

ČESKÝ  
HYDROMETEOROLOGICKÝ  
ÚSTAV

# STATISTICKÁ PŘEDPOVĚĎ ZNEČIŠTĚNÍ OVZDUŠÍ, NÁVRH A TESTY METODY

Josef Keder

*ČHMÚ ÚOČO, Observatoř Tušimice*

[keder@chmi.cz](mailto:keder@chmi.cz)



# Proč statistická předpověď – motivace (1)

Možnost předpovědět úroveň znečištění ovzduší na určité lokalitě s předstihem alespoň jednoho dne je předmětem zájmu orgánů, zodpovědných za ochranu kvality ovzduší a zdraví, i široké veřejnosti

Prognóza počasí pomocí numerických předpovědních modelů je provozována v operativním režimu již řadu let a stále se zlepšuje díky inovacím v měřicích metodách, fyzikálně-numerických postupech a používané výpočetní technice

Imisní koncentrace znečišťujících látek jsou z rozhodující části determinovány stavem a vývojem atmosférických podmínek

Postupy předpovědi znečištění ovzduší proto logicky navazují na meteorologické modely a používají jejich prognostické produkty jako vstupní údaje

# Proč statistická předpověď – motivace (2)

Zkušenosti, získané s deterministickým přístupem, t.j. předpověď typu  
**„zítra bude v místě X dosažena denní průměrná koncentrace  $C_x$ “**  
ukázaly, že kvalita takto pojaté předpovědi je často značně neuspokojivá

Velkou měrou se na tom podílí chaotický, stochastický charakter systému  
atmosféra-znečišťující látka

Podobně jako v případě meteorologických předpovědí se proto pozornost upírá  
i na vývoj a využití metod statistických předpovědí

Poskytují výsledek ve tvaru

**„zítra bude v místě X překročena denní průměrná koncentrace  $C_x$   
s pravděpodobností  $P(C_x)$ “**

Možná je rovněž formulace typu

**„koncentrace, která bude překročena s pravděpodobností  $P$ , je rovna  $C_p$ “**

# Statistická předpověď

Jak může vzniknout statistická předpověď

Použije se vhodná statistická metoda (regresní vztahy, neuronové sítě, fuzzy postupy, GAMs..), stanoví se nejistoty výsledku

Ansámblová předpověď - opakovaný běh deterministického modelu s variací vstupních parametrů

Jiný přístup, demonstrován dále

# Návrh jednoduché metody

Podstata postupu – učení z historických dat, soubor koncentrací dostatečného rozsahu pro určitou lokalitu

- Soubor se rozdělí do podmnožin (kategorií) podle vhodných kritérií
- Musí být předpověditelné, která podmnožina se vyskytne
- Pro každou kategorii se stanoví statistické rozložení koncentrací, které umožňuje odhad pravděpodobnosti výskytu určité hodnoty koncentrace nebo překročení prahových hodnot
- Máme-li předpověď určité kategorie, můžeme predikovat pravděpodobnosti překročení

# Jak stanovit kategorie

Použijí se meteorologické veličiny, které významně ovlivňují imisní koncentrace: výška směšovací vrstvy, rychlost větru, teplota vzduchu...

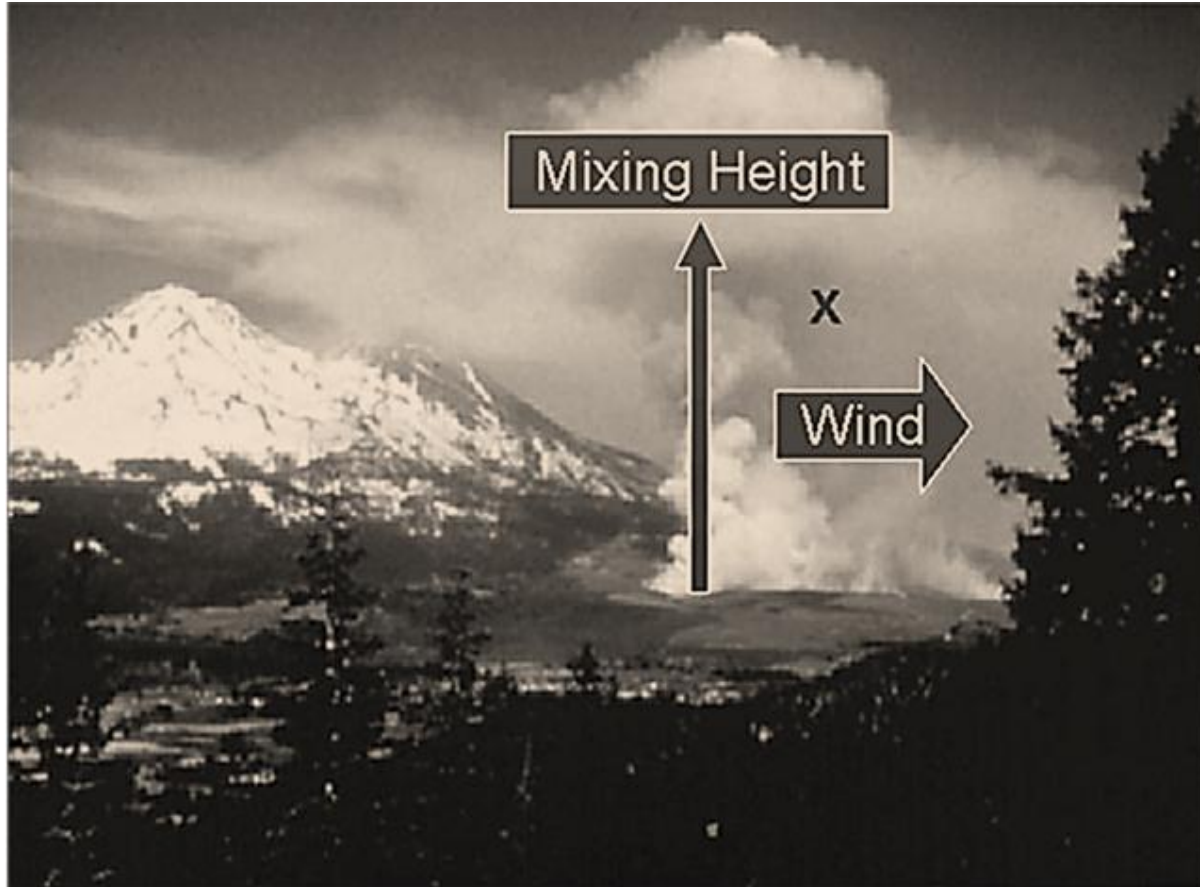
Osvědčené z jiných studií

- Ventilační index VI – zahrnuje výšku MH a rychlost větru
- Teplota vzduchu T

Jsou vcelku dobře predikované metrologickými modely v provozním režimu – ALADIN

V této fázi postup použít na denní průměry koncentrace  $PM_{10}$ , měřené 06 – 06 UTC (aby bylo možno využít i manuálních stanic)  
VI a T rovněž jako denní průměry 06 – 06 UTC

