

# K možnostem krátkodobé předpovědi úrovně znečištění ovzduší statistickými metodami

Josef Keder

# Motivace

- Předpověď budoucí úrovně znečištění ovzduší s předstihem v řádu alespoň několika hodin má význam zejména pro provoz SVRS
- Hlavní směr vývoje v této oblasti představují numerické modely šíření a rozptylu znečišťujících látek ve spojení s numerickými modely předpovědi počasí.
- V příspěvku jsou představeny alternativní přístupy, které mohou uvedené doplnit a rozšířit – zejména umělé neuronové sítě.
- Jsou shrnuty základní principy metody a představeny výsledky prvních testů na reálných datech

## Současný stav problematiky

- Oba přístupy – rozptylové modelování a statistické metody – se rozvíjely v podstatě paralelně
- Zdokonalení a zrychlení výpočetní techniky umožnilo prudký rozvoj v obou směrech
- U statistických metod se jedná o výzkum chování časových řad koncentrací a predikci jejich vývoje
- Je možné těžit z výsledků zejména ekonometrických studií (vývoj cen akcií a komodit)
- Objevily se metody tzv. soft-computingu a data-miningu („dolování“ informací z dat)

## Některé známé postupy, mimo jiné ...

- Pojednány v řadě knih, článků a učebních textů včetně webu, paleta je nepřeborná
- Vícerozměrná regrese, logistická regrese
- ARIMA, exponenciální vyhlazování (Holt, Winters, Brown), filtry KAMA (Kaufman), TEMA, Kalmánův filtr
- Klasifikační a regresní stromy, shluková analýza
- Umělá inteligence - umělé neuronové sítě (ANNs)
- Provedeny pokusy s filtrací a hlazením časových řad, poté pozornost upřena na ANNs

# Co jsou ANNs (stručný úvod) ?

StatSoft

## Úvod do neuronových sítí

Vzhledem k vzrůstající popularitě neuronových sítí jsme se rozhodli Vám je v tomto článku představit a říci si něco o jejich využití. Co si tedy představit pod pojmem neuronová síť?



# Co jsou ANNs?

Velice hezká práce v češtině (a řada dalších)



MASARYKOVA UNIVERZITA  
PŘÍRODOVĚDECKÁ FAKULTA  
STUDIJNÍ PROGRAM:  
EXPERIMENTÁLNÍ BIOLOGIE

---



## Predikce časových řad pomocí umělých neuronových sítí

Bakalářská práce

**Roman Vyškovský**



# Co jsou ANNs?

Velice hezká práce v češtině (a řada dalších)



MASARYKOVA UNIVERZITA  
PŘÍRODOVĚDECKÁ FAKULTA  
STUDIJNÍ PROGRAM:  
EXPERIMENTÁLNÍ BIOLOGIE

---



## Predikce časových řad pomocí umělých neuronových sítí

Bakalářská práce

**Roman Vyškovský**



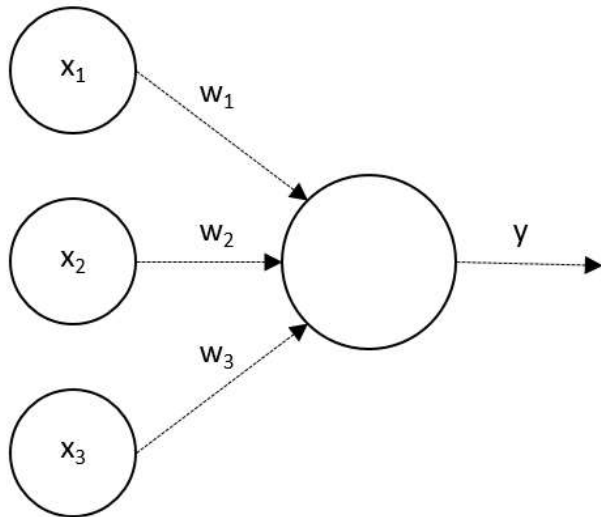
## Jak to tedy funguje?

- ANN = algoritmus, který modeluje činnost lidského mozku
- Mozek se skládá z obrovského množství vzájemně propjených buněk – neurony
- Komunikují navzájem pomocí elektrických impulzů
- **Neuron (perceptron)** - základní prvek přirozené i umělé neuronové sítě
- Každý neuron může mít více vstupů, ale jen jeden výstup – tento může být napojen na více než jeden další neuron



# Model neuronu

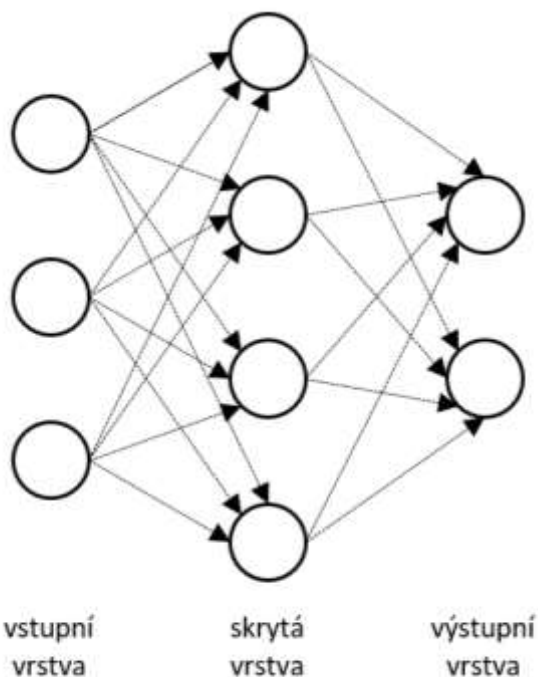
- **Vstupem** neuronu může být buď výstup z jiného neuronu nebo informace z vnějšku (koncentrace)
- Každý vstup má přiřazenou **váhu**.



- Neuron přijme vstupy
- Jejich hodnoty vynásobí váhami a součiny sečte
- Je-li výsledek větší než stanovený **práh**, transformuje se předem danou **přenosovou funkcí**, pošle se na **výstup**.

# Propojení neuronů do vícevrstvé sítě

- Příklad - síť typu 3-4-2,
- 3 vstupy, 4 neurony ve skryté vrstvě, 2 výstupy



- **Vstupní vrstva** zopakuje hodnoty vstupů a předá je všem neuronům ve **skryté vrstvě**.
- Jejich výstupy jsou vstupem pro dva neurony ve **výstupní vrstvě** - kódují výstup celé neuronové sítě.

