

Predikce volebních výsledků z průběžného sčítání

Jan Tuček a Daniel Prokop



O co jde v predikci

Exit-polls

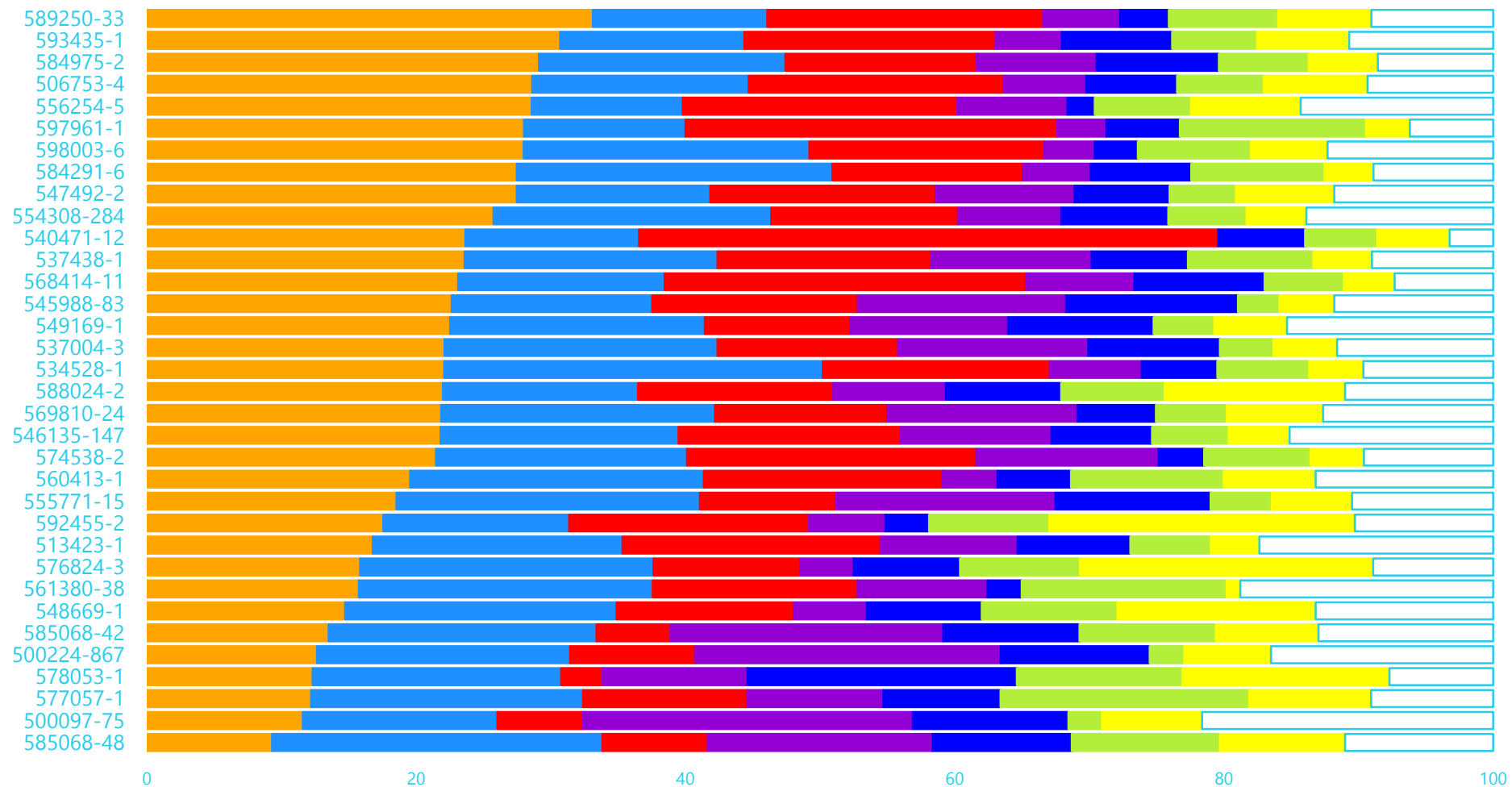
- publikace 14:00
- průměrné odchylky cca 1 p.b. a maximální 2,5 p.b.
- 15-25 tisíc respondentů
- cena v řádech milionů korun

Predikce ze sčítání

- model finálního výsledku z průběžného sčítání
- publikace prvních výsledků 14:30-15:00
- cíle
- přesnost do 1,5 p.b. na stranu
- zpřesnění sčítání: redukce odchylky na cca 1/3
- průběžnost: ideálně real-time
- více možných metod (AUS, FR, CZ)

Ukázka výsledků v jednotlivých VO (2013)

Kód obce-okrsek, cca 14.800 okrsků



ANO



TOP 09

ODS



Metodika STEM/MARK (kolektiv autorů😊)

Výsledky z roku 2013

- agregace okrsků do shluků (kontrola na výsledky 2010)
- zjištění velikosti shluků

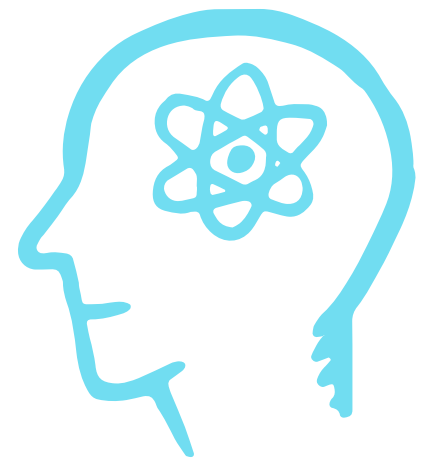


Průběžné výsledky 2017

- zařazení okrsků do shluků (dle 2013)
- převážení podle očekávané velikosti shluku (dle 2013)