# K definici ventilačního indexu



Hardy, Peterson, Seamon, P., eds. 2001

Keder, Škáchová 2011

# Výběr hranic kategorií

Podle denních průměrů VI a T

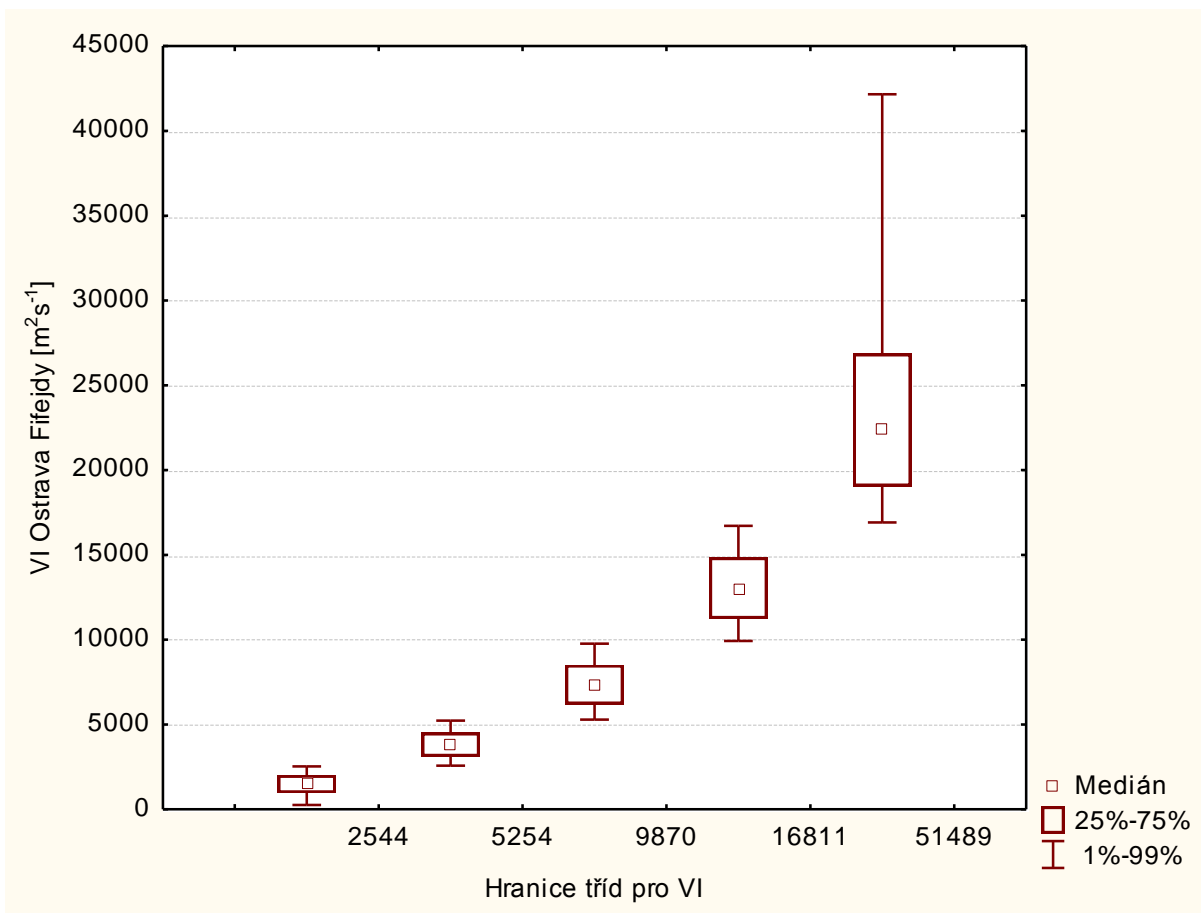
Hranice stanoveny jako 2., 4., 6., 8. a 10. decil souborů VI a T