## Učení neuronové sítě

- Síti se předkládají vstupy a výstupy, které těmto vstupům odpovídají – z historických dat
- Převod vstupů na výstupy závisí na hodnotě vah a prahů
- Stanovují se během **učení**
- Vstupy se pustí do sítě, výstup se porovná s požadovaným výstupem a následně se provede korekce (změna vah a prahů) tak, aby byl rozdíl mezi skutečným a požadovaným výstupem co nejmenší
- Síť se tedy učí ze svých chyb, postupně je snižuje

## Výhody ANNs

- Schopnost učit se = zapamatovat si kombinace, které vedly k požadovanému výstupu
- U neznámých vstupů se schopnost obracet na „svou“ paměť a na základě zkušeností odhadovat nový výsledek = generalizace jako zárodek umělé inteligence
- Řeší i silně nelineární, kde není dostatek apriorních znalostí o tom, co způsobuje konkrétní variabilitu zkoumané proměnné
- Neznámé vlivy pak odvozujeme na základě znalosti historických naměřených dat

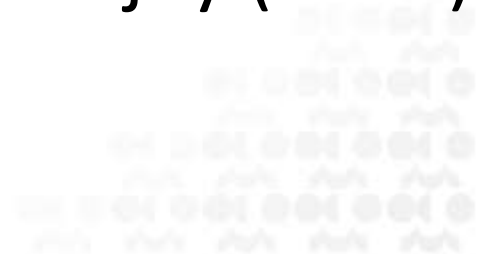
## Kritizované nevýhody ANNs

- Neinterpretovatelnost některých parametrů sítě
- „Black box“, neumíme jako u klasických modelů říci, k jakým výpočtům uvnitř systému dochází



## Období, data, postupy (1)

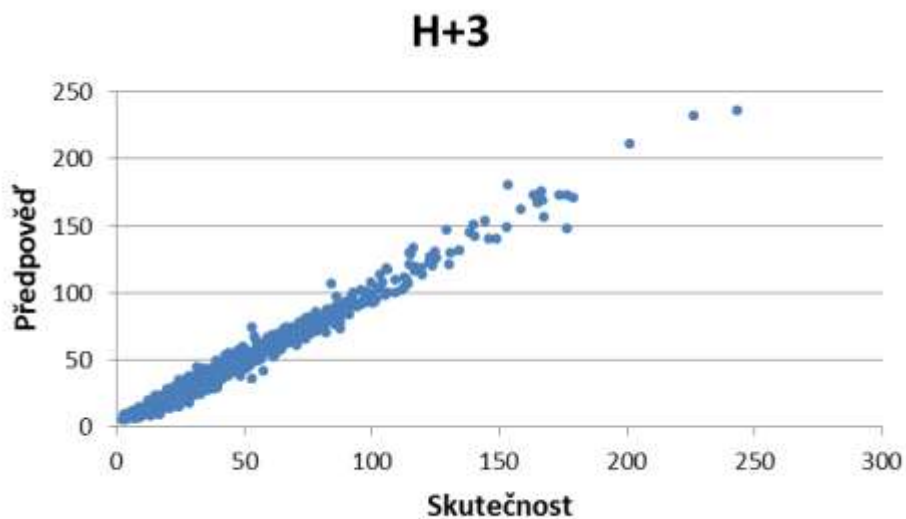
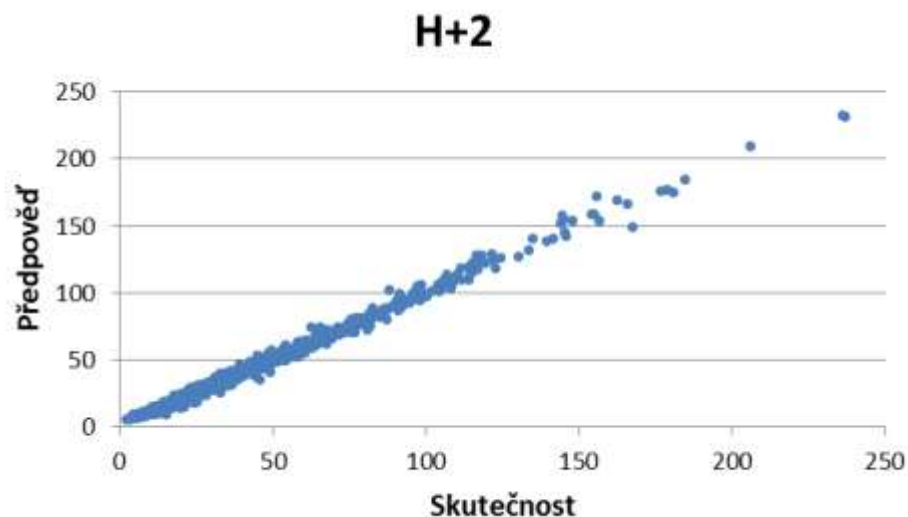
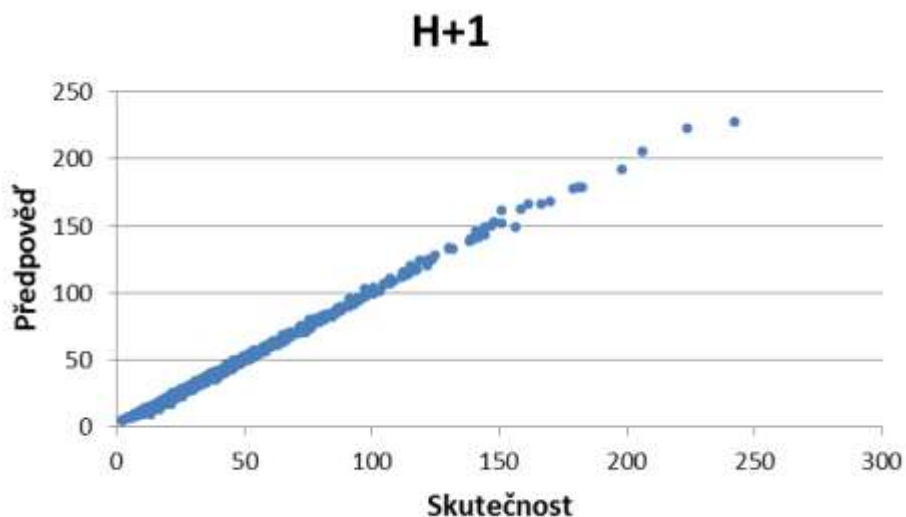
- Časové řady koncentrací PM10 ze stanic MS kraje
- Roky 2015 – 2018
- Vyhodnoceny klouzavé 12 hodinové průměry (časové řady) jako vstupy pro ANNs
- Použit nástroj Automatická tvorba neuronových sítí, součást programu STATISTICA
- Učení sítě na datech 9/2015 až 3/2016
- Prezentace testů na stanici Ostrava - Fifejdy (TOFF)



## Období, data, postupy (2)

- Pro každou hodinu  $H$  na vstup zavedeno 6 hodnot 12h průměrné koncentrace  $c(H)$ ,  $c(H-1)$  ...  $c(H-5)$
- Jako „kýžené“ výstupy předloženy hodnoty z hodin  $c(H+1)$ ,  $c(H+2)$  ...  $c(H+6)$
- 25% dat nebylo pro učení sítě použito, síť je neznala, použita k testování úspěšnosti modelu v období 2015/16
- Trénováno celkem 20 sítí, pro predikce použito 5 nejúspěšnějších, pracovaly jako kolekce
- Provozní test na datech z dalšího zimního období 9/2016 – 3/2017, zcela neznámá pro ANNs