Model



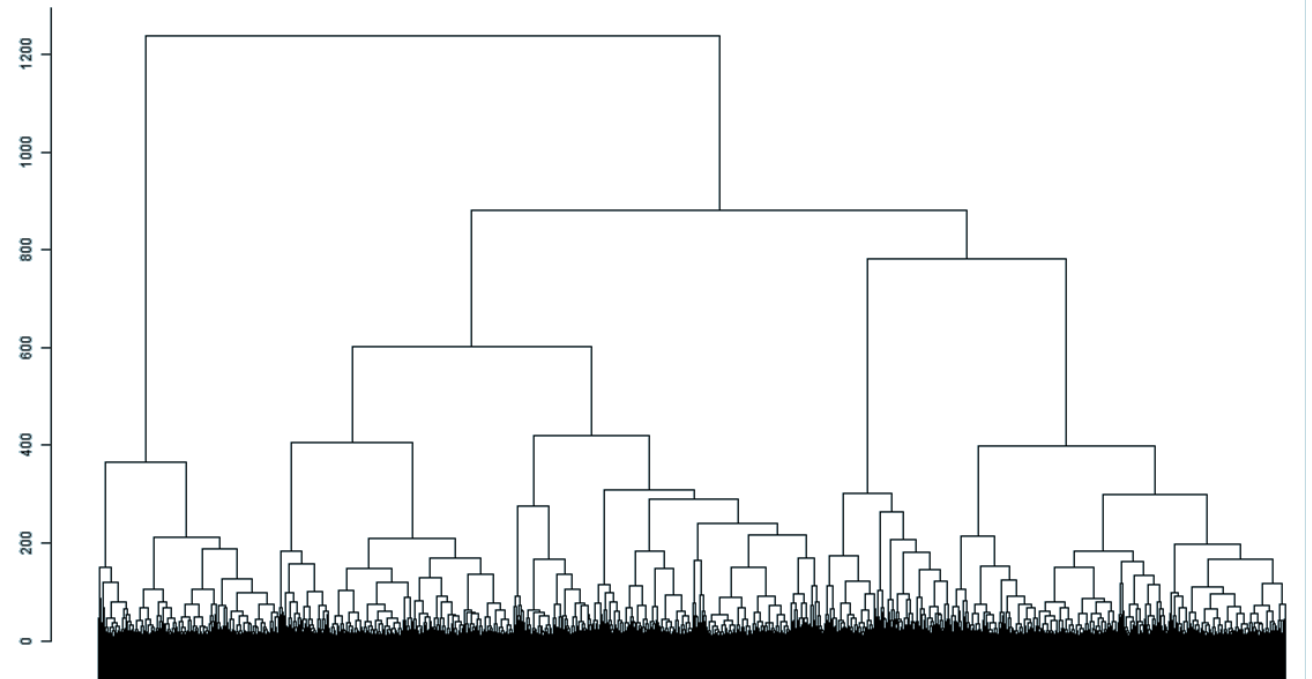
Shluková analýza

Strom podobnosti okrsků

- rozdělit co nejpodrobněji
- i nejmenší cluster musí být dostatečně velký

Vhodný počet skupin = 13

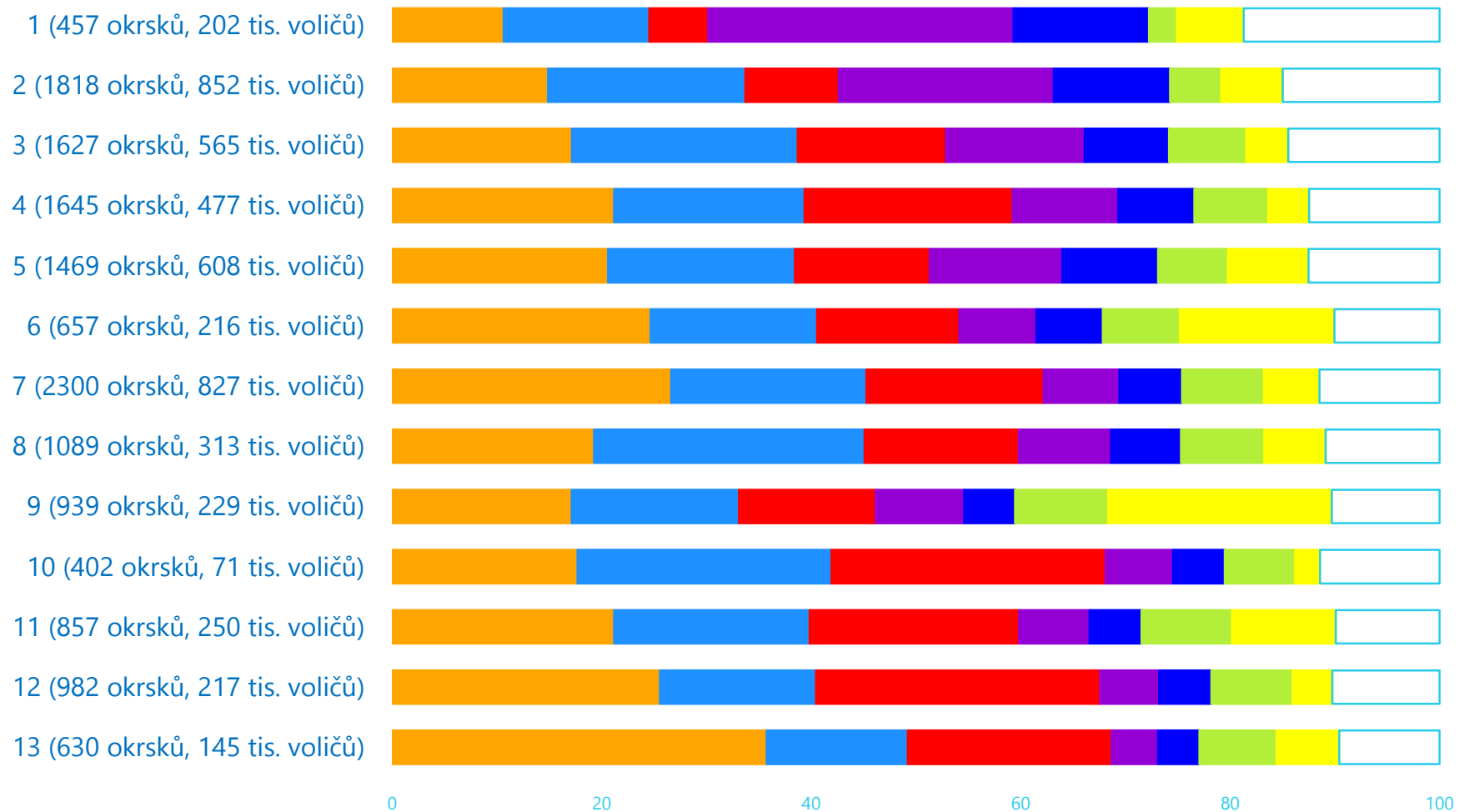
- $k = 9$ velikost nejmenšího = 400
- $k = 13$ velikost nejmenšího = 400
- $k = 14$ velikost nejmenšího = 105