V každé kategorii je stejný a dostatečný počet hodnot

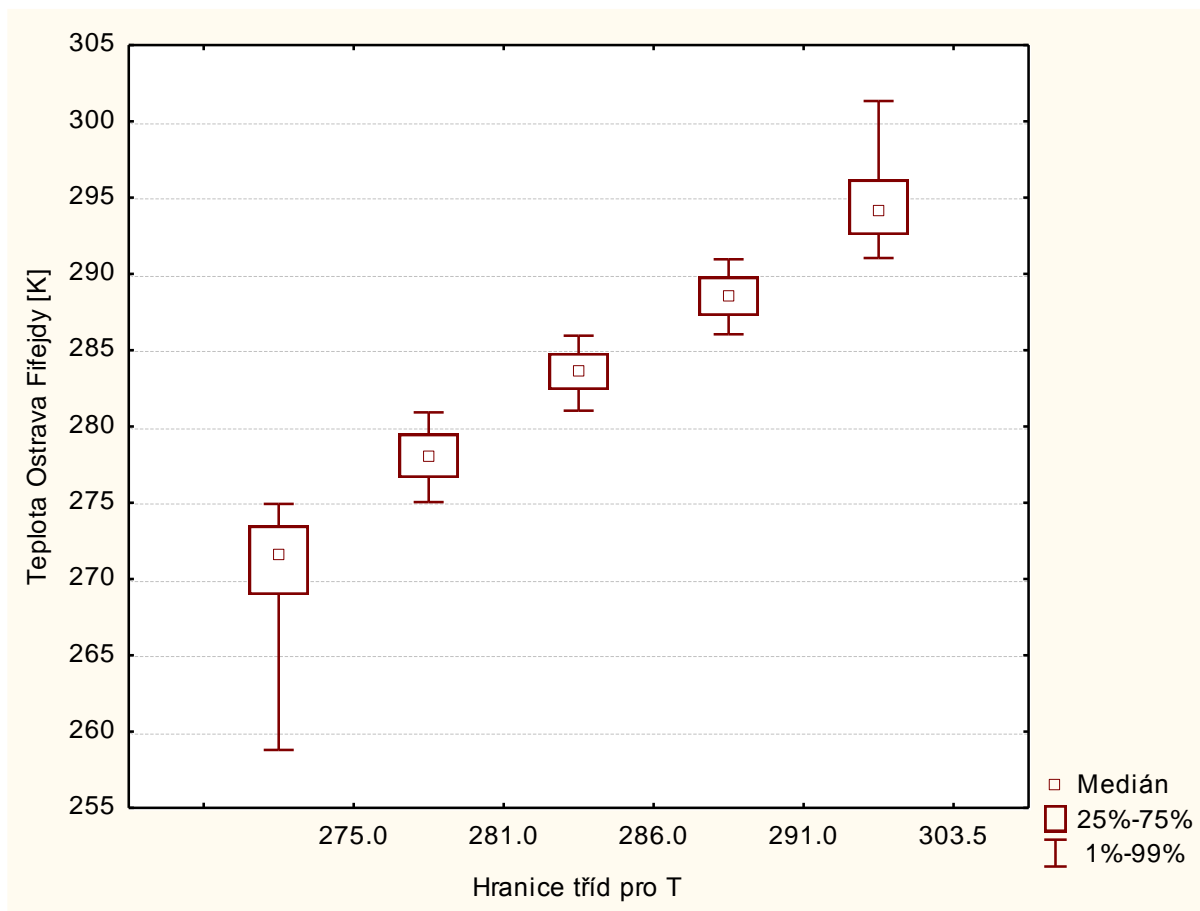
Kategorie se nepřekrývají



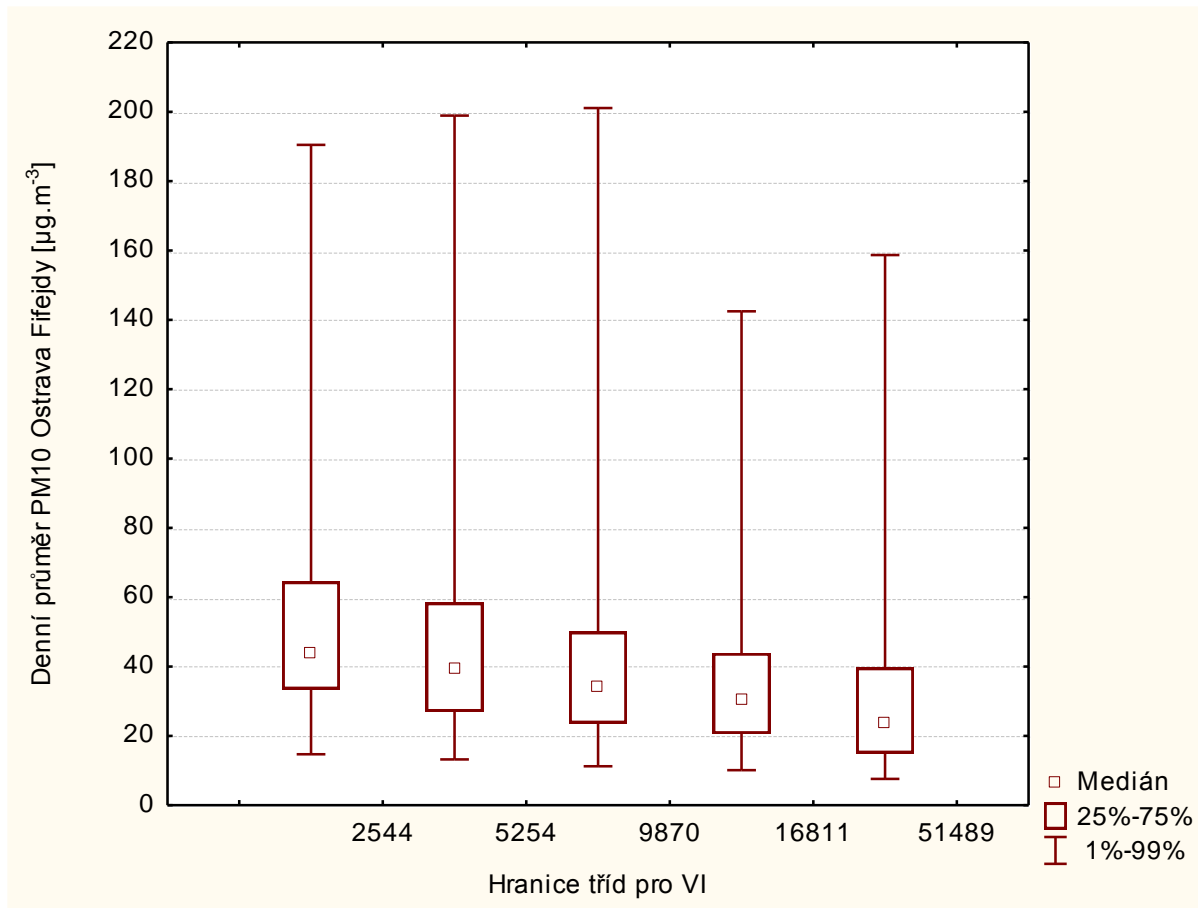
# Lokalita O. Fifejdy – rozložení VI pro hranice, stanovené jako sudé decily



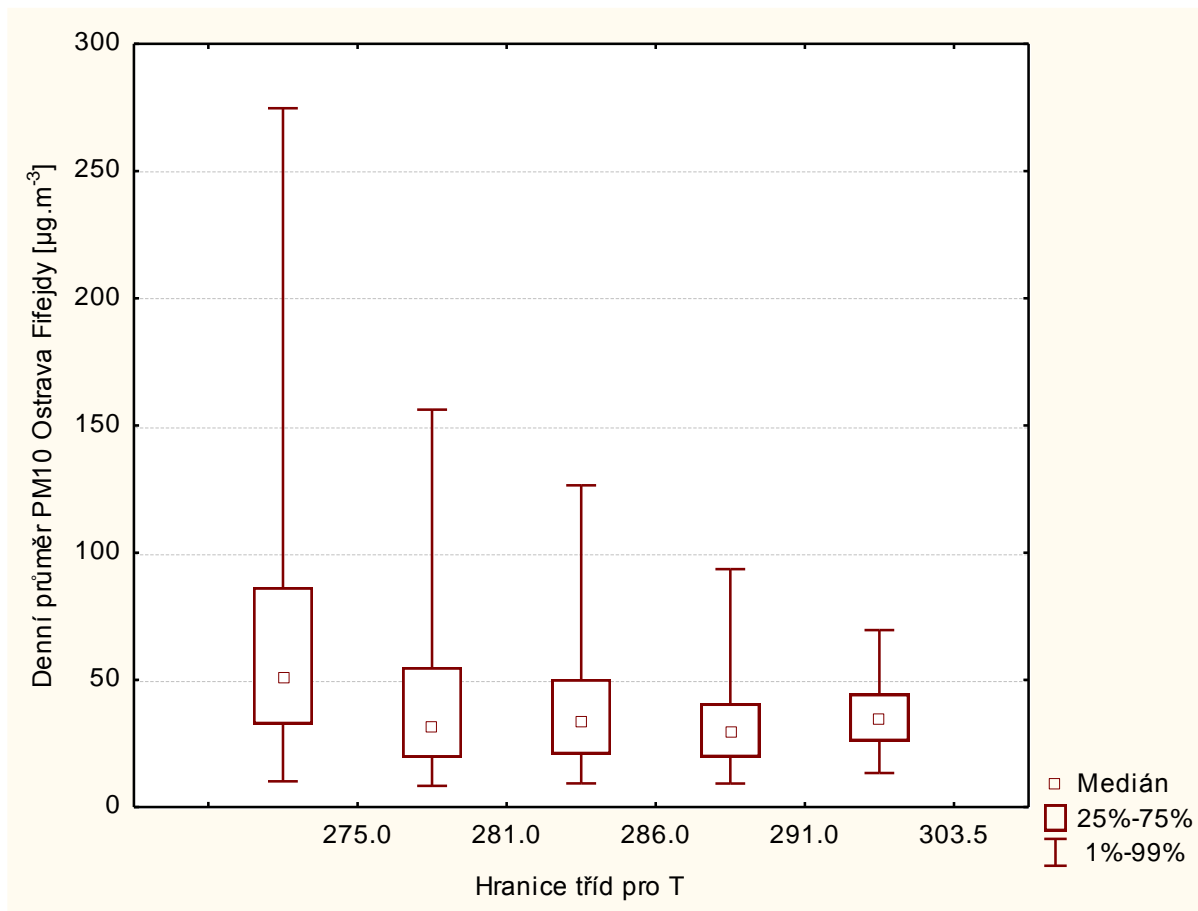
# Lokalita O. Fifejdy – rozložení T pro hranice, stanovené jako sudé decily



# Lokalita O. Fifejdy – rozložení $PM_{10}$ pro hranice VI, stanovené jako sudé decily



# Lokalita O. Fifejdy – rozložení $PM_{10}$ pro hranice T, stanovené jako sudé decily



# Vývoj pokusného modelu (1)

Použitá data: koncentrace  $PM_{10}$ , teplota vzduchu a  
ventilační index pro lokalitu Ostrava-Fifejdy  
a další místa za roky 2004 - 2013  
(denní průměru 06-06 UTC)

