FAKULTA MECHATRONIKY,
INFORMATIKY A MEZIOBOROVÝCH
STUDIÍ TUL

Bakalářská práce

Predikce hodnot dat v energetice budov

Studijní program: B0613A140005 Informační technologie
Studijní obor: Aplikovaná informatika

Autor práce: **Jakub Šilhán**
Vedoucí práce: Ing. Jan Kraus, Ph.D.
Ústav mechatroniky a technické informatiky

Liberec 2024

FAKULTA MECHATRONIKY,
INFORMATIKY A MEZIOBOROVÝCH
STUDIÍ TUL

Zadání bakalářské práce

Predikce hodnot dat v energetice budov

Jméno a příjmení:	**Jakub Šilhán**
Osobní číslo:	M21000140
Studijní program:	B0613A140005 Informační technologie
Specializace:	Aplikovaná informatika
Zadávající katedra:	Ústav mechatroniky a technické informatiky
Akademický rok:	2023/2024

Zásady pro vypracování:

1. Proveďte důkladnou rešerši metod a nástrojů pro modelovaní a predikci hodnot veličin měřených v chytrých budovách.
2. Pro platformu .NET implementujte s využitím existujících knihoven vlastní experimentální aplikaci pro predikce na zadaných datech.
3. Prakticky ověřte správnou funkci aplikace a na vybraných konkrétních příkladech vhodnými metrikami změřte konkrétní parametry predikcí.
4. V závěru práce stručně a přehledně shrňte dosažené výsledky a diskutujte možnosti jejich dalšího uplatnění v praxi.

Rozsah grafických prací:	dle potřeby dokumentace
Rozsah pracovní zprávy:	30 až 40 stran
Forma zpracování práce:	tištěná/elektronická
Jazyk práce:	čeština

Seznam odborné literatury:

[1] MOLINA-SOLANA, Miguel et al. Data Science for Building Energy Management: A Review. Renewable and Sustainable Energy Reviews. 2017, roč. 70, s. 598–609. Dostupné z doi: 10.1016/j.rser.2016.11.132.

[2] FILDES, Robert. The evaluation of extrapolative forecasting methods. International Journal of Forecasting [online]. 1992, 8(1), 81-98 [cit. 2023-09-13]. ISSN 01692070. Dostupné z: doi:10.1016/0169-2070(92)90009-X

[3] ZHANG, Yang, Tao HUANG a Ettore Francesco BOMPARD. Big data analytics in smart grids: a review. Energy Informatics [online]. 2018, 1(1) [cit. 2023-09-13]. ISSN 2520-8942. Dostupné z: doi:10.1186/s42162-018-0007-5

Vedoucí práce:	Ing. Jan Kraus, Ph.D. Ústav mechatroniky a technické informatiky
Datum zadání práce:	12. října 2023
Předpokládaný termín odevzdání:	14. května 2024

L.S.

prof. Ing. Zdeněk Plíva, Ph.D.
děkan

doc. Ing. Josef Černohorský, Ph.D.
vedoucí ústavu

V Liberci dne 12. října 2023

Prohlášení

Prohlašuji, že svou bakalářskou práci jsem vypracoval samostatně jako původní dílo s použitím uvedené literatury a na základě konzultací s vedoucím mé bakalářské práce a konzultantem.

Jsem si vědom toho, že na mou bakalářskou práci se plně vztahuje zákon č. 121/2000 Sb., o právu autorském, zejména § 60 – školní dílo.

Beru na vědomí, že Technická univerzita v Liberci nezasahuje do mých autorských práv užitím mé bakalářské práce pro vnitřní potřebu Technické univerzity v Liberci.

Užiji-li bakalářskou práci nebo poskytnu-li licenci k jejímu využití, jsem si vědom povinnosti informovat o této skutečnosti Technickou univerzitu v Liberci; v tomto případě má Technická univerzita v Liberci právo ode mne požadovat úhradu nákladů, které vynaložila na vytvoření díla, až do jejich skutečné výše.

Současně čestně prohlašuji, že text elektronické podoby práce vložený do IS/STAG se shoduje s textem tištěné podoby práce.

Beru na vědomí, že má bakalářská práce bude zveřejněna Technickou univerzitou v Liberci v souladu s § 47b zákona č. 111/1998 Sb., o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách), ve znění pozdějších předpisů.

Jsem si vědom následků, které podle zákona o vysokých školách mohou vyplývat z porušení tohoto prohlášení.

17. března 2024 Jakub Šilhán

Predikce hodnot dat v energetice budov

Abstrakt

V posledních letech je silně podporován rozvoj chytrých sítí. Jelikož tyto sítě poskytují mnoho dat, je nutné zajistit jejich vhodné využití. Zde se může jednat o pouhé monitorování, ale také je skvělým nápadem tato data využít pro predikování budoucích hodnot pomocí různých metod. Tato práce se zabývá porovnáním několika těchto metod a stanovením jejich výhod a nevýhod. Nejlepší metodou byla při práci ARIMA, která však stále vykazovala problémy spojené s prací s velmi dynamickými daty. Struktura takových dat byla velmi vlivná na výsledné predikce.

Klíčová slova: Chytré Sítě, Strojové Učení, Predikce

Forecasting of data values in smart buildings

Abstract

There has been a lot of support for the development of smart grids in the last few years. Due to these grids providing large amount of data, it is important to ensure their appropriate use. That can either be just monitoring, or it can be a good idea to use the data to forecast future values using different methods. This work deals with a comparison of these methods and the establishment of their advantages and disadvantages. The best method used during the work was ARIMA, which, however, still exhibited issues related to handling very dynamic data. The structure of such data significantly influenced the resulting predictions.

Keywords: Smart Grids, Machine Learning, Forecasting

Poděkování

Děkuji vedoucímu bakalářské práce Ing. Janu Krausovi, Ph.D. za odbornou pomoc při zpracování mé práce.

Obsah

Seznam zkratek 12

1 Úvod **13**

2 Chytré sítě a jejich analýza **14**
2.1 Chytré sítě 14
2.2 Predikce dat 14
2.2.1 Lineární regrese 15
2.2.2 Klouzavý průměr 16
2.2.3 Exponenciální vyhlazování 16
2.2.4 Singular Spectrum Analysis 18
2.2.5 ARIMA 19
2.2.6 Další metody 20
2.3 Metriky pro měření přesnosti 21
2.3.1 RMSE 21
2.3.2 MAE 21
2.3.3 MAPE 22
2.4 Optimalizační Algoritmy 22
2.4.1 Nelder-Mead 22

3 Výsledky analýzy metod **24**
3.1 Data 24
3.1.1 Explorační analýza dat 25
3.2 Metodika měření 28
3.3 Analýza lineární regrese 29
3.3.1 Rozsah vstupních dat v lineární regresi 29
3.3.2 Délka predikovaného období v lineární regresi 30
3.4 Analýza klouzavého průměru 31
3.4.1 Velikost průměrovacího okna v klouzavém průměru 32
3.4.2 Délka sezóny v klouzavém průměru 33
3.4.3 Délka predikovaného období v klouzavém průměru 34
3.5 Analýza metody Holt-Winters 35
3.5.1 Volba vyhlazovacích parametrů Holt-Winters 35
3.5.2 Typ metody Holt-Winters 37
3.5.3 Rozsah vstupních dat v metodě Holt-Winters 37
3.5.4 Parametr sezónnosti v metodě Holt-Winters 39

3.5.5 Délka predikovaného období v metodě Holt-Winters 40
3.6 Analýza metody Singular Spectrum Analysis 41
3.6.1 Rozsah vstupních dat v metodě SSA 41
3.6.2 Parametr sezónnosti v metodě SSA 43
3.6.3 Délka predikovaného období v metodě SSA 43
3.7 Analýza metody ARIMA 44
3.7.1 Rozsah vstupních dat v metodě ARIMA 45
3.7.2 Parametr sezónnosti v metodě ARIMA 46
3.7.3 Délka predikovaného období v metodě ARIMA 47
3.8 Globální porovnání metod 48
3.8.1 Porovnání metod dle přesnosti 48
3.8.2 Porovnání metod dle časové náročnosti 51
3.8.3 Porovnání metod dle složitosti jejich využití 51
3.8.4 Analýza metod na jiné datové sadě 52

Závěr **54**

Použitá literatura **55**

A Přílohy **57**
A.1 Slovník pojmů 57
A.2 Používaná data 58
A.3 Dodatečné grafy 59
A.4 Testovací aplikace 60

Seznam obrázků

3.1 Autokorelace pro celá data . . . 25
3.2 Autokorelace pro týdenní shodu . . . 26
3.3 Hodinová agregace dat . . . 27
3.4 Denní agregace dat . . . 27
3.5 Analýza rozsahu vstupních dat u Lineární Regrese . . . 29
3.6 Analýza časové náročnosti dle rozsahu vstupních dat u Lineární Regrese 30
3.7 Analýza délky predikce u Lineární Regrese . . . 31
3.8 Analýza průměrovacího okna pro týdenní sezónnost . . . 32
3.9 Analýza průměrovacího okna pro denní sezónnost . . . 33
3.10 Analýza parametru sezónnosti u metody klouzavého průměru . . . 34
3.11 Analýza délky predikce u klouzavého průměru . . . 35
3.12 Analýza volby vyhlazovacích parametrů u modelu Holt-Winters . . . 36
3.13 Analýza typů výpočtu u modelu Holt-Winters . . . 37
3.14 Analýza rozsahu vstupních dat u modelu Holt-Winters . . . 38
3.15 Analýza menšího rozsahu vstupních dat u modelu Holt-Winters se statickými parametry . . . 38
3.16 Analýza parametru sezónnosti u modelu Holt-Winter . . . 39
3.17 Analýza délky predikce u modelu Holt-Winters . . . 40
3.18 Analýza rozsahu vstupních dat u modelu SSA . . . 41
3.19 Analýza menšího rozsahu vstupních dat u modelu SSA . . . 42
3.20 Analýza časové náročnosti dle rozsahu vstupních dat u modelu SSA . 42
3.21 Analýza parametru sezónnosti u modelu SSA . . . 43
3.22 Analýza délky predikce u modelu SSA . . . 44
3.23 Analýza rozsahu vstupních dat u modelu ARIMA . . . 45
3.24 Analýza časové náročnosti dle rozsahu vstupních dat u modelu ARIMA 46
3.25 Analýza parametru sezónnosti u modelu ARIMA . . . 47
3.26 Analýza délky predikce u modelu ARIMA . . . 48
3.27 Porovnání chyby MAPE jednotlivých metod vzhledem ke způsobu doplnění chybějících dat . . . 50

A.1 Zobrazení týdenního intervalu ve francouzských datech . . . 58
A.2 Zobrazení srpnové anomálie ve francouzských datech . . . 58
A.3 Zobrazení týdenního intervalu v londýnských datech . . . 59
A.4 Porovnání chyby MAPE jednotlivých metod ve francouzských datech 59
A.5 Porovnání chyby MAPE jednotlivých metod v londýnských datech . . 60

A.6 Zobrazení vytvořené aplikace 60

Seznam tabulek

2.1 Varianty modelu ARIMA 20

3.1 Schéma použití dat 24
3.2 Průměrné chyby zkoumaných metod 49
3.3 Průměrné časové náročnosti zkoumaných metod 51
3.4 Průměrné chyby zkoumaných metod v londýnských datech 52

Seznam zkratek

ARIMA	Autoregressive Integrated Moving Average
FM	Fakulta mechatroniky, informatiky a mezioborových studií Technické univerzity v Liberci
SSA	Singular Spectrum Analysis
SVD	Singular Value Decomposition
EOF	Empirická Ortogonální Funkce
RNN	Rekurentní Neuronové Sítě
LSTM	Long Short-Term Memory
CNN	Konvoluční Neuronové Sítě
ADF	Augmented Dickey-Fuller test
MAPE	Mean Absolute Percentage Error
MAE	Mean Absolute Error
RMSE	Root Mean Square Error

1 Úvod

Cílem této práce je využití metod pro predikce dat v chytrých sítích a zvážení jejich výhod a nevýhod. Jednotlivé metody, využívané k provedení předpovědí, je však nutné správně nastavit a použít. Vliv tohoto nastavení je dalším zkoumaným aspektem této práce.

První část se zabývá především shrnutím chytrých sítí a statistickou analýzou. Nejprve je zde popsáno, co to vlastně ta chytrá síť je, a důvod jejího vzniku. Následně je zde uvedeno, jak se s daty z takové sítě pracuje, a nakonec je zmíněno, jak lze tato data použít k predikování budoucnosti. V této části jsou dále popsány jednotlivé metody, které k těmto predikcím lze využít a je prozkoumáno jejich fungování. Jedná se především o metody SSA, Holt-Winters a ARIMA. Je zde také zmíněn jeden z algoritmů využitelných pro optimalizaci parametrů metod a jelikož je nutné změřit přesnost jednotlivých predikcí, jsou zde popsány používané chybové metriky RMSE, MAE a MAPE.

Druhá část se zabývá praktickou částí této práce. Nejprve jsou zde popsána používaná data, která jsou následně dále prozkoumána. Následně je zde vysvětlena metodika práce s jednotlivými metodami a jsou stanoveny výchozí hodnoty parametrů potřebných pro práci s nimi. Je provedena definice jednotlivých prováděných průzkumů. Zbytek této kapitoly se zabývá zmíněnými měřeními jak individuálními, tak i celkovými. V sekcích jednotlivých metod je také zmíněn způsob jejich implementace.

V závěru práce jsou shrnuty výsledky získané z provedených měření. Na bázi těchto výsledků je diskutováno vhodné využití zkoumaných metod a jsou uvedeny jejich nedostatky. Také je zde shrnuta práce v prostředí .NET.

2 Chytré sítě a jejich analýza

2.1 Chytré sítě

Často probíraným problémem dnešní doby je vysoká spotřeba fosilních paliv. To je navíc spojováno se škodlivými uhlíkovými emisemi. Proto je vyvíjena spousta způsobů jak tuto zátěž na životní prostředí snížit. Jedním z těchto pokusů je například využívání chytrých sítí. Zatímco klasické sítě sbírají data o spotřebě energie hlavně pro vyúčtování, chytré sítě poskytují i informace o stavu sítě a tuto obrovskou spoustu dat lze využít k různým analýzám zajišťujícím lepší provoz. [20]

Právě práce s velkým objemem dat je jednou z náročnějších disciplín v takových projektech. Pro získání užitečných informací je potřeba provést několik kroků, které pomohou data vyčistit a vyvodit z nich různé závěry. Prvním krokem je předzpracování dat, kdy je provedena jejich agregace z různých zdrojů se snahou zabránit duplikaci. Také je potřeba se postarat o chybějící hodnoty, což je často prováděno pomocí interpolace.[3] Po úspěšné přípravě dat lze konečně začít s jejich analýzou. Zde jsou často používány různé matematické algoritmy a strojové učení. [20]

Při takovéto analýze lze provést spoustu výpočtů. Je možné hledat vzájemné vztahy mezi různými veličinami, kterými může být například spotřeba plynu a lokální teplota. Tyto vztahy mohou zdůraznit různé nedostatky používané infrastruktury. U používaného příkladu by to mohl být únik tepla z domu. Historická data lze také používat k provedení předpovědí budoucích hodnot. Takovéto předpovědi mohou být využity při práci s energiemi, kdy lze předem očekávat špičkovou poptávku. Kdyby dodavatel energií takovou poptávku nečekal, mohly by poté nastat výpadky, které jsou velmi nechtěné. Právě tyto a další postupy jsou důvodem velkého rozvoje tohoto odvětví a důvodem vypracování této práce. [14]

2.2 Predikce dat

Tato část se zaměřuje především na predikování dat, které bylo zmíněno v sekci 2.1. Zde se jedná o použití historických informací k předpovědi budoucích hodnot. Jelikož však data mohou mít různé vlastnosti, je nutné použít algoritmy, které je zvládnou správně zpracovat.

Nejdříve je nutné definovat jaké predikce je potřeba dělat. Volba je mezi kvalitativním a kvantitativním předpovídáním. Kvalitativní přístup je používán především při nedostatku historických dat. Zde jsou použity oborové znalosti a informace z podobných situací. Preferovanou možností jsou však samozřejmě kvantitativní predikce, kdy jsou použity statistické metody společně s historickými daty. [11]

Při snaze provádět co nejpřesnější predikce je nutné vědět, že existují modely jednorozměrné, které používají pouze jednu historickou proměnnou k předpovědi budoucnosti, a modely vícerozměrné, které takových proměnných používají více. U vícerozměrných modelů je pak také zkoumán vztah mezi jednotlivými vlastnostmi dat a pokud se jedná o vlastnosti, které s jistotou známe, například státní svátky, tak je lze do modelu dodat při provádění předpovědi a zmenšit tak výslednou chybu. [11]

Většina kvantitativních problémů se nakonec dělí mezi práci s časovými řadami nebo s průřezovými daty. Jelikož jsou data o spotřebě energie měřeny v pravidelných intervalech, dají se tato data nazvat časovými řadami. [11]

Protože se jedná o velmi užitečnou disciplínu, existuje již mnoho modelů, které se předpověďmi zabývají. Existují také neuronové sítě, které se z historických dat nejprve co nejvíce naučí a pak tyto znalosti použijí k provedení predikcí. Tato práce se však zabývá především tradičními prediktivními metodami, jelikož je práce provedena s jednorozměrnými časovými řadami.

V této části jsou nejprve probrány jednodušší algoritmy pro základní modelování dat. Těmi jsou metody klouzavého průměru, autokorelace a lineární regrese. Následně jsou projednány prediktivní metody jako Holt-Winters, SSA a ARIMA. Holt-Winters a ARIMA jsou proslulé jejich využitím v předvídání dat, ale metoda SSA byla objevena při práci v prostředí .NET a prokázala schopnost provádět predikce. Jelikož je cílem predikce ověřit a prozkoumat jejich přesnost, jsou zde zmíněny metriky RMSE, MAE a MAPE. Navíc se u jednoho modelu vyskytla nutnost použití optimalizačního algoritmu pro odhad vstupních parametrů a je zde tedy nakonec popsána i metoda Nelder-Mead. [6]

2.2.1 Lineární regrese

Lineární regrese se používá k odhadu vztahu mezi náhodnou (závislou) veličinou a dalšími nezávislými proměnnými. K tomu je potřeba nejdříve zjistit, zda jsou veličiny vůbec vzájemně závislé. Hlavním využitím lineární regrese je predikce hodnot.

Existuje několik metod výpočtů. Ty se dělí na lineární a nelineární. Problematikou lineárních metod je jejich nepřesnost a z toho důvodu se využívají metody nelineární. Toto jsou techniky založené na klasifikaci jako je metoda nejbližších sousedů či metoda podpůrných vektorů (support vector). [14]

Metoda nejmenších čtverců (Ordinary least squares), dále Ols, je jedním z hlavních algoritmů používaných pro výpočet jednoduché lineární regrese. Zatímco u lineární regrese je cílem najít parametry α a β pro vzorec $y_i = \alpha + \beta x_i + \epsilon_i$ tak, že je celková chyba ϵ minimální, Ols tento přístup upravuje tak, že se snaží minimalizovat čtvercovou chybu. To nás zbavuje obav z toho, že by záporná chyba mohla negovat chybu pozitivní a naopak. Zatímco se jednoduchost této optimalizační strategie

může zdát jako hlavní důvod její popularity, existuje ještě jeden závažnější důvod. Tím je to, že výsledné parametry jsou nejlepšími nezaujatými odhady opravdových hodnot α a β. Nezaujatý odhad je takový, který při několika opakováních vrací v průměru hodnotu, která odpovídá reálné odhadované hodnotě. [1]

2.2.2 Klouzavý průměr

Klouzavý průměr je výpočtem, který pomocí průměrování různých výběrů z datové sady provádí datovou analýzu. Jedná se také o filtr s konečnou impulzní odezvou. Výpočet pro jednoduchý klouzavý průměr lze zapsat pomocí vzorce.

$$MA = \frac{(A_1 + A_2 + ... + A_N)}{N} \tag{2.1}$$

kde A jsou reálné hodnoty z N posledních měření. Tento algoritmus je často používán při práci s časovými řadami. Zde je užitím vyhlazování dat pro odstranění krátkodobých výkyvů a zvýraznění dlouhodobějších trendů a cyklů. To však není jediným možným využitím. Nejedná se sice o nejvhodnější přístup, ale výpočet je možné také využít k provedení krátkodobých odhadů. Ty jsou například užitečné při doplňování chybějících hodnot v datové sadě. [17]

Jak již bylo zmíněno, tak tato metoda má několik různých variant, které jsou vhodné pro různé situace. Jednou z těchto variant je například vážený klouzavý průměr, který každé hodnotě v průměrovaném okně dodává určitou váhu. Díky tomu lze například dát větší váhu novějším hodnotám nebo naopak. Dále je tento výpočet i základem exponenciálního vyhlazování a modelu ARIMA. Více o těchto metodách je napsáno dále v sekcích 2.2.3 a 2.2.5. [11, 17]

V této práci je jednoduchý klouzavý průměr použit především při náhradě chybějících měření v používaných datech a také v porovnání s dalšími prediktivními metodami.