14 872 okrsků

Typologie volebních okrsků

Vážený průměr volebních výsledků 2013



ANO



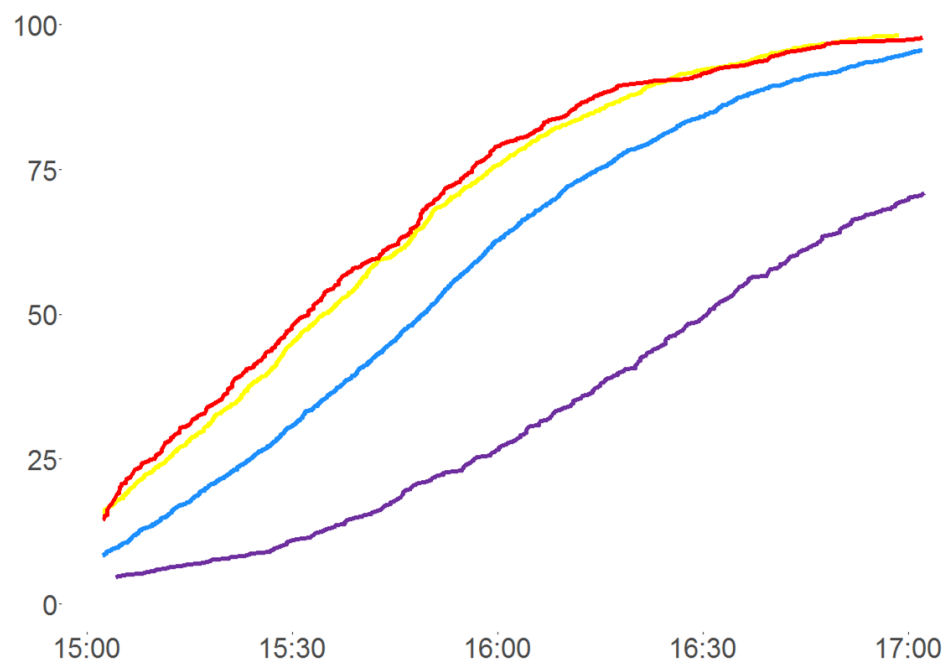
TOP 09

ODS



Rychlost sčítání skupin okrsků 2013

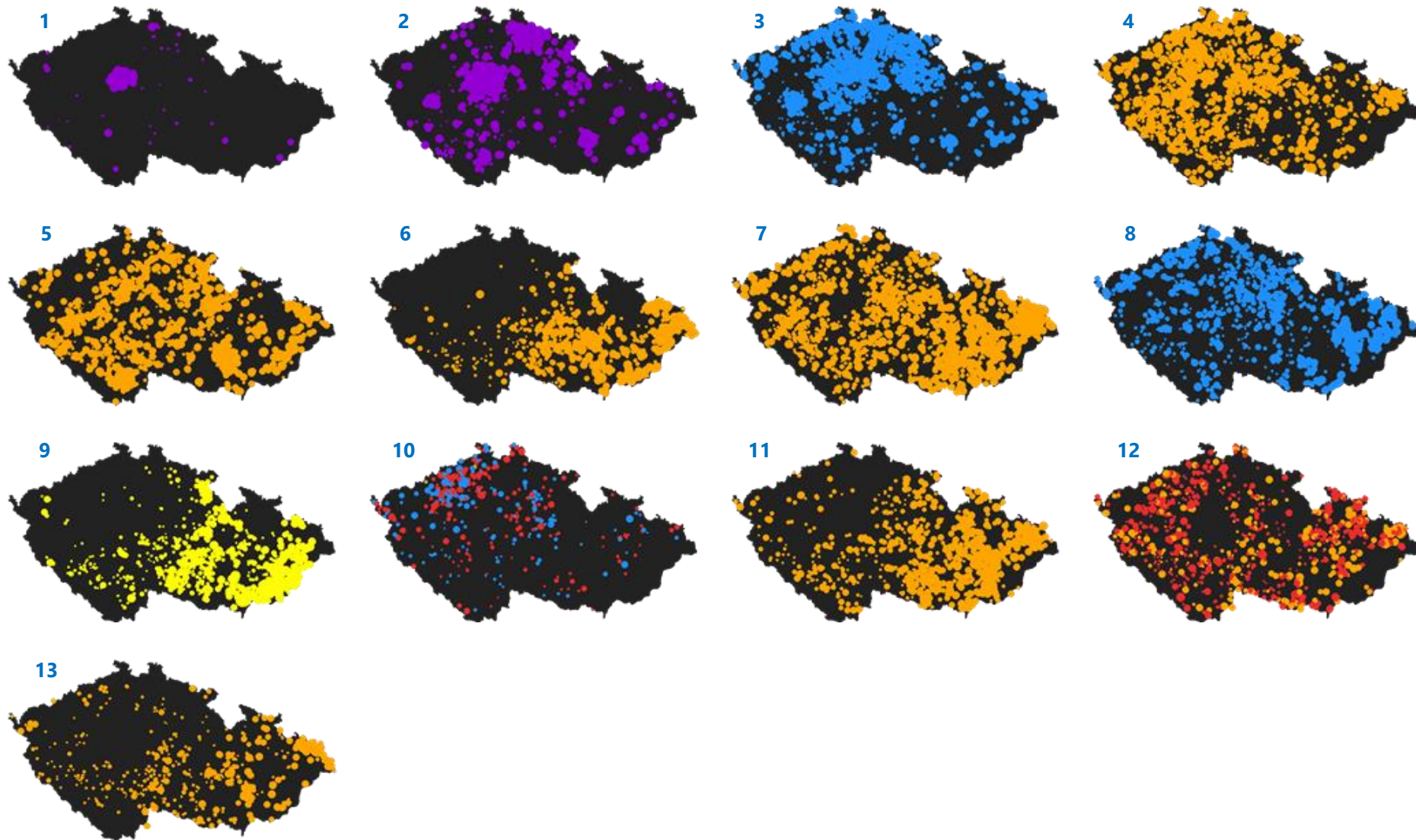
Shluk č.1 se sčítá výrazně pomaleji



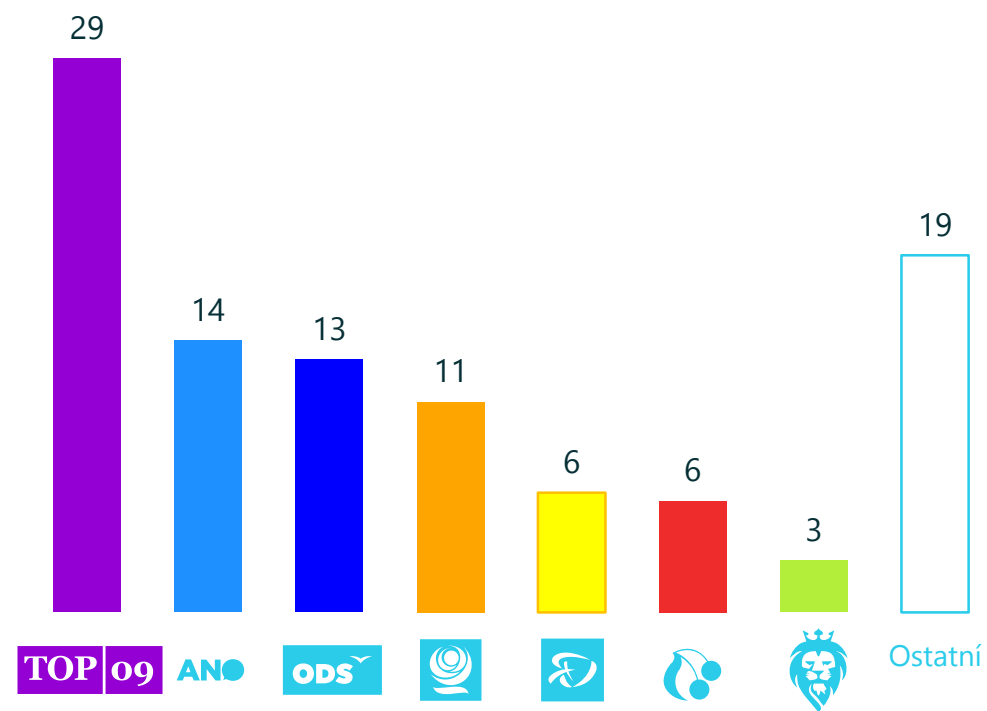
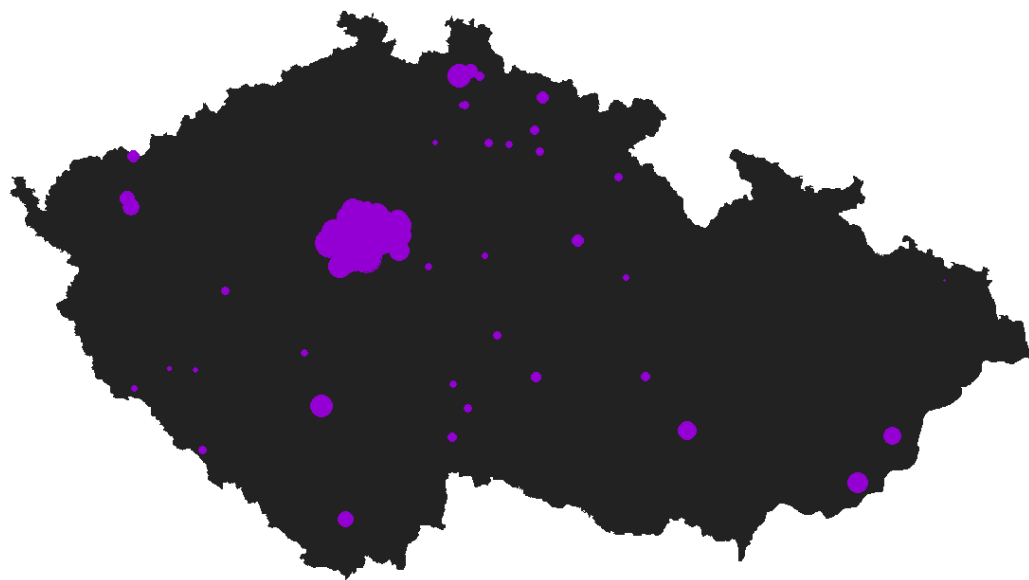
Shluk	Sečteno 10%	Sečteno 25%	Sečteno 50%	Sečteno 75%	Sečteno 90%	Sečteno 100%
1	15:28	15:56	16:30	17:06	17:35	19:54
2	15:17	15:38	16:04	16:33	17:01	19:38
3	15:04	15:23	15:49	16:14	16:42	19:02
4	14:59	15:16	15:38	16:04	16:31	19:10
5	15:07	15:25	15:47	16:10	16:33	19:18
6	15:00	15:17	15:39	16:04	16:29	18:48
7	15:05	15:21	15:42	16:04	16:28	19:02
8	14:58	15:12	15:34	16:00	16:26	18:40
9	14:55	15:11	15:34	15:58	16:23	18:31
10	14:59	15:09	15:32	15:56	16:21	19:43
11	15:02	15:18	15:40	16:05	16:30	17:49
12	14:57	15:11	15:30	15:55	16:19	17:39
13	14:54	15:08	15:27	15:51	16:11	19:03

Geografické rozložení okrsků podle skupin

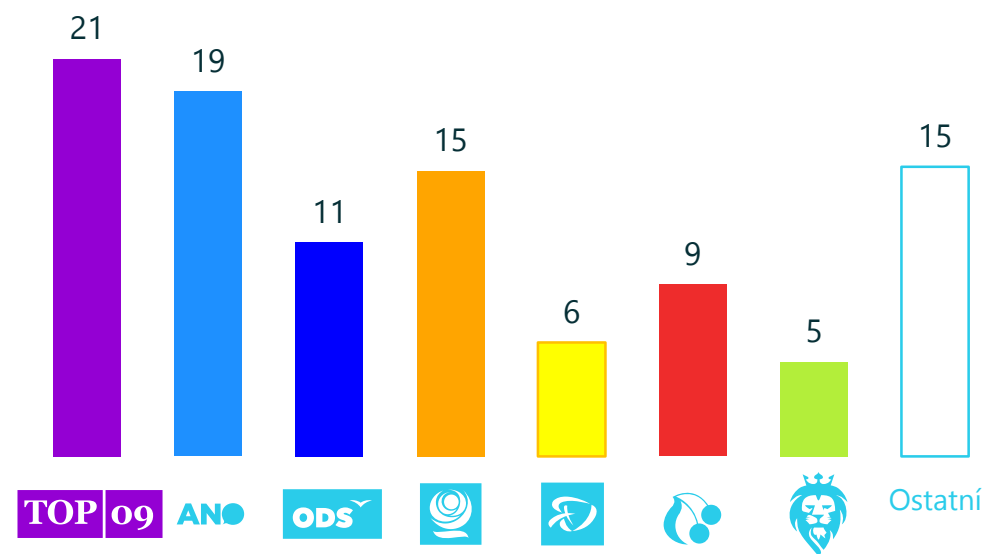
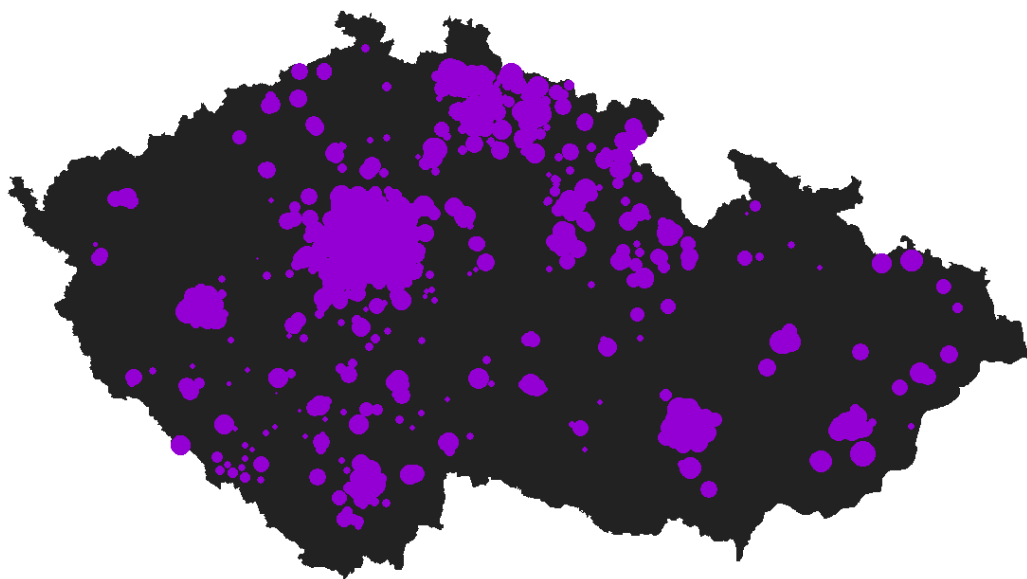
Barva odpovídá vítězi voleb v daném okrsku



