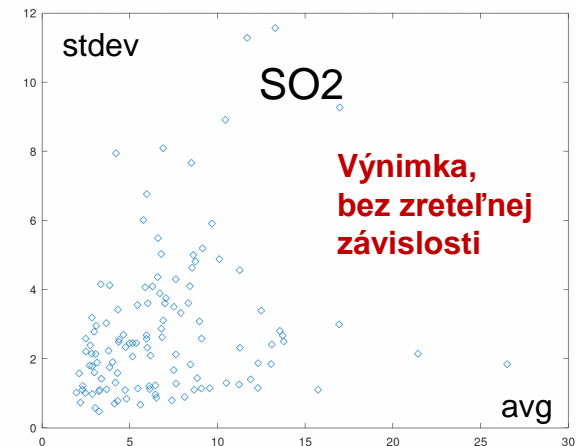
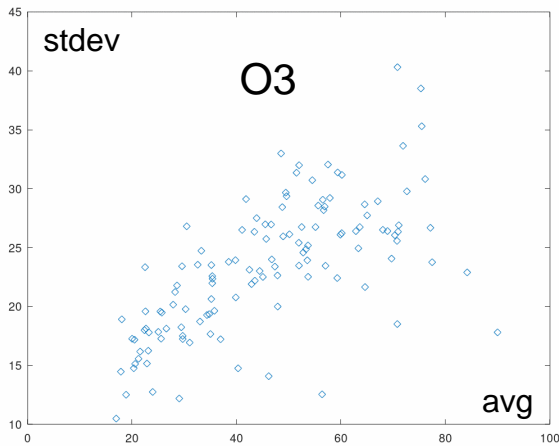
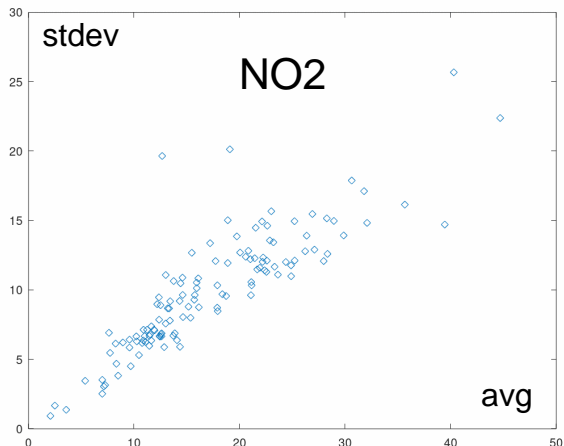
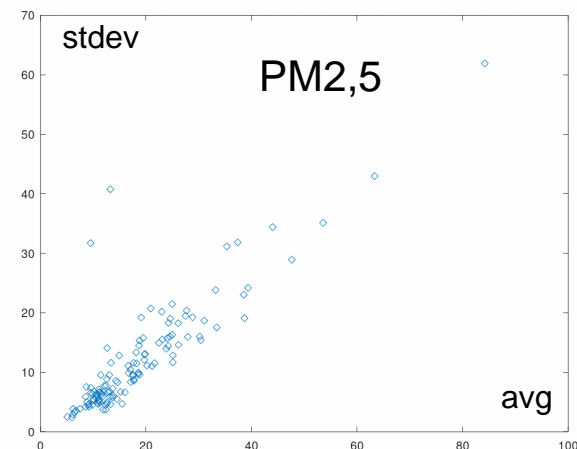
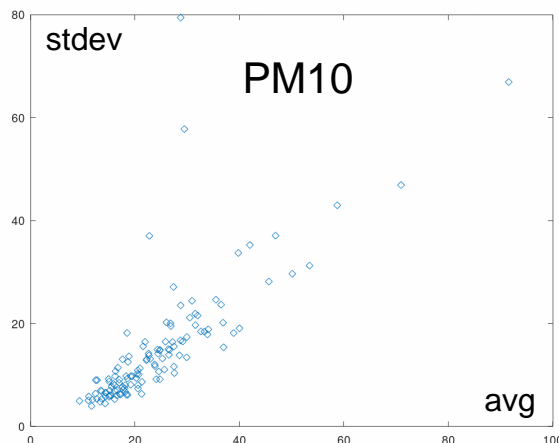
Problém: **Nízka časová granularita dát** – iba ročné priemery  
(na určenie horného odhadu veličín treba vziať do úvahy aj ich  
variabilitu v čase, nielen ich priemer);

- napr. množiny {4, 6} a {1, 9} majú obe priemer 5, ale rôzne maximá:  
6 verzus 9.

# Vychádza pomerne jasná lineárna závislosť medzi priemerom (avg) a stdev SHMÚ veličín:



Scatterplot diagramy mesačných priemerov a ich smerodajných odchýlok pre hlavné veličiny od SHMÚ



- Dobrá správa: okrem SO<sub>2</sub>, lineárna funkčná závislosť cez bod  $\sim (0, 0)$
- Zlá správa: modely ju už vzali do úvahy, teda nevylepší ich presnosť
- Do budúcnosti: v nových dátach môžeme odhaliť aj užitočnejšie závislosti



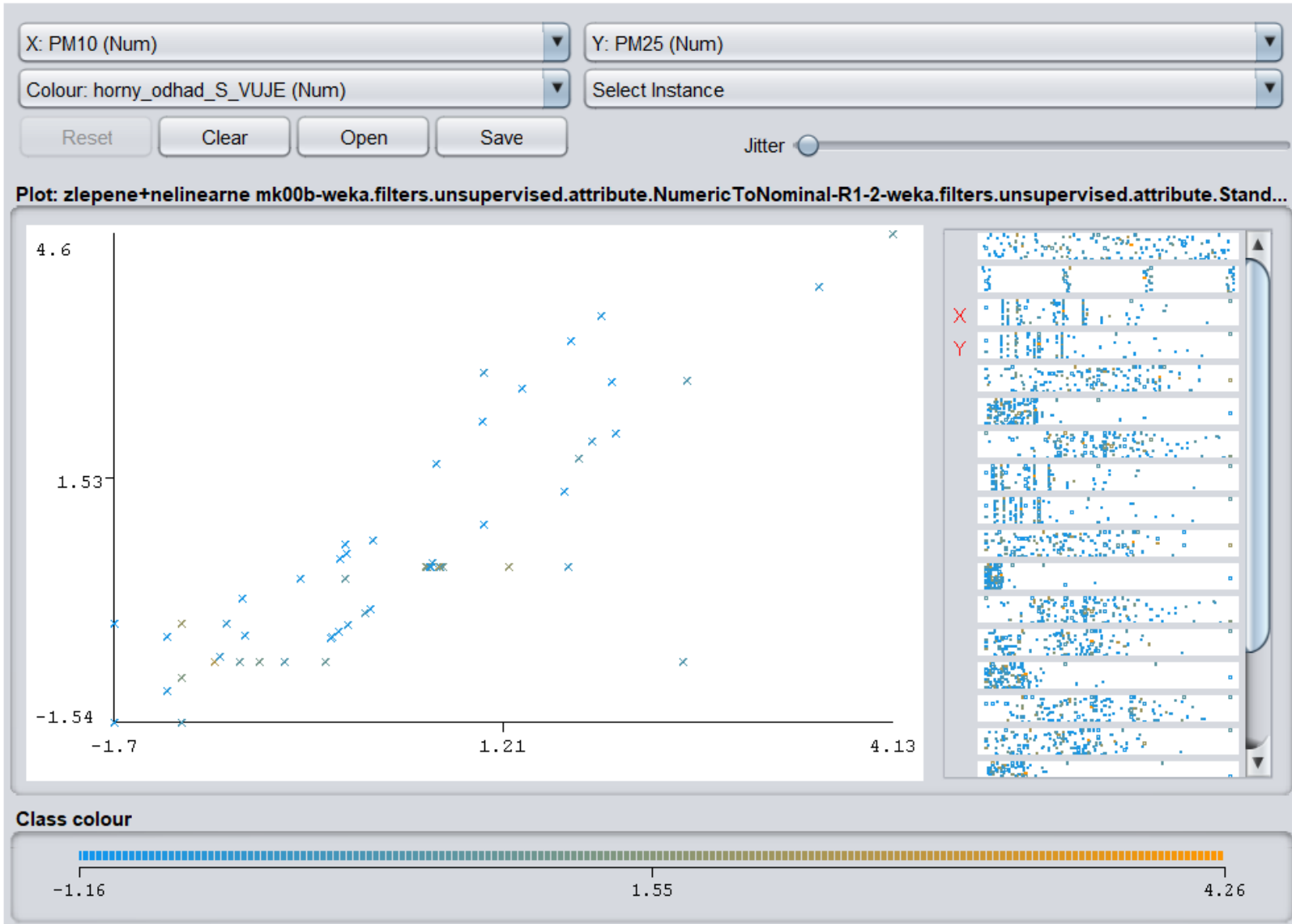
# Hlavné príčiny neuspokojivej presnosti modelov