2.2.3 Exponenciální vyhlazování

Exponenciální vyhlazování vzniklo v 50. letech a stalo se základem některých nejúspěšnějších prediktivních metod. Tyto predikce jsou zpravidla vytvářeny pomocí vážených průměrů historických pozorování, kde jednotlivé váhy exponenciálně klesají. To znamená, že čím novější hodnota je, tím větší má váhu na výslednou předpověď. Existuje několik typů tohoto výpočtu, kde každý je vhodný pro jinou situaci. [11]

První metodou je jednoduché exponenciální vyhlazování. Je vhodné především při práci s daty, které nevykazují žádný trend či sezónnost. Tuto základní metodu lze pak zapsat takto:

$$\ell_t = \ell_{t-1} + \alpha(y_t - \ell_{t-1}) \tag{2.2}$$

$$\hat{y}_{t+h|t} = \ell_t \tag{2.3}$$

kde α je vyhlazovací faktor a vyhlazená statistika ℓ_t, která se aktualizuje na základě nového pozorování y_t a předchozí statistiky ℓ_{t-1}. [11]

Druhou metodou je dvojité exponenciální vyhlazování, neboli Holt lineární metoda. To je rozšířením pro předchozí výpočet, které umožňuje práci s daty, které vykazují trend. Tento výpočet zahrnuje dvě rovnice.

$$\begin{aligned} \ell_t &= \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\ b_t &= \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1} \\ \hat{y}_{t+h|t} &= \ell_t + hb_t \end{aligned} \tag{2.4}$$

kde l_t označuje odhad úrovně řady v čase t a b_t značí odhad trendu v čase t. α je vyhlazovacím faktorem úrovně a β je nyní vyhlazovacím faktorem trendu. Právě tento faktor trendu je zde vypočítán pomocí aktuální a předchozí hodnoty faktoru úrovně ℓ a předchozí hodnoty faktoru trendu b_{t-1}. [11]

Třetí a v této práci využitou variantou exponenciálního vyhlazování je metoda Holt-Winters neboli trojité exponenciální vyhlazování. Cílem tohoto rozšíření bylo zachytit sezónnost v časových řadách. Jak je již pochopitelné z názvu, tak tato metoda využívá tří vyhlazovacích rovnic, kde se každá zaměřuje na jiný parametr - úroveň, trend a sezónnost. Tato metoda má formu aditivní a multiplikativní. Zatímco aditivní je používána v případě, že velikost sezónních variací je stálá, tak multiplikativní je preferována především když jsou tyto variace proporcionální. To znamená, že pokud bude přes zimu spotřeba energie v průměru vyšší, a jednotlivé nárůsty a poklesy skrz den budou také vyšší, tak se jedná o data vhodná pro použití multiplikativního výpočtu. Pokud si však tyto krátkodobé změny zachovají konstantní velikosti bez závislosti na okolních datech, tak je lepší volbou výpočet aditivní. Samotné výpočty vypadají takto.

Aditivní:

$$\begin{aligned} \ell_t &= \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\ b_t &= \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1} \\ s_t &= \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m} \\ \hat{y}_{t+h|t} &= \ell_t + hb_t + s_{t+h-m(k+1)} \end{aligned} \tag{2.5}$$

Multiplikativní:

$$\begin{aligned} \ell_t &= \alpha\frac{x_t}{s_{t-m}} + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\ b_t &= \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1} \\ s_t &= \gamma\frac{y_t}{(\ell_{t-1} - b_{t-1})} + (1-\gamma)s_{t-m} \\ \hat{y}_{t+h|t} &= (\ell_t + hb_t)s_{t+h-m(k+1)} \end{aligned} \tag{2.6}$$

Nyní k jednotlivým parametrům. K α a β se nyní přidává γ, která je vyhlazovacím faktorem sezónnosti. Také přibyla rovnice sezónnosti s_t a proměnná m. Ta vyznačuje frekvenci sezónnosti, takže lze například u hodinových měření tuto hodnotu nastavit

na dvacet čtyři, což znamená, že algoritmus má pracovat se dny. Také lze vidět, že se jedná o rekurzivní výpočet, jelikož každý odhad je počítán pomocí všech předchozích odhadů. Proto je metoda vhodná především pro krátkodobé predikce. [11]

Nakonec je důležité zmínit, že volba vyhlazovacích faktorů silně ovlivňuje výkon těchto metod a je užitečné znát omezení daných faktorů, kde $0 \leq \alpha \leq 1$, $0 \leq \beta \leq 1$ a $0 \leq \gamma \leq (1 - \alpha)$. [11]

Metoda Holt-Winters byla využita například při provádění predikcí poptávky energie ve Filipínách. Zde byla metoda použita pro práci s daty z rezidenčních domů ve franšízové oblasti CEPALCO a prokázala se jako spolehlivá. [18]

2.2.4 Singular Spectrum Analysis

Singulární spektrální analýza (Singular Spectrum Analysis), dále SSA, je další metodou pro analýzu časových řad, kde je aplikováno větší množství statistických technik. Cílem je zde rozklad původní řady na několik menších a čitelnějších komponentů, kterými mohou být trend, oscilační složka a šum. SSA má spoustu různých využití jako extrakce trendu, detekce periodicity, vyhlazování, redukce šumu a predikce. Velkou výhodou této metody je, že není nutné znát parametrický model používané časové řady, avšak tato řada musí vyhovovat vzorci lineární rekurence. To znamená, že lze členy řady vyjádřit lineární kombinací předchozích členů. Jednoduchým příkladem takové řady je Fibonacciho posloupnost. Při práci je nutné také správně zvolit parametr délky okna tak, aby rozklad dat vytěžil co nejvíce užitečných informací. Samotné fungování této metody se skládá ze dvou fází s několika kroky. [7, 9]

První fází je rozklad, kdy je časová řada rozdělena na již zmíněné komponenty. Prvním krokem tohoto rozkladu je takzvaně vkládání (embedding), kdy je počáteční jednorozměrná řada mapována do řady dvojrozměrné tak, že vznikne matice trajektorie X. V této matici je každý řádek posunutou verzí originální řady. Dalším krokem je Singular Value Decomposition (SVD), neboli singulární rozklad. Zde je matice trajektorie rozložena na sumu elementárních matic. Tyto elementární matice se skládají ze tří částí. První je vlastní číslo λ_i pro XX^T , které vyznačuje podíl vlivu elementární matice na X. Dále jsou zde vlastní vektory U_i pro XX^T. Takový vektor je často nazýván faktorová empirická ortogonální funkce neboli EOF. Třetí částí jsou vektory V_i, které jsou transformací vektorů U_i dle vzorce $V_i = X^T U_i / \sqrt{\lambda_i}$. Právě Vi je často označováno jako hlavní komponenta. Nakonec lze SVD matice trajektorie zapsat jako $X = X_1 + ... + X_d$, kde d je hodností matice X s $\lambda_i > 0$. Jednotlivé komponenty jsou výsledkem $X_i = \sqrt{\lambda_i} U_i V_i^T$. [8, 9]

Druhou fází této analýzy je rekonstrukce. Zde je hlavním cílem převést informace získané z předchozí fáze zpět do formátu časových řad. Prvním krokem je „grouping", kdy jsou výsledky SVD rozděleny do několika skupin. Tyto výsledky jsou následně sečteny, čímž vzniká jedna matice pro každou skupinu. Vlastní čísla jsou také sečtena, takže význam jednotlivých matic je stále zachován. Nakonec je provedeno diagonální průměrování, kdy jsou tyto skupinové matice převedeny zpět do formátu časových řad. To je provedeno průměrováním prvků na diagonálách. Nakonec vznikne několik časových řad, kde každá odpovídá jiné komponentě. Těmito

komponentami mohou být trend, šum, sezónnost a další. Tyto požadované zjednodušené komponenty jsou poté použity k provedení predikcí, které jsou nakonec sečteny dohromady pro vytvoření rekonstruované řady. [8, 9]

Metoda SSA byla například využita při předvídání poptávky po zemním plynu v Indii. Vysoká přizpůsobivost této metody zde umožnila provést velmi kvalitní předpovědi. [12]

2.2.5 ARIMA

ARIMA, neboli Autoregressive Integrated Moving Average, je společně s exponenciálním vyhlazováním jedním z nejpoužívanějších modelů při předvídání časových řad. Zde jsou k popisu dat použity autokorelace. Tento model je složen z diferenciace, autoregrese a klouzavého průměru, který již byl zmíněn. [11]

Je potřeba vědět, že ARIMA vyžaduje po vstupních datech, aby byla stacionární. To znamená, že průměr a rozptyl celé řady zůstává stále stejný a tato data tak vypadají stejně bez ohledu na to, kde se na ně pozorovatel podívá. Pokud tomu tak není, je nutné provést transformaci, která tuto stacionaritu zajistí. Takovou transformací je diference, která je v názvu modelu skrytá ve slově „Integrated". [11]

Autoregrese z názvu je podobná modelu vícenásobné regrese. Zde stále platí že predikce jsou prováděny pomocí lineární kombinace prediktorů. Rozdílem je, že místo jiných proměnných jsou tentokrát prediktorem historické hodnoty předvídaných dat. Takový model **AR(p)** lze zapsat tímto způsobem:

$$y_t = c + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + ... + \alpha_p y_{t-p} + \epsilon_t \tag{2.7}$$

kde, c je úrovní dat a ϵ je šum. Koeficienty α_i je nutné odhadnout z dat. K tomu lze použít již zmiňovanou metodu nejmenších čtverců. Nakonec p je řádem této autoregrese. To znamená, že model používá k predikcím data s p zpožděním. [11]

Zatímco klouzavý průměr již byl zmíněn, jednalo se o formu vyhlazování, která používala pouze předchozí hodnoty. Zde klouzavý průměr místo historických hodnot používá minulé chyby predikcí. Takový model **MA(q)** lze následně zapsat takto:

$$y_t = c + \epsilon_t + \beta_1 \epsilon_{t-1} + \beta_2 \epsilon_{t-2} + ... + \beta_q \epsilon_{t-q} \tag{2.8}$$

kde, c je opět úrovní dat, ϵ je chybou v čase t. Koeficienty β_i jsou opět odhadovány z dat a tentokrát je řádem modelu q, které funguje stejně jako u modelů autoregresních. [11]

Jak již bylo řečeno tak ARIMA je kombinací všech těchto metod a modelů. Celý tento model **ARIMA(p,d,q)** lze zapsat pomocí vzorce:

$$y'_t = c + \alpha_1 y'_{t-1} + ... + \alpha_p y'_{t-p} + \beta_1 \epsilon_{t-1} + ... + \beta_q \epsilon_{t-q} + \epsilon_t \tag{2.9}$$

kde y' je diferencovaná řada. Parametry p, d a q určují počet zpoždění v autoregresním modelu, řád diferencování a počet zpoždění v modelu klouzavého průměru. Proto existuje několik speciálních verzí tohoto modelu. [11]

Tabulka 2.1: Varianty modelu ARIMA

Název	Vstup
Bílý Šum	ARIMA(0,0,0)
Random Walk	ARIMA(0,1,0) bez konstanty
Random Walk s driftem	ARIMA(0,1,0) s konstantou
Autoregrese	ARIMA(p, 0, 0)
Klouzavý průměr	ARIMA(0, 0, q)

Jednotlivé vstupní parametry p a q pro celkovou metodu je někdy možné určit analýzou vstupních dat. Tato analýza používá autokorelační funkci a částečnou autokorelační funkci. Nakonec tento model nejdříve vypočítá jednotlivé části a poté pomocí lineární kombinace tyto části použije k popisu zadané časové řady a k provedení predikcí. [11]

Pro modelování sezónních dat je možné rozšířit model ARIMA o sezónní parametry. $\mathbf{ARIMA(p,d,q)(P,D,Q)}_m$ je metoda, která využívá sezónních parametrů P,D a Q s délkou sezóny m k modelování takových dat a lze ji zapsat takto:

$$(1 - \sum_{i=1}^{p} \alpha_i L^i)(1 - \sum_{i=1}^{P} A_i L^{is})(1 - L)^d (1 - L^s)^D y_t = c + (1 + \sum_{i=1}^{q} \beta_i L^i + \sum_{i=1}^{Q} B_i L^{is})\epsilon_t \tag{2.10}$$

kde L je operátorem zpoždění. α a β si zachovávají svůj význam a d je řádem diferencování, přičemž s stanovuje délku sezóny. Proměnné označené velkými písmeny pak náleží sezónním parametrům. [11]

Jako příklad využití metody ARIMA lze uvést předpověď poptávky energie v Turecku. Tato metoda se osvědčila při zpracování energetických dat, a to jak v její základní podobě ARIMA, tak i v sezónně upravené variantě SARIMA. [5]

2.2.6 Další metody

Existuje spousta dalších modelů pro práci s časovými řadami. Jednou z novějších a velmi často využívaných technik je použití neuronových sítí. První variantou jsou rekurentní neuronové sítě (RNN) nazývané Long short-term memory (LSTM). Jedná se o síť s několika vrstvami, která pomocí bran rozhoduje jakou informaci zahodit, přidat a předat do dalšího kroku. LSTM sítě jsou používány především kvůli jejich schopnosti pracovat s dlouhodobými závislostmi. Druhou variantou těchto sítí jsou konvoluční neuronové sítě (CNN). Ty jsou používány především k extrakci jednotlivých vlastností těchto řad. Tyto dva modely lze propojit tak, že CNN nejdříve vytěží nejdůležitější vlastnosti vstupních dat a LSTM je využije k budoucím predikcím. Modely využívající neuronových sítí a hlubokého učení jsou často označovány jako „black box" modely. To znamená, že uživatel vidí pouze vstupy a výstupy, ale není vidět jak přesně byly tyto výstupy vytvořeny. Právě tato vlastnost je jedním z důvodů, proč tyto modely nejsou v této práci dále použity pro testování. [2]

2.3 Metriky pro měření přesnosti

Práce s prediktivními modely často vyžaduje užití metody pokus-omyl. To je způsobeno tím, že jednotlivá data, které jsou k předpovědím používána, vyžadují individuální přístup a vhodnou volbu modelu. Tuto volbu lze často zjednodušit analýzou dat, ale někdy je potřeba rozhodnout mezi teoreticky stejně vhodnými modely. Jelikož tyto modely také používají různé vstupní parametry, je nutné je také zjistit a ověřit. K tomu lze využít různé metriky, které vypočítají chybu e_t předpovědí. Pod pojmem chyba je myšlen rozdíl mezi předpovědí a reálnou hodnotou, přičemž tento rozdíl může být vypočítán i pro více dat najednou. Klasickým využitím je vytvoření modelu na stavebních datech, následně použití testovacích dat ke zjištění jeho přesnosti a nakonec použití nebo přetrénování daného modelu. V této práci jsou tyto metriky také použity k porovnání jednotlivých modelů v různých situacích. Častými metrikami bývá Root Mean Square Error (RMSE), Mean Absolute Error (MAE) a Mean Absolute Percentage Error (MAPE). Jednotlivé metriky mají různé vlastnosti, které lze využít ke zvýraznění chování testovaných modelů. [11]

2.3.1 RMSE

RMSE je metrika, která při výpočtu využívá chyby závislé na měřítku původních dat. To znamená, že pokud původní data zobrazují spotřebu energie v kW, tak je výsledná chyba také v kW. Samotný výpočet lze zapsat takto:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (e_i)^2}{N}} \tag{2.11}$$

kde e_i je odchylkou mezi reálnou a predikovanou hodnotou a N je počtem predikcí. Použití takovéto metriky je možné pouze při porovnávání metod použitých na stejných datech nebo na datech se stejným měřítkem. Pokud tomu tak není, jsou tato porovnání naprosto nevhodná. RMSE patří mezi nejčastěji používané chybové metriky i přes lehce horší interpretovatelnost. Z důvodu mocnin je u tohoto výpočtu dána větší váha větším chybám a toto chování vykazuje exponenciální trend. [11, 19]

2.3.2 MAE

Druhou často využívanou metrikou v této oblasti je Mean Absolute Error. Jedná se opět o výpočet chyby závislé na měřítku dat, takže oblast využití je velmi podobná RMSE. MAE lze matematicky zapsat takto:

$$MAE = \frac{\sum_{i=1}^{N} |e_i|}{N} \tag{2.12}$$

kde e_i je opět změřená chyba a N je počet měření. Na rozdíl od RMSE sleduje rozložení váhy chyb lineární trend. To znamená, že pokud jsou nalezeny chyby pět a deset, tak hodnota deset má dvojnásobnou váhu. [11, 19]

2.3.3 MAPE

Metrika Mean Absolute Percentage Error je velmi podobná MAE. Hlavní rozdíl je tentokrát v chybě se kterou počítá, jelikož používá procentuální chybu, která je relativní k reálné hodnotě. Tentokrát je tato metrika tedy vhodná i při porovnání metod s daty s různými měřítky. Lze ji zapsat takto:

$$MAPE = \frac{\sum_{i=1}^{N} \frac{|A_i - B_i|}{A_i} * 100}{N} \tag{2.13}$$

kde A_i je reálná hodnota, F_i je predikovaná hodnota a N je počet predikcí. Problém nastává v případě pokud $A_i = 0$, jelikož výsledná hodnota takové metriky je nekonečno nebo nedefinováno. Už i v situaci kdy y_t je blízko nule nastávají komplikace, jelikož výsledek metriky bude nabývat velmi vysokých hodnot. Celkově lze říci, že zatímco je tato hodnota velmi dobře čitelná, existuje velké množství situací, kdy může tato metrika být negativně ovlivněna vlastnostmi testovaných dat. [11, 15]

2.4 Optimalizační Algoritmy

Tyto algoritmy se zabývají matematickou optimalizací. To je disciplína, která se zabývá minimalizací, nebo maximalizací výstupu účelové funkce $f(x)$. Příkladem takové optimalizace je minimalizace chybové metriky, která je využita v této práci. V tomto případě je účelovou funkcí výpočet hodnoty MAPE pro trénovací data pro predikce a využití optimalizačních algoritmů k nalezení těch nejlepších parametrů pro prediktivní modely ke snížení chyby. Během takové optimalizace lze také zvolit různá omezení a cílovou hodnotu. Pro tuto disciplínu existují různé algoritmy s různými požadavky, proto je nutné před jejich použitím provést rešerši těch nejvhodnějších. [4]

2.4.1 Nelder-Mead

Nelder-Mead je jedním z nejpopulárnějších algoritmů pro vícerozměrnou optimalizaci. Jedním z velkých důvodů této popularity je to, že zde nejsou užity derivace. Právě tato vlastnost je vhodná při práci s funkcemi, které nejsou hladké. Hlavním využitím je zde především odhad parametrů pro jiné funkce a další podobné statistické problémy, kde účelová funkce obsahuje šum a nejisté hodnoty. [16]

Hlavním stavebním blokem tohoto algoritmu je takzvaný simplex. Simplex je pojem popisující nejjednodušší polytop v daném prostoru. Například v jednorozměrném prostoru je takovým polytopem linie, zatímco ve dvojrozměrném prostoru by jím byl trojúhelník. Prvním krokem této metody je tvorba tohoto simplexu s $n+1$ vrcholy, kde n je počet odhadovaných parametrů. Klasicky je nejprve zvolen jeden vrchol pomocí počátečních odhadů a následně jsou od tohoto bodu odvozeny další vrcholy. [16]

Druhým krokem je iterativní transformace simplexu, která je prováděna několika způsoby. Nejdříve jsou jednotlivé vrcholy seřazeny od nejhoršího po nejlepší. Poté je nalezeno těžiště t pro všechny vrcholy až na ten nejhorší. S těmito dvěma kroky hotovými je čas provést transformaci. Ta se může zabývat náhradou pouze jednoho vrcholu, nebo je možné zmenšit celý simplex směrem k nejlepšímu vrcholu. Celkové zmenšení je prováděno pouze v případě, že se náhrada jednoho vrcholu prokázala nedostatečnou. Jsou definovány tři způsoby náhrady vrcholu a v následujícím vysvětlení je nejhorší vrchol značen x_h a nový vrchol x_r. První možností je zrcadlení, kdy je x_h promítnuto skrz těžiště t pro vznik nového vrcholu. Pokud jsou výsledné parametry lepší, lze zkusit rozšíření. V tomto případě je nový vrchol x_e výsledkem zrcadlení, které je od těžiště vzdáleno dvakrát tak daleko. Podle toho jaký z těchto vrcholů x_r a x_e je provedena kontrakce. V tomto případě je zvolen bod mezi těžištěm a vrcholem x_h nebo x_r. Pokud nenastalo zlepšení výsledků, je provedeno již zmenšení simplexu. Nakonec nastává další krok iterace. [16]

Výsledek těchto transformací je vrácen ve chvíli, kdy je zlepšení mezi jednotlivými transformacemi zanedbatelné, nebo je dosažen maximální počet iterací. [16]

3 Výsledky analýzy metod

3.1 Data

Jednou z důležitých částí této práce bylo najít vhodný datový soubor zabývající se energetikou v budovách. Nalezena byla data naměřená v rodinném domě v rozmezí čtyř let. Tento dům se nachází v oblasti Sceaux, která je vzdálena sedm kilometrů od Paříže. Měření začala v prosinci 2006 a skončila v listopadu 2010. Data tedy obsahují 47 měsíců měření, což je perfektní pro analýzu modelů pro predikce. Frekvence měření je zde minutová, avšak pro optimalizaci výpočtů byl zvolen hodinový průměr hodnot. [10]

Jelikož tento datový soubor obsahoval několik naměřených veličin (datum, globální činný výkon, globální jalový výkon a další dílčí měření), bylo nutné vyjmout jen vyžadované hodnoty. V případě práce s celkovou spotřebou energie je touto veličinou globální činný výkon. Ostatní veličiny zde byly pro predikce nevhodné. Práce s těmito daty probíhala tak, že byla zvolena trénovací datová sada, která byla použita ve výpočtech metod. Dále byla zvolena evaluační datová sada, která byla použita pro změření přesnosti predikcí. To bylo provedeno provedením predikcí pro dané období a porovnáním těchto predikcí s reálnými hodnotami. Jelikož se pracovalo s časovými řadami, musela evaluační sada navazovat na trénovací sadu. Schéma těchto dat lze vidět v tabulce 3.1, přičemž datum používala pouze lineární regrese.