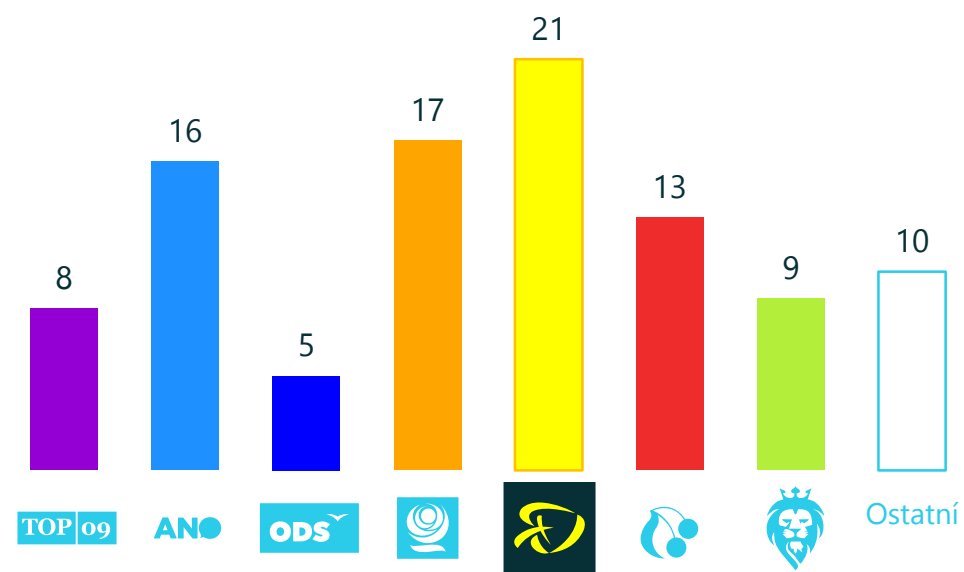
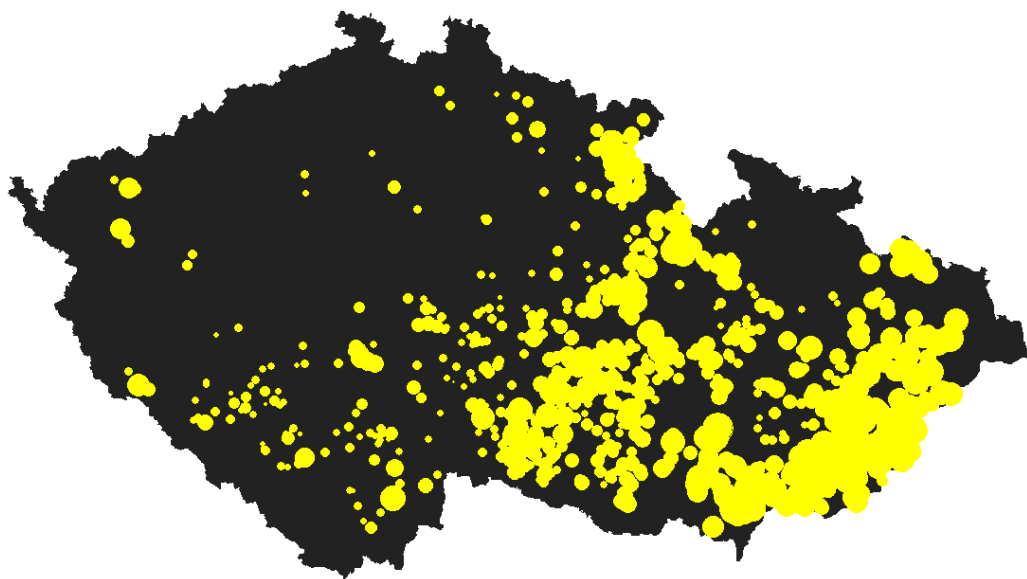
Shluk 1: Praha a velká města



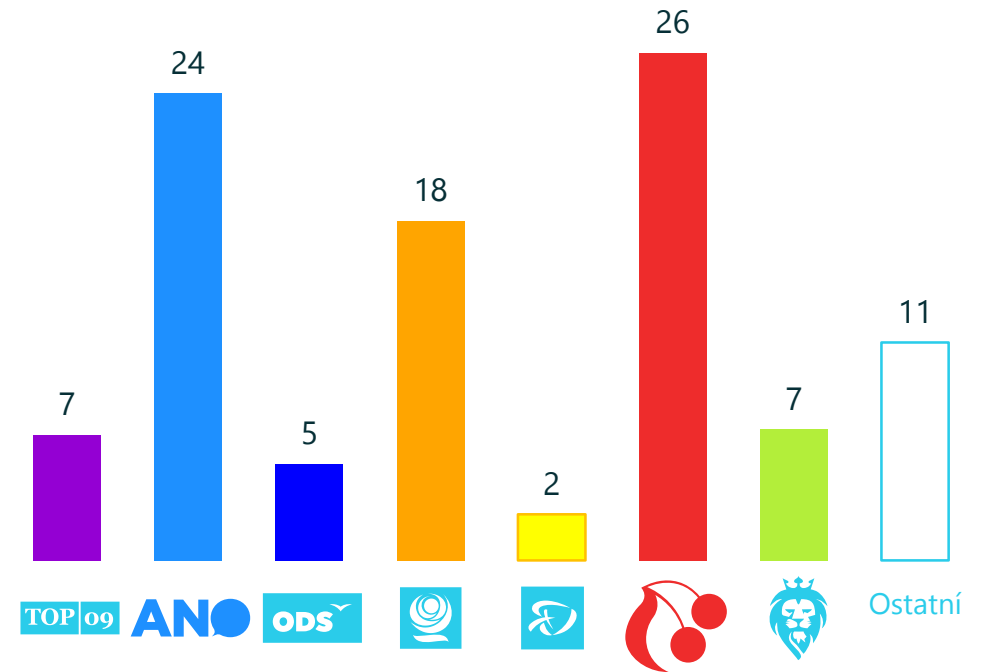
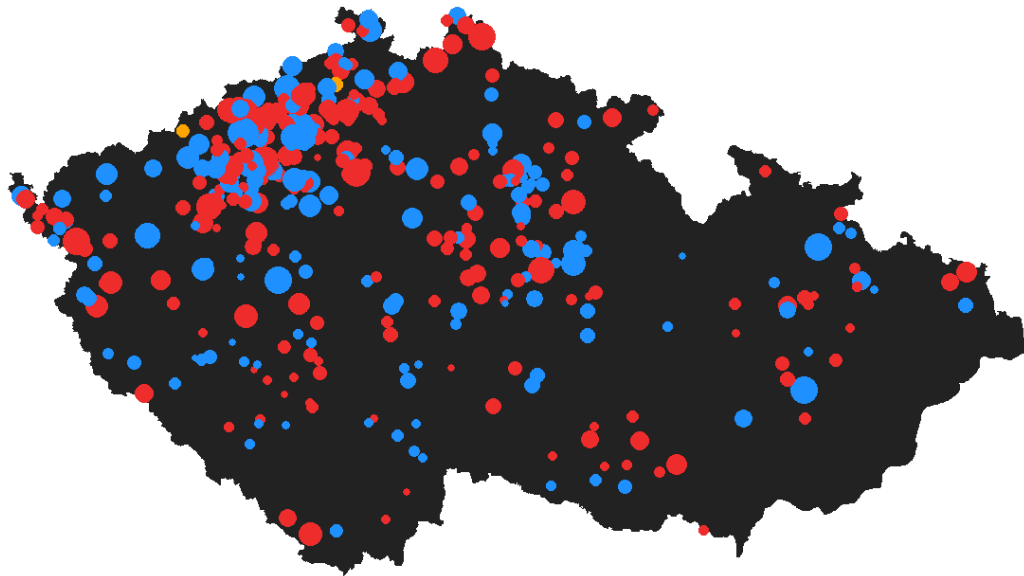
Shluk 2: Velká města a satelity



Shluk 9: Moravské vesničky



Shluk 10: Ústecko a podobné



Metodika MEDIAN (autor: Tomáš Hovorka)

Odhad výsledku strany v okrsku

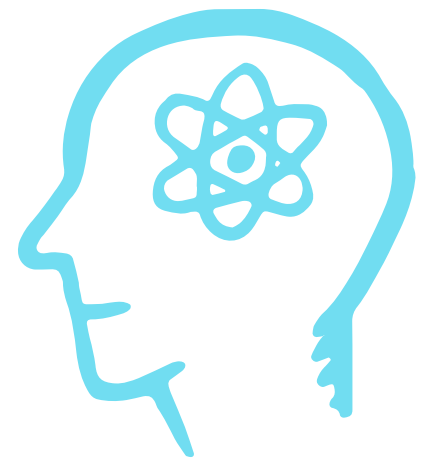
- odhadovaná jednotka = STRANA x OKRSEK
- odhad u každé strany (P) ze všech sečtených okrsků (N) a do všech nesečtených okrsků (M)
 - *Příklad: 1000 sečtených okrsků. Každá ze 13 stran má v každém z 13 000 nesečtených okrsku 1000 separátních odhadů výsledku.*