Roky 2004 – 2012 učební soubor

Rok 2013 soubor pro validaci (tato data model „nezná“)

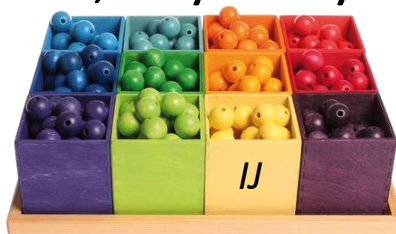
Vytvoří se 25 „krabiček“ kombinací 5 kategorií pro VI  
a 5 kategorií pro T

# Vývoj pokusného modelu (2)

25 „krabiček“ – tříd pro kombinace kategorií VI a T,  
označeno 11 až 55

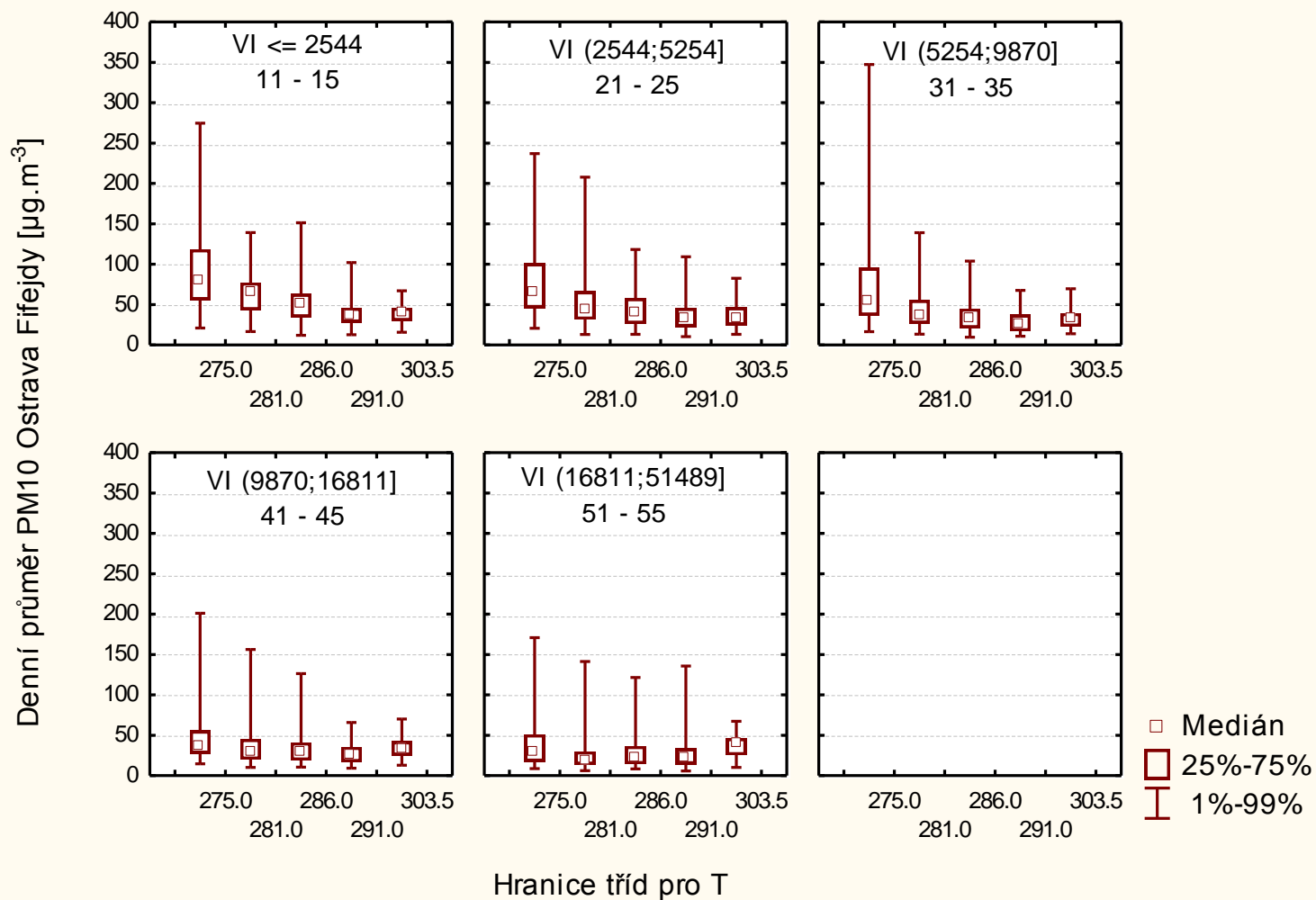
T	1	<275	11	12	13	14	15
	2	275-281	21	22	23	24	25
	3	281-286	31	32	33	34	35
	4	286-291	41	42	43	44	45
	5	291-303.5	51	52	53	54	55
		<2544	2544-5254	5254-9870	9870-16811	16811-51489	
		1	2	3	4	5	
				VI			

Do „krabiček“ *IJ* se nasypou denní koncentrace  $PM_{10}$ , které byly naměřeny ve dnech, kdy se vyskytla *I-tá* katgorie VI a *J-tá* kategorie T



Zjistí se statistická rozložení  $PM_{10}$  v jednotlivých krabičkách a jejich percentily

# Rozložení $PM_{10}$ v „krabičkách“



# Příklad výstupu modelu (1)

Pravděpodobnosti překročení prahových hodnot pro denní koncentrace PM<sub>10</sub> kategoriích 11 a 12, odvozené z percentilů rozložení

		Kategorie									
		11	12	13	14	15	21	22	23	24	25
Pravděpodobnost překročení prahové koncentrace pro denní průměr	25	0.995	0.993	0.966	0.879	0.804	0.952	0.909	0.927	0.730	0.612
	50	0.772	0.775	0.540	0.218	0.117	0.649	0.610	0.393	0.143	0.093
	75	0.509	0.309	0.233	0.041	0.000	0.359	0.309	0.132	0.028	0.000
	100	0.327	0.198	0.078	0.000	0.000	0.235	0.220	0.051	0.004	0.000
	125	0.207	0.081	0.015	0.000	0.000	0.170	0.129	0.005	0.000	0.000
	150	0.076	0.048	0.009	0.000	0.000	0.116	0.053	0.000	0.000	0.000
	175	0.047	0.019	0.000	0.000	0.000	0.067	0.038	0.000	0.000	0.000
	200	0.039	0.008	0.000	0.000	0.000	0.037	0.014	0.000	0.000	0.000
	225	0.023	0.000	0.000	0.000	0.000	0.028	0.000	0.000	0.000	0.000
	250	0.016	0.000	0.000	0.000	0.000	0.013	0.000	0.000	0.000	0.000
	275	0.011	0.000	0.000	0.000	0.000	0.008	0.000	0.000	0.000	0.000
	300	0.008	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
	325	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	350	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	375	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	400	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	425	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	450	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	475	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000



# Příklad výstupu modelu (2)

Hodnoty denní koncentrace PM<sub>10</sub> kategoriích 11 a 12, které mohou být překročeny se zadanou pravděpodobností.