Tabulka 3.1: Schéma použití dat

Datum	Hodnota	Typ dat
...	...	Trénovací data
16.12.2006 17:00:00	4,22	
16.12.2006 18:00:00	5,36	
16.12.2006 19:00:00	5,38	
16.12.2006 20:00:00	5,42	
16.12.2006 21:00:00	5,48	Předvídaná data
16.12.2006 22:00:00	5,53	
...	...	

Dalším problémem byly chybějící hodnoty, které jsou v takových datech normální. Pro doplnění takových dat existuje množství různých technik. Zde byla vyzkoušena interpolace a vyhlazování klouzavým průměrem. Interpolace data vyplnila

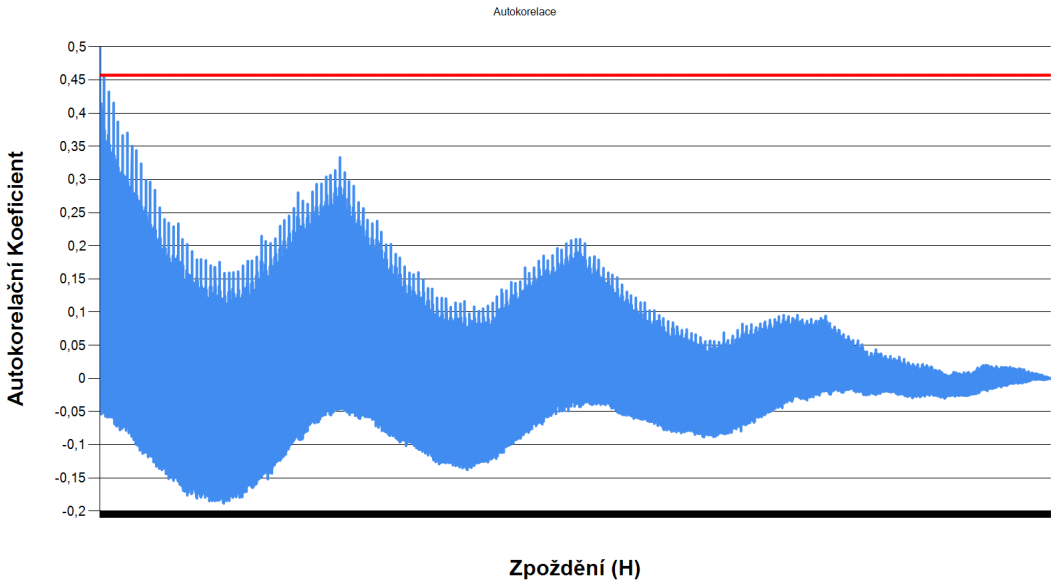
pouhou přímkou, která nijak nevyjadřovala pravděpodobné chování skutečné spotřeby, a proto byl zvolen klouzavý průměr předchozích pěti dnů. Pět dnů bylo zvoleno tak, aby doplněná data neobsahovala příliš velké zpoždění.

Pro lepší práci s daty byla provedena měření různých metrik. Nejprve byla nalezena absolutní minima a maxima 0,076 kW a 11,122 kW. Zde však byly započítány i extrémy, které se mohly v datech vyskytnout i pouze jednou. Proto byly pro tento výběr využity percentily. Konkrétně zde byl využit 1. percentil pro nalezení minima v hodnotě 0,112 kW a 99. percentil pro nalezení maxima v hodnotě 4,84 kW. Takový výběr umožňoval vynechat extrémy. Rozptyl hodnot v celé datové sadě byl 1,11 kW^2 a průměr se rovnal 1,09 kW.

3.1.1 Explorační analýza dat

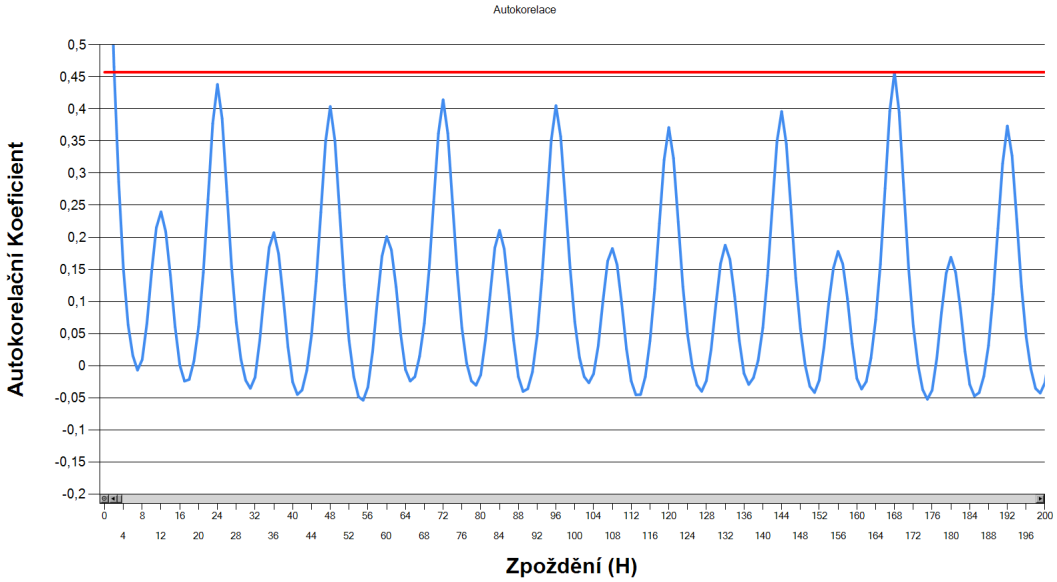
Jednou z vlastností dat, vyžadovaných modelem ARIMA, je stacionarita. Ta vyjadřuje vliv času na datovou řadu. Pokud je tato řada stacionární, tak se její vlastnosti (průměr a rozptyl) nemění. Tuto vlastnost lze ověřit několika způsoby. Nejjednodušší možností je vykreslení těchto dat a manuální posouzení. Poté také existuje Augmented Dickey-Fuller test (ADF). Na časovou řadu byla tedy nejprve využita metoda ADF, která ji označovala za stacionární. Avšak z autokorelační funkce na obrázku 3.1 lze vypozorovat, že data obsahují opakující se vzor a tím pádem je stacionarita zpochybnitelná. Proto byla pro jistotu provedena sezónní diferenciace pro 168 hodin a autokorelační funkce z takto upravených dat již podporovala předpoklad stacionarity.

Protože data vykazovala sezónnost, bylo nutné nalézt nejvhodnější délku tohoto parametru. Autokorelace je často využívána pro tuto analýzu a byla vhodná i zde.



Obrázek 3.1: Autokorelace pro celá data

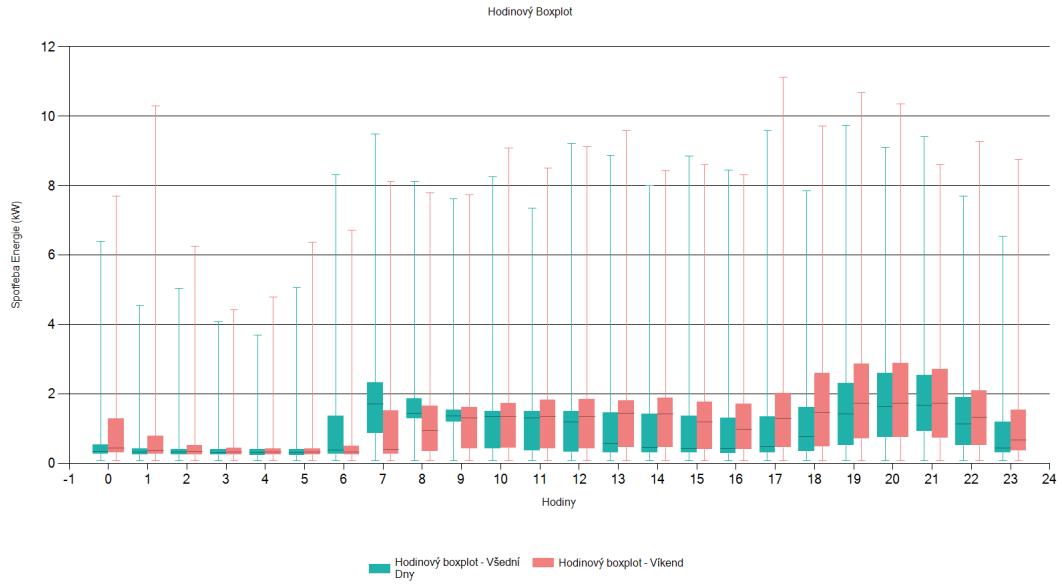
Na grafu v obrázku 3.1 je vyobrazena autokorelační funkce pro celou datovou sadu. Je viditelné, že s rostoucím zpožděním klesá autokorelační koeficient, ale přece jen se v datech vyskytují opakující vzory. Dokonce je viditelné i roční opakování. Červená linie zobrazuje maximální hodnotu hledaného koeficientu a jelikož z tohoto grafu není příliš dobře vidět, jaké zpoždění tohoto maxima nabývá, bylo nutné přiblížení.



Obrázek 3.2: Autokorelace pro týdenní shodu

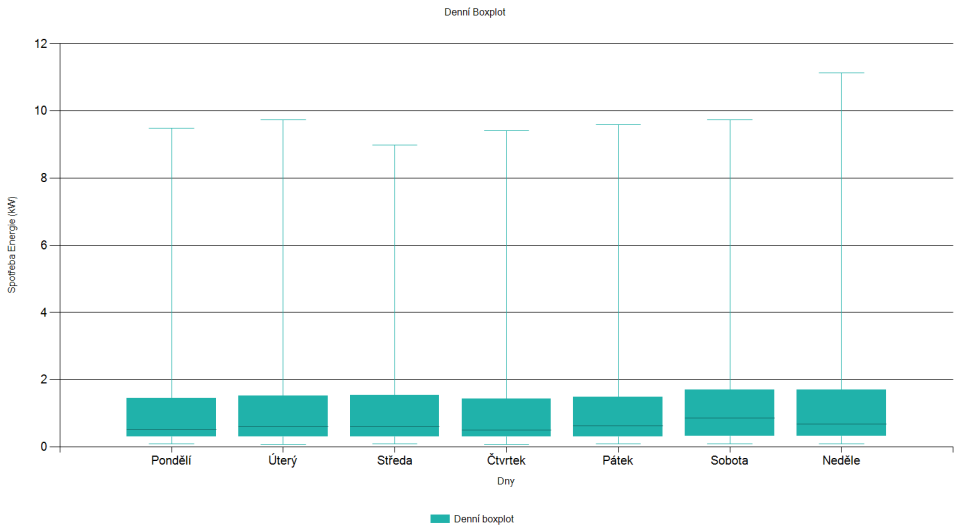
Po přiblížení na obrázku 3.2 je rozeznatelné, že nejvyšších hodnot autokorelace dosahují zpoždění 12, 24 a 168 hodin. V případě 12 hodin lze chápat, že dopoledne a odpoledne mají podobný průběh, kdy je spotřeba zprvu nízká, v polovině tohoto intervalu stoupne (když se obyvatelé vzbudí nebo vrátí z práce) a poté spotřeba opět klesne. Z vyššího skóre autokorelace, dosahujícího u 24 hodinového zpoždění hodnoty 0,44, lze vyvodit, že jednotlivé dny mají podobnější průběh než kdyby byly děleny do dvou intervalů. U zpoždění 168 hodin nastává střet autokorelační funkce a linie zobrazující maximální hodnotu 0,46 autokorelačního koeficientu, což znamená že data s tímto zpožděním si jsou nejvíce podobná s původními daty. Jedná se o nejlepší hodnotu pro nastavení parametru sezónnosti, jelikož je zde zakomponována individualita jednotlivých dnů v týdnu. Protože je však rozdíl mezi 24 a 168 hodinovým zpožděním malý, bude tento parametr ještě u jednotlivých modelů prozkoumán.

Jelikož už byl zmíněn opakující se průběh jednotlivých dnů, bylo provedeno pozorování hodnot pro jednotlivé hodiny napříč celými daty.



Obrázek 3.3: Hodinová agregace dat

Obrázek 3.3 zobrazuje krabicový graf pro hodinovou agregaci skrz den pro celá data. Měření pro všední dny jsou zde zeleně, zatímco víkendy jsou červené. Tento graf perfektně zvýrazňuje průběh klasického dne v pozorovaném domě. V noci majitelé spí a spotřeba je nízká. Ráno se obyvatelé probudí a spotřeba stoupne, jelikož rozsvítí a začnou používat domácí spotřebiče. Spotřeba se pak během dne udržuje na mírně nižší úrovni. Pokud jde o všední dny, spotřeba je obecně nižší než ráno. O víkendech spotřeba zůstává na podobné úrovni jako ráno, protože obyvatelé domu pravděpodobně zůstávají doma. Spotřeba vzroste když se lidé vrátí domů z práce a je nutné začít svítit, a skrz celý večer zůstává vysoko, dokud není čas jít spát.



Obrázek 3.4: Denní agregace dat

Na obrázku 3.4 bylo provedeno pozorování dnů v týdnu. Jedná se o další krabicový graf, tentokrát pro denní agregaci, ze kterého je vidět důvod vyššího autokorelačního koeficientu u týdenního zpoždění. Jak již bylo možné očekávat, víkendy vykazují vyšší spotřebu, zatímco všední dny jsou úspornější.

3.2 Metodika měření

Při snaze prozkoumat chování požadovaných metod bylo důležité vytvořit prostředí, které by to umožňovalo. Problémem však bylo, že jakékoliv změny parametrů těchto metod a jejich dat způsobovaly příliš velké odchylky v predikcích. Proto byly stanoveny výchozí hodnoty těchto parametrů, které byly mezi zkoumanými metodami sdíleny a měnily se pouze pokud to bylo u daného měření zmíněno.

Nastavení výchozích parametrů vypadalo takto. Frekvence jednotlivých měření spotřeby byla obecně nastavená na jednu hodinu. Pomocí datové analýzy bylo také zjištěno, že nejlepší hodnotou sezónnosti byl jeden týden, což odpovídalo 168 měřením. Predikce byly prováděny především pro jeden celý den a vstupní data obsahovala 400 dní historických měření. Velikost vstupních dat byla v této práci uvažována ve dnech, které obsahovaly 24 měření pro jednotlivé hodiny.

Ve většině měřeních byla především zkoumána metrika MAPE, jelikož je mnohem lépe čitelná než RMSE a MAE. Navíc se jedná o hodnotu relativní k predikované oblasti, což bylo dokonce u některých průzkumů nutností. Chyba byla tedy převážně zobrazena jako procentuální chyba relativní k opravdu naměřené hodnotě. Metriky RMSE a MAE byly však využity ve výsledném porovnání přesností metod k zobrazení průměrné chyby v původních jednotkách.

Již byl zmíněn cíl prozkoumat chování sledovaných metod v různém nastavení. U všech metod byl tedy zkoumán vliv parametru sezónnosti s hodnotami jednoho dne a jednoho týdne, přičemž očekáváním byly lepší předpovědi u týdenní sezónnosti, i když autokorelační koeficient zde narostl pouze o 0.02. Dále byla pozorování zaměřena na vliv různě velkých trénovacích dat, kde bylo očekáváno zlepšení při využití většího množství dat. Nakonec byl prozkoumán rozdíl mezi různě dlouhými predikcemi s očekáváním růstu chyb při dlouhodobějších predikcích. Jednotlivé metody měly samozřejmě i svá další individuální nastavení, která byla v některých případech také vyzkoušena.

Druhým cílem práce bylo porovnání použitých metod, které bylo provedeno za použití výchozích parametrů stanovených v této sekci a případného dalšího nastavení zjištěného z měření u jednotlivých metod. V tomto porovnání byla pozorována jejich celková přesnost, časová náročnost a složitost využití.

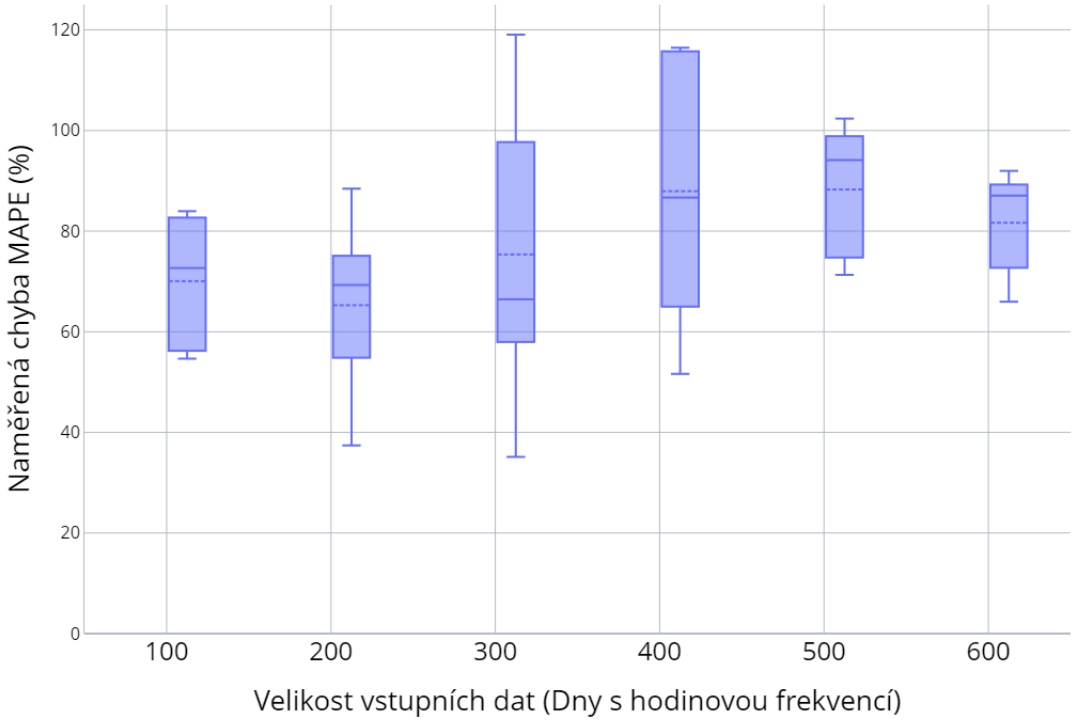
3.3 Analýza lineární regrese

Tato metoda byla zprovozněna pomocí knihovny ML.NET. Zde byl k predikcím využit vztah mezi spotřebou energie a datem měření. Problémem však byl požadavek modelu pro použití datového typu float ve vstupních parametrech. Proto byla data měření transformována na čas, který uplynul od 1.1.2000 v sekundách. To tento problém vyřešilo a metoda již fungovala bez jakýchkoliv problémů.