1. Málo dát (meracích miest, t.j. riadkov databázy);
2. Nízka časová granularita dát;

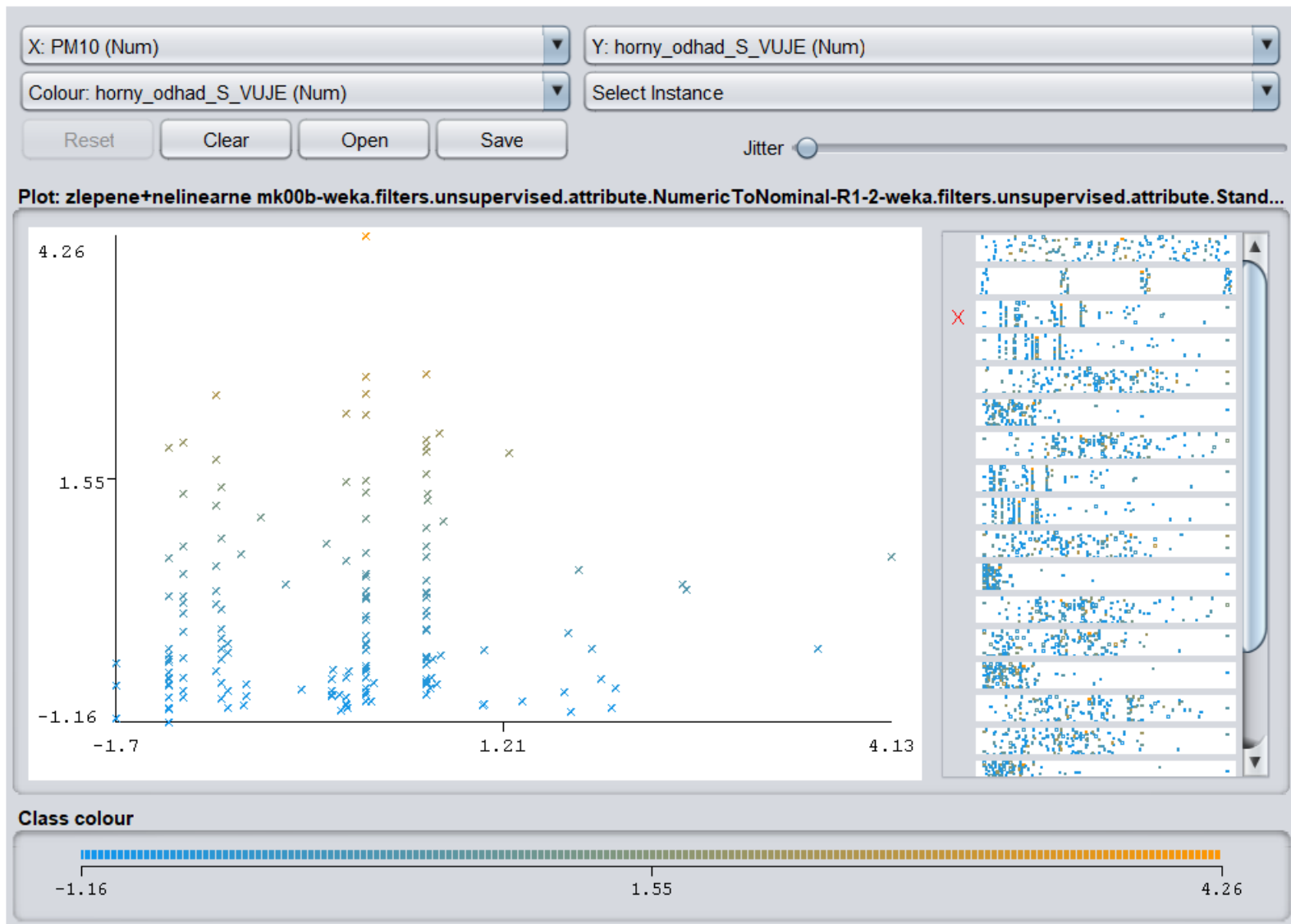
... a ďalšie problémy:



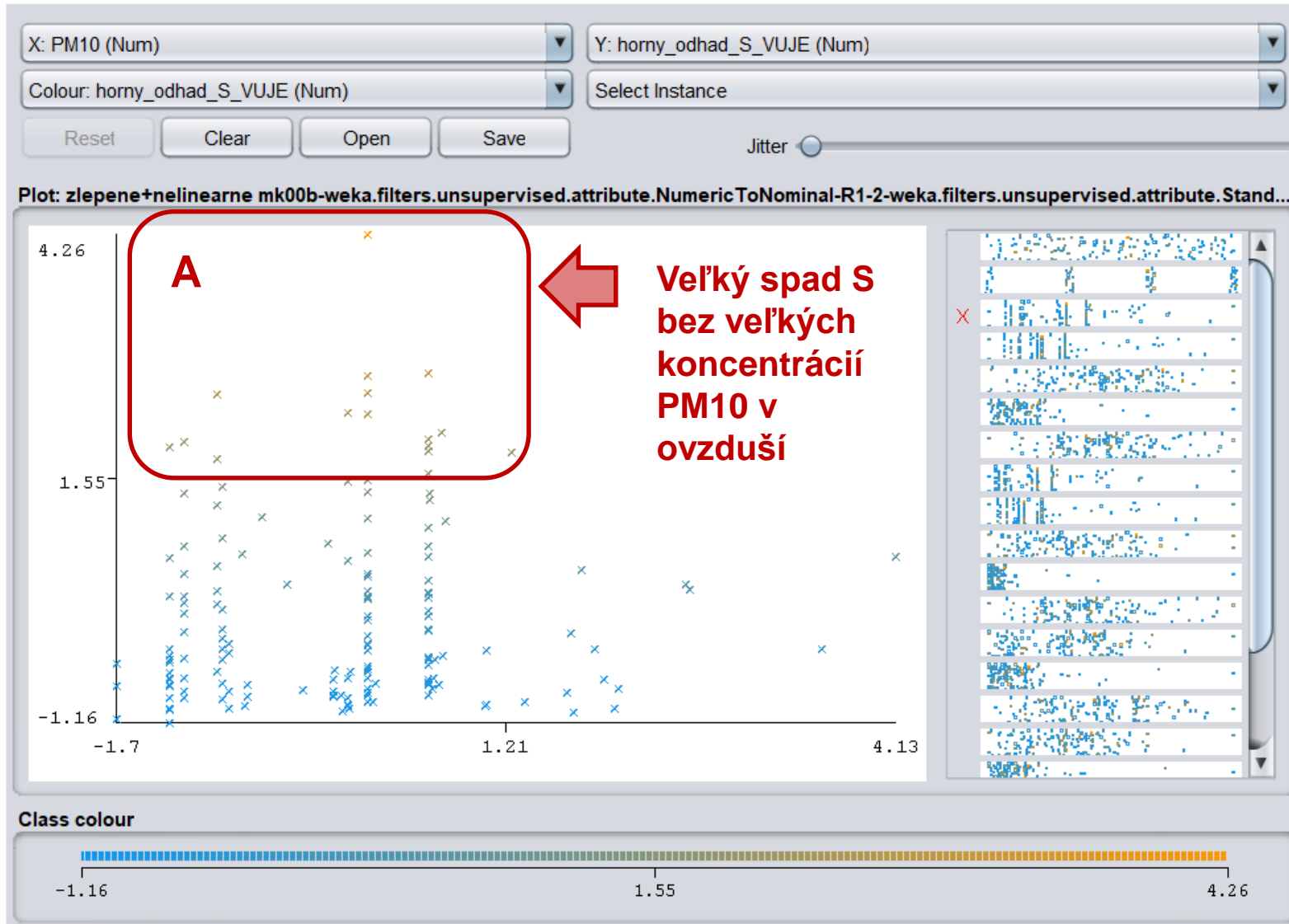
# Terajšie dáta SHMÚ: Pekný vzťah PM10 vs. PM2.5:



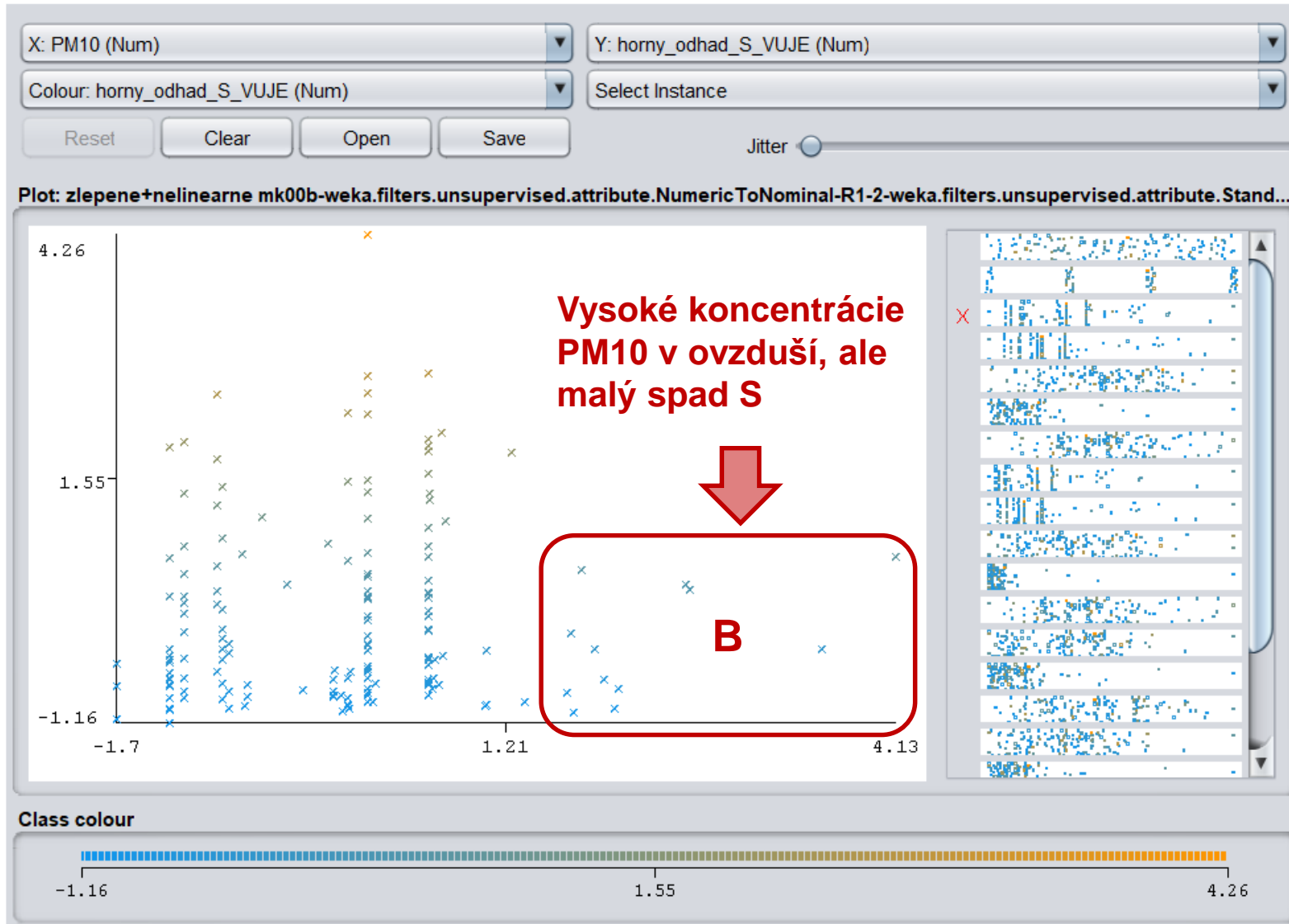
# Problém 3: chýba vzťah medzi PM10 a S\_VUJE:



# Problém 3: prejav 1

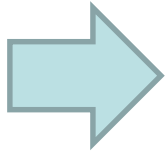


# Problém 3: prejav 2



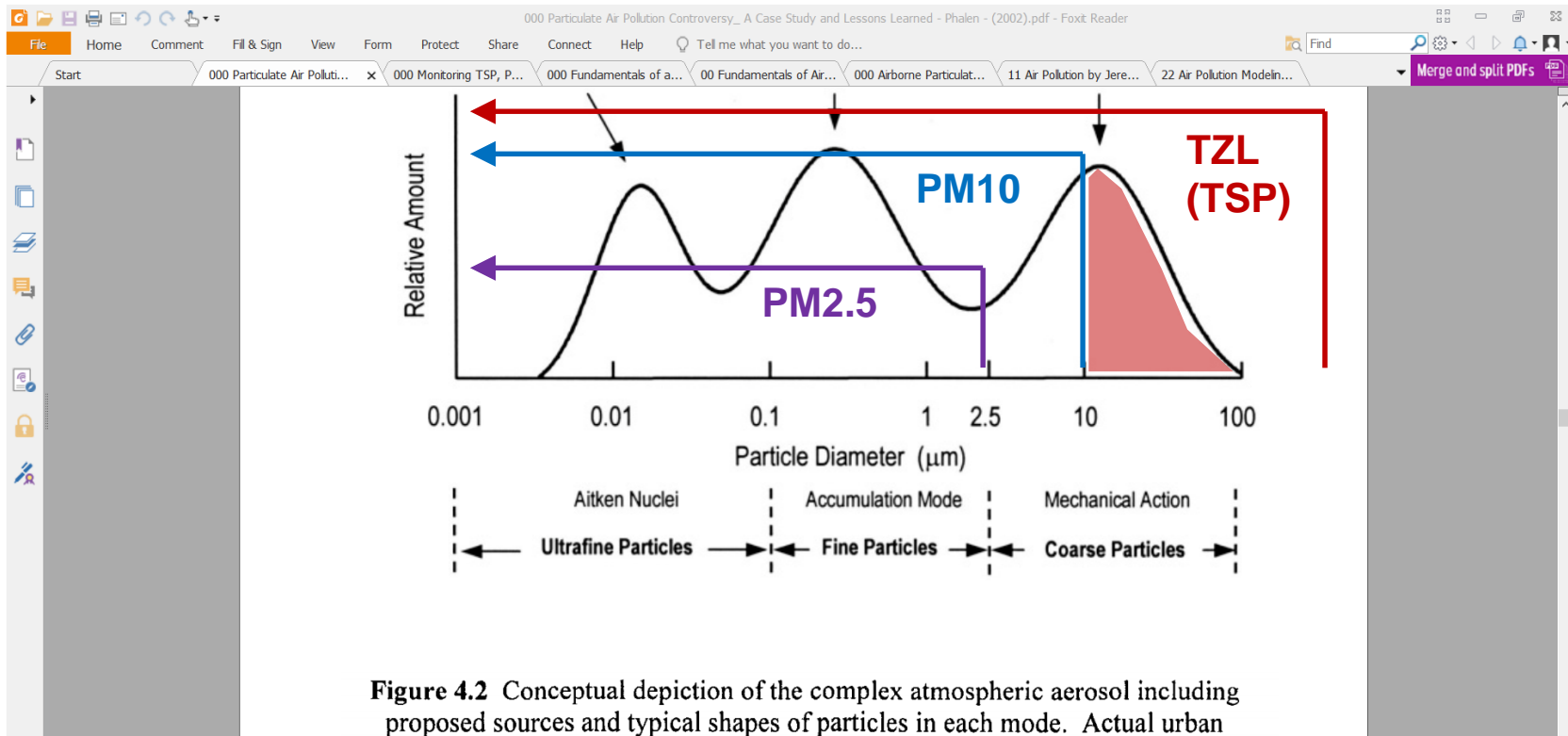


Chýbajúci funkčný vzťah medzi PM10 a horným odhadom množstva spadu (S) zisteným zo 6-týždňových meraní sa z logického hľadiska javí ako **zásadný problém**, lebo až keď vieme správne odhadnúť množstvo spadu (S), má zmysel skúmať, koľko z neho sa vo vode rozpustí ( $S_r$ ) a aká bude vodivosť výsledného roztoku ( $g_{0.2}$ ).



**To nás viedlo k štúdiu publikovanej literatúry v snahe nájsť možné vysvetlenie tejto anomálie.**

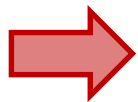
# Definícia PM2.5 a PM10 ( $PM_{2.5} \leq PM_{10}$ )



**Figure 4.2** Conceptual depiction of the complex atmospheric aerosol including proposed sources and typical shapes of particles in each mode. Actual urban

## „Histogram“ pre komplexný aerosol

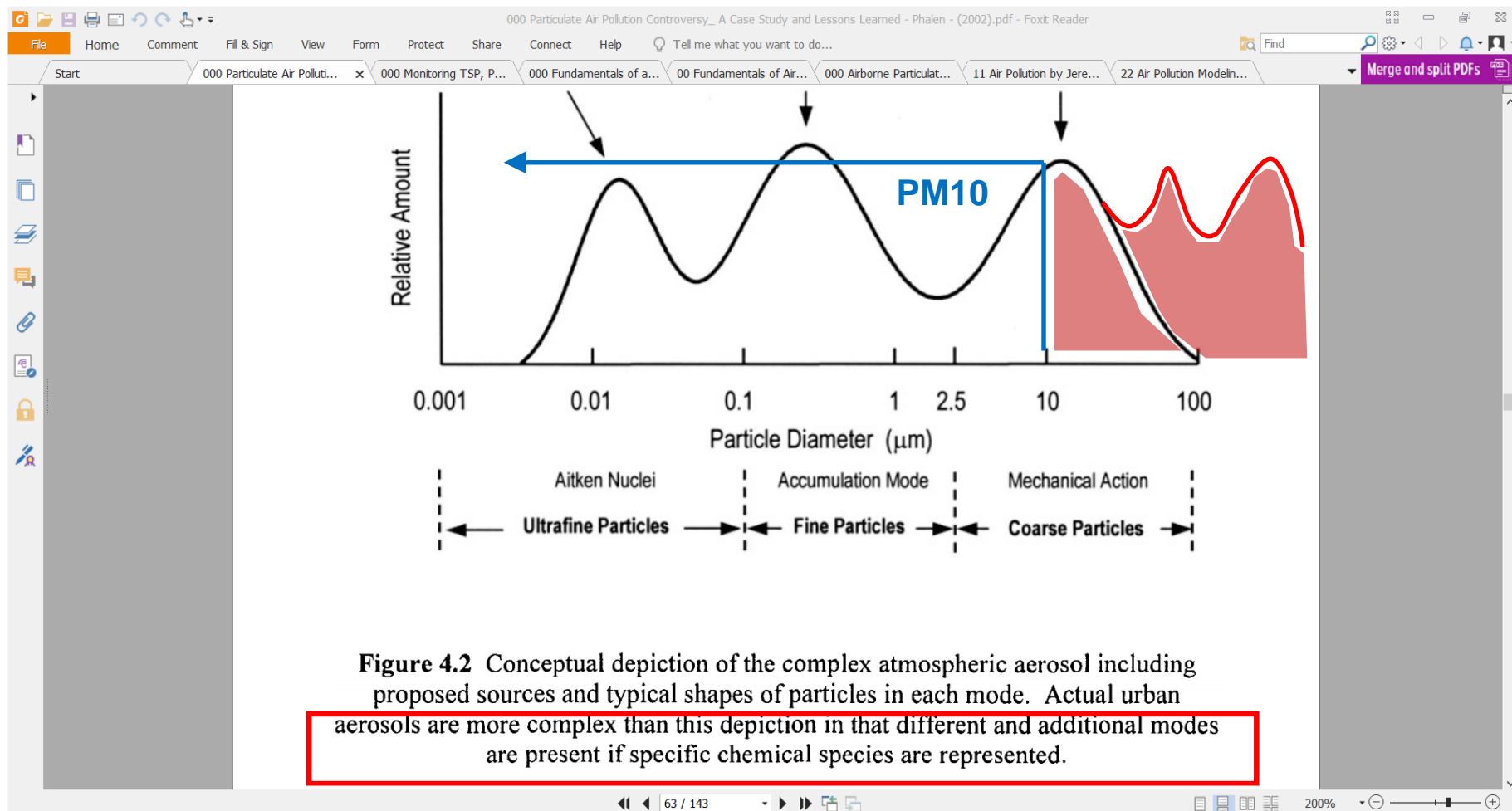
(tento aj nasledujúce obrázky sú prevzaté a adaptované z publikovanej literatúry)



- PM10 predstavuje iba časť prachových častíc vo vzduchu. Tie sa merajú, lebo sú nebezpečné pre ľudí (preniknú do pľúc.)
- Celok reprezentuje veličina TZL (Tuhé znečisťujúce látky), resp. TSP (Total Suspended Particulate). Ten sa dnes meria pomerne málo.