# Testování zima 2015/16, horizont predikce 1 až 3h

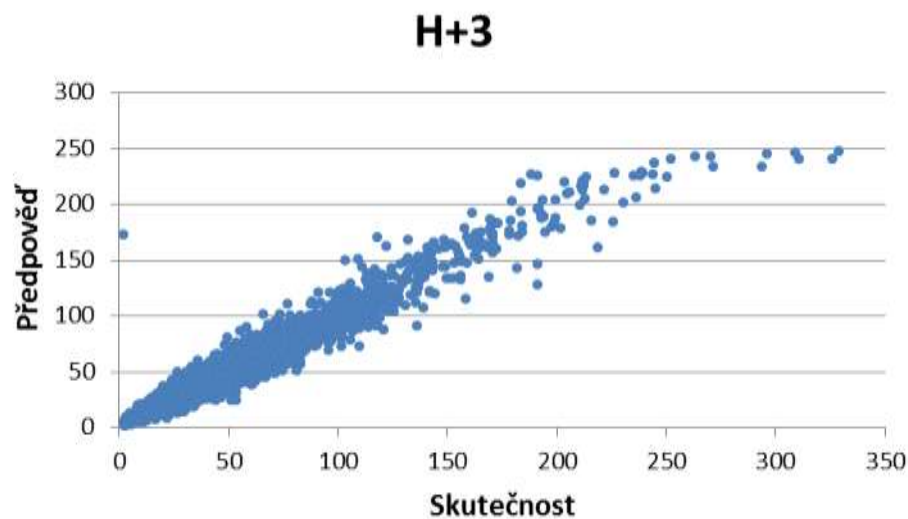
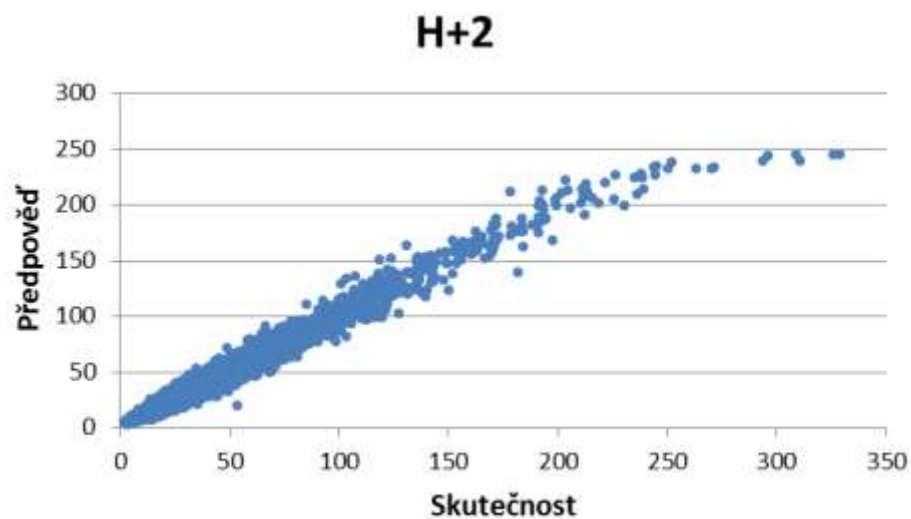
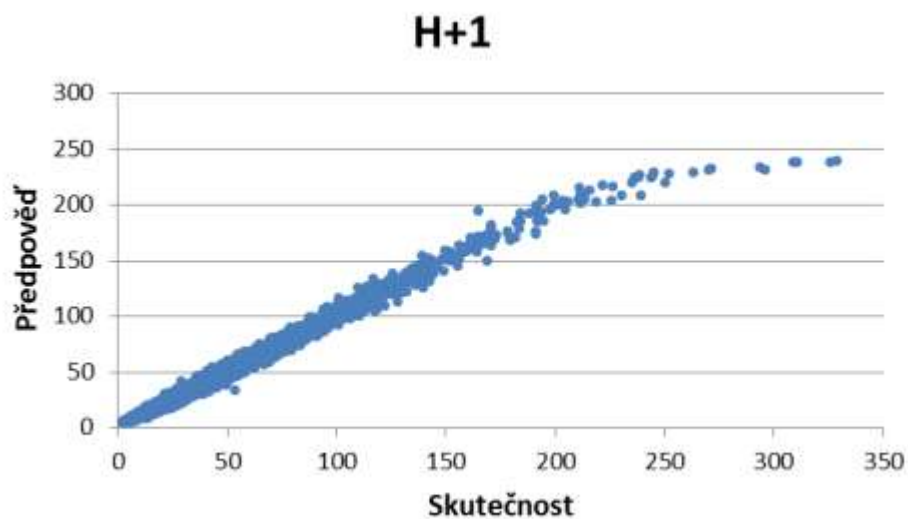


H+1	H+2	H+3	
0.023	2.523	4.014	RMS
0.999	0.996	0.991	CORR

**Klouzavé průměry PM10, 12 hodin,  
Ostrava-Fifejdy**



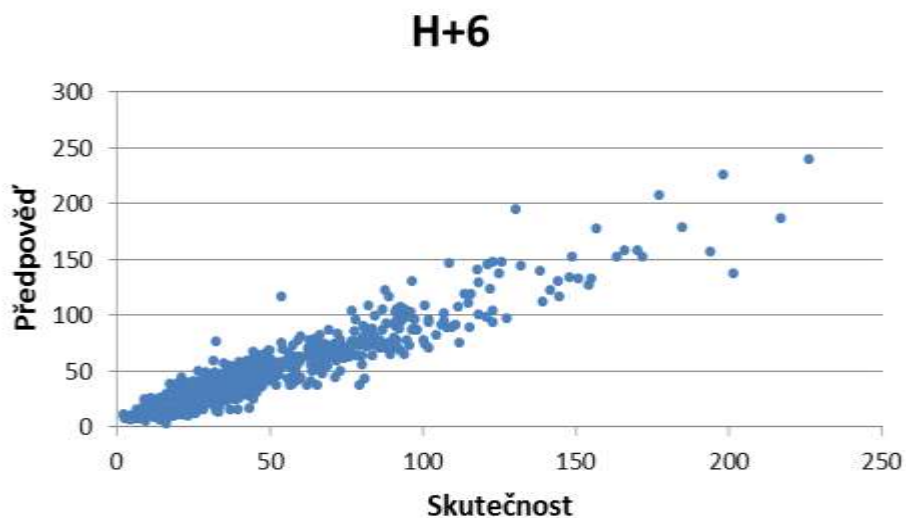
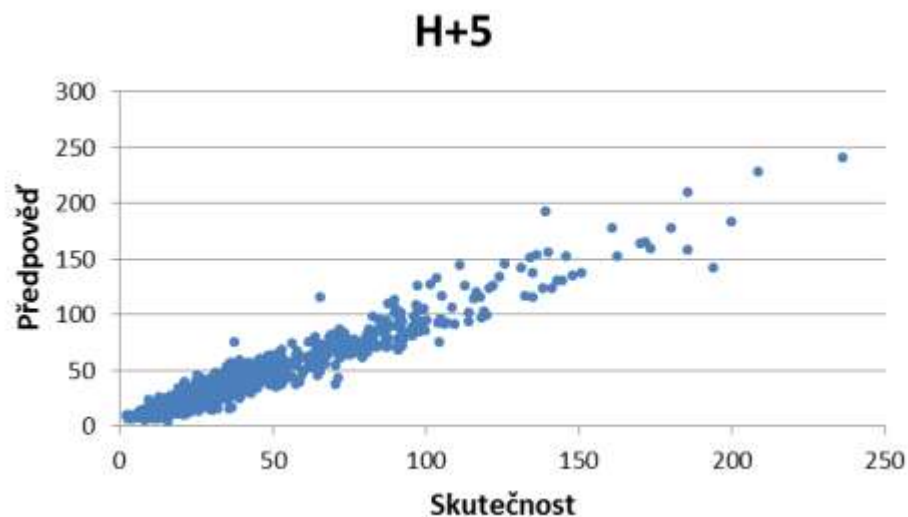
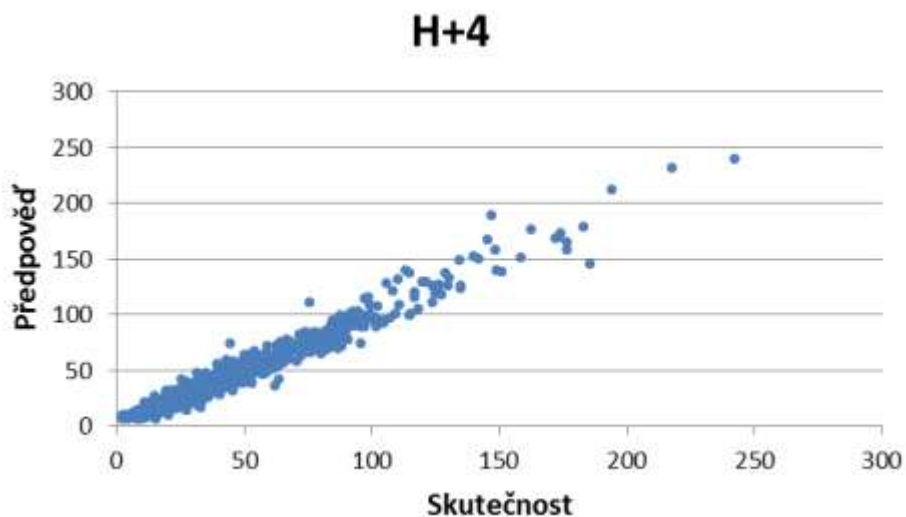
# Testování zima 2016/17, horizont predikce 1 až 3h