Důležité je zmínit, že tato metoda nebyla příliš ovlivněna parametry z kapitoly 3.2. Zkoumán zde byl tedy pouze vliv délky predikce a velikosti vstupních dat.

3.3.1 Rozsah vstupních dat v lineární regresi

Byl zde nejprve proveden plánovaný průzkum vlivu rozsahu vstupních dat. Parametry byly mimo rozsah vstupních dat nastaveny na výchozí hodnoty, což u lineární regrese znamenalo pouze hodinovou agregaci měření. Zkoumané rozsahy vstupních dat se zde pohybovaly v rozmezí 100 až 600 dní s inkrementací po 100 dnech a byla zde měřena chyba MAPE.

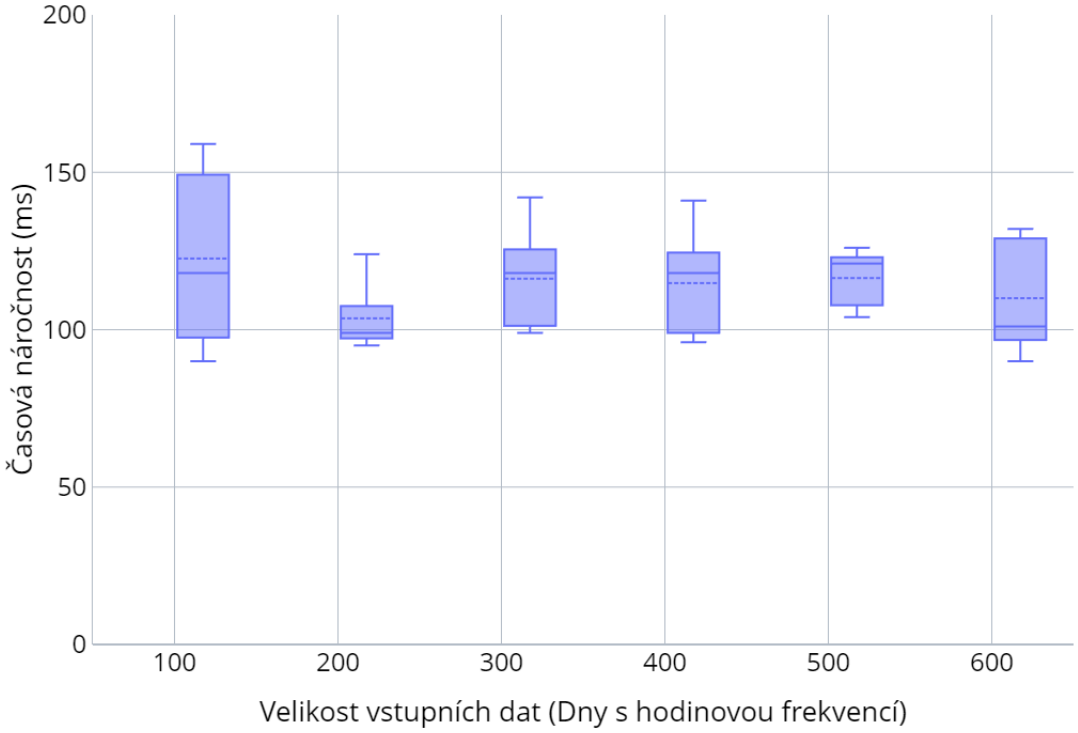


Obrázek 3.5: Analýza rozsahu vstupních dat u Lineární Regrese

Z obrázku 3.5 lze vidět růst chyby s rostoucím množstvím vstupních dat. Jedná se o očekávané chování, jelikož starší data mají u lineární regrese stejnou váhu jako nová data a zhoršují tak výsledky predikcí. Je možné si také povšimnout většího

rozptylu chyb u rozsahů dat 300 a 400 dní, což by mohlo být způsobeno blížícím se koncem jedné roční periody. Ta sice v autokorelační funkci nevykazovala příliš významný vliv, ale stále se tam lehce vyskytovala.

Jelikož existovalo očekávání změny časové náročnosti výpočtu pro různě velká data, byl zde změřen i čas potřebný k dokončení výpočtu.



Obrázek 3.6: Analýza časové náročnosti dle rozsahu vstupních dat u Lineární Regrese

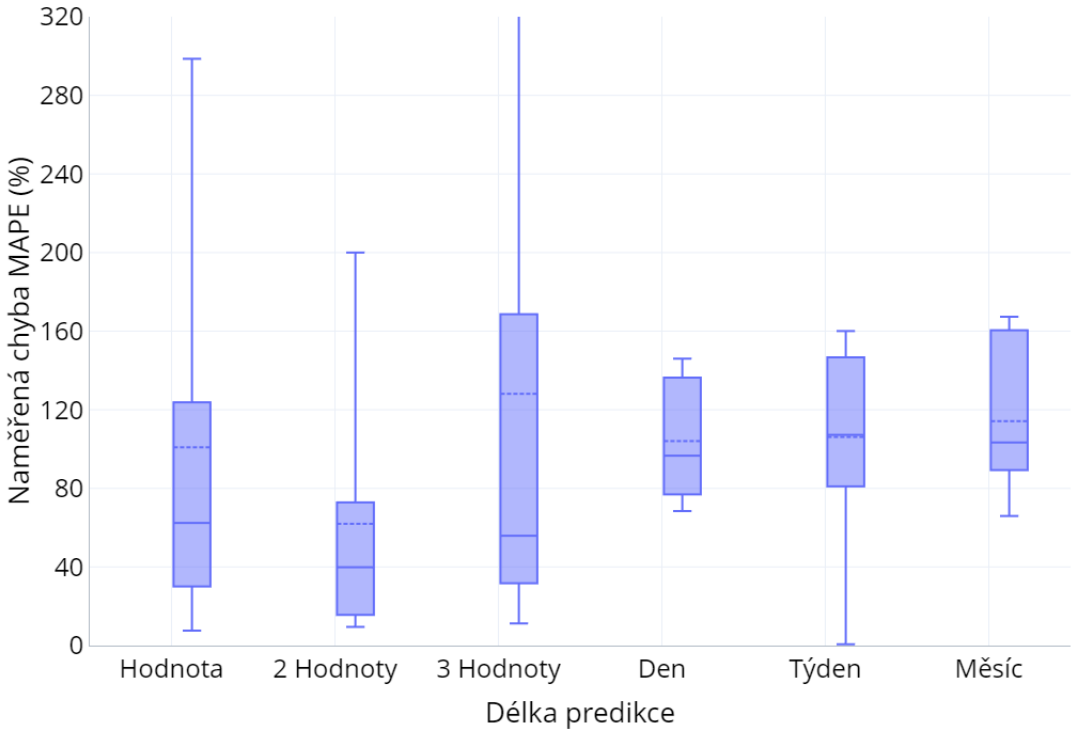
V grafu na obrázku 3.6 jsou vidět pozorování této náročnosti pro 100 až 600 dní dat. Tato měření byla provedena v milisekundách a jelikož se jednalo o jednoduchou metodu, bylo očekáváno, že trvání výpočtu nebude dosahovat vysokých hodnot. Právě to je z grafu dobře vidět, jelikož čas potřebný pro predikci se zde u všech rozsahů dat pohyboval okolo 110 milisekund a žádné z měření nevykazovalo příliš velkou odchylku od této hodnoty.

3.3.2 Délka predikovaného období v lineární regresi

Jelikož se jedná o metodu cílenou na nalezení přímky s nejmenší odchylkou od původních dat, nebyl zde očekáván příliš zajímavý trend.

Jak je však vidět z obrázku 3.7, který zobrazuje metriku MAPE pro různě dlouhá období predikcí, vyskytuje se zde alespoň jedna zajímavá informace. Tou je velký rozptyl chyb u velmi krátkých predikcí, který byl však s delšími predikovanými úseky zúžen. To bylo způsobeno tím, že u delších úseků byly dohromady zprůměrovány

menší a větší chyby. Proto lze takovéto chování předpokládat i u obdobného měření u ostatních metod. Navíc se zde přece jen ukázal trend lehkého růstu chyby s rostoucí délkou predikce.



Obrázek 3.7: Analýza délky predikce u Lineární Regrese

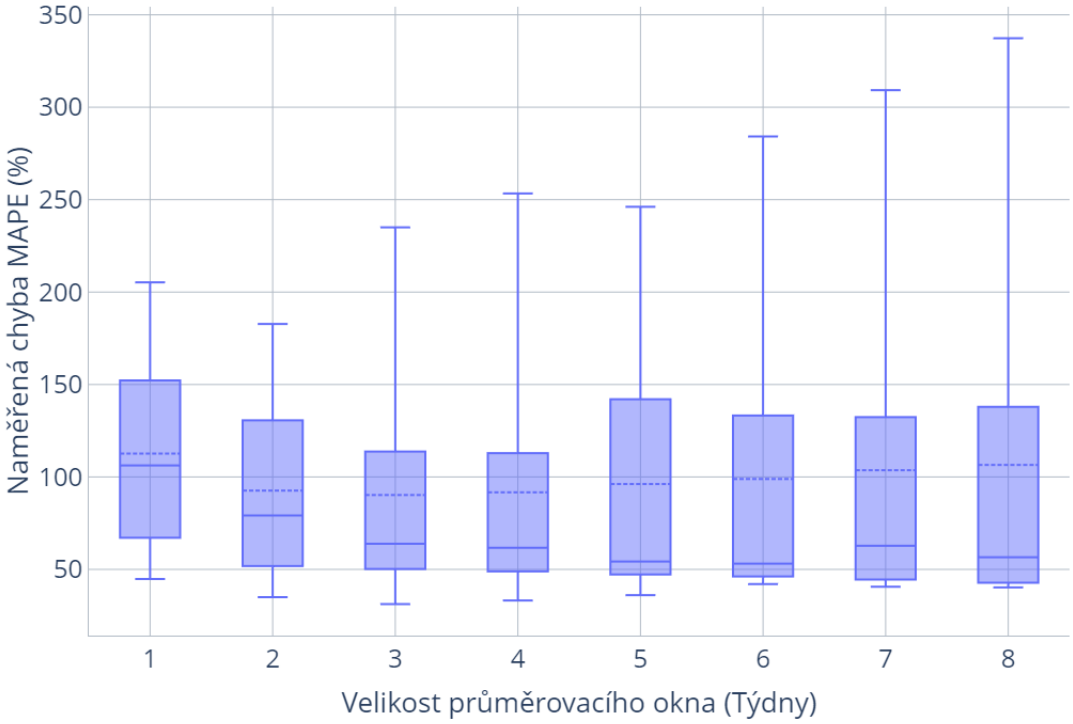
3.4 Analýza klouzavého průměru

Klouzavý průměr byl pro potřeby této práce implementován ručně. Hlavním důvodem k tomuto rozhodnutí byla jeho jednoduchost a lepší kontrola nad jeho vnitřním fungováním. Je nutné však zmínit, že princip fungování této metody neumožňoval provádění dlouhodobějších predikcí za použití krátkodobé sezónnosti, jelikož nebylo využito rekurzivních predikcí.

Implementace této metody odpovídala spíše sezónnímu klouzavému průměru, jelikož zde nebylo používáno pouze posledních n hodnot dat. Například s denní sezónností zde byla predikce na další den vytvořena průměrováním hodnot ve stejný čas z n předchozích dnů. V případě práce s týdenní sezónností se jednalo o průměr hodnot ze stejného dne v týdnu ve stejném čase z n posledních týdnů. Tato úprava byla provedena po experimentech se základním klouzavým průměrem, jelikož ten se po předpovědi několika hodnot ustálil a výsledkem byla pouhá přímka.

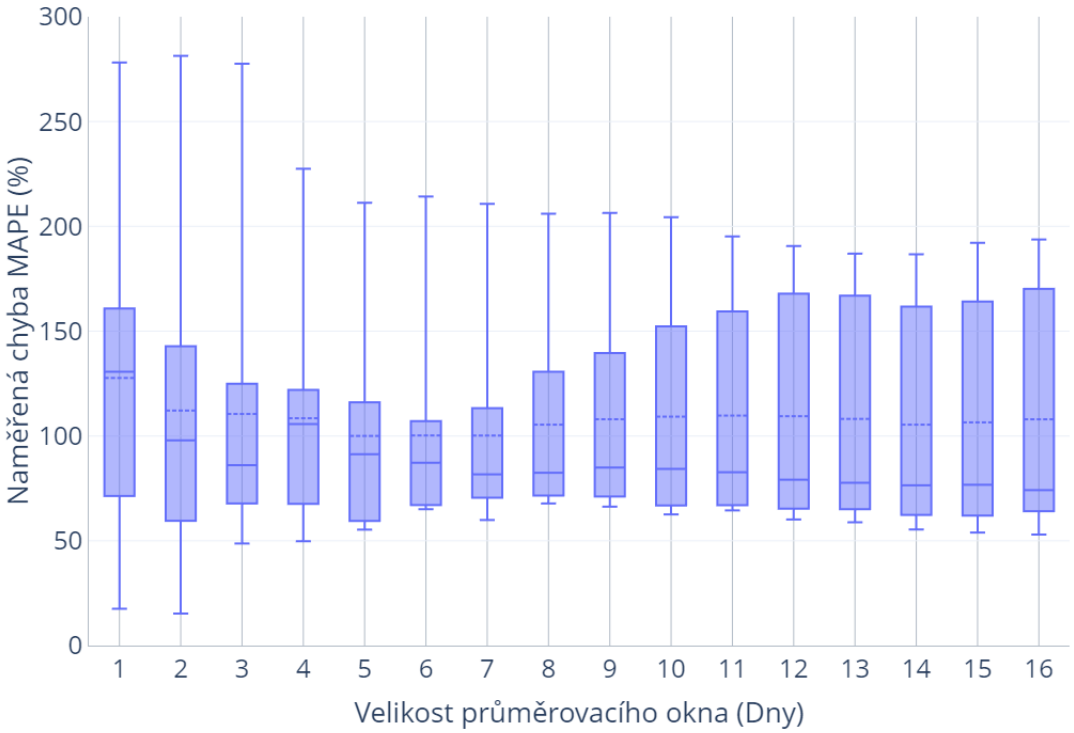
3.4.1 Velikost průměrovacího okna v klouzavém průměru

Při práci s touto metodou bylo nutné zvolit správnou délku průměrovacího okna. Ta říkala kolik předchozích hodnot bude použito pro výpočet další hodnoty. Zde bylo však velmi důležité pochopit, že pokud byla používána verze této metody s parametrem sezónnosti, tak se délka sezóny a velikost průměrovacího okna navzájem ovlivňovaly. To bylo způsobeno tím, že u týdenní sezónnosti měla vstupní data zpoždění minimálně sedmi dní, zatímco u denní sezónnosti to byl pochopitelně jen jeden den. Proto bylo nutné nalézt vhodné okno jak pro denní, tak i pro týdenní sezónnost.



Obrázek 3.8: Analýza průměrovacího okna pro týdenní sezónnost

Pro graf na obrázku 3.8 bylo pokaždé provedeno osm měření pro několik velikostí průměrovacího okna. Tyto velikosti se vyskytovaly v rozmezí od jedné do osmi a v tomto případě byla délka sezóny nastavena na jeden týden. To znamená, že v případě předpovědi čtvrtka bylo provedeno průměrování měření z n předchozích čtvrtků. V grafu byla tato měření vyjádřena ve formě chyby MAPE. Bylo zde vcelku dobře vidět, že délka průměrovacího okna zde neměla takový vliv, jak bylo nejprve očekáváno. Nejhůře na tom bylo okno délky jedna, ale ostatní si byla dosti podobná. Nejlépe na tom podle průměru měření bylo okno délky tři, což odpovídá skoro jednomu měsíci. V budoucích měřeních, kdy byla délka sezóny stanovena na jeden týden, byla díky tomuto průzkumu hodnota délky průměrovacího okna nastavena na poslední tři týdny.

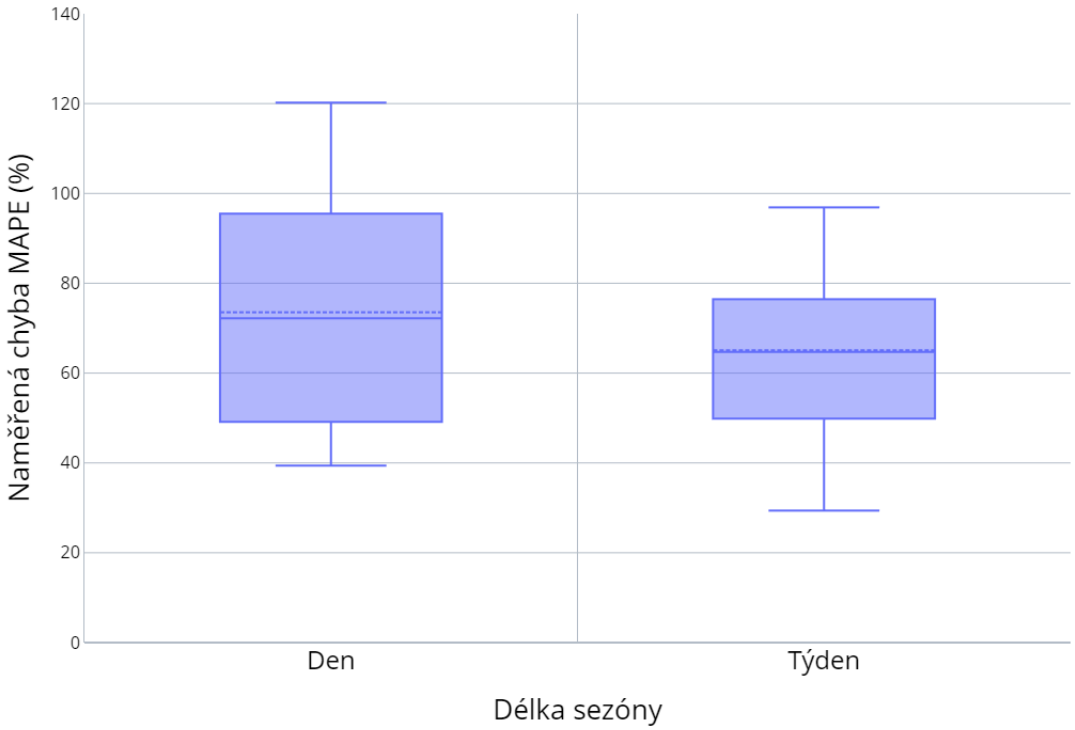


Obrázek 3.9: Analýza průměrovacího okna pro denní sezónnost

Naopak v grafu na obrázku 3.9 byla délka sezóny nastavena na jeden den, což znamenalo pouhé průměrování hodnot z n předchozích dnů bez jakékoliv jiné specifikace. Ostatní parametry zde byly stejné jako u předchozího měření a výsledkem byl opět graf zobrazující MAPE. Velmi krátká okna zde vykazovala největší chybu. Zajímavostí bylo to, že u oken, která byla násobkem sedmi, byla chyba lehce menší oproti svému okolí. To odpovídalo nejlepší odhadované hodnotě pro délku sezóny z autokorelace. Nejlépe na tom v tomto případě byla délka průměrovacího okna 7 dní. Délka okna zde měla navíc vliv i na rozptyl jednotlivých chyb, kde větší velikosti vykazovaly větší rozptyl. To bylo způsobeno zpožděním, které bylo v těchto případech do predikce dodáno staršími daty.