Odvozeno z percentilů rozložení

		Kategorie									
		11	12	13	14	15	21	22	23	24	25
Denní průměr, který bude překročen se zadanou pravděpodobností	0.99	31.2	30.4	20.2	16.2	13.7	19.1	14.3	18.8	11.6	13.1
	0.95	34.1	33.2	26.8	19.4	18.2	25.3	19.9	21.8	16.4	15.4
	0.9	36.9	39.2	30.0	23.4	20.0	30.5	27.0	27.0	19.8	18.1
	0.85	41.0	42.7	34.6	26.9	21.9	35.5	30.5	30.9	21.4	18.9
	0.8	45.1	46.3	35.7	28.5	25.2	37.8	33.6	32.0	23.0	19.9
	0.75	50.6	51.2	37.9	31.0	28.2	41.8	37.1	33.6	24.5	21.7
	0.7	58.9	53.9	41.2	32.0	29.5	46.3	42.9	36.4	25.7	22.6
	0.65	65.3	55.4	43.4	33.8	30.8	49.9	46.4	38.0	27.4	24.0
	0.6	69.8	56.3	45.2	35.3	32.5	54.5	51.6	40.1	30.2	25.1
	0.55	72.4	59.1	49.7	37.1	34.7	58.5	56.5	41.9	31.9	26.8
	0.5	76.0	62.0	52.5	39.0	36.5	60.8	61.5	44.5	33.3	28.2
	0.45	82.5	66.2	53.6	40.2	37.8	63.5	64.5	46.0	34.5	29.6
	0.4	90.8	67.8	56.7	41.5	39.6	69.5	69.8	49.7	36.4	30.6
	0.35	94.3	71.9	60.7	43.9	40.8	75.2	71.8	51.9	38.1	33.0
	0.3	105.4	78.9	65.6	45.9	42.9	87.6	75.4	56.7	40.2	35.8
	0.25	115.0	89.7	73.0	47.4	44.6	96.7	93.0	59.8	43.9	40.6
	0.2	126.8	99.9	79.5	51.0	45.8	111.1	107.3	63.6	47.7	44.8
	0.15	133.6	110.6	84.7	55.2	47.8	129.7	118.3	73.8	49.8	47.4
	0.1	144.1	122.0	91.9	61.1	52.2	155.5	130.3	81.0	54.1	49.5
	0.05	162.0	144.9	104.4	67.3	56.5	190.7	152.4	100.1	62.0	56.2
0.01	276.4	192.2	143.3	92.5	65.6	261.6	202.5	116.0	82.9	60.9	
0.001	351.3	221.5	158.9	98.9	70.2	298.0	207.3	133.5	110.5	62.7	
0	359.6	224.8	159.9	99.9	70.8	301.7	207.8	136.0	115.1	62.9	

# Postup predikce

Z NWP modelu (ALADIN) se získají předpovědi denních průměrů ventilačního indexu a teploty pro lokalitu

Zjistí se, mezi které meze pro VI a T předpovězené hodnoty padnou, určí se krabička *IJ*

Stanoví se pravděpodobnosti překročení jednotlivých prahových jednotlivých prahových hodnot pro predikovaný den

Prahové hodnoty je možno stanovit i jinak, než je v popisu uvedeno

# Testování modelu

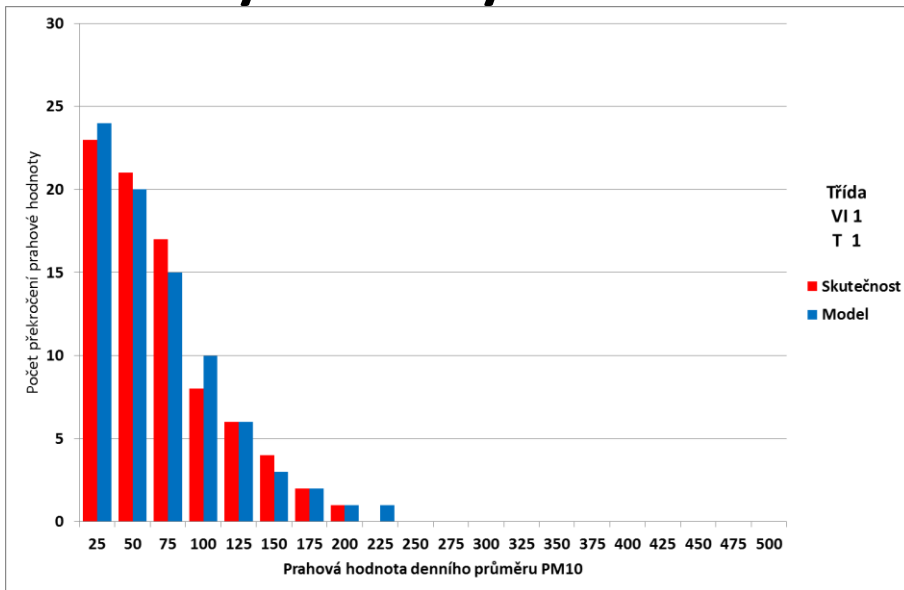
Na datech VI a T z roku 2013, predikovaných modelem ALADIN pro stanici Ostrava Fifejdy na 24 hodin dopředu

Ze predikovaných VI a T pro každý den pro jednotlivé kategorie stanoveny počty překročení prahových hodnot z naměřených dat a teoretické počty překročení podle modelu

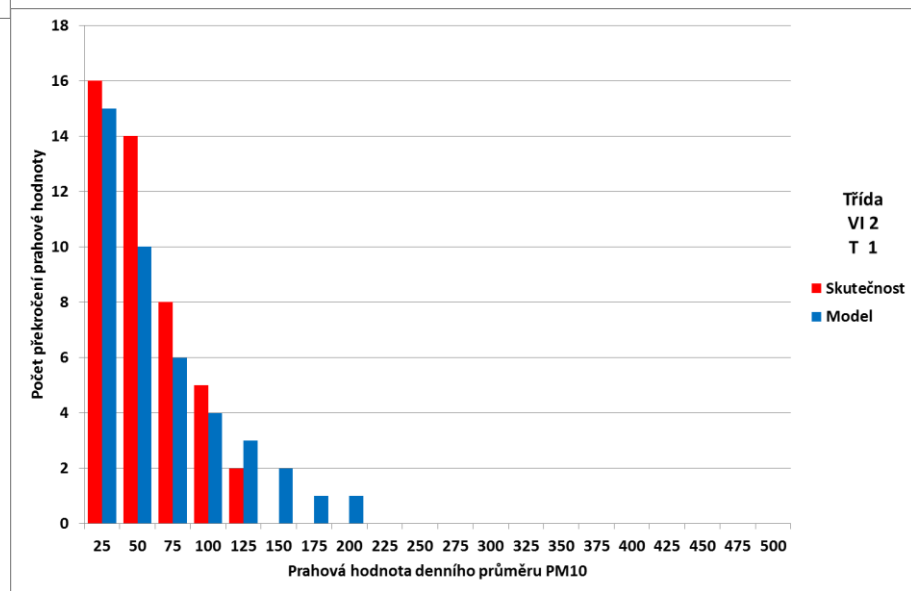
*= celkový počet výskytu dané kategorie \* pravděpodobnost překročení prahu*