H+1	H+2	H+3	
4.392	5.992	8.304	RMS
0.993	0.986	0.973	CORR

**Klouzavé průměry PM10, 12 hodin,  
Ostrava-Fifejdy**

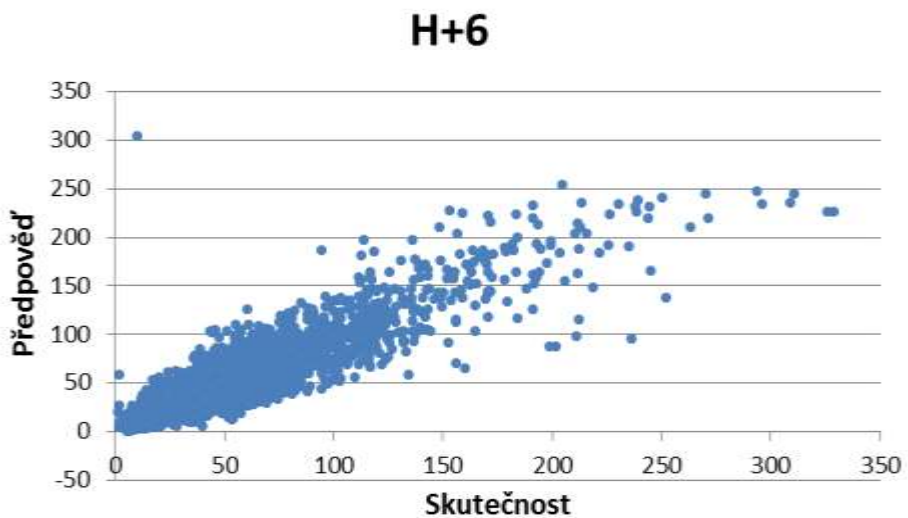
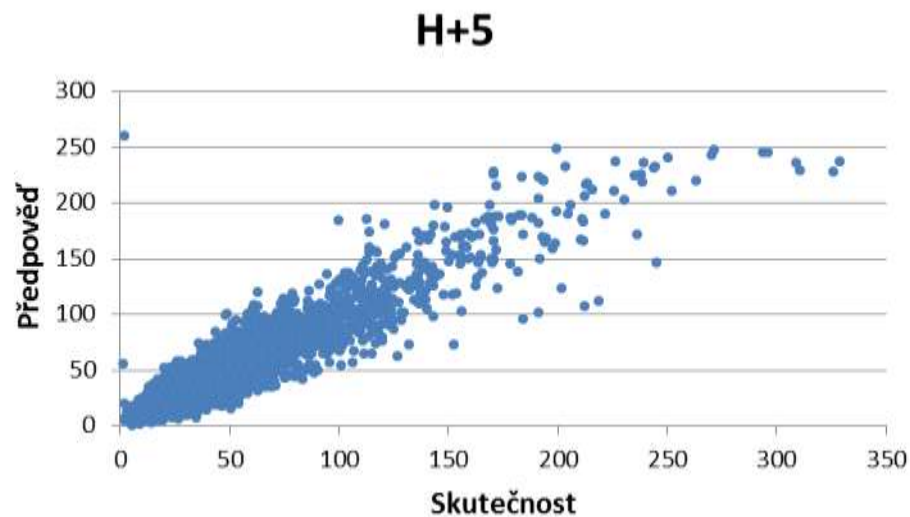
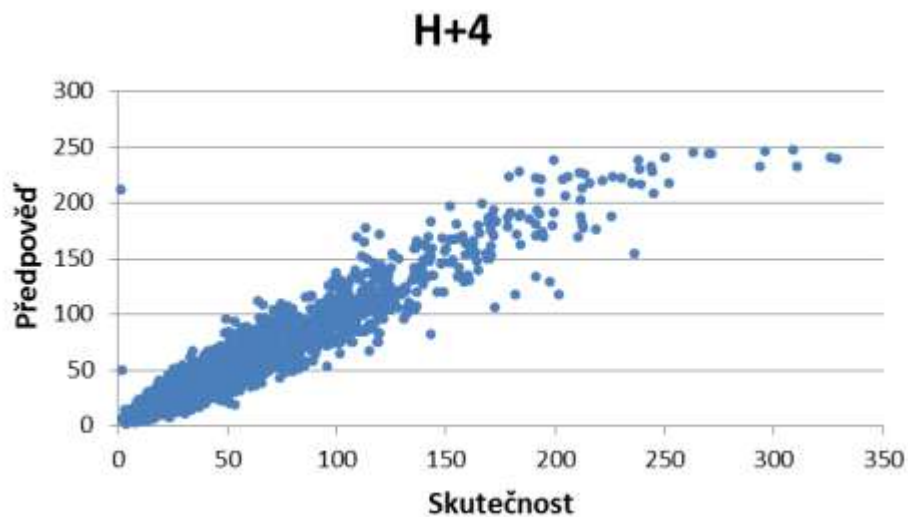
# Testování zima 2015/16, horizont predikce 4 až 6h