Agregování odhadů

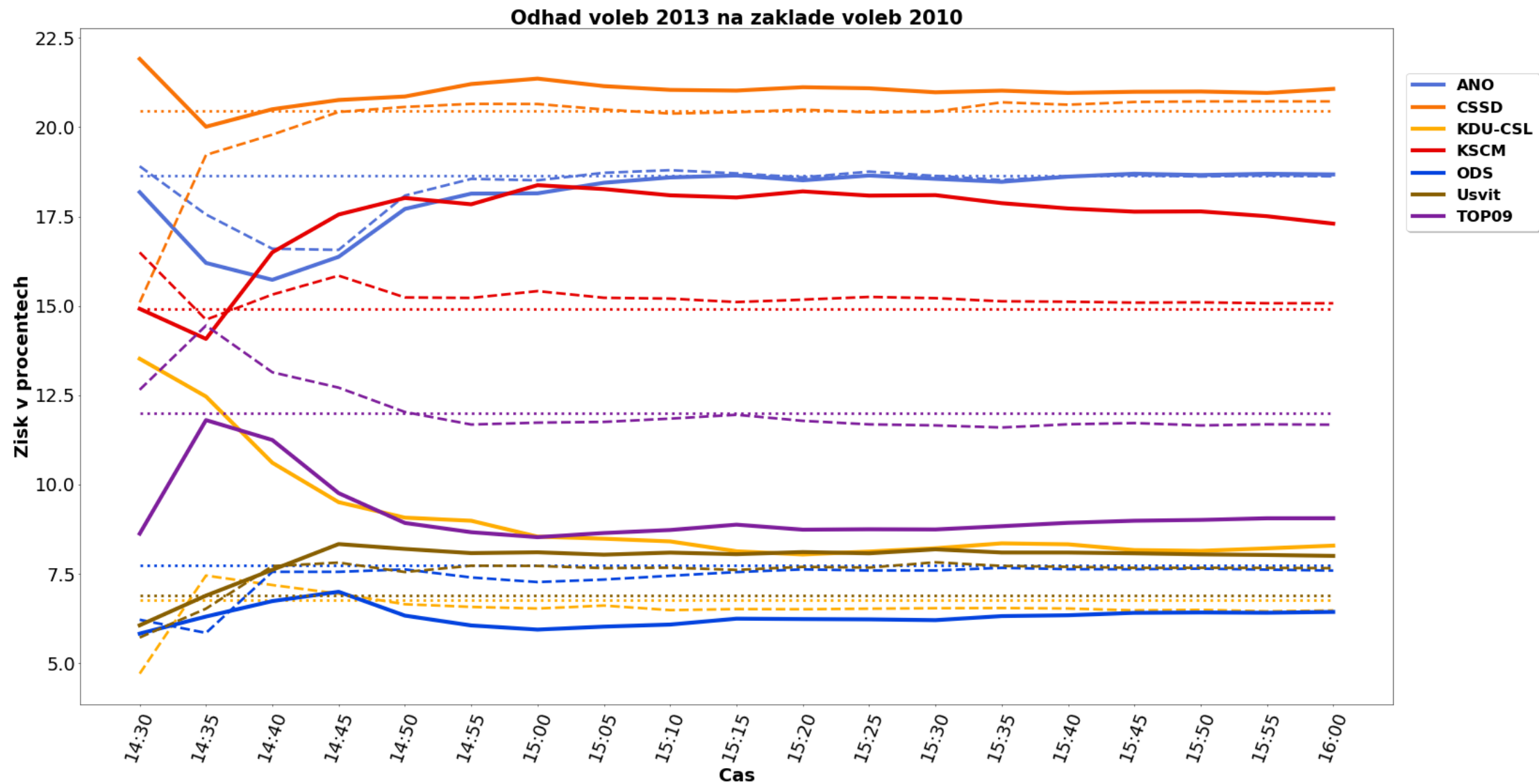
- První agregace
 - agregace (N) odhadů pro danou stranu (P) v okrsku (M)
 - vážená agregace – zohledněna např. korelace okrsků v minulých volbách
 - spojitá metoda (místo dichotomizace segmentací)
- Druhá agregace
 - sečtení výsledků v okrscích pro celou ČR

Naučení modelu

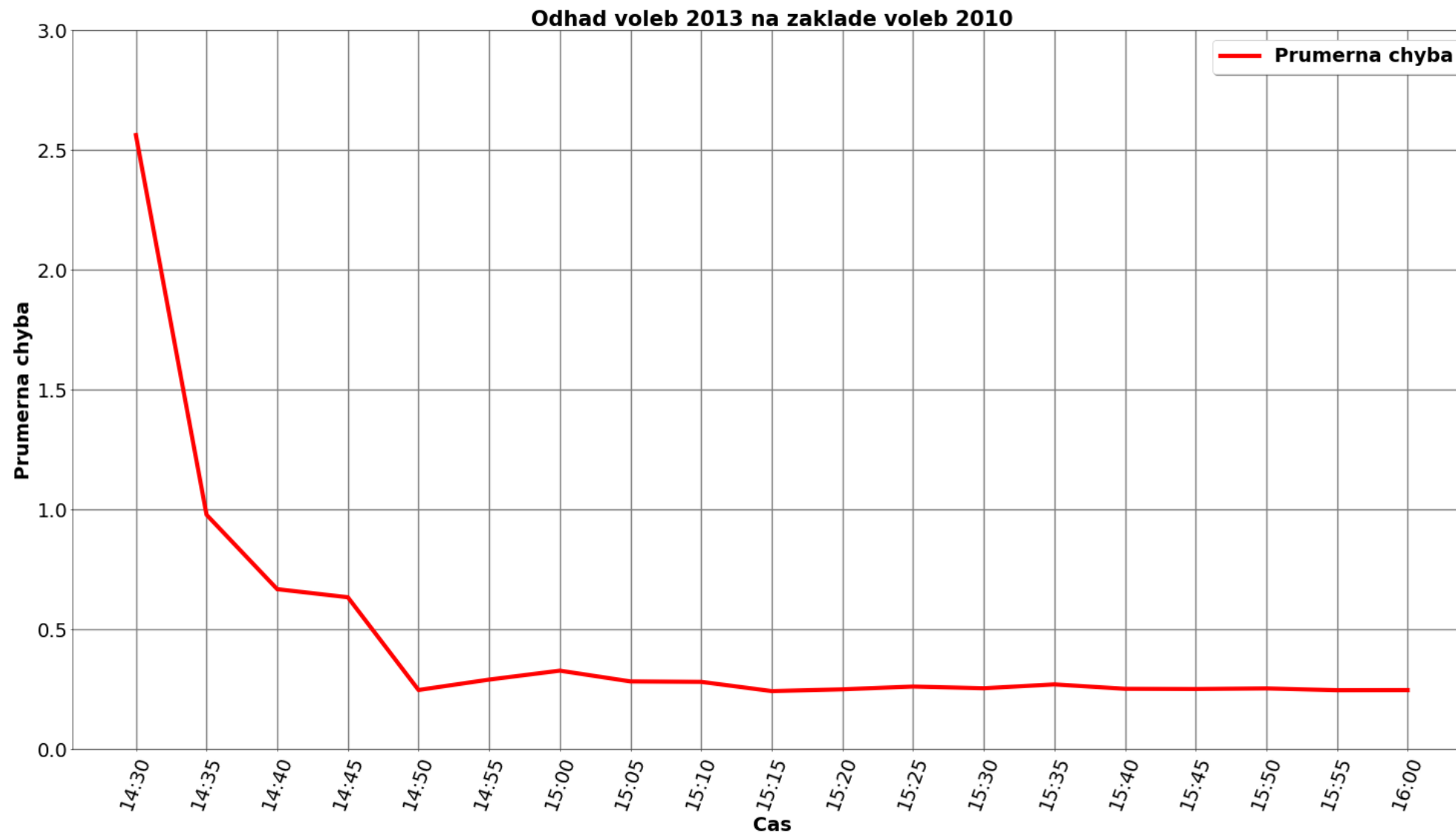
- Vychází ze srovnání dvou voleb stejného typu
- Možná verifikace v posledních 20 letech voleb



Predikce a průběžné sčítání v čase – volby 2013 na základě 2010



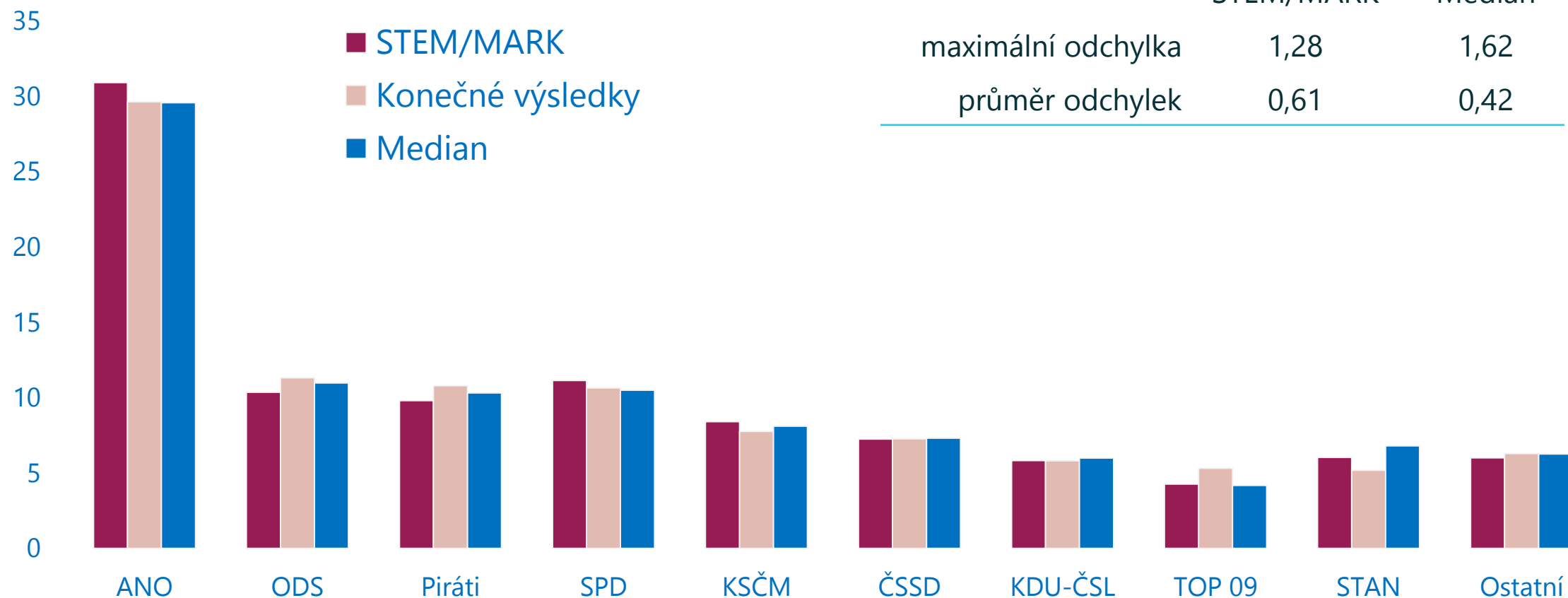
Snižování průměrné odchylky – volby 2013 na základě 2010