3.4.2 Délka sezóny v klouzavém průměru

V sekci 3.4.1 již byl vysvětlen vzájemný vliv průměrovacího okna a sezónnosti. Proto bylo průměrovací okno během měření zvoleno pro každou délku sezóny individuálně dle předchozího průzkumu. Pro týdenní sezónnost to byla hodnota tři a pro denní jí byla hodnota čtrnáct.

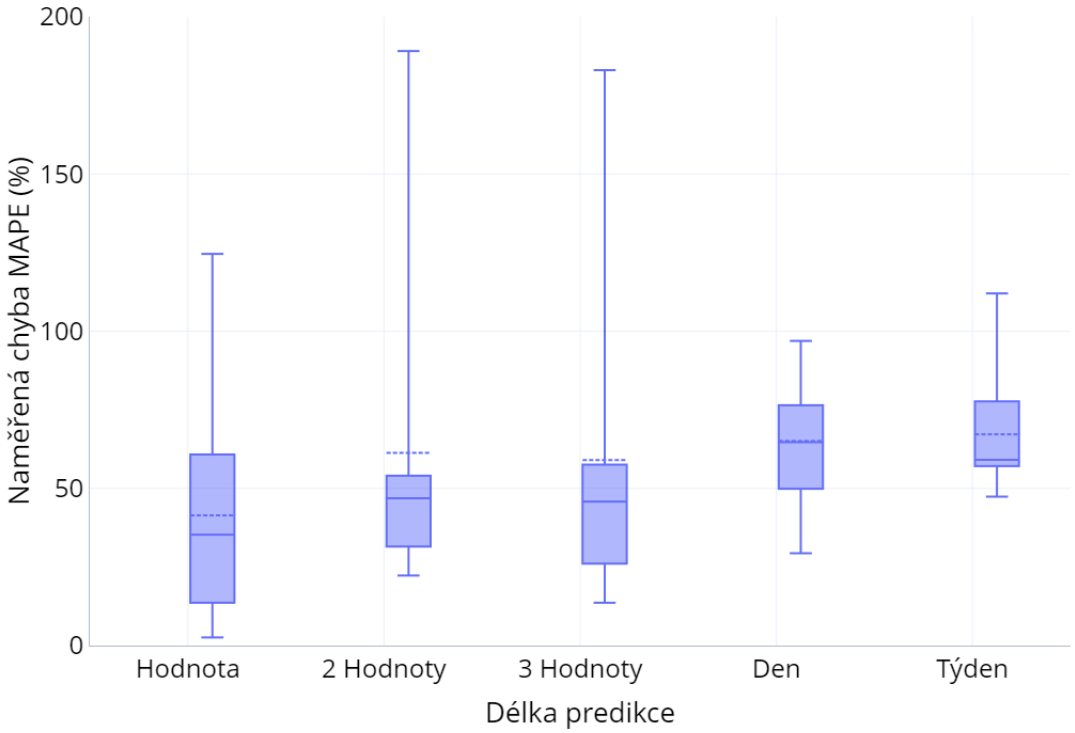


Obrázek 3.10: Analýza parametru sezónnosti u metody klouzavého průměru

Zde se poprvé potvrdila informace z autokorelační funkce o vhodném parametru sezónnosti. Na obrázku 3.10 je velmi dobře vidět rozdíl mezi délkou sezóny jednoho dne a jednoho týdne. Právě týdenní délka sezóny zde vykazovala průměrný pokles chyby o 8 procentních bodů, což je v rámci práce významný rozdíl. Je však nutné připomenout, že i predikce s délkou sezóny pouze jednoho dne nesla informaci o týdenním cyklu spotřeby ve formě velikosti průměrovacího okna 7 dní.

3.4.3 Délka predikovaného období v klouzavém průměru

Nakonec bylo u této metody provedeno i měření přesnosti dle délky predikce. Kvůli principu fungování zkoumaného algoritmu však bylo možné pozorovat predikce maximálně jednoho týdne, jelikož byla zvolena pouze týdenní sezónnost. Pokud by byla délka sezóny však nastavena až na jeden rok, bylo by možné predikovat pro celý další rok. Navíc již bylo pomocí lineární regrese stanoveno očekávání většího rozptylu chyb u kratších období.



Obrázek 3.11: Analýza délky predikce u klouzavého průměru

V obrázku 3.11 se tedy opět promítla lehká nestabilita svázaná s predikováním velmi krátkých období. Jinak zde byl vidět mírný růst chyb pro delší predikovaná období. Tento růst mohl být způsoben zpožděním, které bylo určeno délkou průměrovacího okna.

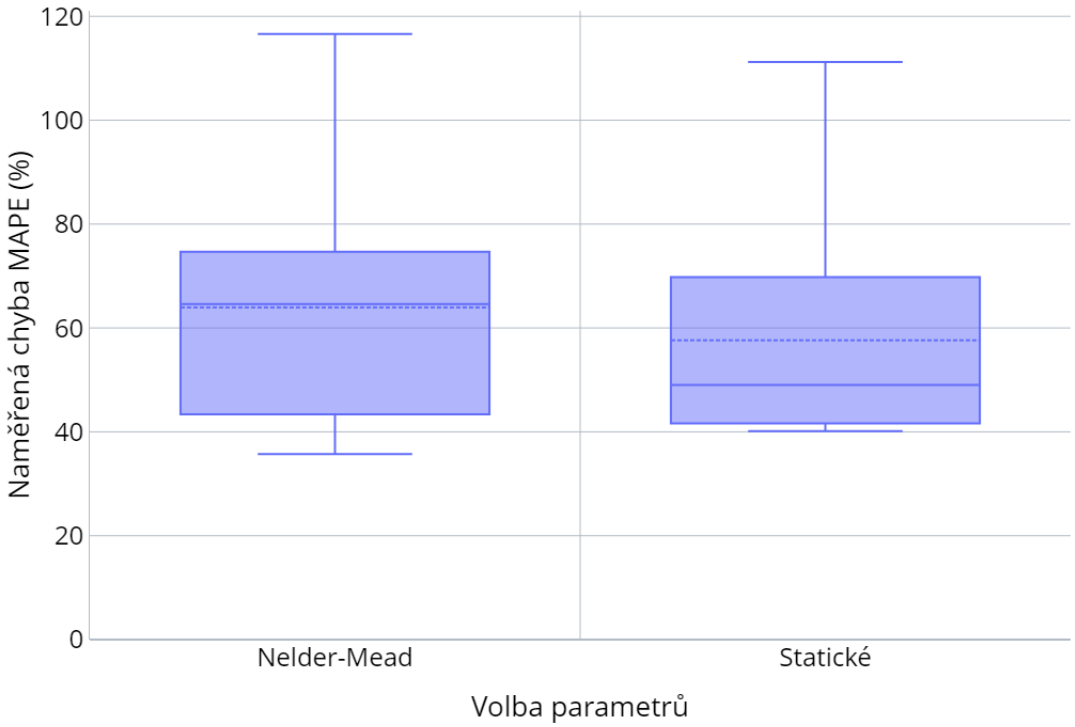
3.5 Analýza metody Holt-Winters

Metoda Holt-Winters byla stejně jako vyhlazování klouzavým průměrem implementována ručně. Jedním z důvodů pro tuto implementaci byla opět lepší kontrola nad fungováním této metody. Dalším důvodem bylo to, že spousta knihoven v prostředí .NET poskytovala pouze jednoduché exponenciální vyhlazování a někdy dvojité. Jelikož data však vykazovala sezónnost, byla potřeba trojitá verze tohoto vyhlazování, která tak dostupná nebyla. Implementace byla provedena pro multiplikativní verzi této metody, jelikož rozptyl dat byl proporcionální k jejich okolí.

3.5.1 Volba vyhlazovacích parametrů Holt-Winters

Jak již bylo zmíněno v teoretické části, tak tato metoda vyžadovala stanovení tří parametrů alfa, beta a gama. Toto nastavení je často prováděno pomocí různých vyhledávacích algoritmů. Zde byl zvolen optimalizační algoritmus Nelder-Mead, který

minimalizoval MAPE pro predikci posledního týdne v trénovacích datech. Navíc byly manuálním hledáním zvoleny počáteční odhady 0,01, 0,01, 0,01. V průběhu práce však tato optimalizace způsobila množství anomálií, které práci s metodou spíše zkomplikovaly. Z tohoto důvodu bylo provedeno porovnání předpovědí při použití optimalizačního algoritmu a při využití statických hodnot. V případě statických hodnot se jednalo o již zmíněné počáteční odhady.

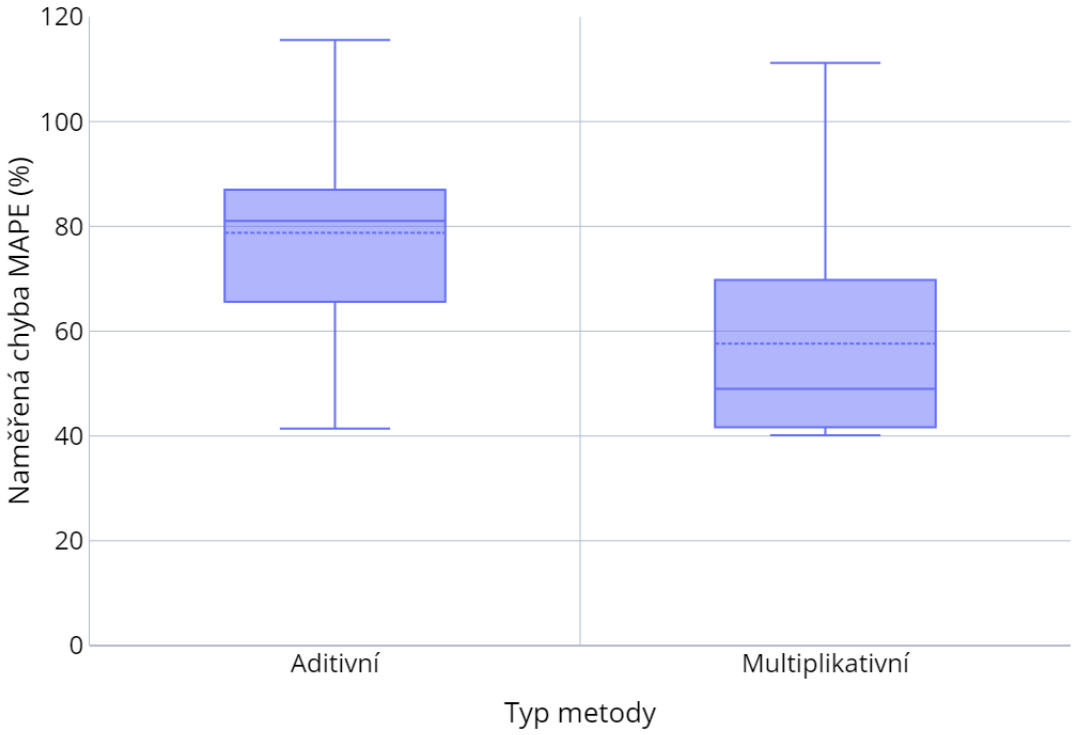


Obrázek 3.12: Analýza volby vyhlazovacích parametrů u modelu Holt-Winters

Na obrázku 3.12 je vidět malé zlepšení predikcí při použití statických hodnot. Algoritmus Nelder-Mead nedokázal prozkoumat celý možný interval optimalizovaných parametrů a navíc prováděl overfitting. To znamená, že nastavil metodu spíše pro trénovací data, než pro budoucí predikce. Dále byl problém s navýšením časové náročnosti výpočtu při provádění optimalizací. Místo instantního výpočtu zde metoda trvala průměrně tři vteřiny. To je sice málo, ale tento algoritmus se mohl zaseknout v lokálním minimu a způsobit tak neočekávané chování. Proto byl nakonec k dalším měřením použit přístup s manuálně nastavenými hodnotami (0,01, 0,01, 0,01) a Nelder-Mead byl pouze volitelným nastavením v používané testovací aplikaci.

3.5.2 Typ metody Holt-Winters

Zatímco všechna další měření s touto metodou využívali její multiplikativní verzi, zde bylo provedeno porovnání mezi tímto a aditivním výpočtem. Až na typ metody byly všechny parametry nastaveny na výchozí hodnoty.



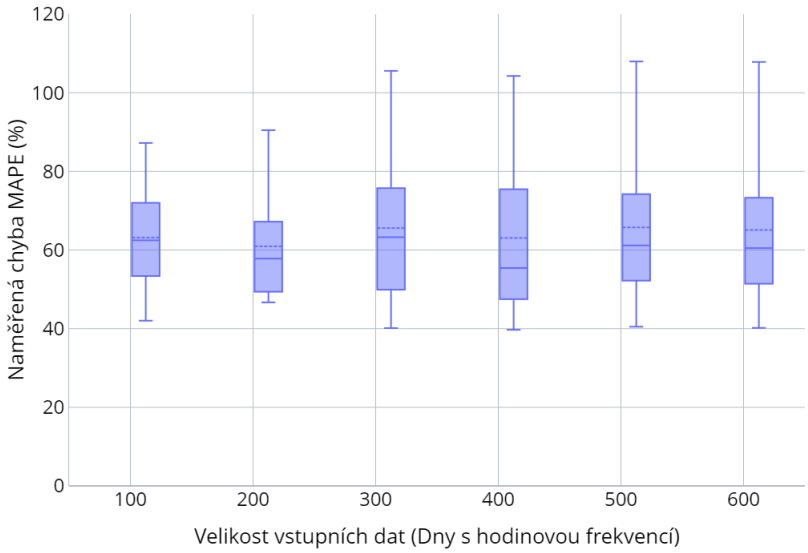
Obrázek 3.13: Analýza typů výpočtu u modelu Holt-Winters

Rozdíl zde byl významný. Jak lze vidět z chybové metriky MAPE v grafu na obrázku 3.13, tak metoda multiplikativní vykazovala zlepšení od metody aditivní v průměru o 21 procentních bodů. Multiplikativní výpočet však vykazoval trochu větší náchylnost k provádění anomálních predikcí. Tato vlastnost je viditelná i v dalších prováděných měřeních s touto metodou.

Zde je důležité zmínit, že z počátku byl využíván především výpočet aditivní, ale tento průzkum pomohl najít chybu v tomto přístupu. Proto byla metoda upravena tak, aby ve všech měření používala výpočet multiplikativní.

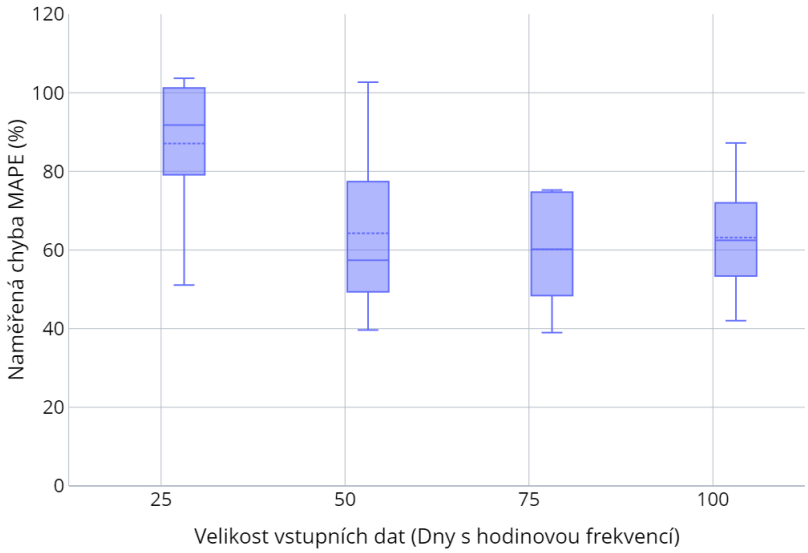
3.5.3 Rozsah vstupních dat v metodě Holt-Winters

Pro trojité exponenciální vyhlazování byl opět proveden průzkum dle rozsahu vstupní datové sady jako u lineární regrese. Opět byly vyzkoušeny velikosti od 100 do 600 dnů s inkrementací po 100 dnech a byly použity výchozí parametry s multiplikativním výpočtem.



Obrázek 3.14: Analýza rozsahu vstupních dat u modelu Holt-Winters

Velikost vstupních dat zde neměla téměř žádná vliv. Jak lze pozorovat na obrázku 3.14, tak všechny velikosti vstupních dat měli velmi podobnou průměrnou chybu. Nejlépe na tom byly překvapivě velikosti 100 a 200 dnů, které měly lehce užší rozptyl chyb. To bylo způsobeno exponenciálním přiřazováním vah pro jednotlivá měření, které velmi stará data v podstatě ignorovalo.

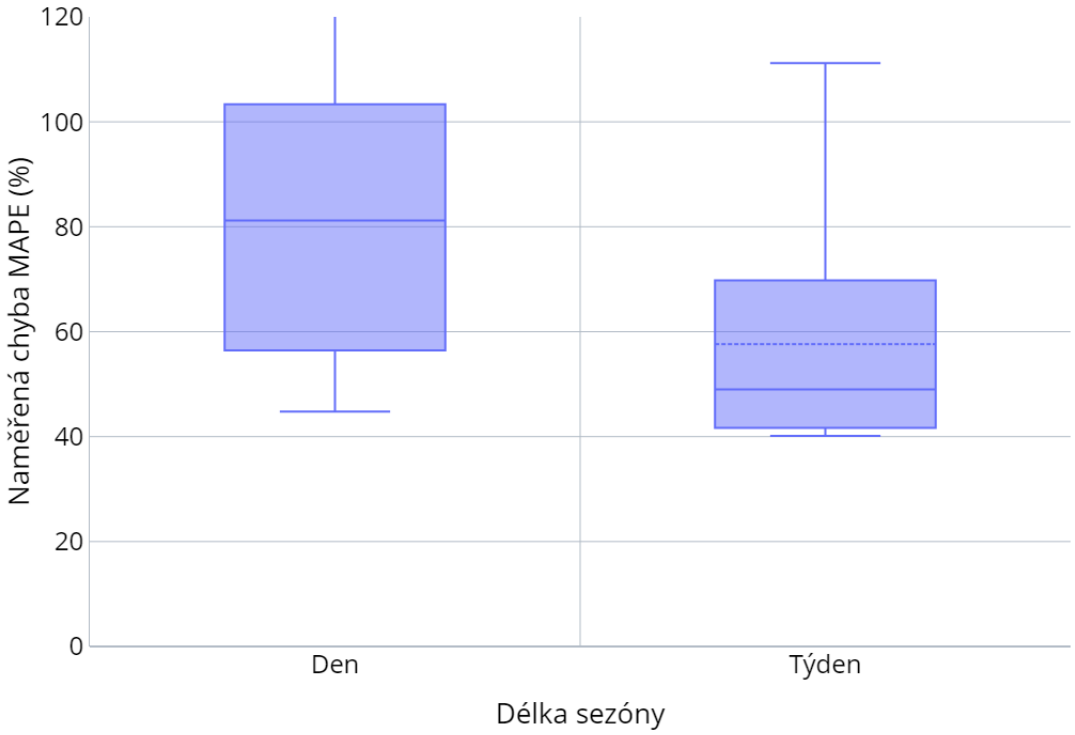


Obrázek 3.15: Analýza menšího rozsahu vstupních dat u modelu Holt-Winters se statickými parametry

Pro lepší vhled bylo provedeno pozorování pro ještě menší rozsahy 25 až 100 dní. Graf na obrázku 3.15 dobře ukazuje významné zhoršení predikcí při použití nedostatečných vstupních dat v rozsahu 25 dní. Chyba MAPE zde byla větší v průměru o 23 procentních bodů. To byl již velký rozdíl, který naznačoval velký pokles v kvalitě predikcí při malém množství trénovacích dat.

3.5.4 Parametr sezónnosti v metodě Holt-Winters

Jako u klouzavého průměru i zde byl proveden průzkum parametru sezónnosti. Zde byly opět všechny ostatní parametry nastaveny do výchozích hodnot a byl použit multiplikativní výpočet. Z provedené autokorelační analýzy bylo očekáváno zlepšení predikcí u týdenní sezónnosti stejně jako u klouzavého průměru.