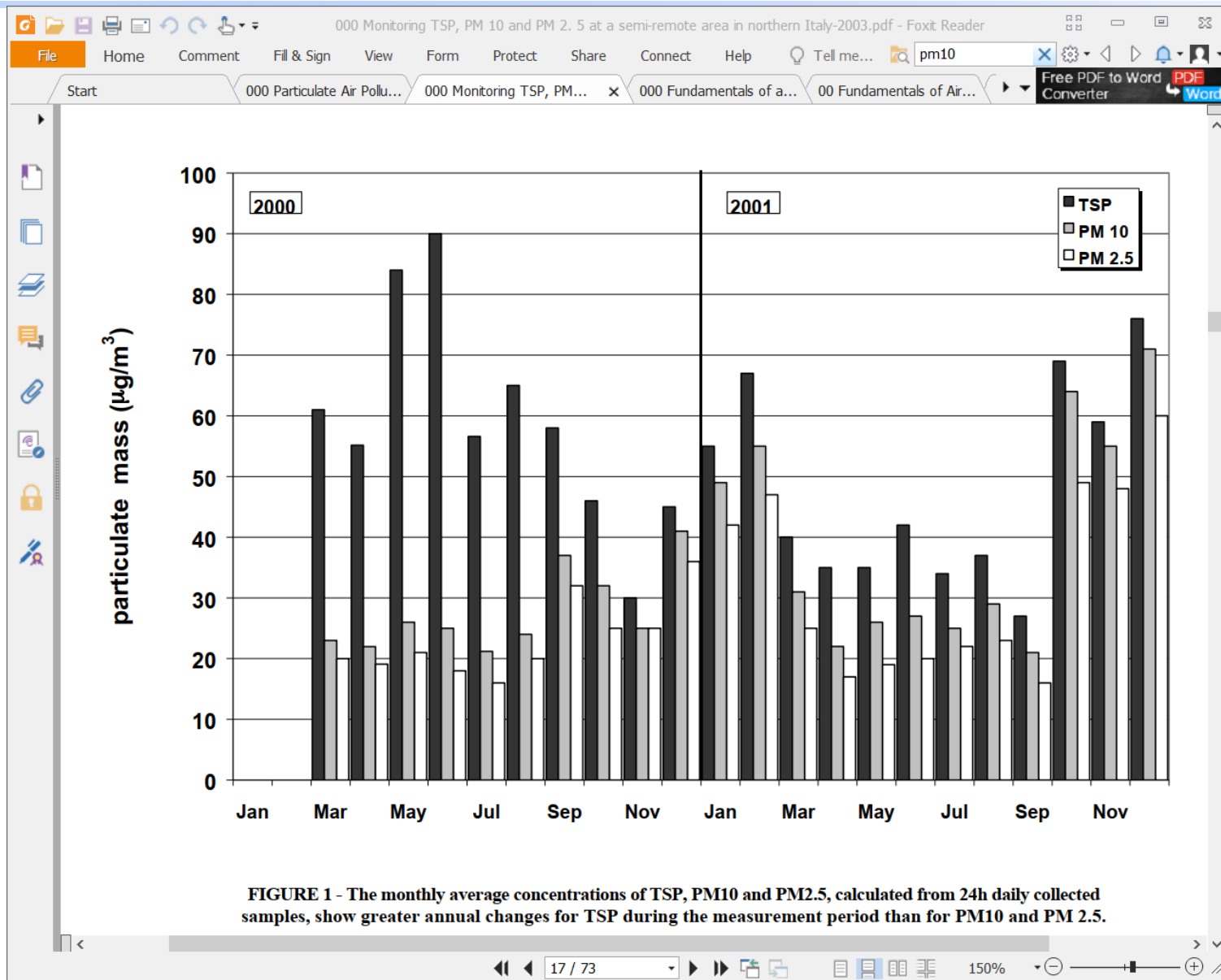
# Možné príčiny Problému 3 (1):



- Častice TZL/TSP môžu presahovať 100 mikrometrov (veľké sadze, organické častice a pod. v blízkosti lokálnych zdrojov bez odlučovačov)

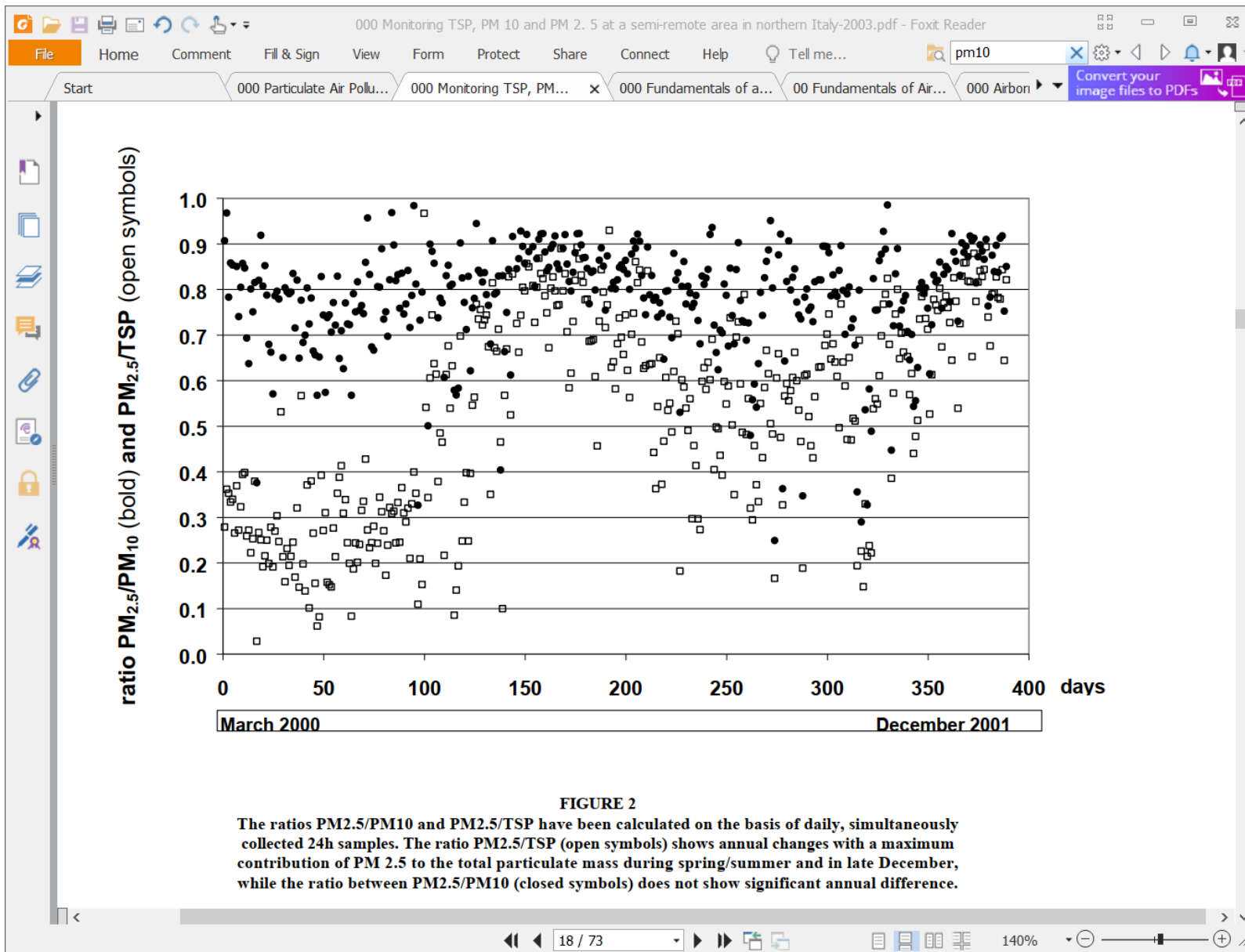
# Možné príčiny Problému 3 (2)