Grafické porovnání pro jednotlivé kategorie i pro celý soubor roku 2013

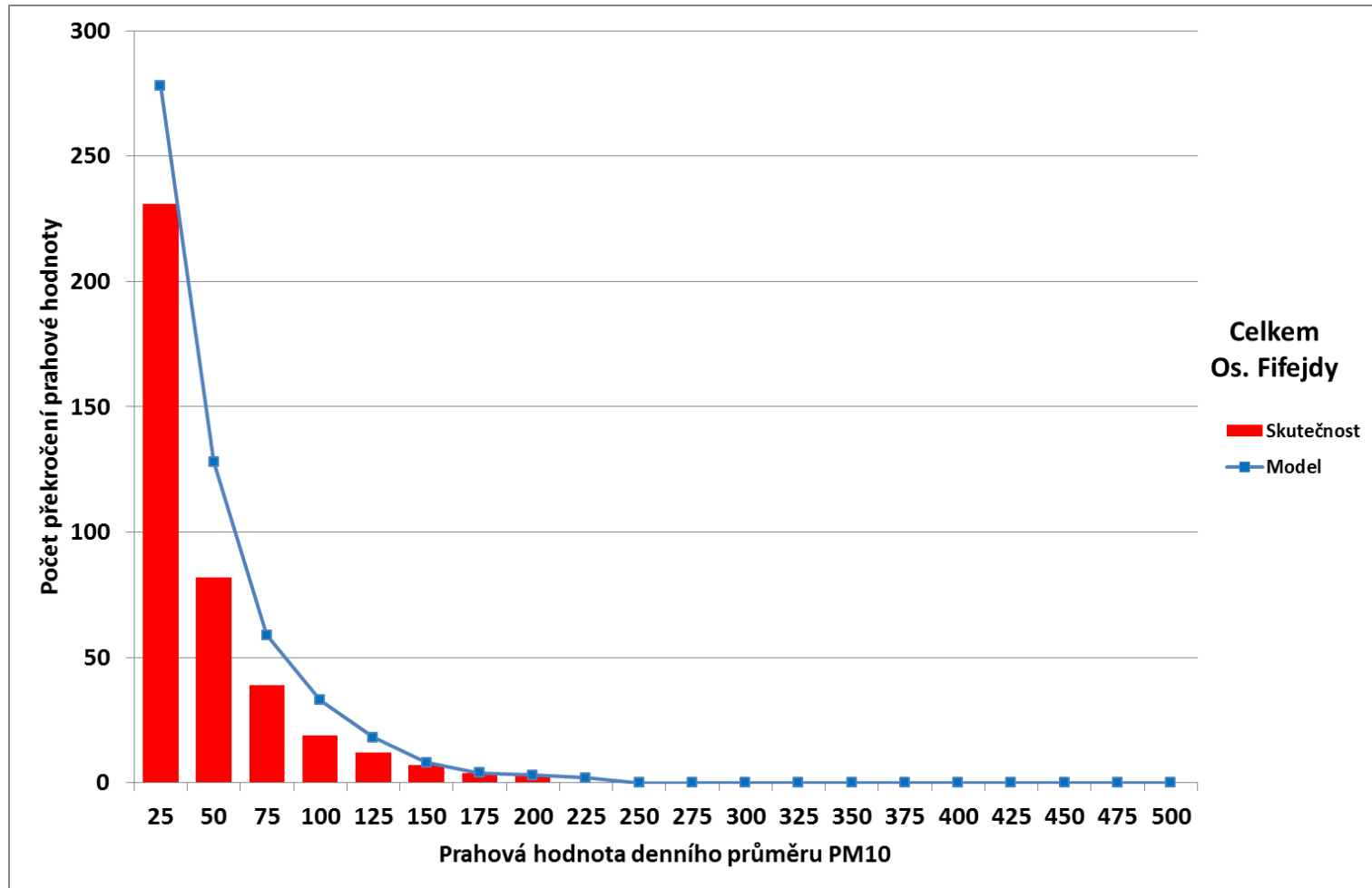
# Výsledky testování modelu - kategorie



## Ostrava - Fifejdy



# Výsledky testování modelu – celý soubor



# Výsledky testování modelu – Fifejdy

Z predikce hodnot VI a T modelem ALADIN se stanoví predikované hodnoty percentilů

Časová řada predikovaných percentilů se porovná s časovou řadou denních průměrů koncentrace

Použity 90tý (P90) a 99tý (P99) percentil

Interpretace:

Předpovídá se, že denní průměr PM10

*překročí hodnotu P90 v 1 případě z 10*

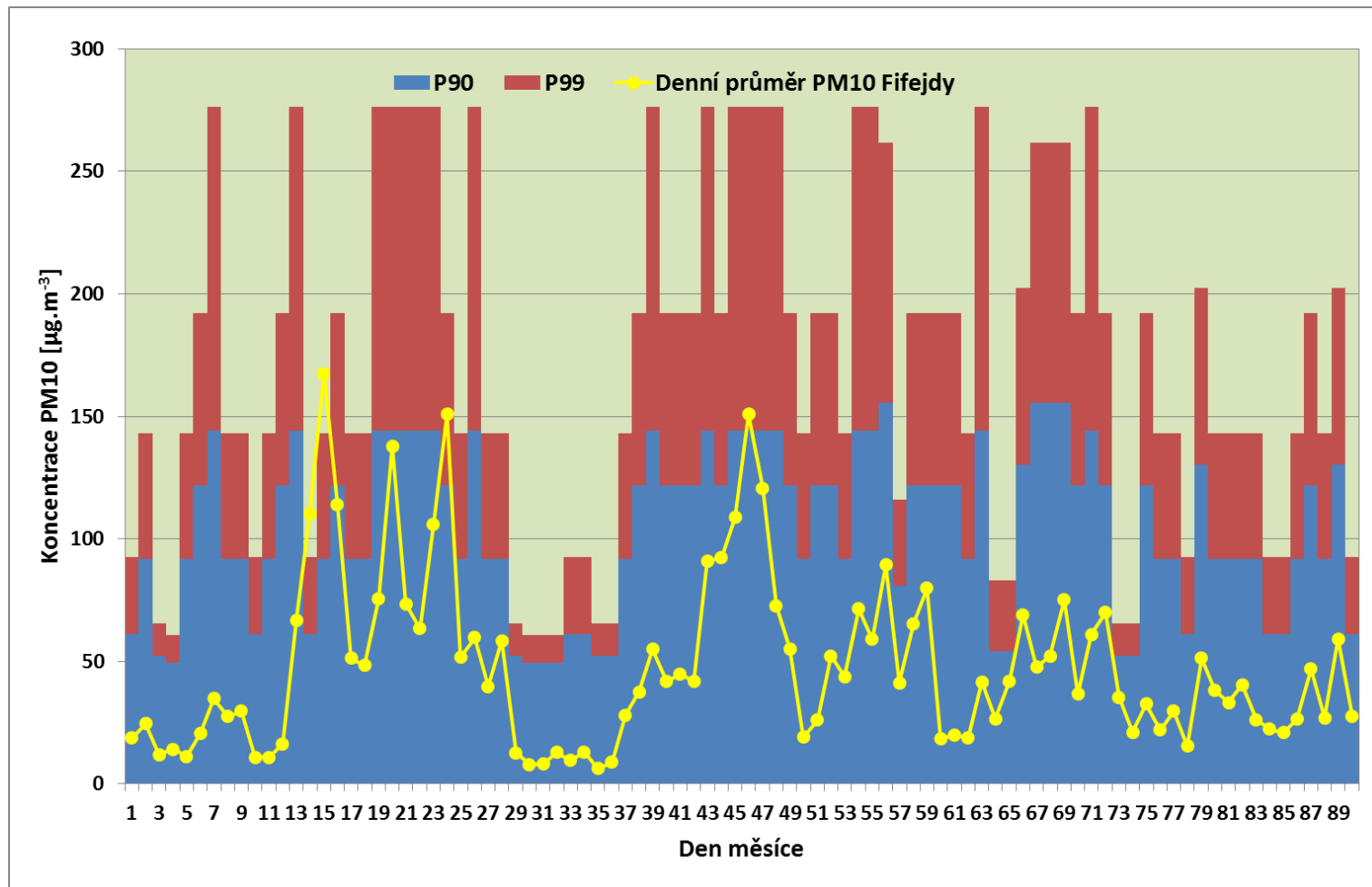
a současně

*v 99 případech ze 100 bude menší než hodnota P99*

# Výsledky testování modelu leden-březen 2013

## Ostrava Fifejdy

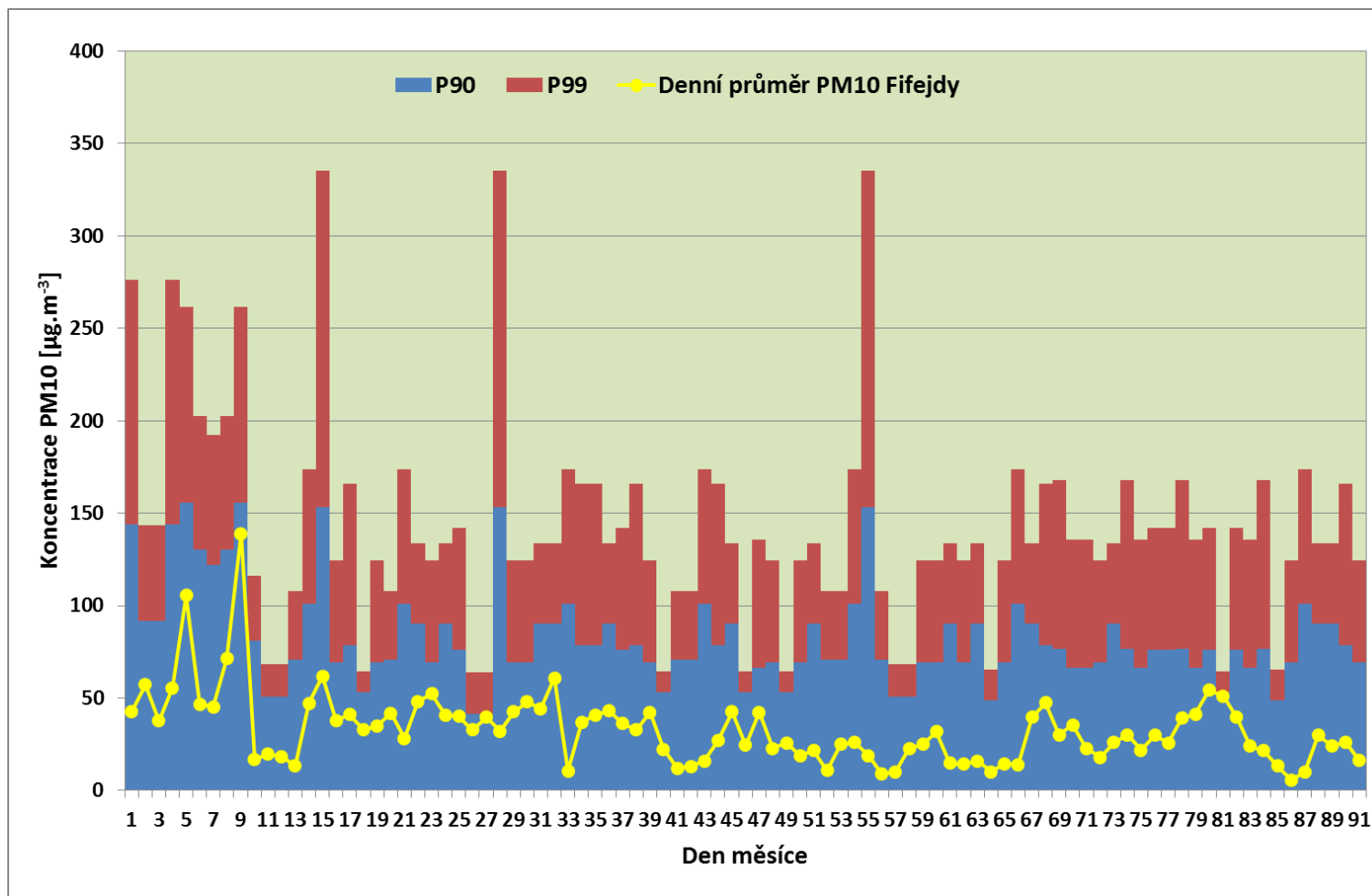
Denní průměr překročí P90 v 1 případě z 10 a je menší než P99 v 99 případech ze 100



# Výsledky testování modelu duben-červen 2013

## Ostrava Fifejdy

Denní průměr překročí P90 v 1 případě z 10 a je menší než P99 v 99 případech ze 100