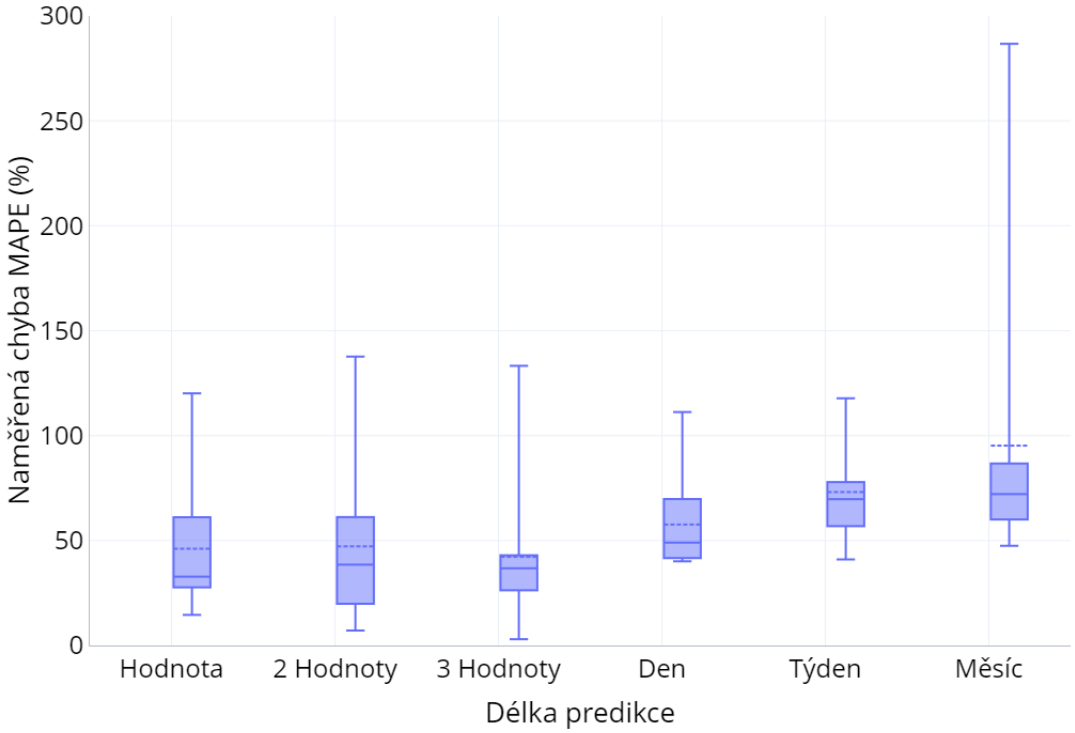


Obrázek 3.16: Analýza parametru sezónnosti u modelu Holt-Winter

Délka sezóny zde byla opět významným parametrem, který prokazoval významně větší přesnost při užití týdenní sezónnosti. Měření u denní délky zde vykazovaly vysokou nestabilitu a chyba dosahovala až 2761,7 %, což naznačovalo kompletní selhání algoritmu. Výskyt těchto nečekaně vysokých chyb znemožnil provedení porovnání průměrných chyb mezi sezónnostmi, proto byl alespoň porovnán medián jednotlivých měření. Týdenní délka sezóny u této metody snížila chybu MAPE o 32 procentních bodů. Opět se tedy prokázala důležitost provedení datové analýzy před prováděním predikcí.

3.5.5 Délka predikovaného období v metodě Holt-Winters

Stejně jako u předchozích modelů bylo provedeno pozorování predikcí vzhledem k jejich délce.



Obrázek 3.17: Analýza délky predikce u modelu Holt-Winters

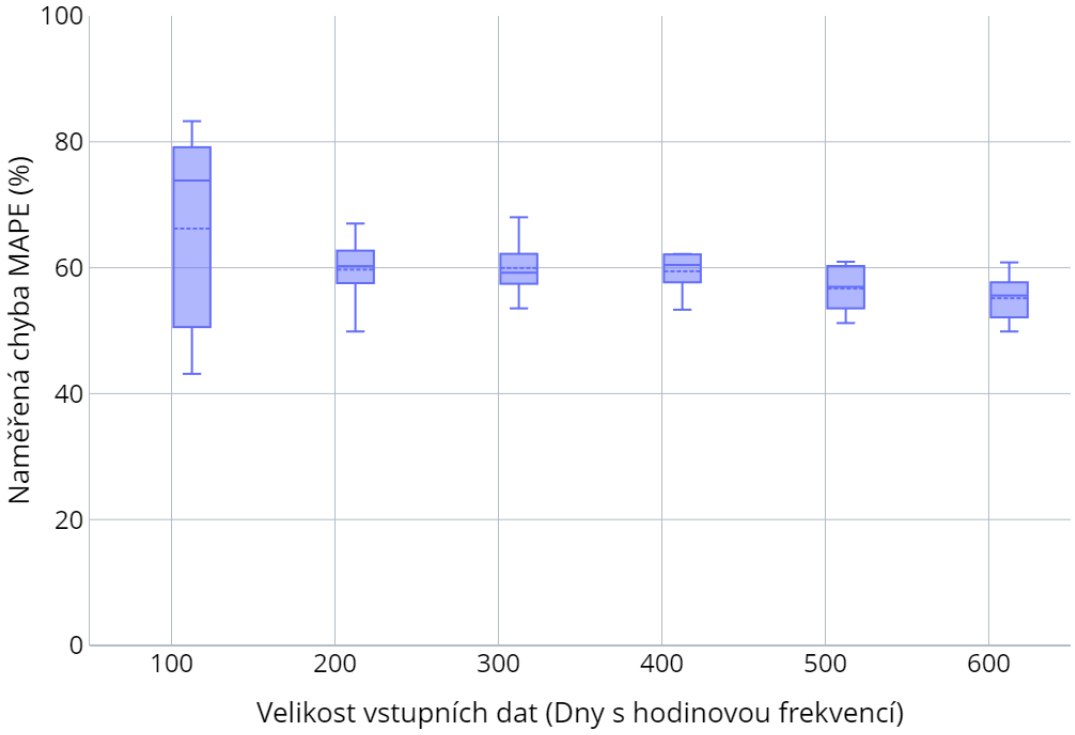
V grafu na obrázku 3.17 jsou vidět chyby MAPE pro jednotlivá období. Je zde opět vidět, že v případě kratších období byl rozptyl chyb často větší, ale již se nejednalo o takový rozdíl, jaký byl pozorován u předchozích metod. Nárůst chyby s rostoucí délkou predikovaného období zde byl rychlejší, pomocí čehož lze potvrdit již předpokládanou vhodnost metody pro provádění krátkodobějších predikcí. Tento rychlejší růst chyby byl způsoben především využitím vlastních predikcí k předvídání dalších dat a již vzniklé chyby se tak dostávali do budoucích predikcí.

3.6 Analýza metody Singular Spectrum Analysis

Metoda SSA byla opět implementována pomocí knihovny ML.NET. Její využití bylo v celku jednoduché, jelikož bylo nutné pouze nastavit délku sezóny, predikce a celkovou velikost trénovacích dat. To jsou vše parametry, které již byly dobře známy a celková práce s knihovnou byla podobná jako u lineární regrese.

3.6.1 Rozsah vstupních dat v metodě SSA

Byl proveden průzkum vlivu rozsahu vstupních dat s rozsahem 100-600 dnů. Metoda však zvládala i mnohem větší rozsahy dat, pokud se nejednalo o příliš velká data.

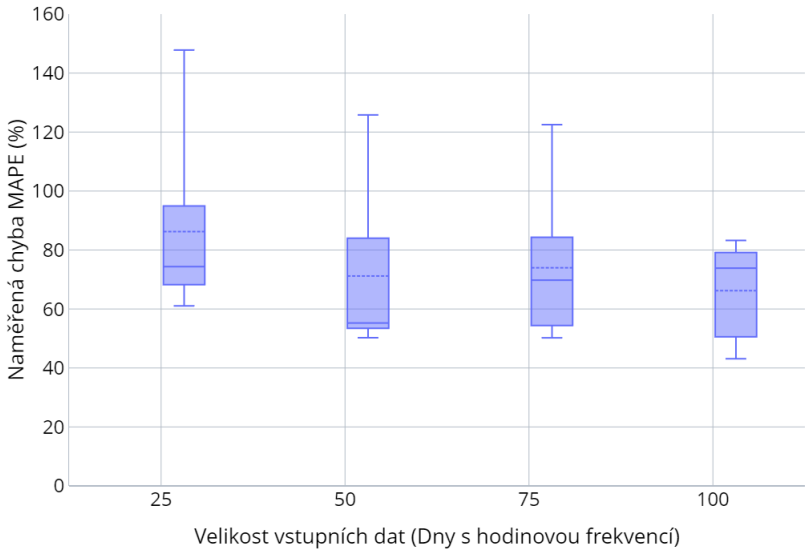


Obrázek 3.18: Analýza rozsahu vstupních dat u modelu SSA

Z grafu na obrázku 3.18 je vidět pomalý pokles chyby s větším množstvím vstupních dat. Navíc bylo zaznamenáno i veliké zúžení rozptylu chyb mezi rozsahy 100 a 200 dnů. Právě tato velká změna si vyžádala další prozkoumání menších rozsahů 25-100 dnů.

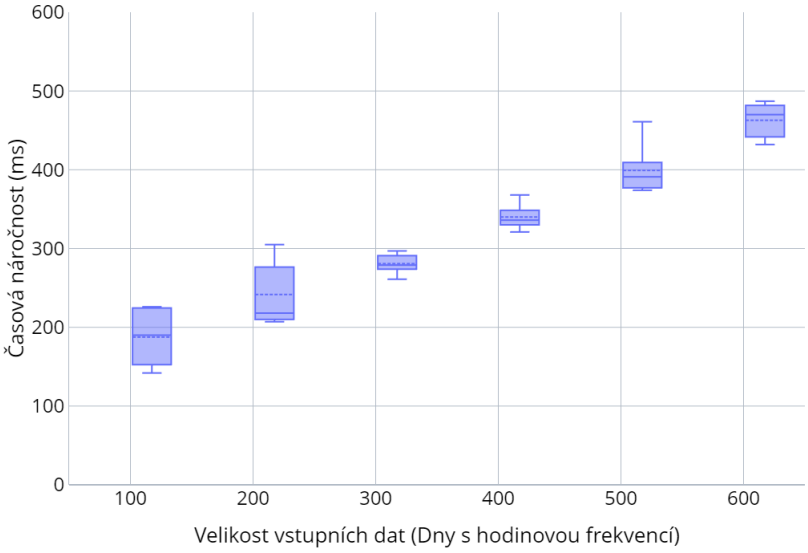
Z obrázku 3.19 lze opravdu vypozorovat, že pokud byla použita příliš malá datová sada, tak rychle klesala efektivita této metody. Tyto dva grafy dokazují, že čím více dat je SSA dodáno, tím lepší budou predikce. Tento trend byl navíc viditelný i při

poskytnutí většího objemu dat v jednotkách 500 a 600 dnů, což bylo pro zkoumané metody unikátní. Promítá se zde tedy schopnost zachycení dlouhodobějších vzorců ze vstupních dat.



Obrázek 3.19: Analýza menšího rozsahu vstupních dat u modelu SSA

Také byla pro jednotlivé velikosti vstupních dat naměřena časová náročnost, která však u tohoto modelu nedosahovala příliš vysokých hodnot.

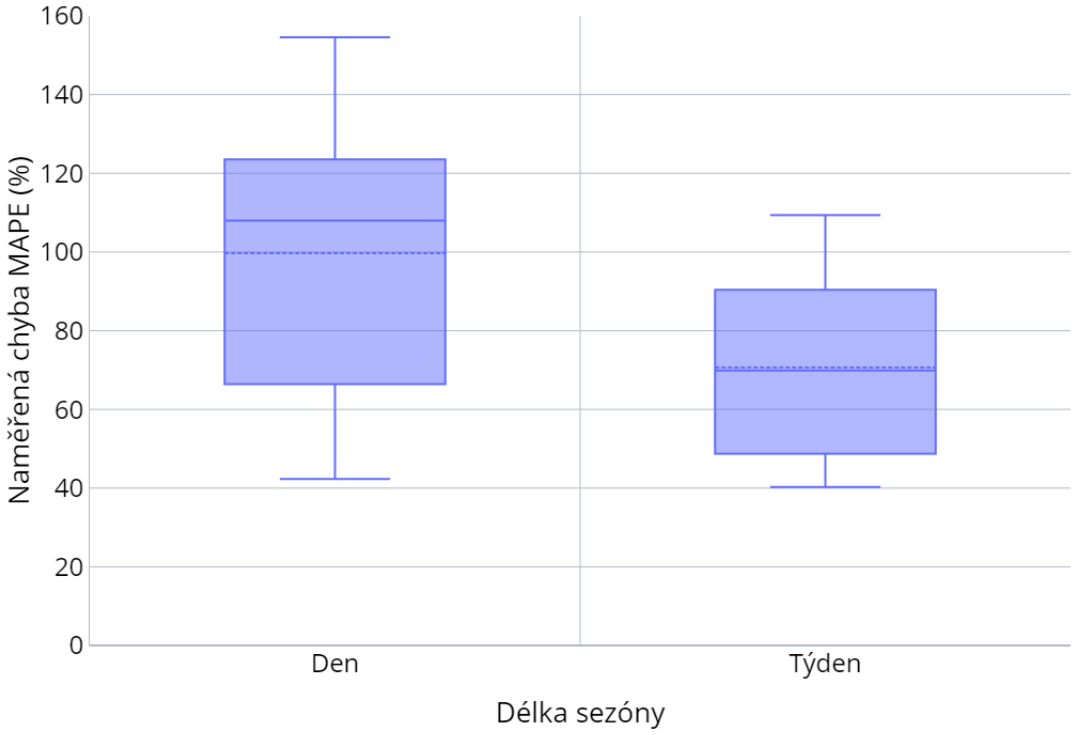


Obrázek 3.20: Analýza časové náročnosti dle rozsahu vstupních dat u modelu SSA

Zde u modelu SSA vykazoval vztah časové náročnosti a velikosti vstupních dat lineární průběh, jak lze vidět z grafu na obrázku 3.20. Jelikož však metoda pro malé množství dat, v tomto případě pro 100 dní, vyžadovala v průměru pouze 187 milisekund, trval výpočet i pro větší data krátkou dobu.

3.6.2 Parametr sezónnosti v metodě SSA

Jako u předchozích metod, i zde byl proveden průzkum vlivu délky sezóny. Protože opravdu všechny parametry zde byly stejné, na rozdíl od délky okna v klouzavém průměru, bylo zde jisté očekávání výrazně lepších výsledků u týdenní sezónnosti.

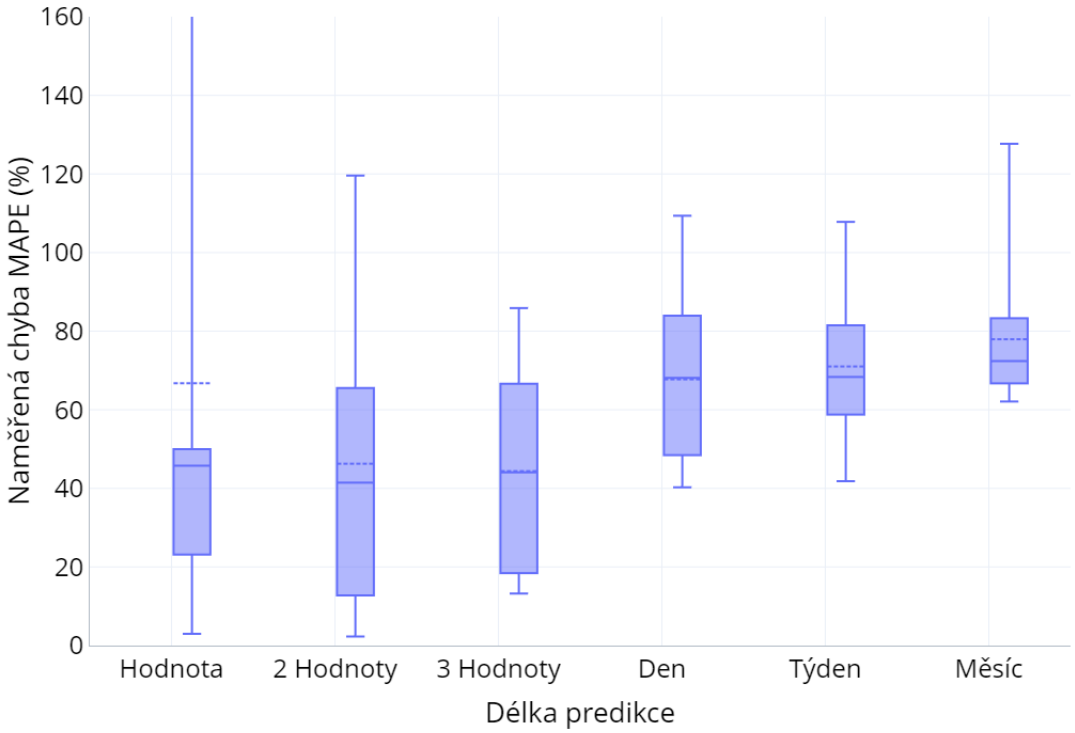


Obrázek 3.21: Analýza parametru sezónnosti u modelu SSA

Tentokrát byl rozdíl opravdu dobře vidět, jak lze vypozorovat z obrázku 3.21. Denní sezónnost vykazovala významně vyšší chybu MAPE v podobě jejího průměrného nárůstu o 32 procentních bodů. To již byl velmi velký rozdíl, který znovu podporoval výsledky získané z autokorelační analýzy a předchozích metod.

3.6.3 Délka predikovaného období v metodě SSA

Nakonec byl proveden stejný průzkum dle délky predikovaného období jako u předchozích metod.



Obrázek 3.22: Analýza délky predikce u modelu SSA

Z obrázku 3.22 lze opět vypozorovat již známé chování rozptylu chyb u kratších období. Navíc se zde opět vyskytl trend růstu chyby při provádění delších predikcí. Tento trend byl však velmi slabý, takže metoda dokázala provádět i delší predikce s velmi malým růstem chyby. Metoda byla tedy vhodná pro provádění i střednědobých predikcí.

3.7 Analýza metody ARIMA

ARIMA byla naopak od SSA jednou z komplikovanějších metod na zprovoznění, jelikož byla příliš složitá pro vlastní implementaci a žádná dostupná knihovna ji neposkytovala. Proto byl zvolen vcelku nezvyklý přístup s použitím jazyka Python a jeho knihovny statsmodels. Sice zde byl pokus využít metodu auto_arima z knihovny pmdarima, ale ta se ukázala jako ještě hardwarově náročnější a při testování netvořila vhodné predikce.