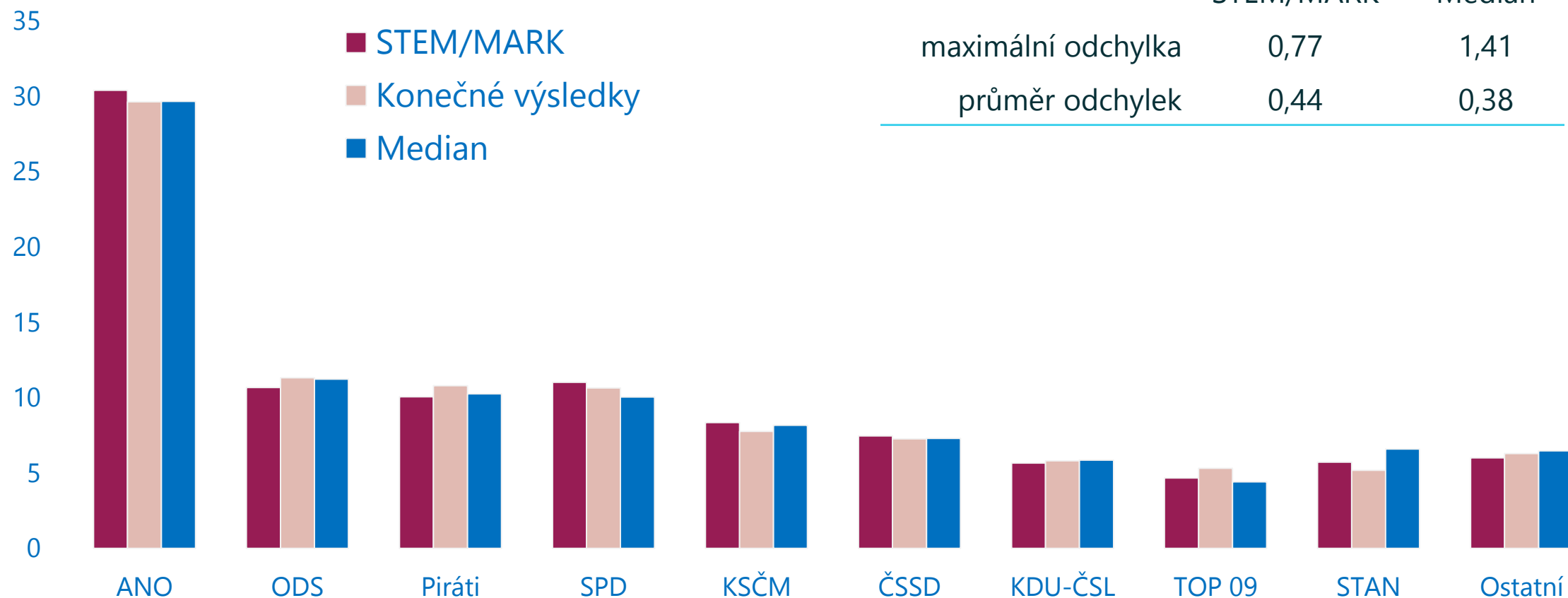
Výsledky MEDIAN od roku 2010

Jaké volby z jakých odhadujeme	Průměrná chyba v 14:40	Průměrná chyba v 14:50	Maximální chyba v 14:40	Maximální chyba v 14:50
2010 z 2006	1,67	0,68	3,62	1,87
2013 z 2006	0,85	0,94	2,07	2,04
2013 z 2010	0,67	0,25	2,05	0,66
2017 z 2006	1,10	0,74	3,79	1,06
2017 z 2010	0,67	0,52	2,64	1,04
2017 z 2013	0,47	0,44	1,09	1,18
	0,95	0,61	2,50	1,38

15:00 - Median lepší v průměru odchylek, STEM/MARK v maximální

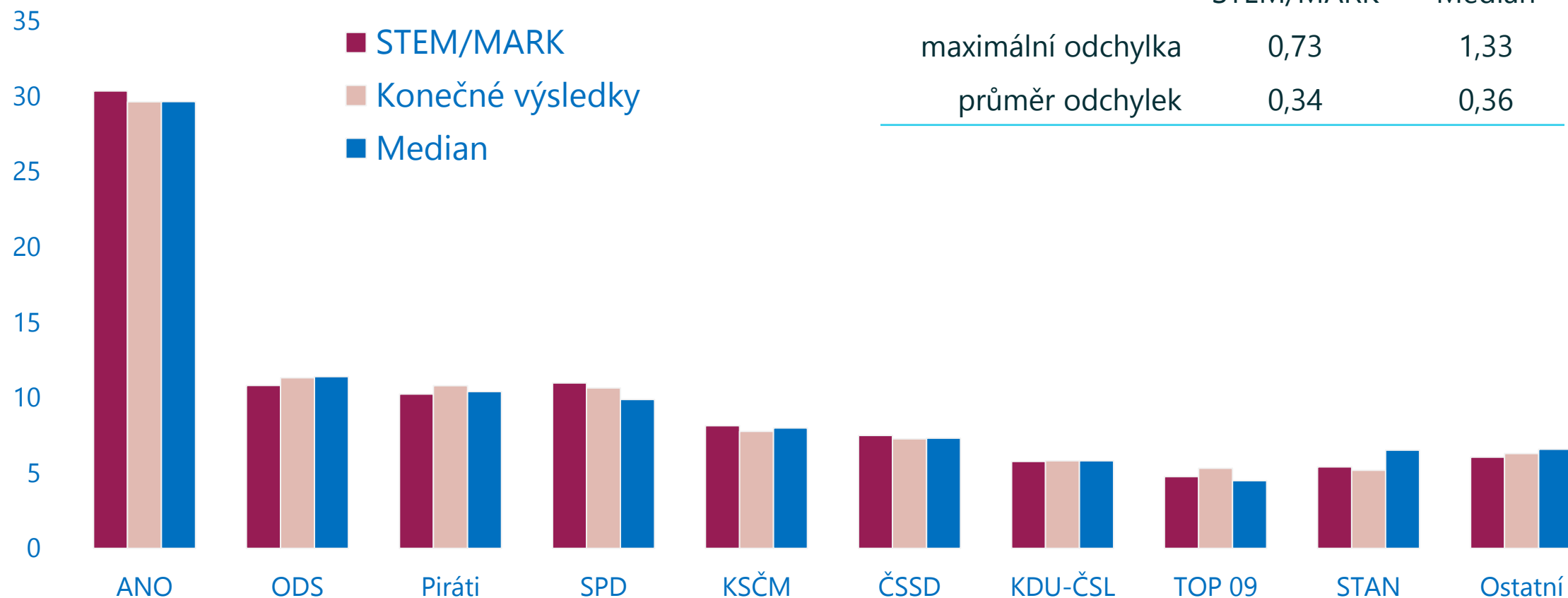


15:15 - Predikce se zpřesňují...



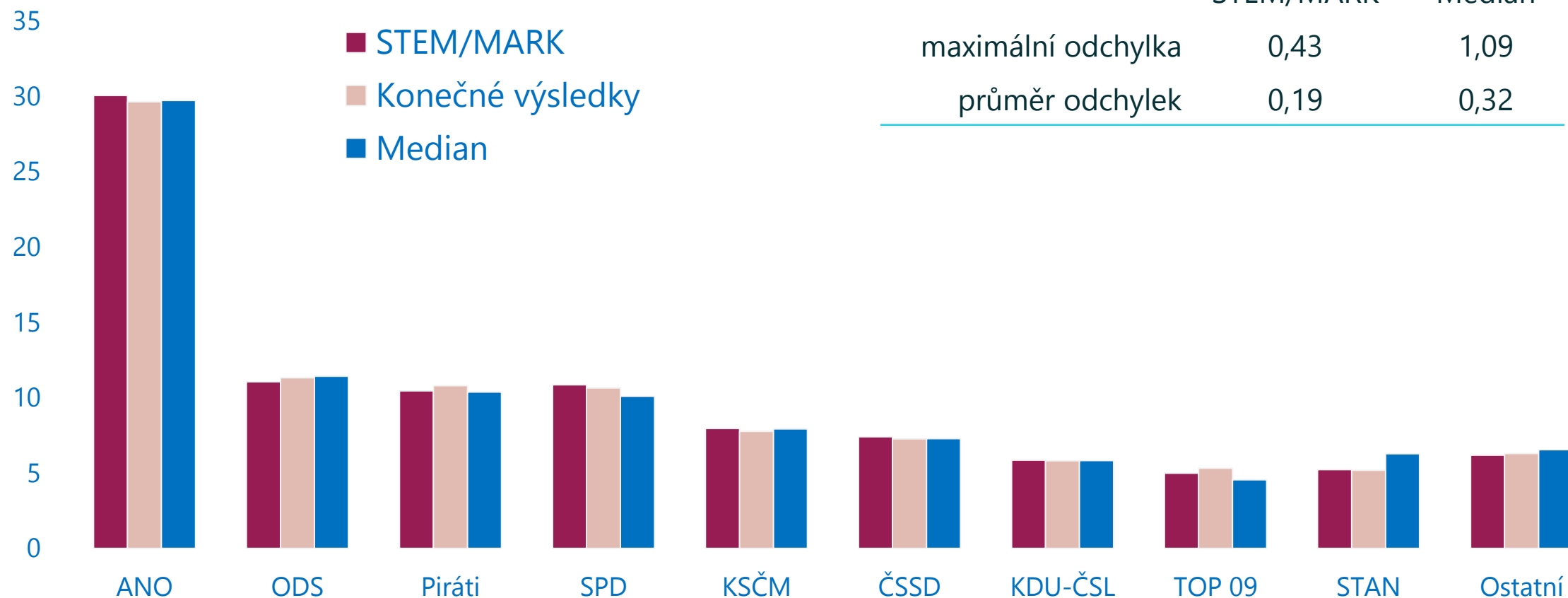
	STEM/MARK	Median
maximální odchylka	0,77	1,41
průměr odchylek	0,44	0,38