## Polo-odľahlá lokácia v severnom Taliansku – mesiace:



# Možné príčiny Problému 3 (3)

## Polo-odľahlá lokácia v severnom Taliansku - dni:



# 3. Možné vysvetlenie Problému 3 (Pôvodná predstava)

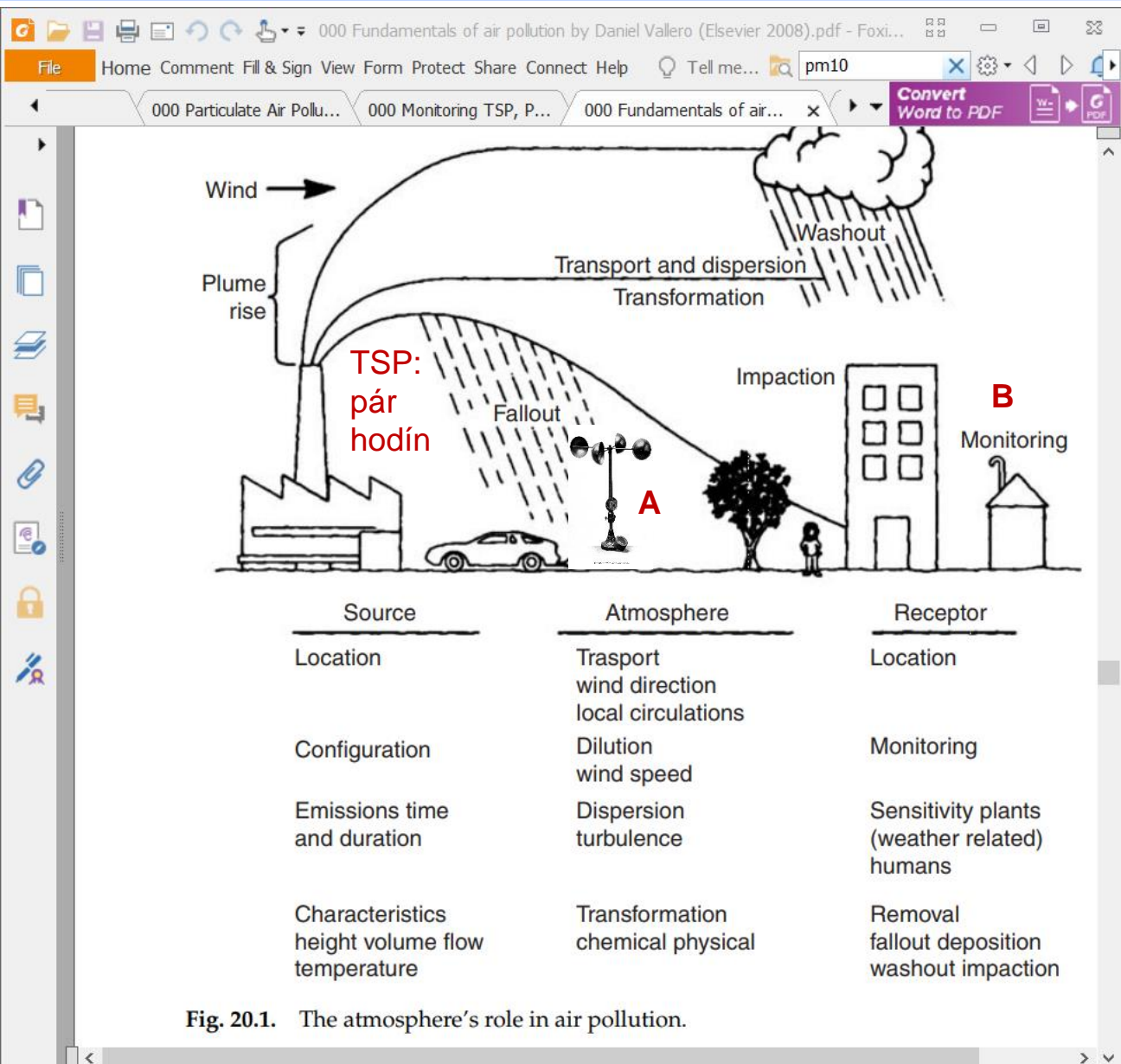


Fig. 20.1. The atmosphere's role in air pollution.

## Možné zdroje TZL (TZL > PM10)

**silne lokálne:**

1. Okolie komínov **bez odlučovačov**
2. Orba, žatva
3. Stavebné práce
4. Povrchová ťažba
5. Vietor z oráčiny
6. Vykurovanie drevom
7. Prach na cestách
8. Horiace skládky odpadu
9. ...

# Problém 3: Aktualizované hypotézy pre kvadrant A (po konzultácii so SHMÚ)



TZL-PM10

PM10

**Hypotéza 1:**  
Pre SHMÚ (NEIS)  
neznámy, lokálny zdroj  
tuhých znečisťujúcich  
látok (TZL)

**Hypotéza 2:**  
Gravitácia takmer  
nepôsobí na PM10:

PM10 mimo dažďa  
a rosy zväčša  
nezachytené

**Dôsledok:** Ak nie je lokalita v dosahu  
pre SHMÚ známeho silného zdroja PM10:

- Nesprávny nízky odhad PM10 v danej lokalite
- Veľký nameraný spad  $S \approx \text{TZL-PM10}$

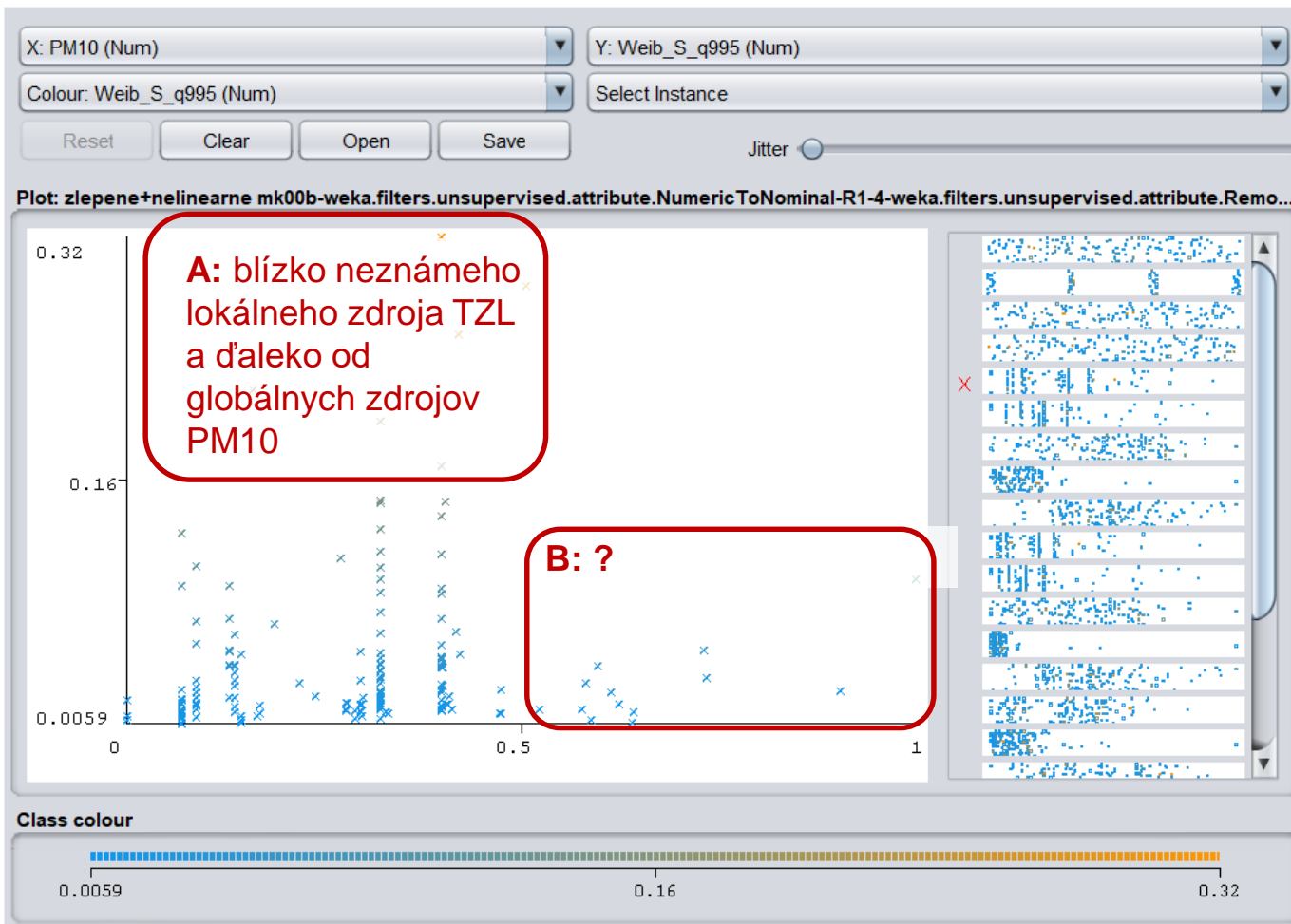
- => Kvadrant A v grafe  
(malé PM10, veľký spad S)