H+3	H+5	H+6	
5.773	7.559	9.476	RMS
0.981	0.968	0.950	CORR

**Klouzavé průměry PM10, 12 hodin,  
Ostrava-Fifejdy**

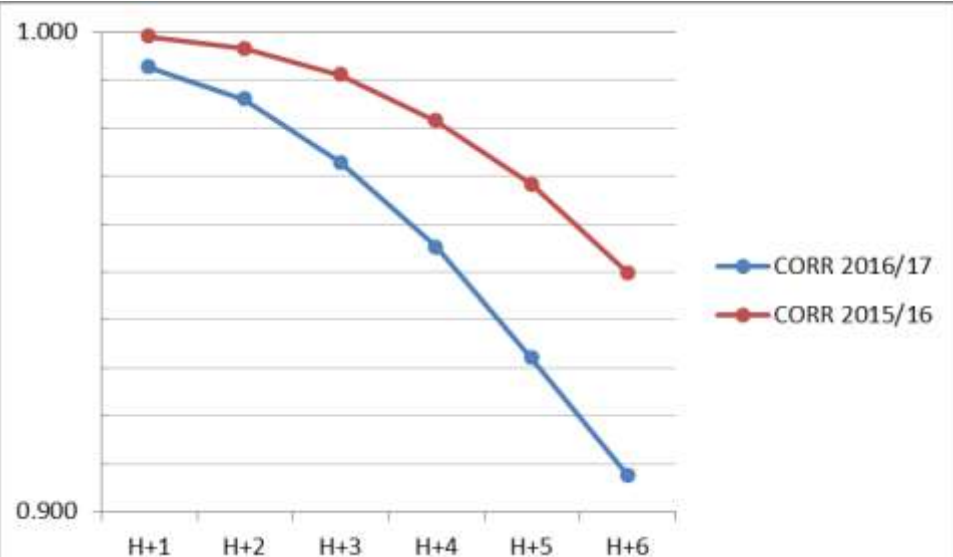
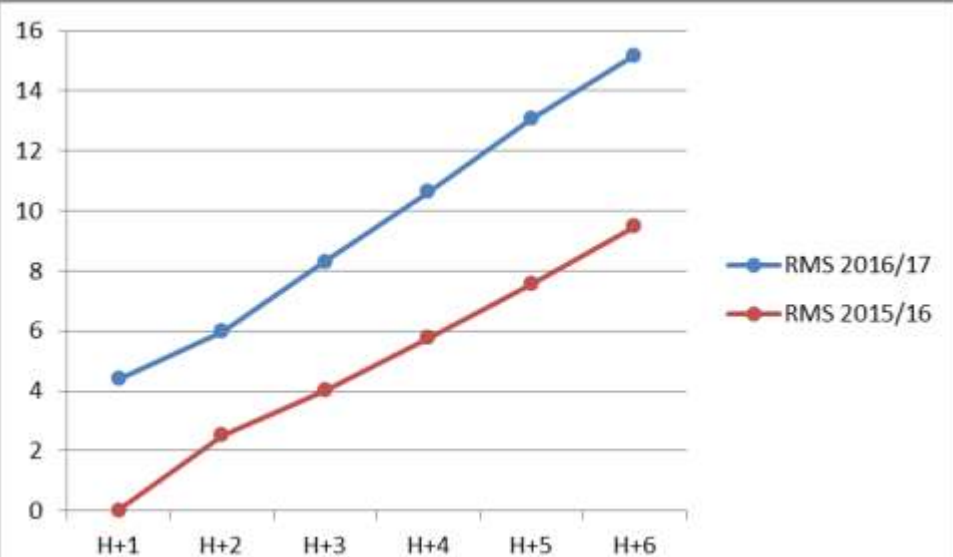
# Testování zima 2016/17, horizont 4 až 6h



H+4	H+5	H+6	
10.641	13.068	15.174	RMS
0.955	0.932	0.908	CORR

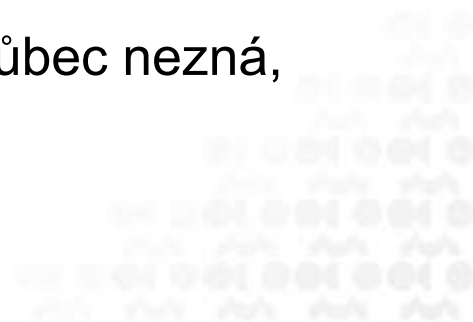
Klouzavé průměry PM10, 12 hodin,  
Ostrava-Fifejdy