15:30 - ... v průměru nastejno



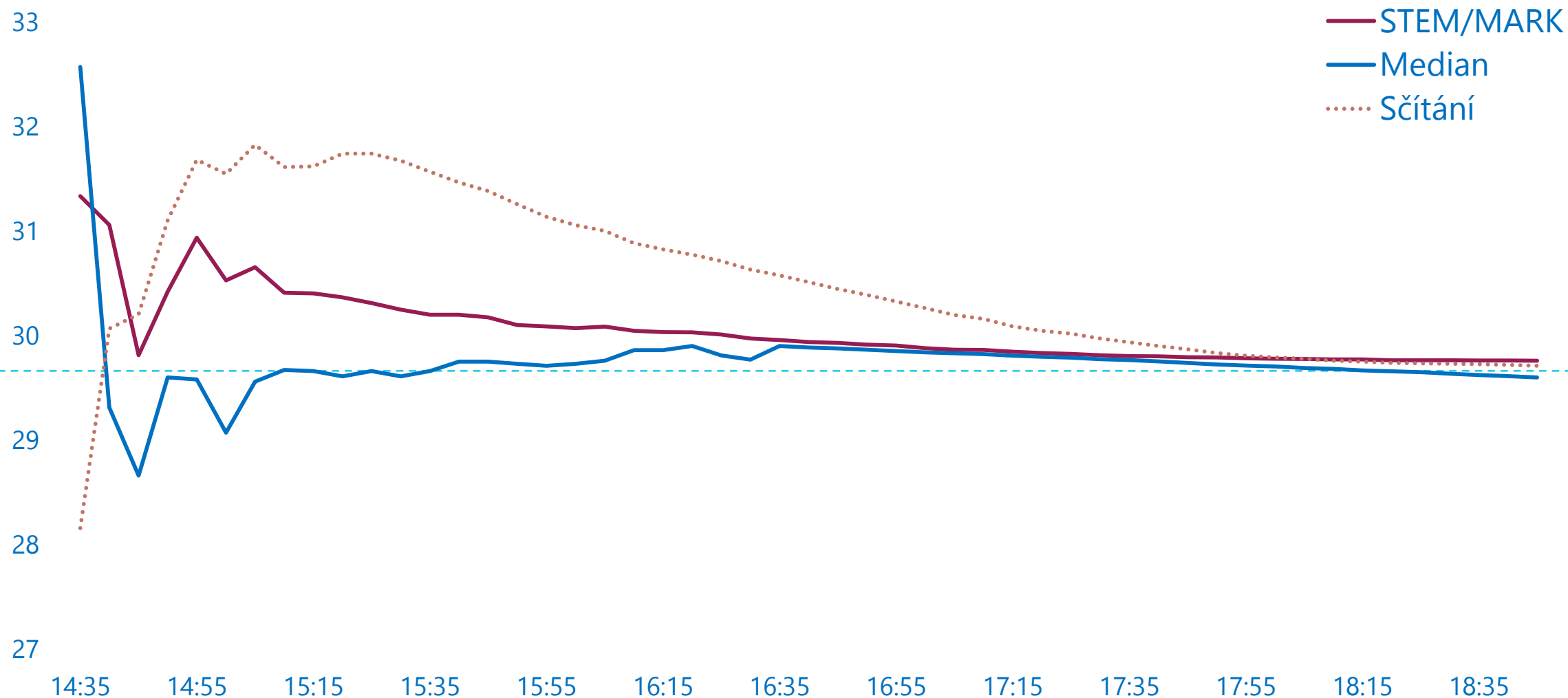
	STEM/MARK	Median
maximální odchylka	0,73	1,33
průměr odchylek	0,34	0,36