**Hypotéza 3:**  
spad  $S \approx \text{TZL-PM10}$



# Problém 3: Zhrnutie pre kvadrant A (aktualizovaná verzia)



# Problém 3: Aktualizované hypotézy pre kvadrant B



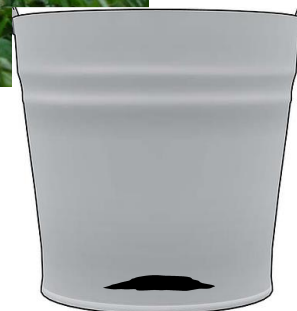
## Hypotéza 1:

Lokalita v dosahu SHMÚ (NEIS) známeho silného zdroja PM10, ale bez lokálneho zdroja TZL (napr. Košice-Šaca):

## Hypotéza 2:

Gravitácia takmer nepôsobí na PM10:

PM10 mimo dažďa a rosy zväčša nezachytené

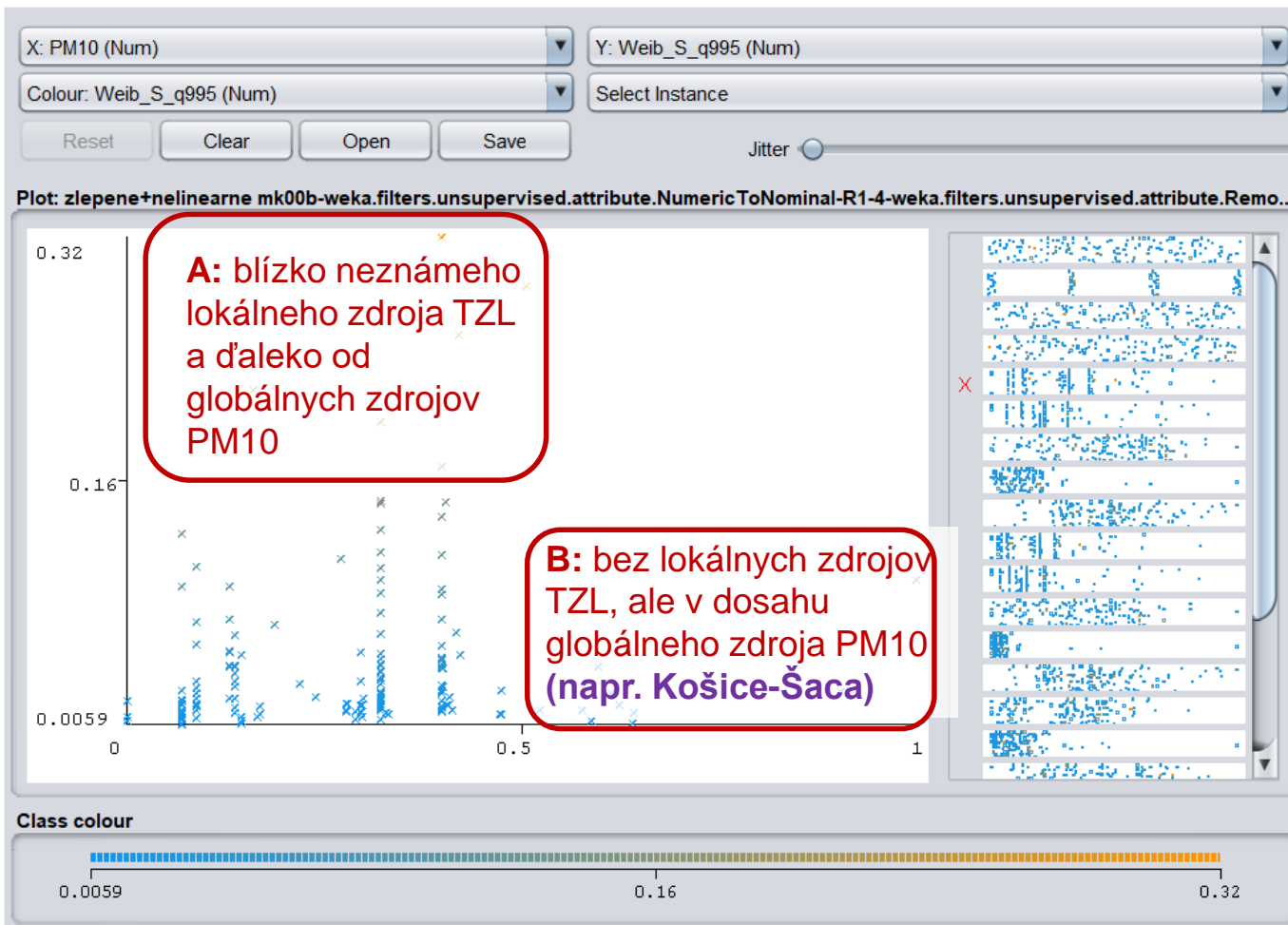


## Dôsledok:

=> Kvadrant B v grafe (veľké PM10, malý spad S)

Hypotéza 3:  
spad S  $\approx$  TZL-PM10

# Problém 3: Zhrnutie pre kvadrant B (aktualizovaná verzia)



**Dôsledok: PM10 a PM2.5 nie sú vhodné prediktory spadu S**



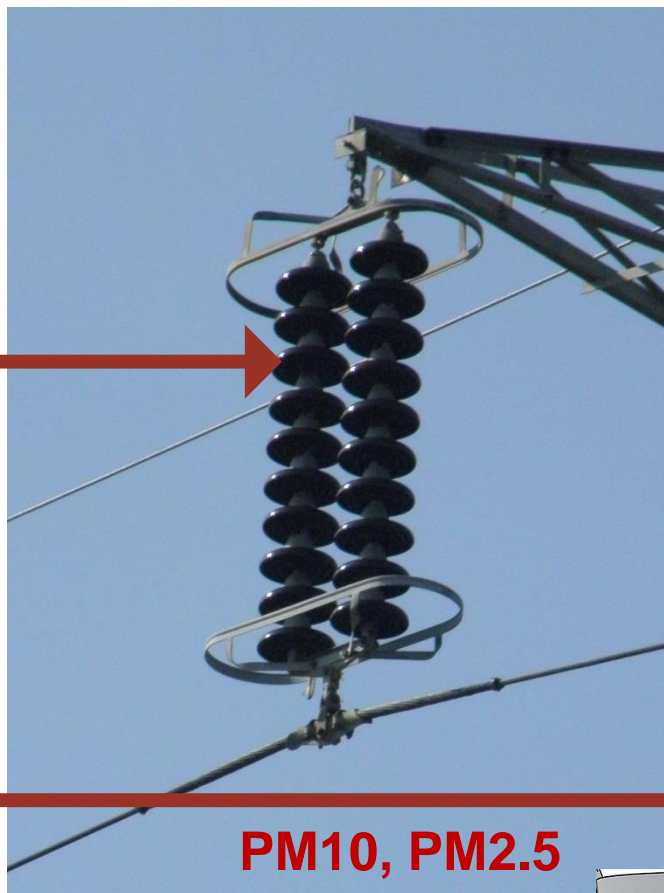
# Problém 3: Overenie Chí-kvadrát testom na vzorkách nameraných VUJE 2008-13:



REAL	PM10 low	PM10 high	<i>sum row</i>
S high	15	0	15
S low	139	14	153
<i>sum col</i>	154	14	168
EXPECTED	PM10 low	PM10 high	
S high	13.75	1.25	
S low	140.25	12.75	
	<b>0.221083</b>	(Chi-squared p-value)	
	<b>potvrdenie nezávislosti PM10 a S</b>		