# Závislost RMS a CORR na horizontu předpovědi

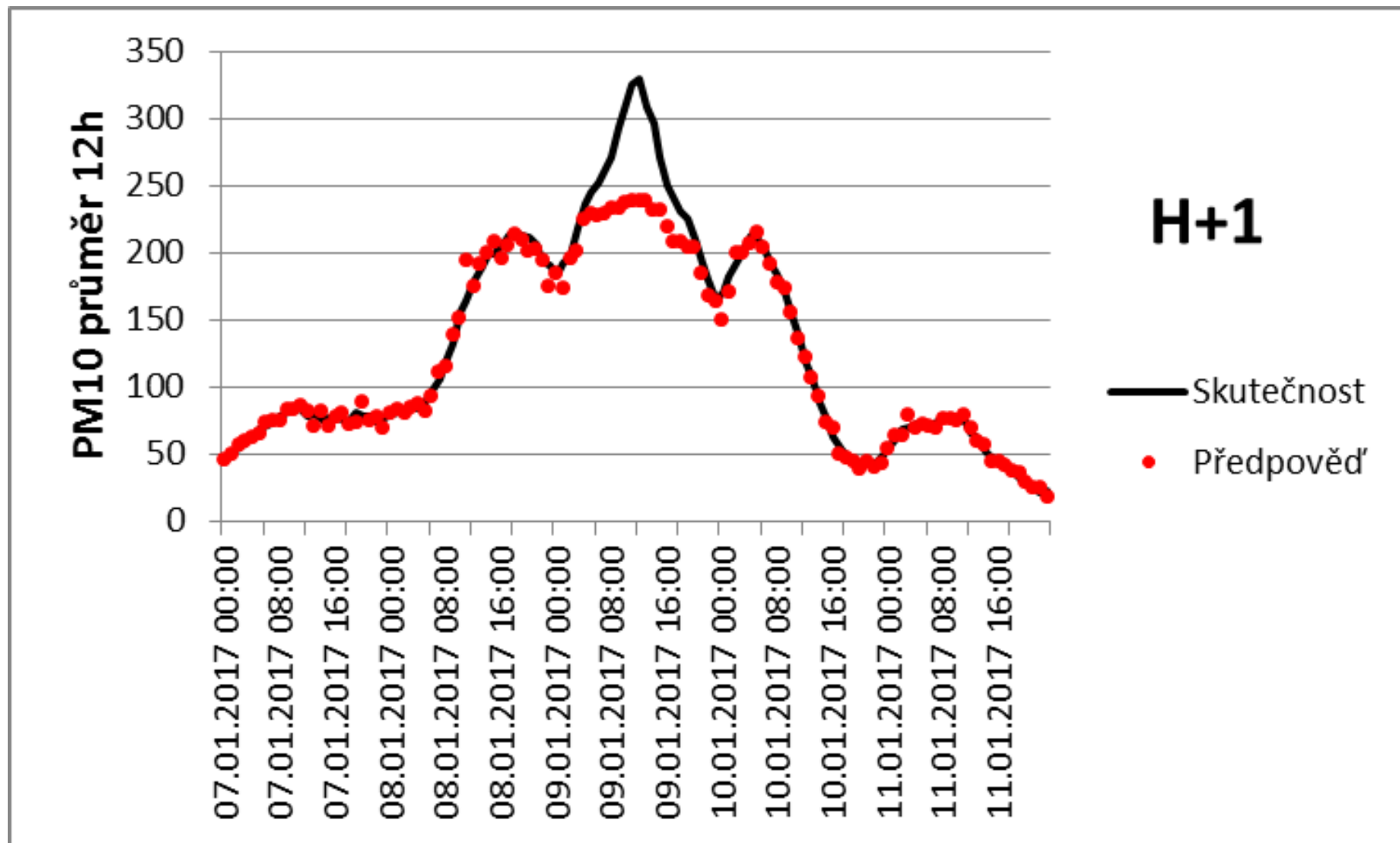


## Podle očekávání

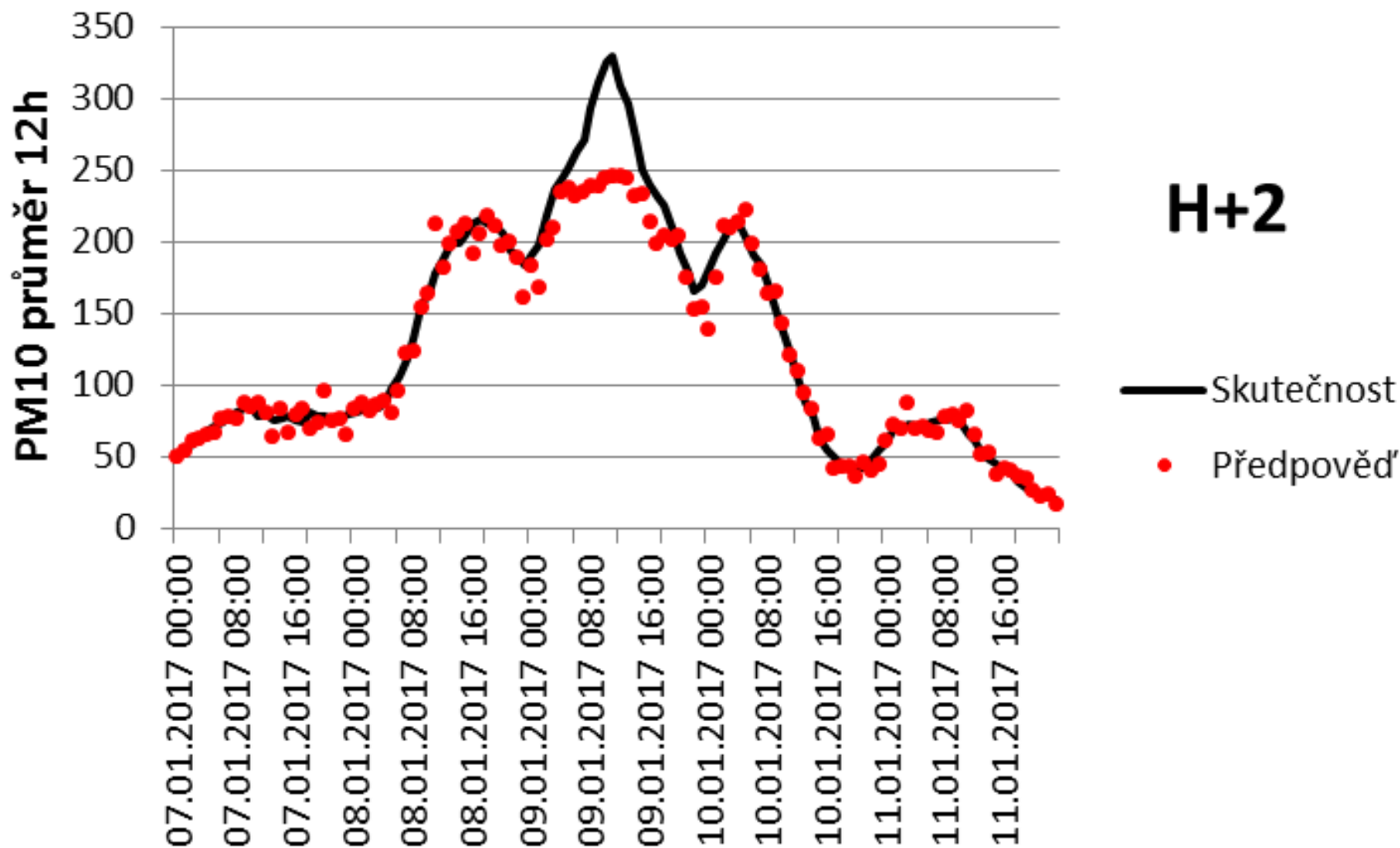
- RMS roste s horizontem předpovědi
- CORR naopak klesá
- Výsledky pro sezónu 2016/17, jejíž data ANN vůbec nezná, jsou horší než pro sezónu 2015/16



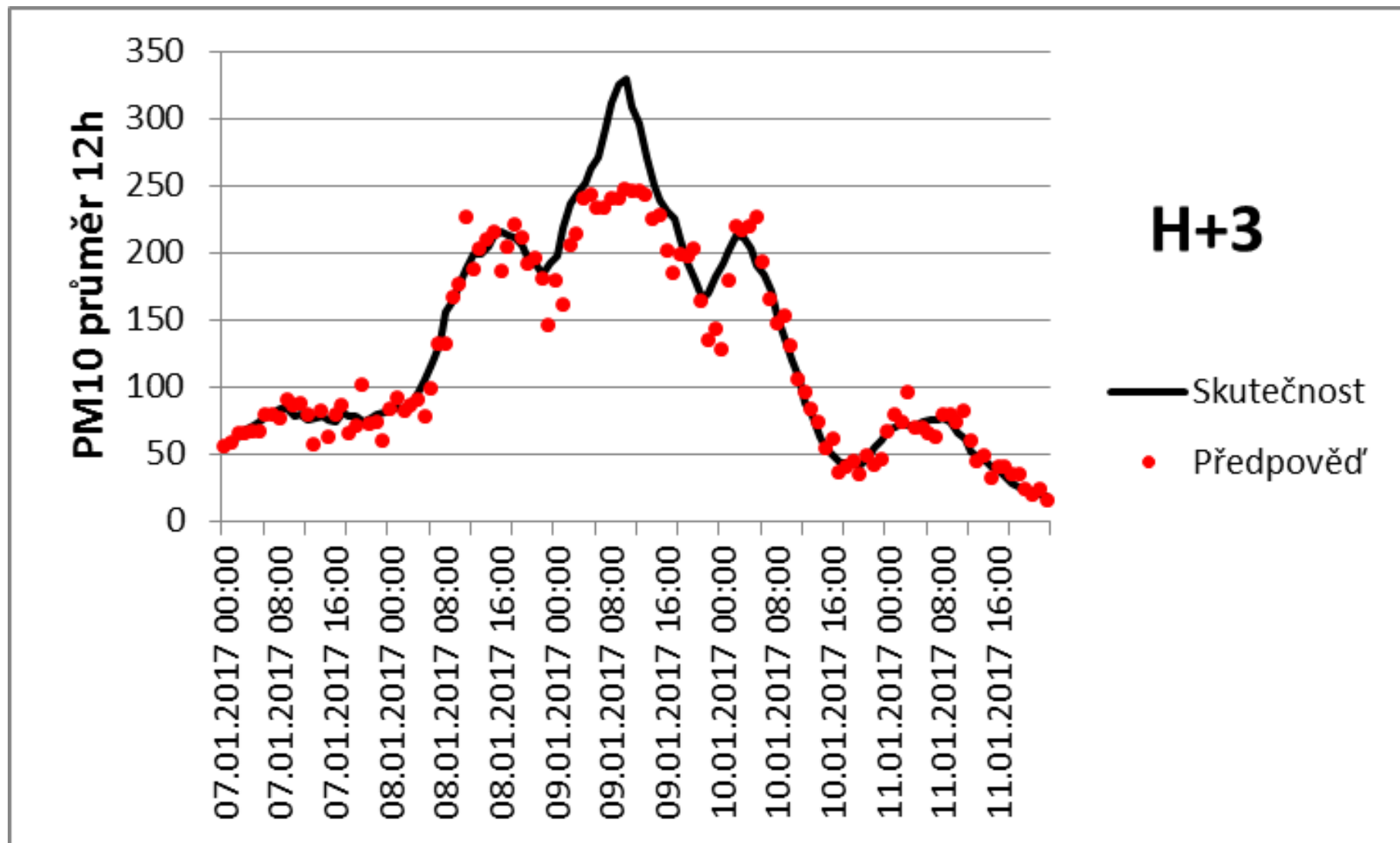
# Smogová situace 2017, úspěšnost predikce, O.-Fifejdy