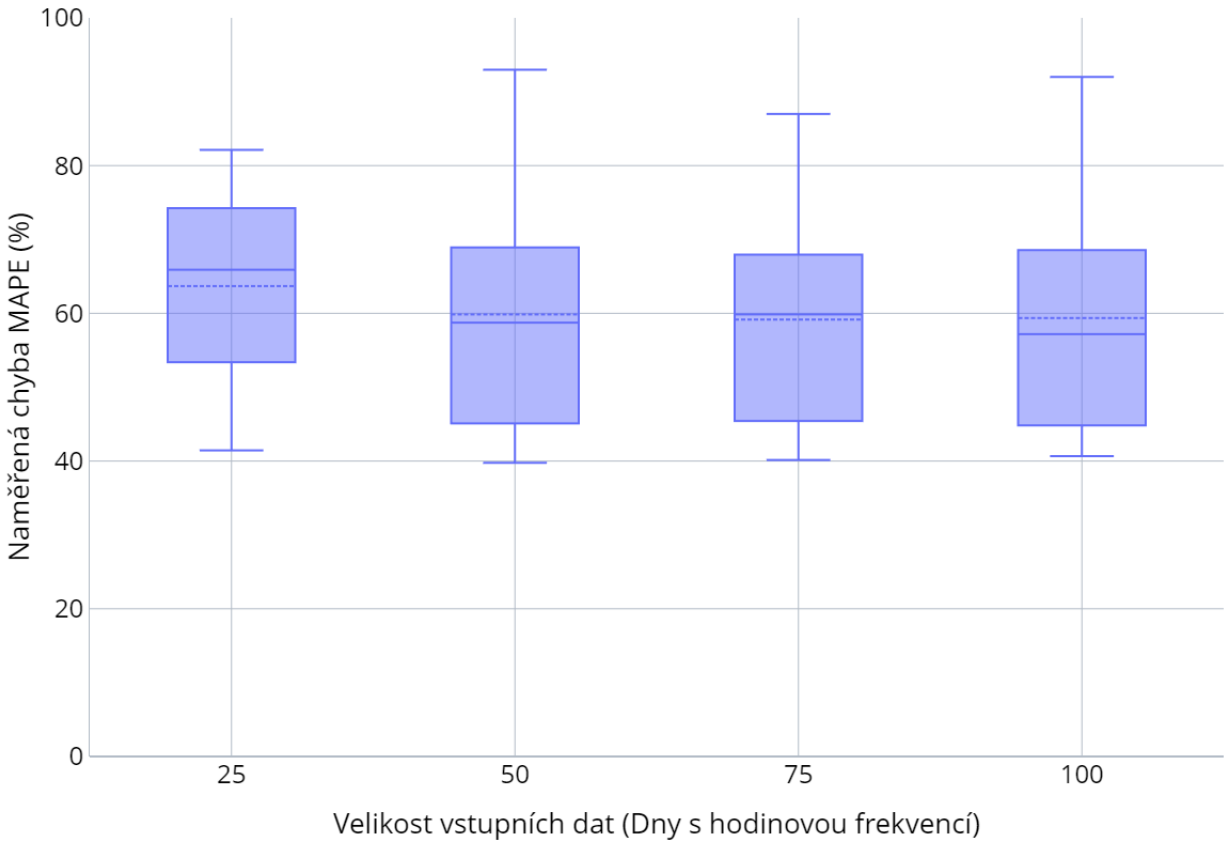
Největším problémem bylo nastavení parametrů tohoto modelu. To bylo provedeno pomocí autokorelační a částečné autokorelační funkce. Jelikož tento přístup nebylo možné využít pro nastavení všech parametrů v jednom společném modelu, bylo dodatečné nastavení nutno provést iterativním testováním a manuálním výbě-

rem. Dále zde byl problém s velikostí trénovacích dat, kdy bylo možné využít pouze posledních 100 dnů měření. Toto omezení bylo způsobeno poměrně dlouhou délkou sezóny 168, se kterou má ARIMA často problémy.

Model byl používán ve tvaru SARIMA(1,0,1)(0,1,1,168). To znamená, že ve výchozím nastavení byla používána délka sezóny 168, což odpovídalo jednomu týdnu. Následně byla prováděna sezónní diference, která byla potřebná pro transformaci vstupní časové řady do stacionární podoby. Nakonec byl prováděn klouzavý průměr pro sezónní i nesezónní část modelu a nesezónní autoregrese. Jednalo se o model vybraný pomocí autokorelace, částečné autokorelace a manuálního testování. Dosahoval nejlepších výsledků v rozumném čase.

3.7.1 Rozsah vstupních dat v metodě ARIMA

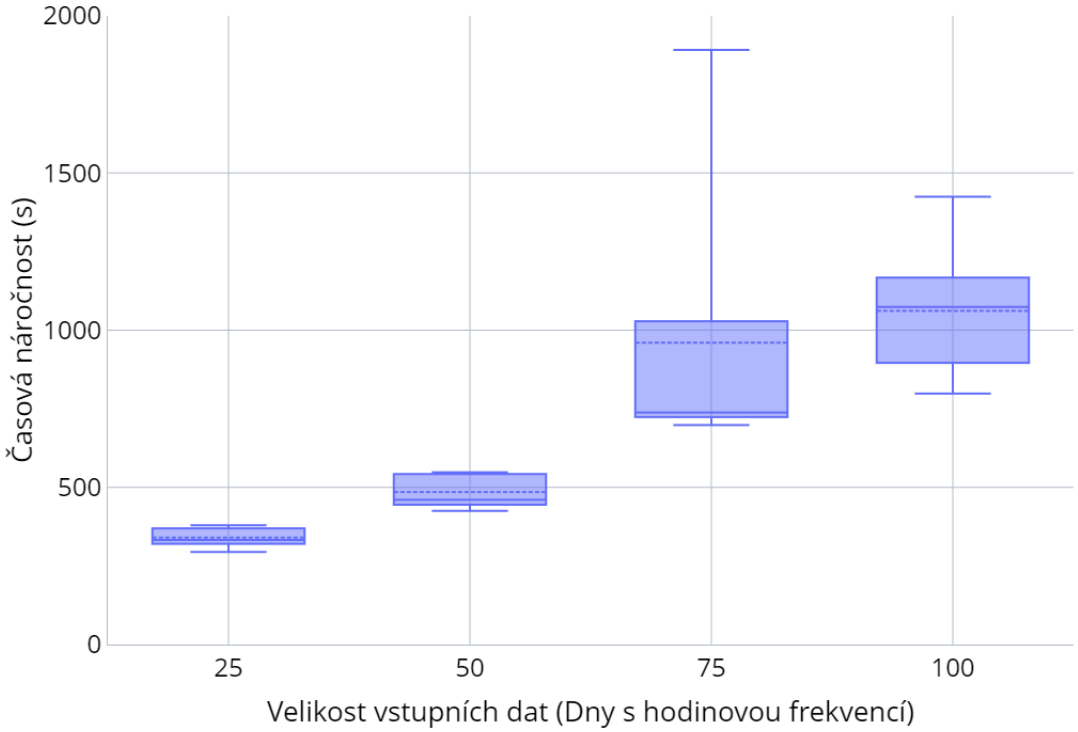
Jako u ostatních metod, i zde byl zkoumán vliv rozsahu vstupních dat na predikce. Rozsah dat zde byl omezen na rozmezí 25 až 100 dní z důvodu vysoké časové náročnosti a častých selhání při využití větších dat. Jinak byly využity výchozí parametry.



Obrázek 3.23: Analýza rozsahu vstupních dat u modelu ARIMA

V grafu na obrázku 3.23 lze vidět jen malý pokles chyby mezi rozsahy 25 a 50 dnů a absence dalšího poklesu u větších rozsahů. Metodě zde stačilo velmi malé množství dat k provádění dobrých predikcí, avšak kvůli výpočetní náročnosti nešlo vyzkoušet větší rozsahy a nebylo možné zjistit, zda mnohem větší trénovací data tyto predikce ještě zpřesní. V dalších měření byl rozsah trénovacích dat nastaven na 50 dní, jelikož se jednalo o nejméně časově náročnou velikost s dobrými výsledky.

Časová náročnost byla samozřejmě opět změřena a tentokrát se jednalo o důležitou vlastnost, jelikož tato metoda byla zcela nejpomalejší ze všech.

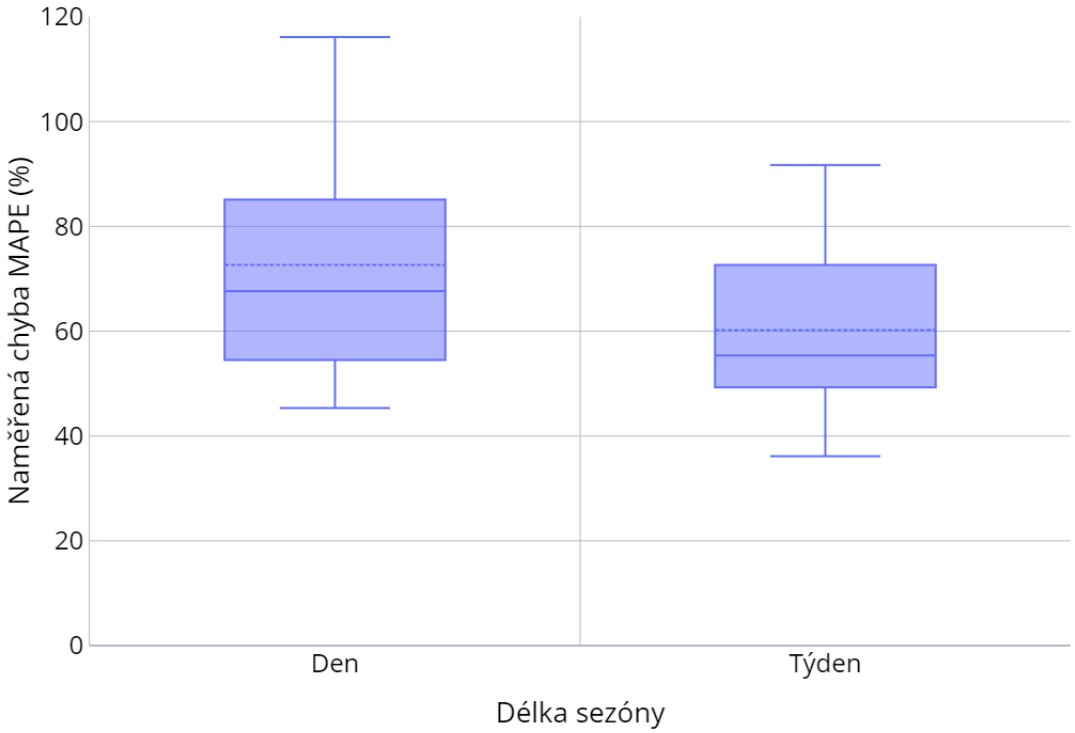


Obrázek 3.24: Analýza časové náročnosti dle rozsahu vstupních dat u modelu ARIMA

Jak lze z grafu na obrázku 3.24 velmi dobře vidět, dosahovalo trvání výpočtu velmi vysokých hodnot. Navíc tato časová náročnost rychle rostla s přibývajícím množstvím vstupních dat a v případě přesnějšího nastavení parametrů modelu na vyšší hodnoty tento výpočet zpomalil ještě více. Celkově se jednalo o velmi náročnou metodu, která byla ještě více zkomplikována implementací přes propojení s jazykem Python.

3.7.2 Parametr sezónnosti v metodě ARIMA

Přestože byl vliv tohoto parametru již ověřen, byl stále prozkoumán i u této metody. Ostatní parametry byly nastaveny na výchozí hodnoty a bylo použito 50 dní dat.

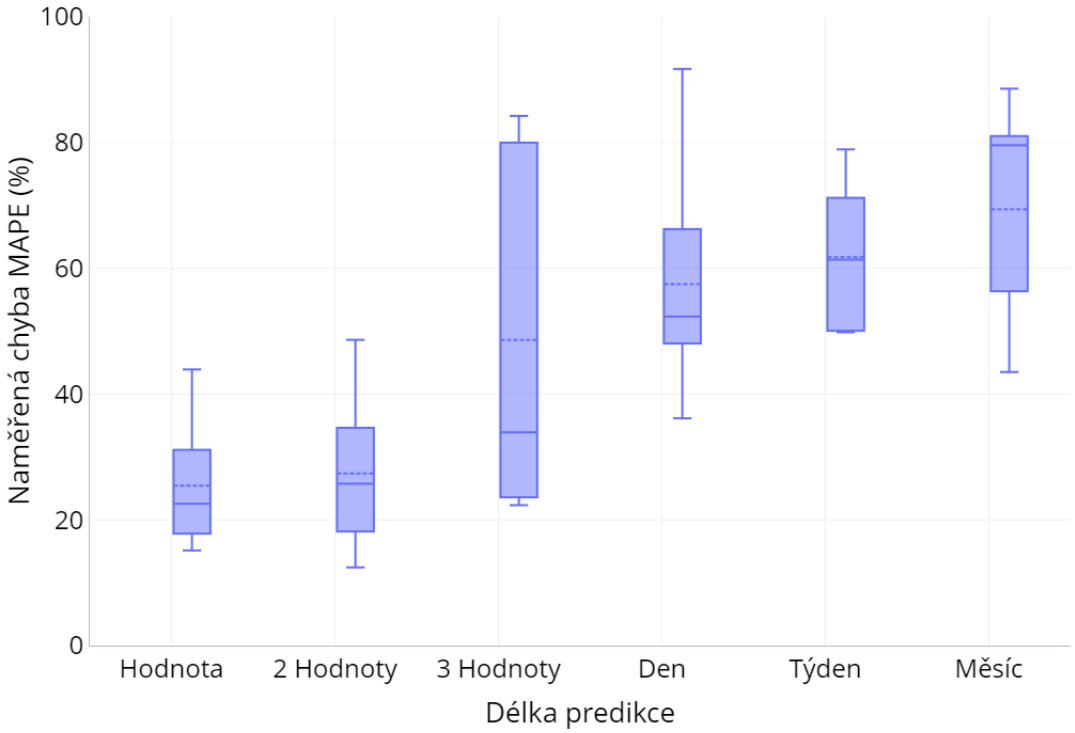


Obrázek 3.25: Analýza parametru sezónnosti u modelu ARIMA

Z obrázku 3.25 je vidět, že vliv zde tento parametr měl a dokonce i takový, jaký byl předpokládán, avšak nejednalo se o tak veliký rozdíl, jaký byl pozorován u SSA a Holt-Winters. Týdenní sezónnost zde vykazovala menší chybu. V tomto případě to bylo snížení MAPE o pouze 15 procentních bodů.

3.7.3 Délka predikovaného období v metodě ARIMA

Stejně jako u všech předchozích metod byla i zde prozkoumána tato charakteristika se stejnými očekáváními.



Obrázek 3.26: Analýza délky predikce u modelu ARIMA

Z grafu na obrázku 3.26 lze pozorovat stejné chování rozptylu jako u předchozích metod a vyskytoval se zde trend růstu chyby při provádění delších predikcí. Tento trend však nebyl příliš silný a metoda dokázala provádět i střednědobé predikce.

3.8 Globální porovnání metod

S hotovou individuální analýzou jednotlivých metod bylo možné provést jejich vzájemné porovnání. Byla zvážena jejich přesnost, časová náročnost a složitost jejich využití

3.8.1 Porovnání metod dle přesnosti

Nejpožadovanější kvalitou metod pro predikce dat je jejich přesnost. Ta sice již byla uvažována v individuální analýze, ale pouze k prozkoumání vlivu specifických parametrů. Tato sekce se naopak zabývá samotnou přesností při nejlepším nastavení metod za použití výchozích parametrů ze sekce 3.2 . Jsou zde využity metriky RMSE, MAE a MAPE k porovnání jednotlivých metod.

Tabulka 3.2: Průměrné chyby zkoumaných metod

Metoda	RMSE	MAPE	MAE
Sarima (1,0,1)(0,1,1,168)	0,61	57,47	0,45
Holt-Winters	0,62	57,63	0,47
SSA	0,66	67,75	0,50
Klouzavý Průměr	0,75	65,08	0,55
Lineární Regrese	1,00	104,08	0,78

V tabulce 3.2 je vidět porovnání jednotlivých metod. Grafické zobrazení lze nalézt v grafu A.4 .

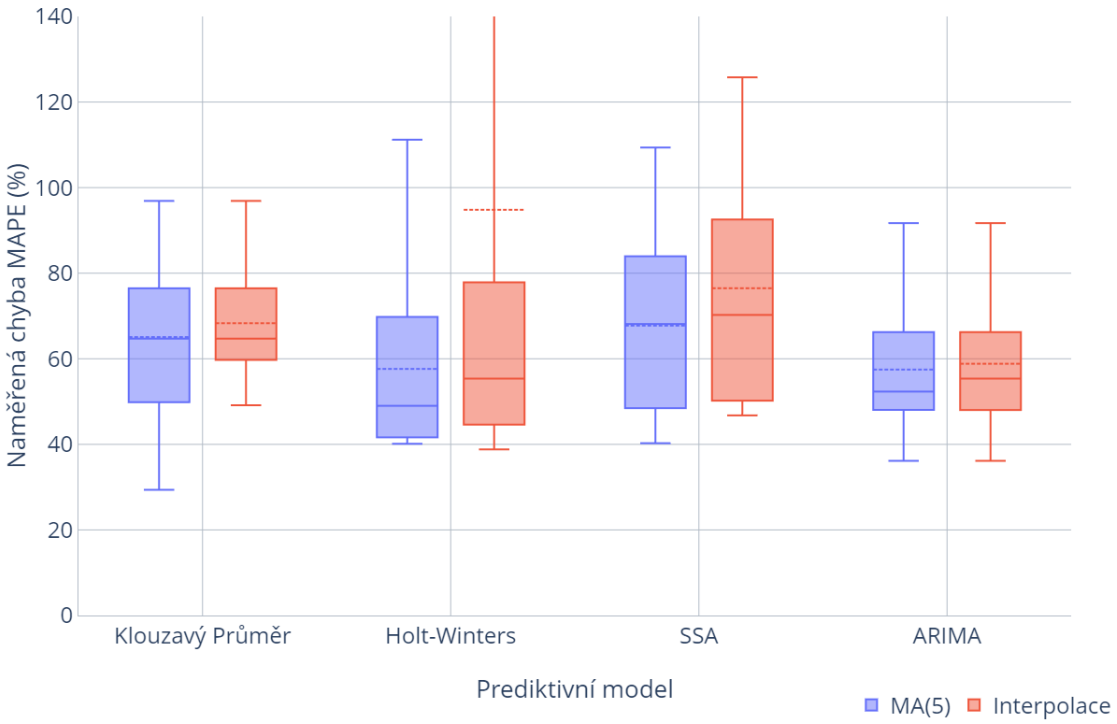
Podle očekávání na tom byla opravdu nejhůře metoda lineární regrese, jelikož všechny chybové metriky u ní dosahovaly nejvyšších hodnot. Ta sice má své využití, ale v oblasti predikcí se jedná pouze o jednoduchou metodu k nalezení trendu časové řady. Neumí pracovat se samostatnými variacemi ve vstupních datech. Proto byla tato metoda zvolena spíše jako základ k porovnání s ostatními metodami. Jednotlivé chybové metriky zde dosahovaly velmi vysokých hodnot.

Ostatní metody si však již byly mnohem blíže. Nejhůře na tom byla po lineární regresi překvapivě metoda SSA. Ta sice reagovala na různá nastavení očekávatelně, ale nedařilo se jí modelovat specifické chování dat. V porovnání s lineární regresí však stále došlo k výraznému zlepšení, kde se MAPE snížilo o 33 procentních bodů. Navíc dosahovala SSA nižších chyb RMSE a MAE než klouzavý průměr, se kterým zde byla srovnatelná.

Dále následovala již zmíněná metoda klouzavého průměru, která pomocí manuálního nastavení dosahovala nižší chyby MAPE než SSA o téměř 3 procentní body. Jednalo se o malé zlepšení, které se však nepromítlo do chyb RMSE a MAE, které zde byly vyšší než u SSA.

Nakonec zde byl velmi blízky souboj, ve kterém však metoda Holt-Winters skončila druhá. Po dlouhém přemítání zde byly nakonec využity statické parametry, jelikož zatímco Nelder-Mead dokázal někdy vylepšit výsledky při nesprávném nastavení, statické parametry při vhodném nastavení metody vykazovali lepší a stabilnější výsledky. Chyba MAPE se zde již blížila hranici 57,6 %, což bylo vzhledem k používaným metodám dobrým výsledkem.

Nejlépe na tom však byla metoda ARIMA. I přes hromadu problémů, od hardwarové náročnosti až po komplikované nastavení parametrů, vytvářela tato metoda predikce s nejnižší průměrnou chybou. V případě odstranění limitací způsobených implementací pomocí Pythonnet a s využitím exogenních proměnných je však předpokládáno, že by metoda mohla dosahovat ještě lepších výsledků.



Obrázek 3.27: Porovnání chyby MAPE jednotlivých metod vzhledem ke způsobu doplnění chybějících dat

V grafu na obrázku 3.27 je poté vyobrazeno porovnání jednotlivých metod při použití rozdílných způsobů pro doplnění chybějících hodnot ve vstupních datech. Toto porovnání bylo provedeno, jelikož existovalo podezření, že využití klouzavého průměru k doplnění chybějících hodnot mohlo ovlivnit použití stejného algoritmu pro predikce. To se však ukázalo jako nepravdivé, jelikož všechny metody dosahovaly horších výsledků v případě použití interpolace. Holt-Winters dokonce u jednoho měření zcela selhala a vykazovala chybu 405 %, což bylo pravděpodobně způsobeno náhlou změnou povahy dat při doplnění.

Z jednotlivých porovnání lze vidět významné zlepšení přesnosti u složitějších metod. Stále je však nutné chápat, že tyto metody pouze zachycují krátkodobé historické chování a nedokáží proto předpovědět unikátní změny spotřeby, které jsou s touto doménou spojeny. Jedním takovým příkladem by mohl být srpen 2008 z použitých dat, kdy byla spotřeba celý měsíc skoro nulová a metody to zpočátku nečekaly a musely se rychle přizpůsobit. Tato anomálie je viditelná na obrázku A.2. Navíc zde překvapivě dobrého výsledku dosahovala jednoduchá metoda klouzavého průměru.