- Chí-kvadrát test nepriamo potvrdil aj to, že nameraný spad S nezávisí od PM10, čiže PM10 nie je vhodným ukazovateľom pre odhad spadu S
- Nezávislosť spadu S od PM10 potvrdili aj veľmi nízke korelačné koeficienty: Pearsonov (0.03656) a Spearmanov (0.06487, p-value = 0.4034)

# Problém 3: PM10 a PM2.5 môžu stále pomôcť (ide o nedostatok metódy zberných nádob)



PM10, PM2.5

PM10, PM2.5

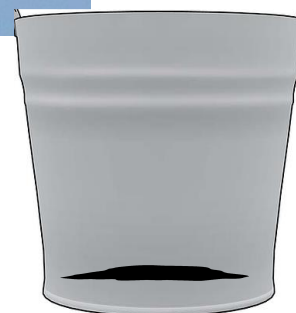
Záchyt PM10 a PM2.5 priamo na izolátore nárazom a prilepením

=> Hypotéza 4: Stery z izolátorov by mali mať väčšiu vodivosť ako spad zo zberných nádob (pri suchom počasí)

PM10 mimo dažďa a rosy zväčša nezachytené zbernou nádobou

Podľa literatúry, PM10 a PM2.5 môžu ovplyvňovať vodivosť:

=> Potreba meraní zvodového prúdu priamo na izolátoroch

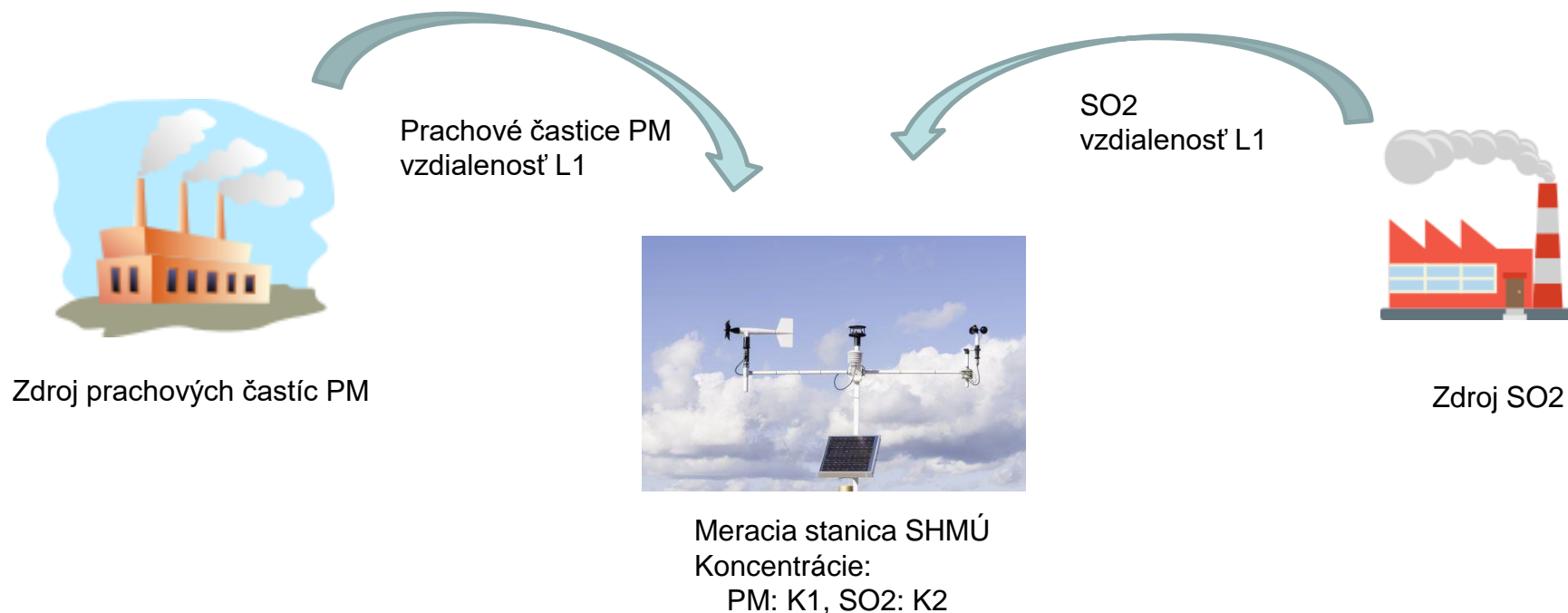


=> Hypotéza 5: g0.2 (S) porastie s množstvom zrážok (Košice-Šaca)

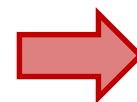
# Problém 4: Nepoznáme chemické zloženie TZL (vplyv na podiel rozpustnej zložky Sr a vodivosť g0.2):



- a) Nepoznáme chemické zloženie častíc TZL, PM10 a PM2.5  
(uhlík s plynmi reaguje inak ako zemné minerály, resp. kovy)



- b) Nepoznáme reakčné podmienky, čas, teplotu, ani koncentráciu



# Problém 4b: Nerozoznateľná alternatíva s oveľa väčšou rozpustnosťou Sr a vodivosťou g0.2

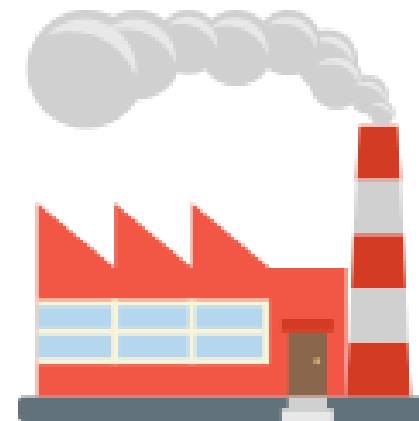


Meracia stanica SHMÚ  
**Nezmenené koncentrácie:**  
**PM: K1, SO2: K2**



SO2 + PM  
vzdialenosť L1  
**max koncentrácia**  
**aj čas a teplota**  
**na reakciu**

Zdroj SO2 + PM



Systematické riešenie pre problémy 3 a 4:

**Získať nové typy dát od SHMÚ alebo z iných zdrojov**

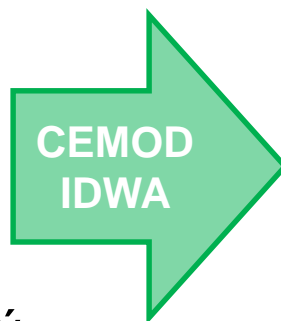




## Zdroje vstupných dát SHMÚ

### 1. Meracie stanice znečistenia

- a) *PM10, PM2.5, TZL (?)*
- b) *Plyny SO2, NO2, O3*
- c) *Ťažké kovy*
- d) *...*



### 2. **Databáza NEIS**

(Národný emisný informačný systém): každý znečisťovateľ kvantifikuje

- a) *Technologické vstupy (palivá, suroviny...)*
- b) *Výsledné produkty (znečisťujúce látky)*

### 3. *Meteo údaje (prevládajúci smer a sila vetra, zrážky, vlhkosť...)*

Modely a mapy časopriestorového rozloženia znečisťujúcich látok na území SR

**IDWA simulácie na zakázku (?)**:

- **TZL,**
- **Ľahké kovy (Ca...),**
- **Cestná soľ/posyp**



# Celkové zhodnotenie a návrh ďalšieho postupu

1. Úspech výskumu závisí od získania nových dát, či už od SHMÚ (databázy znečisťovateľov, pestovaných plodín, mokrá depozícia nečistôt,...) alebo z ďalších verejne dostupných zdrojov (satelitné snímky...). Veríme, že ak sa ich podarí získať, presnosť modelov výrazne vzrastie.
2. Výskum pokračuje v rámci národného projektu APVV-20-0548 ARIEN, ktorý začal 1.7.2021.



**Ďakujem za pozornosť**