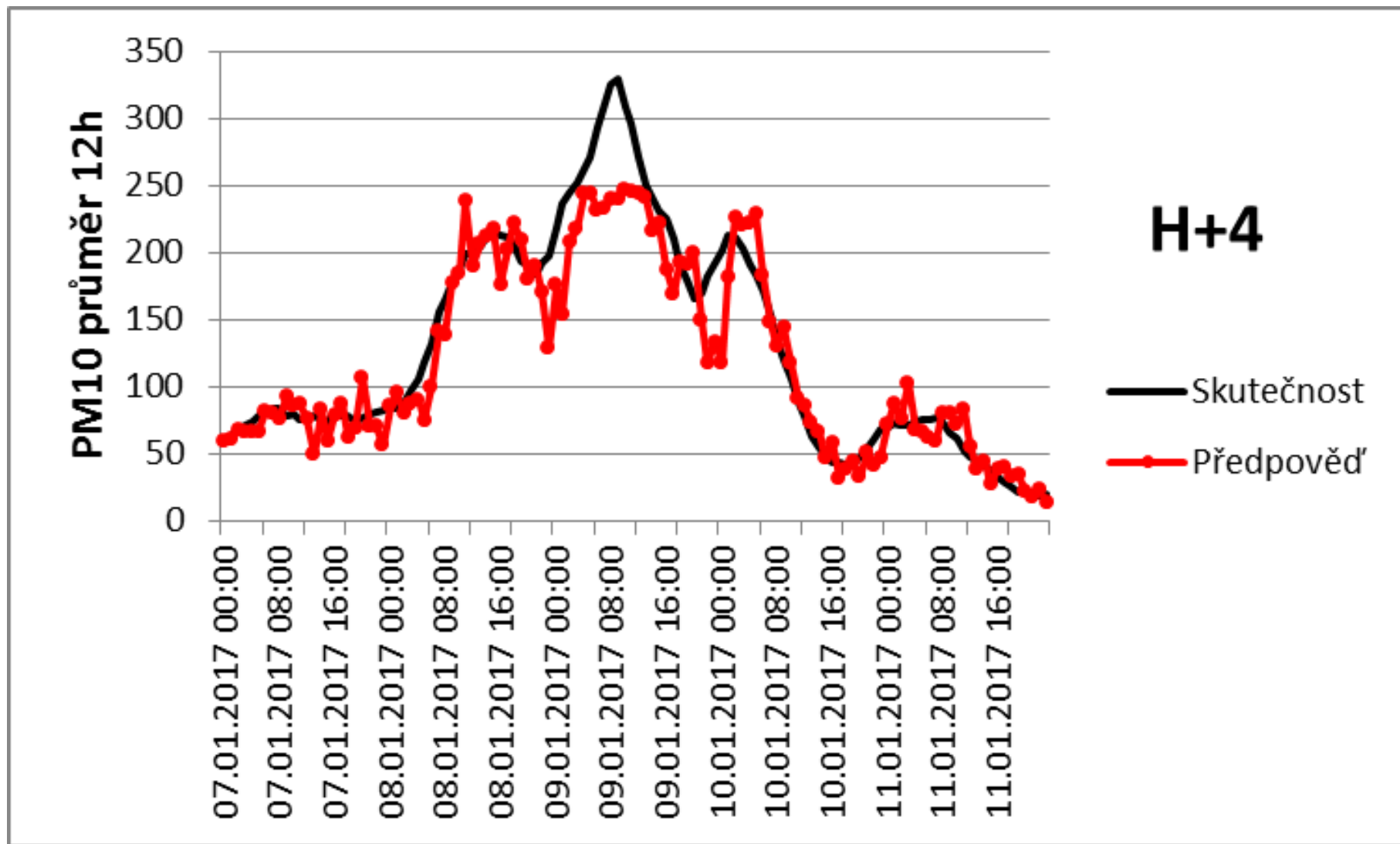
# Smogová situace 2017, úspěšnost predikce, O.-Fifejdy