16:00 - STEM/MARK přesnější v průměru i max. odchylce

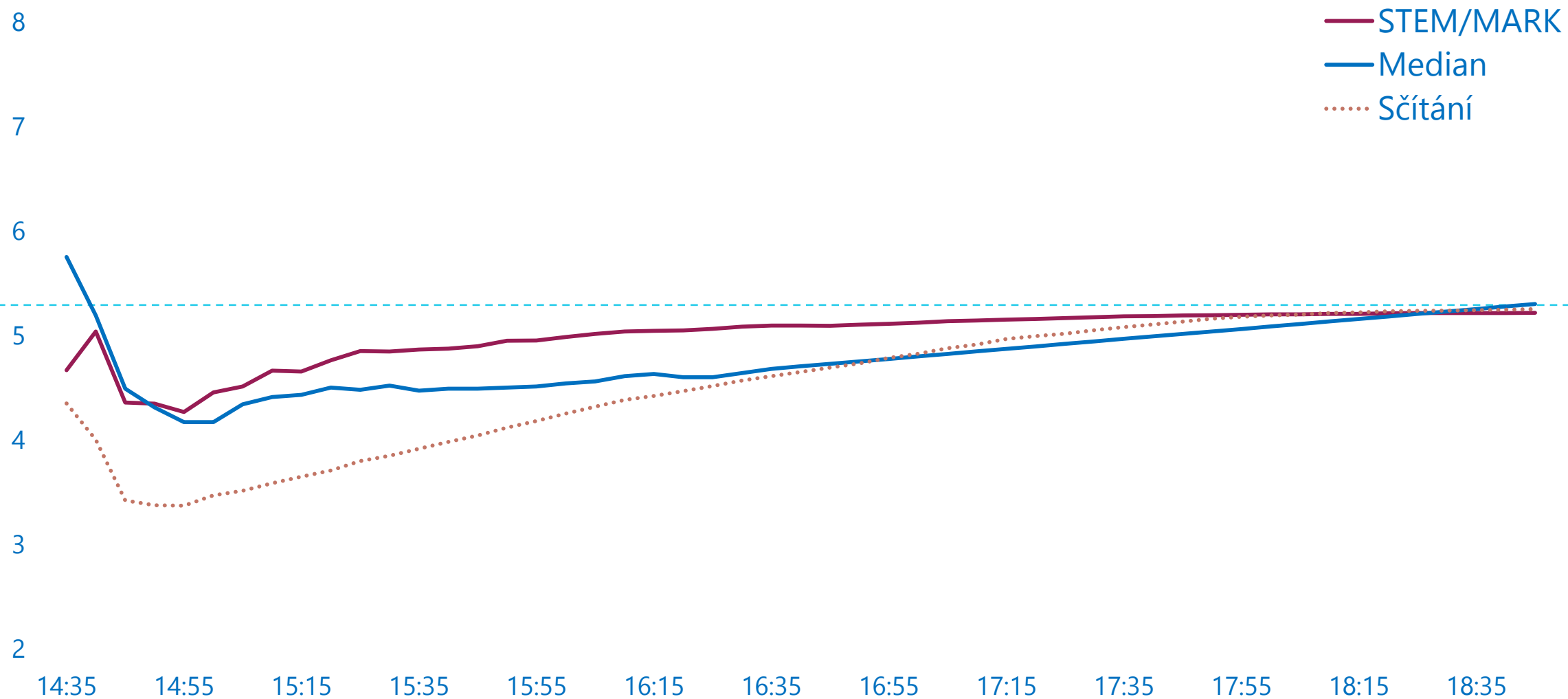


	STEM/MARK	Median
maximální odchylka	0,43	1,09
průměr odchylek	0,19	0,32

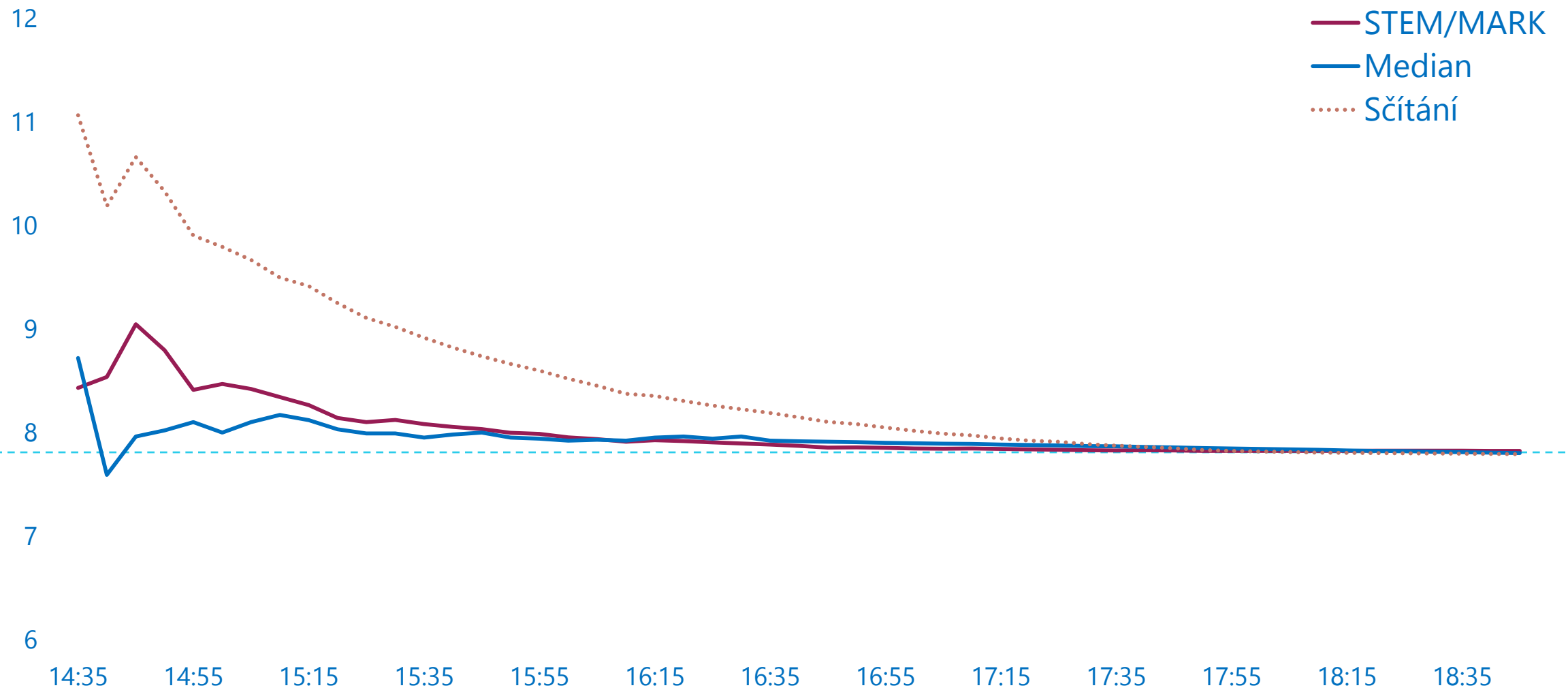
Vývoj modelu a průběžných součtů: ANO



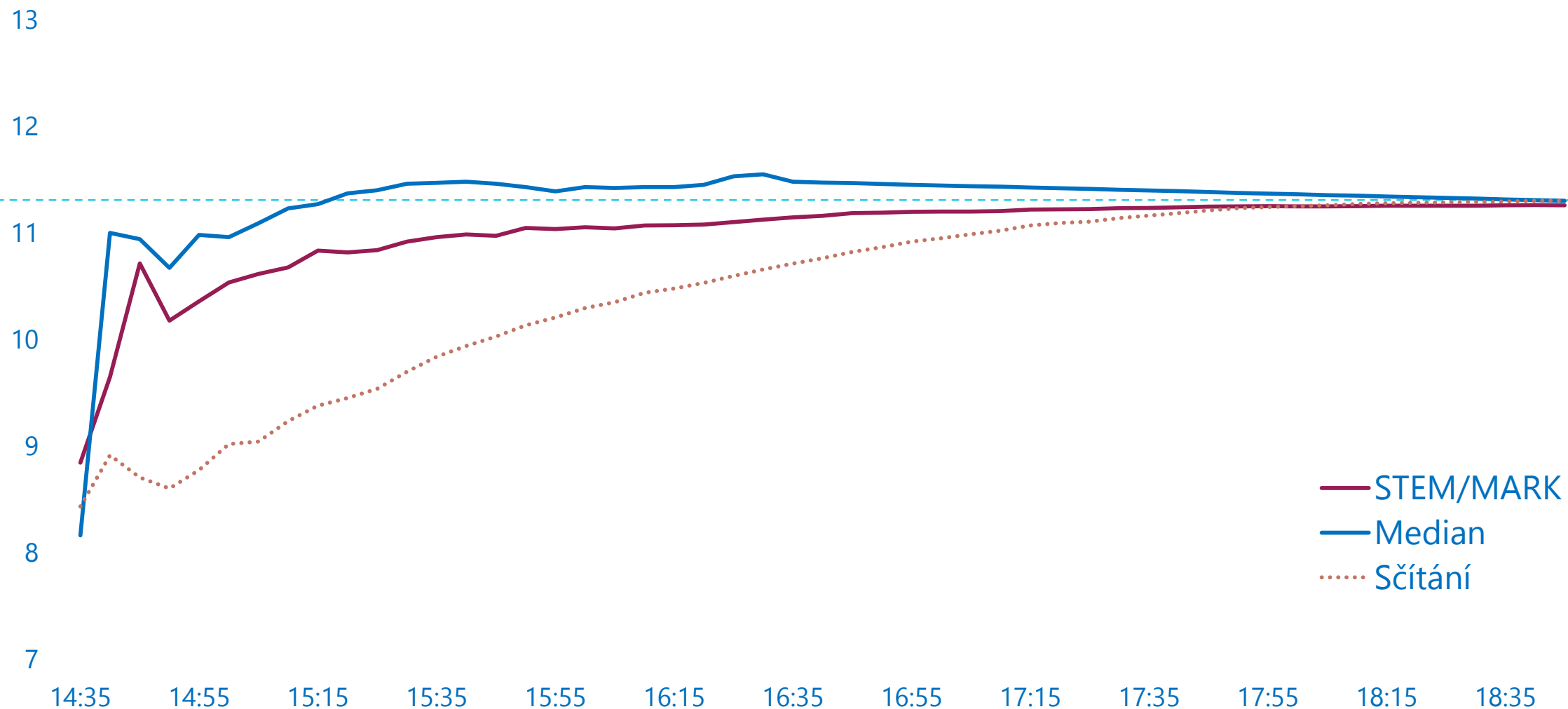
Vývoj modelu a průběžných součtů: TOP 09



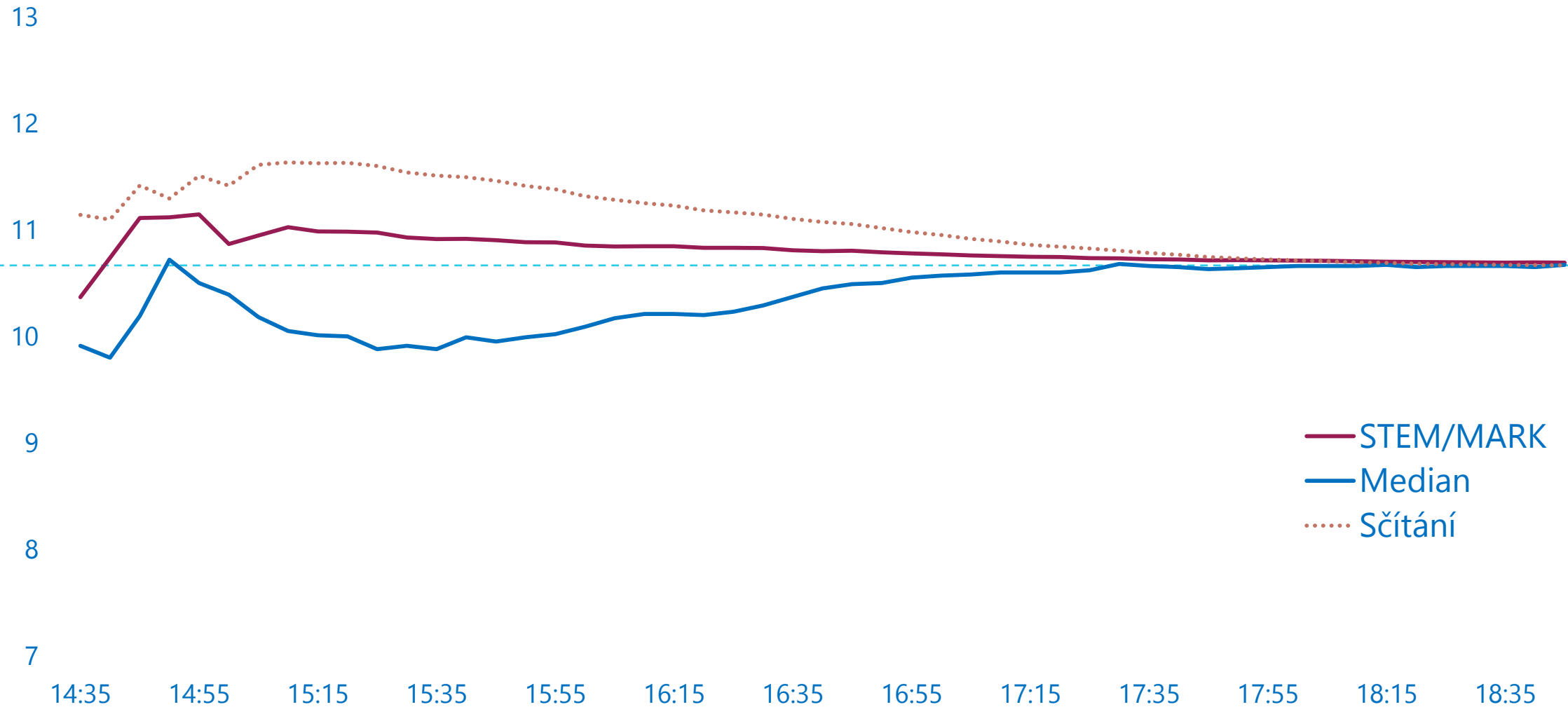
Vývoj modelu a průběžných součtů: KSČM



Vývoj modelu a průběžných součtů: ODS



Vývoj modelu a průběžných součtů: SPD



Srovnání výsledků STEM/MARK a Median

Median

- lepší průměrná chyba na stranu do 15:45
- lepší pořadí stran v prvních odhadech
- přesnější predikce vítěze (ANO pod 30 %)

STEM/MARK

- nižší maximální chyba u odhadů celkově
- lepší přibližování volbám: průměrná chyba na stranu od 16:00
- přesnější predikce TOP09 a STAN

Pro a proti metodik

Median

- + robustnější: vážený odhad místo oddělených odhadů v rámci v segmentu z méně okrsků
- + rychlejší: vyšší přesnost v začátku
- + neměnná a testovatelná napříč volbami minulostí

- více vadil vznik nových stran
- citlivá na proměnu regionální struktury podpory strany (posun zisku TOP09 v Praha / mimo Prahu), mohlo déle udržovat nepřesnost
- potřebuje více voleb stejného typu na naučení

STEM/MARK

- + nevedí vznik nových stran
- + lepší přibližování volbám: ve chvíli nasycení všech segmentů přesnější
- + výpočetně jednodušší a laicky pochopitelnější

- pomalejší: potřebovala zástupce všech typů
- citlivá na zásah nereprezentativního okrsku v rámci pomaleji sčítaných typů
- citlivá na velkou proměnu typologie okrsků mezi volbami

obě: citlivé na velké změny volební účasti v různých typech okrsků