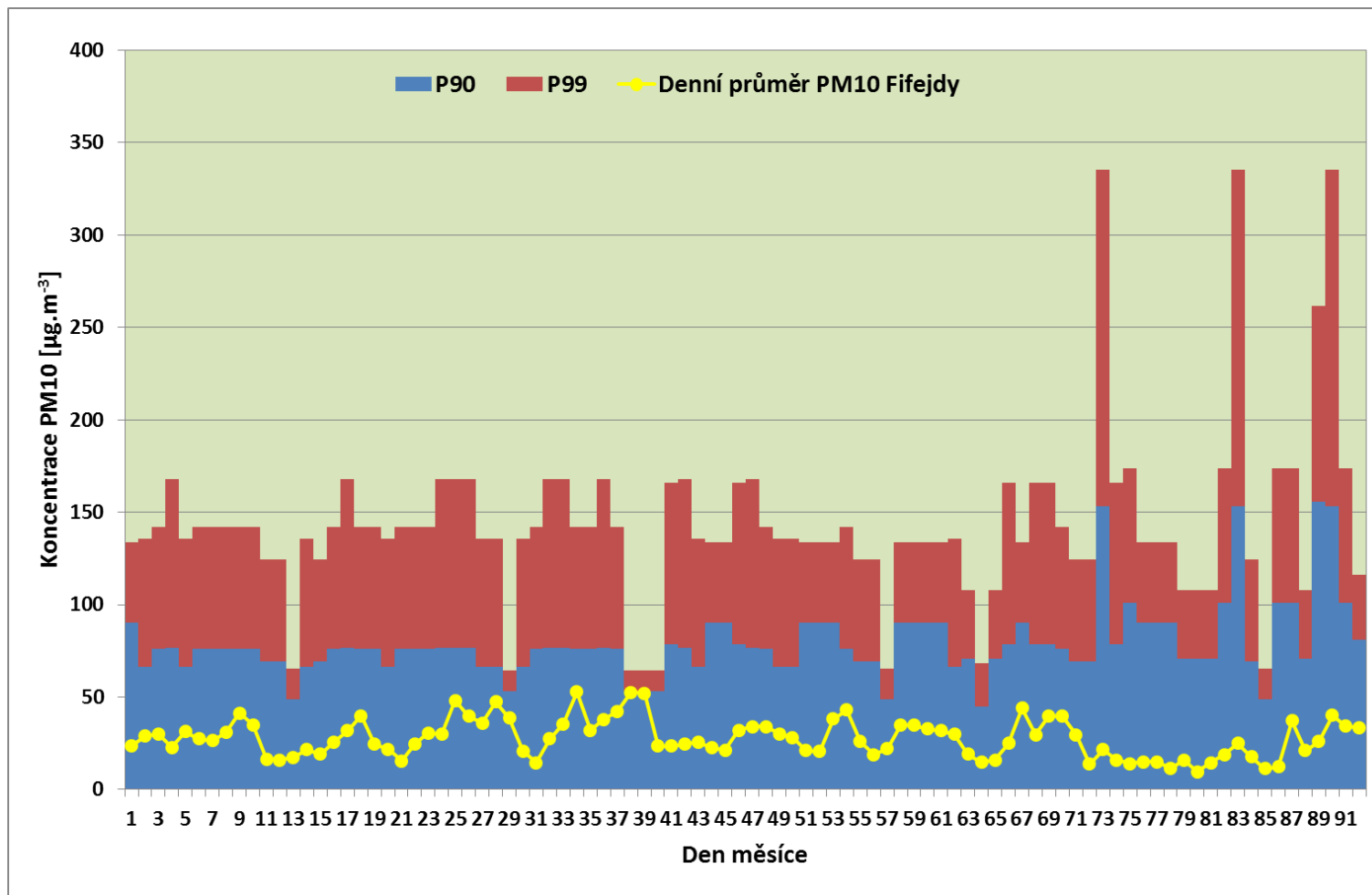




# Výsledky testování modelu červenec – září 2013

## Ostrava Fifejdy

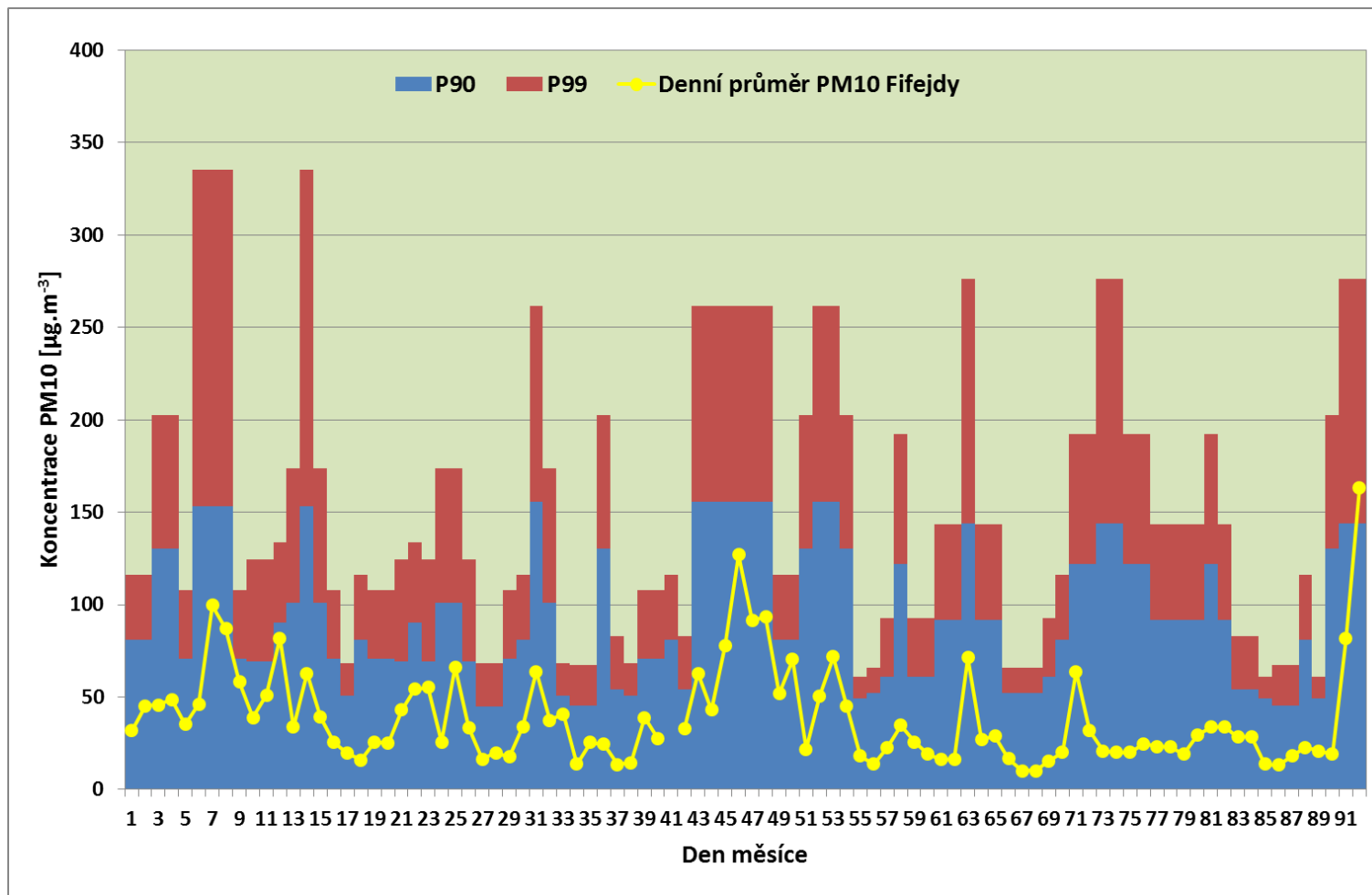
Denní průměr překročí P90 v 1 případě z 10 a je menší než P99 v 99 případech ze 100



# Výsledky testování modelu říjen-prosinec 2013

## Ostrava Fifejdy

Denní průměr překročí P90 v 1 případě z 10 a je menší než P99 v 99 případech ze 100



# Závěry a další možný vývoj

Model si vede docela dobře, nepotvrdily se obavy, že použití predikovaných hodnot VI a T podstatně zhorší úspěšnost modelu

Model v některých třídách (např. VI2T1) podceňuje počet překročení nižších prahových hodnot (pod  $100 \mu\text{g}\cdot\text{m}^{-3}$ )

V součtu pro všechny třídy model nadhodnocuje počet překročení prahových hodnot do  $100 \mu\text{g}\cdot\text{m}^{-3}$ , pro vyšší koncentrace jsou rozdíly mezi modelem a skutečností v řádu jednotek

## **V dalším se plánuje provést**

- Ověření na datech za 2014 a 2015
- Sestavit model pro klouzavé 12ti hodinové průměry – využití pro SVRS
- Provéřit možnost zahrnutí dalšího prediktoru, např, převládajícího proudění – vzroste počet „krabiček“
- Rozšířit predikci na delší horizont, 48 a 72 (?) hodin

Vyvinout modely pro všechny stanice s měřením denních průměrů  $\text{PM}_{10}$

Každý rok modely aktualizovat ze získaných (naměřených) dat  $\text{PM}_{10}$ , VI, T