# Smogová situace 2017, úspěšnost predikce, O.-Fifejdy

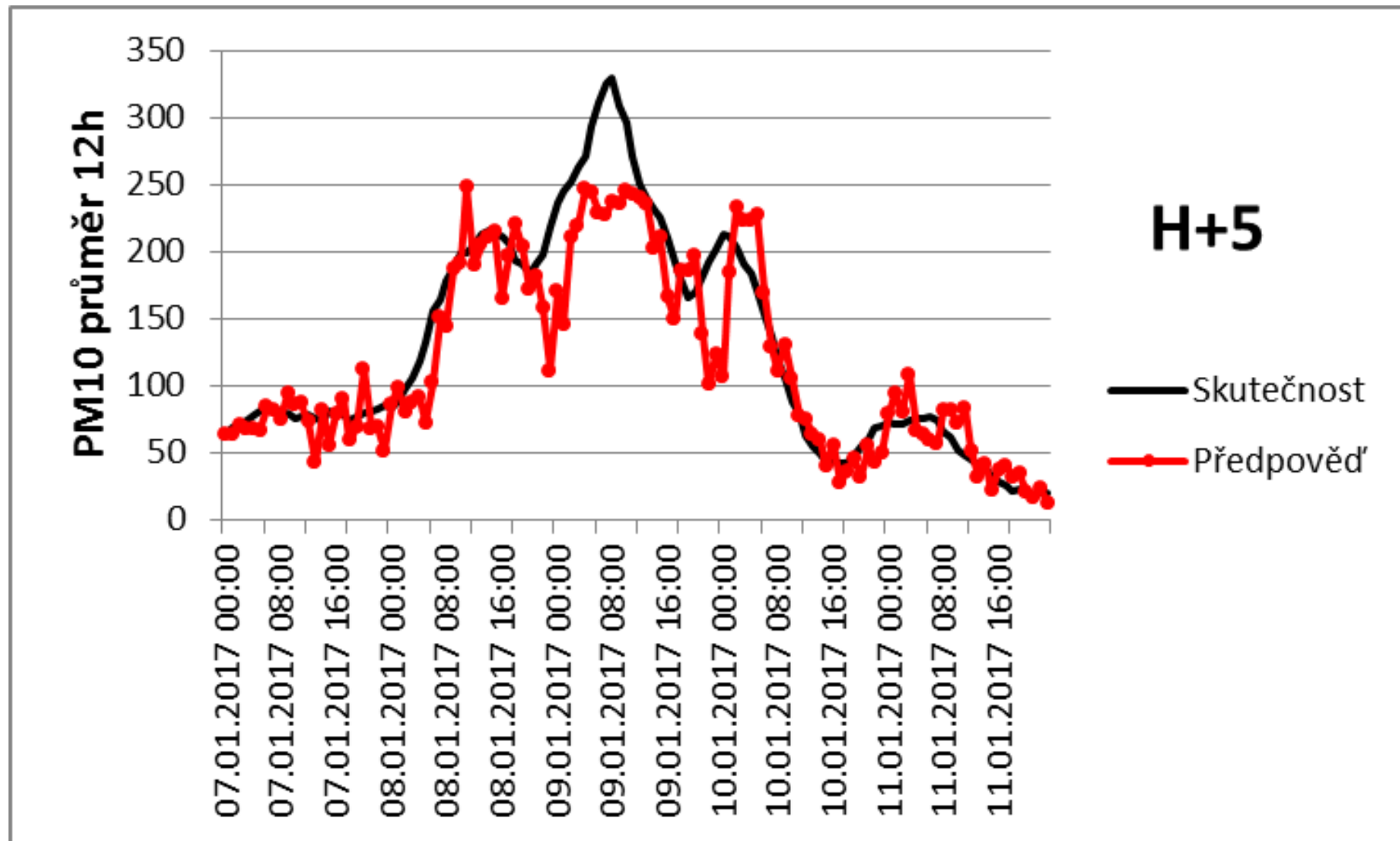


# Smogová situace 2017, úspěšnost predikce, O.-Fifejdy

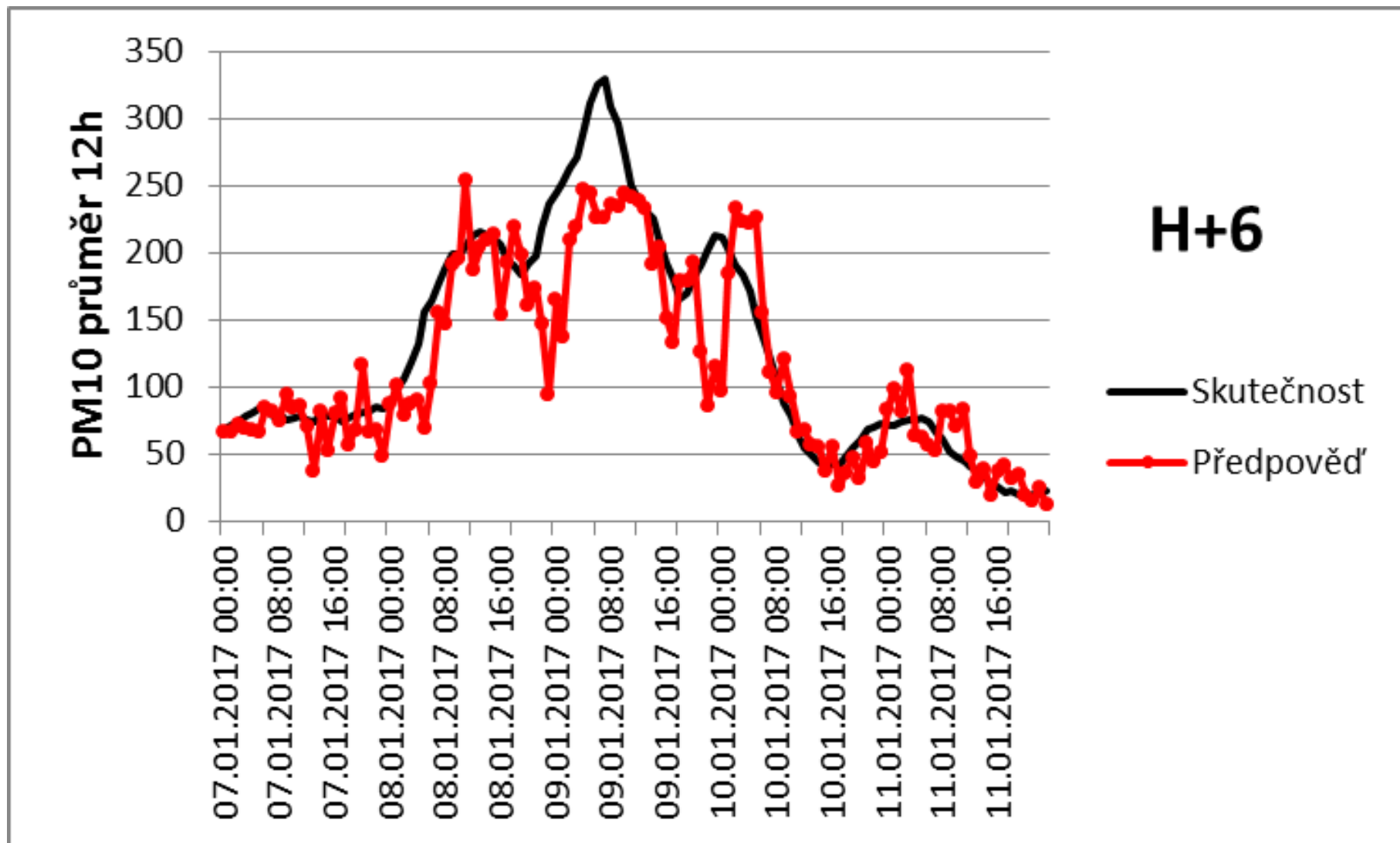




# Smogová situace 2017, úspěšnost predikce, O.-Fifejdy



# Smogová situace 2017, úspěšnost predikce, O.-Fifejdy



## Závěry, možné směry pokračování

- První ověření možností regresních ANNs pro krátkodobou předpověď klouzavých 12h průměrů PM10 vypadá překvapivě nadějně
- Použitelné pro potřeby SVRS?
- Bude-li „společenská objednávka“ ...
- Pro všechny stanice natrénovat ANNs na větším souboru dat – postihnout více variant a kombinací, často aktualizovat
- Zapojit další proměnné – vítr, teplotu, stabilitu ...
- Zkusit kombinaci s metodami vyhlazování časových řad a regresními stromy
- Skombinovat s výstupy NWP a rozptylových modelů

Děkuji, že jste ten  
neurák vydrželi až do  
konce!