3.8.2 Porovnání metod dle časové náročnosti

Tabulka 3.3: Průměrné časové náročnosti zkoumaných metod

Metoda	Časová náročnost (ms)
SARIMA (1,0,1)(0,1,1,168)	1 061 353
Holt-Winters (Nelder-Mead)	2 947
Holt-Winters (Statické Parametry)	0
SSA	340
Klouzavý Průměr	0
Lineární Regrese	114

Bylo provedeno porovnání jednotlivých metod dle jejich časové náročnosti. Již při individuální analýze bylo velmi dobře vidět jaká metoda je na tom lépe a jaká hůře. Jak lze vidět v tabulce 3.8.2, tak nejvíce se náročností na čas provinila metoda ARIMA, která mohla trvat až 17 minut v případě využití výchozích parametrů. Tento čas byl navíc měřen při použití pouze 100 dnů vstupních dat, jelikož metoda data většího objemu moc nezvládala.

Taková náročnost však naštěstí nebyla normou a nejhůře na tom poté byl model Holt-Winters s použitím optimalizačního algoritmu. Zde se časy lišily více, jelikož optimalizace parametrů vyžadovala na různých datech různé množství iterací. Časy výpočtu se zde pohybovaly v rozmezí jednotek vteřin. Při použití statických parametrů byla však metoda v podstatě instantní.

Algoritmus SSA na tom již byl podobně, jelikož jak již bylo zmíněno v individuální analýze, trval výpočet v řádu stovek milisekund. To samé platilo pro lineární regresi.

Jelikož klouzavý průměr používal jen n posledních dnů měření, byl výpočet instantní stejně jako u statického Holt-Winters výpočtu.

3.8.3 Porovnání metod dle složitosti jejich využití

Nejnáročnější metody na využití byly Holt-Winters a Arima, jelikož vyžadovaly manuální nastavení parametrů. To sice bylo vyžadované také u klouzavého průměru, ale tam se jednalo o velice intuitivní hodnoty. Arima byla problematická, jelikož pouhé otestování jedné volby parametrů mohlo v horších případech trvat až 80 minut, nebo zcela shodit běh programu. U Holt-Winters bylo možné parametry nastavit manuálním experimentováním, které dokonce vedlo k nalezení statických parametrů použitých v této práci. Avšak jelikož bylo zpočátku manuální nastavení považováno za nevhodné, byl použit již zmíněný optimalizační algoritmus, který spíše práci zkomplikoval. Bezparametrická metoda SSA a lineární regrese byly nejjednodušší na zprovoznění.

3.8.4 Analýza metod na jiné datové sadě

Pro pořádné ověření informací získaných z analýzy metod v této práci byly zkoumané metody nakonec použity i na jiné datové sadě. Zde se jednalo o data z Londýna, která měla podobné vlastnosti jako hlavní data. To znamená týdenní sezónnost, hodinová měření a nebyla stacionární. Velkým rozdílem však byla změna v rozptylu dat, který se zmenšil téměř na polovinu, neboli 0,59 kW^2. To mělo na měřené chyby nečekaný vliv. [13]

Tabulka 3.4: Průměrné chyby zkoumaných metod v londýnských datech

Metoda	RMSE	MAPE	MAE
Sarima (1,0,1)(0,1,1,168)	0,36	13,41	0,27
Holt-Winters	0,61	19,59	0,42
SSA	0,39	14,20	0,30
Klouzavý Průměr	0,45	17,55	0,37
Lineární Regrese	0,52	21,67	0,43

Jak lze vidět v tabulce 3.8.4, tak všechny metody vykazovaly s těmito daty mnohem menší chyby a i nejhorší lineární regrese se přiblížila průměrné chybě MAPE 21 %. To jednoznačně potvrzuje, že se jednalo o mnohem plošší data. Klouzavý průměr zde vykazoval jen malé zlepšení oproti lineární regresi, avšak metoda SSA dokázala s těmito daty pracovat mnohem lépe než s daty z Francie. Významné zhoršení se však projevilo u metody Holt-Winters, kde ani optimalizační algoritmus nedokázal přizpůsobit vyhlazovací parametry pro lepší fungování. Nejlépe na tom nakonec byla opět metoda ARIMA, která v průměru vykazovala chybu MAPE pouze 13,4 %. Grafické zobrazení těchto výsledků je viditelné v grafu A.5.

Tato měření skvěle ukázala závislost kvality predikcí na používaných datech. Všechny metody jednoznačně fungují lépe na datech s menším rozptylem hodnot. Navíc lze nyní stanovit, že neexistuje jednotná hodnota chyby, která by rozdělila predikce na dobré a špatné. Vše je nutné zpracovat relativně k používaným datům. Jelikož londýnská data dosahovala v průměru vyšších hodnot, byla navíc metrika MAPE obecně menší kvůli větší vzdálenost od nuly, ale snížení chyb RMSE a MAE stále potvrzovalo závislost na homogenitě dat. Hlavní data používaná v této práci byla hůře predikovatelná, ale bylo na nich lépe vidět chování jednotlivých metod v různých nastaveních. Londýnská data byla naopak predikovatelná lépe, avšak rozdíly v nastavení zde nebyly tolik viditelné.

Závěr

Hlavním cílem této práce bylo prozkoumání chování metod pro předpovědi dat v chytrých budovách.

Teoretická část práce se zprvu zabývala popsáním chytrých sítí a jejich využitím. Především zde však byla věnována pozornost jednotlivým metodám, které jsou využívané v oblasti datových predikcí. Mezi tyto metody patří jak jednodušší lineární regrese a klouzavý průměr, tak i složitější metody jako metoda SSA, Holt-Winters a ARIMA. U těchto metod bylo popsáno jejich vnitřní fungování. Dále zde byly zmíněny metriky použité při měření přesnosti jednotlivých metod a nakonec byl popsán optimalizační algoritmus Nelder-Mead.

Praktická část se již zabývala implementací zkoumaných metod. Nejprve však byla popsána používaná data a provedena jejich analýza. Data se prokázala jako nestacionární, což ovlivnilo nastavení metody ARIMA. Také se zde objevilo opakující se chování spotřeby s periodou jednoho dne a jednoho týdne, což vyvolalo potřebu testovat parametr sezónnosti u jednotlivých metod. Hodinová i týdenní agregace dat zobrazila chování, které bylo od rodinného domu očekávané s nižší spotřebou v pracovních hodinách a vyšší spotřebou o víkendech. Výskyt tohoto opakujícího chování umožnil provádět funkční predikce. Dále byly zkoumány vlivy nastavení jednotlivých metod.

Nejdůležitějším parametrem při provádění predikcí byla délka sezóny. Pomocí autokorelační funkce byly zjištěny dvě nejlepší hodnoty, den a týden, a byl prozkoumán rozdíl v přesnosti mezi těmito nastaveními. Týdenní sezónnost sice vykazovala v autokorelaci navýšení autokorelačního koeficientu pouze o 0,02, ale na přesnost predikcí měl tento parametr mnohem větší dopad. U metod SSA a Holt-Winters se jednalo o snížení chyby o 30 procentních bodů, u metody ARIMA bylo toto snížení pouze 15 procentních bodů a u klouzavého průměru to bylo jen 8. Celkově to byl parametr nezbytný pro zachycení týdenního chování.

Délka vstupní datové sady ovlivňovala metody různě a poněkud překvapivě. U lineární regrese bylo méně dat žádoucí. Metoda SSA naopak tvořila s rostoucími vstupními daty lepší predikce. Tento růst pokračoval i při využití mnohem větší datové sady, což bylo u zkoumaných metod jedinečné. Metody ARIMA a Holt-Winters pak potřebovaly pouze omezené množství dat a s rostoucí velikostí vstupní sady nevykazovaly významné zlepšení. ARIMA zde byla nejméně ovlivněná tímto parametrem, jelikož potřebovala opravdu málo dat, avšak také měla největší problémy, pokud byla použita data většího objemu.

Posledním zkoumaným společným nastavením byla délka predikce. Zde bylo očekávání horších výsledků při provádění dlouhodobějších predikcí splněno. U metody SSA a ARIMA se jednalo o velmi pomalý růst chyb a metody šlo považovat za vcelku vhodné pro střednědobé predikce. Naopak Holt-Winters vykazovala rychlejší růst chyb pro delší období a byla vyhodnocena jako vhodnější pro krátkodobé predikce. Navíc zde bylo objeveno zajímavé chování chybové metriky MAPE, jelikož v případě provádění velmi krátkých predikcí v podobě například jedné hodnoty, byly chyby predikcí velmi nestálé. V případě provádění dlouhodobějších predikcí se však tato nestabilita kvůli průměrování ztratila. Lineární regrese vykazovala jen pomalý růst chyb, zatímco u klouzavého průměru nebyly delší predikce vůbec možné.

Některé metody poté vyžadovaly další individuální nastavení, které bylo také prozkoumáno. U metody klouzavého průměru byla postupným testováním nalezena nejvhodnější délka průměrovacího okna. Metoda Holt-Winters vyžadovala specifikaci vyhlazovacích parametrů a byl proveden průzkum vhodnosti využití optimalizačního algoritmu Nelder-Mead, který se ukázal jako nevhodný. Navíc metoda Holt-Winters vyžadovala specifikaci typu výpočtu, kde se jako vhodnější ukázal multiplikativní výpočet.

S hotovým nastavením metod bylo provedeno jejich vzájemné porovnání. Nejhorších výsledků zde dosahovala dle očekávání lineární regrese. Metoda SSA překvapivě dosahovala velmi podobných výsledků jako metoda klouzavého průměru. Ta se ukázala jako velmi funkční. Nejlepší výsledky vykazovaly metody ARIMA a Holt-Winters. I nejlepší metody však na hlavních datech vracely velké chyby v podobě 57 % MAPE. Proto byly metody vyzkoušeny i na jiné datové sadě.

Průzkum metod na nových datech ukázal zajímavé výsledky. Tato data totiž měla mnohem menší rozptyl než hlavní používaná data a metody tak dokázaly provést mnohem lepší predikce. ARIMA zde například dosahovala chyby MAPE pouze 13 %, což bylo veliké zlepšení oproti francouzské datové sadě. Pokles chyby se však vyskytl u všech metod a SSA se dokonce dostala na druhé místo v přesnosti. Z toho bylo ověřeno, že kvalita predikcí silně závisí na struktuře používaných dat. Metody však stále mají problém s proměnlivou povahou dat.

V praxi by byla nejvhodnější metoda ARIMA, která dosahovala nejlepších výsledků. Jelikož je však někdy požadavkem rychlý výpočet, mohla by tato metoda být problematická. V takovém případě by bylo voleno mezi metodami SSA a Holt-Winters, přičemž SSA by byla vhodná v případě práce s homogennějšími daty.

Jednotlivé metody byly implementovány pomocí prostředí .NET a výsledná aplikace umožňuje měření chyb prováděných predikcí. Je však nutné zmínit problematiku způsobenou takovouto prací ve frameworku .NET, jelikož spousta velmi užitečných nástrojů zde neexistovala a bylo nutné hledat různé alternativy. Například hlavní využívaná matematická knihovna MathNet.Numerics byla již dva roky zastaralá. Vzhledem k absenci jiných alternativ ke knihovně statsmodels pro .NET bylo také nezbytné využít propojení s jazykem Python pomocí balíčku Python.NET.

Použitá literatura

[1] ALTO, Valentina. *Understanding ordinary least squares (OLS) regression*. 2023. Dostupné také z: https://builtin.com/data-science/ols-regression.

[2] ATHIYARATH, Srihari, Mousumi PAUL a Srivatsa KRISHNASWAMY. A comparative study and analysis of time series forecasting techniques. *SN Computer Science*. 2020, roč. 1, č. 3. Dostupné z DOI: 10.1007/s42979-020-00180-5.

[3] AWATI, Rahul. *What are extrapolation and interpolation?* TechTarget, 2022. Dostupné také z: https://www.techtarget.com/whatis/definition/extrapolation-and-interpolation.

[4] BROWNLEE, Jason. *How to choose an optimization Algorithm* [online]. 2021-10-11. [cit. 2024-02-08]. Dostupné z: https://machinelearningmastery.com/tour-of-optimization-algorithms/.

[5] EDIGER, Volkan Ş. a Sertaç AKAR. ARIMA forecasting of primary energy demand by fuel in Turkey. *Energy policy*. 2007, roč. 35, č. 3, s. 1701–1708. Dostupné z DOI: 10.1016/j.enpol.2006.05.009.

[6] FILDES, Robert. The evaluation of extrapolative forecasting methods. *International Journal of Forecasting*. 1992, roč. 8, č. 1, s. 81–98. Dostupné z DOI: 10.1016/0169-2070(92)90009-x.

[7] GOLYANDINA, Nina a Anton KOROBEYNIKOV. Basic Singular Spectrum Analysis and forecasting with R. *Computational statistics & data analysis (Print)*. 2014, roč. 71, s. 934–954. Dostupné z DOI: 10.1016/j.csda.2013.04.009.

[8] GOLYANDINA, Nina a Anatoly ZHIGLJAVSKY. *Singular spectrum analysis for time series*. 2013. Dostupné z DOI: 10.1007/978-3-642-34913-3.

[9] HASSANI, Hossein. Singular Spectrum Analysis: Methodology and comparison. *Journal of data science (Print)*. 2021, roč. 5, č. 2, s. 239–257. Dostupné z DOI: 10.6339/jds.2007.05(2).396.

[10] HEBRAIL, Georges a Alice BERARD. *Individual Household Electric Power Consumption* [UCI Machine Learning Repository]. 2012. DOI: https://doi.org/10.24432/C58K54.

[11] HYNDMAN, Rob J a George ATHANASOPOULOS. *Forecasting: principles and practice*. OTexts, 2018.

[12] KUMAR, Ujjwal a V. K. JAIN. Time series models (Grey-Markov, Grey Model with rolling mechanism and singular spectrum analysis) to forecast energy consumption in India. *Energy*. 2010, roč. 35, č. 4, s. 1709–1716. Dostupné z DOI: 10.1016/j.energy.2009.12.021.

[13] MICHAEL, Jean (ed.). *Smart meters in London*. 2022. Dostupné také z: https://www.kaggle.com/datasets/jeanmidev/smart-meters-in-london/data.

[14] MOLINA-SOLANA, Miguel et al. Data Science for Building Energy Management: A Review. *Renewable and Sustainable Energy Reviews*. 2017, roč. 70, s. 598–609. Dostupné z DOI: 10.1016/j.rser.2016.11.132.

[15] ROBERTS, Amber. *Mean Absolute percentage error (MAPE): what you need to know*. 2023. Dostupné také z: https://arize.com/blog-course/mean-absolute-percentage-error-mape-what-you-need-to-know.

[16] SINGER, Sanja a J. A. NELDER. Nelder-Mead Algorithm. *Scholarpedia*. 2009, roč. 4, č. 7, s. 2928. Dostupné z DOI: 10.4249/scholarpedia.2928.

[17] TAYLOR, Sebastian. *Moving average* [online]. 2023-11-21. [cit. 2024-01-13]. Dostupné z: https://corporatefinanceinstitute.com/resources/data-science/moving-average/.

[18] TULANG, Alfeo B. a Alwielland BELLO. Forecasting power load demand using Holt-Winters model. *Social Science Research Network*. 2018. Dostupné z DOI: 10.2139/ssrn.4107011.

[19] VILLALOBOS, Juan Orozco. *Mean Absolute Error vs Root-Mean Square Error*. 2020. Dostupné také z: https://www.brainstobytes.com/mean-absolute-error-vs-root-mean-square-error/.

[20] ZHANG, Yang, Tao HUANG a Ettore Francesco BOMPARD. Big Data Analytics in smart grids: A Review. *Energy Informatics*. 2018, roč. 1, č. 1. Dostupné z DOI: 10.1186/s42162-018-0007-5.

A Přílohy

A.1 Slovník pojmů

ARIMA Autoregresní integrovaný klouzavý průměr (autoregressive integrated moving average) je metoda pro predikci časových řad, která kombinuje principy autoregrese, klouzavého průměru a diferenciace.

Exponenciální vyhlazování Varianta klouzavého průměru založená na přiřazování větší váhy novějším datům.

Holt-Winters Metoda trojitého exponenciálního vyhlazování, která je schopná využít trendu a sezónnosti pro tvorbu predikcí.

Klouzavý průměr Metoda vyhlazování dat využívající průměrování posledních n hodnot.

Lineární regrese Statistická metoda pro stanovení vztahu mezi závislou a nezávislou proměnnou pomocí přímky.

MAE Střední absolutní chyba (mean absolute error) je chybová metrika pro měření průměrné absolutní odchylky mezi predikcí a reálnými daty. Tato odchylka je vyjádřena ve stejných jednotkách jako původní data.

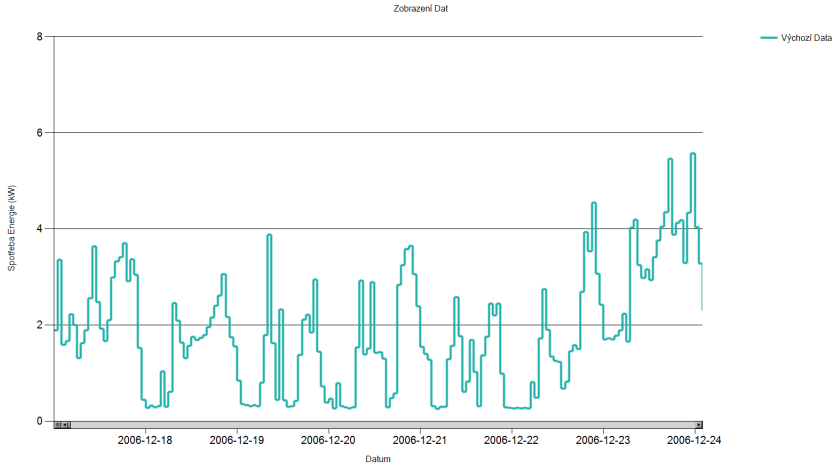
MAPE Střední absolutní procentuální chyba (mean absolute percentage error) je chybová metrika pro měření průměrné absolutní odchylky mezi predikcemi a reálnými daty normalizované podle skutečných hodnot. Tato odchylka je vyjádřena v procentech.

Nelder-Mead Optimalizační algoritmus, který umožňuje minimalizaci a maximalizaci účelové funkce prostřednictvím nastavení jejích vstupních parametrů.

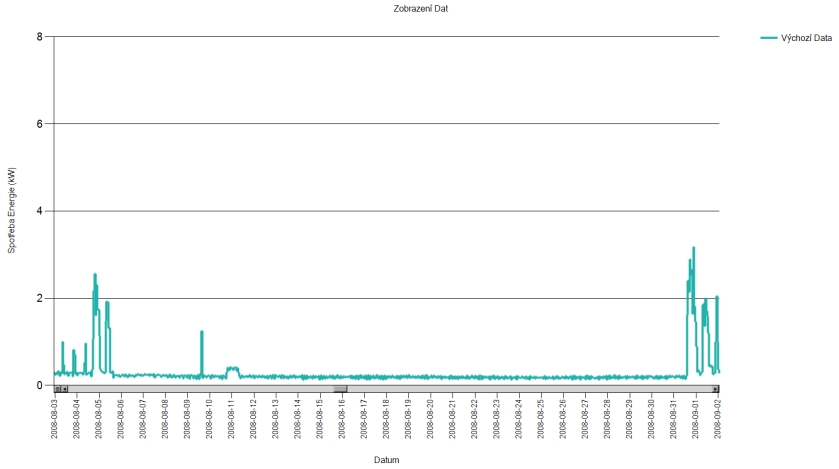
RMSE Střední kvadratická chyba (root mean square error) je chybová metrika pro měření průměrné kvadratické odchylky mezi predikcemi a reálnými daty. Tato odchylka je vyjádřena ve stejných jednotkách jako původní data.

SSA Singulární spektrální analýza (singular spectrum analysis) je technika pro predikci časových řad. Spočívá v rozkladu těchto řad na menší komponenty a následně výběru těch nejdůležitějších. Poté se pomocí těchto komponent rekonstruuje původní časová řada.

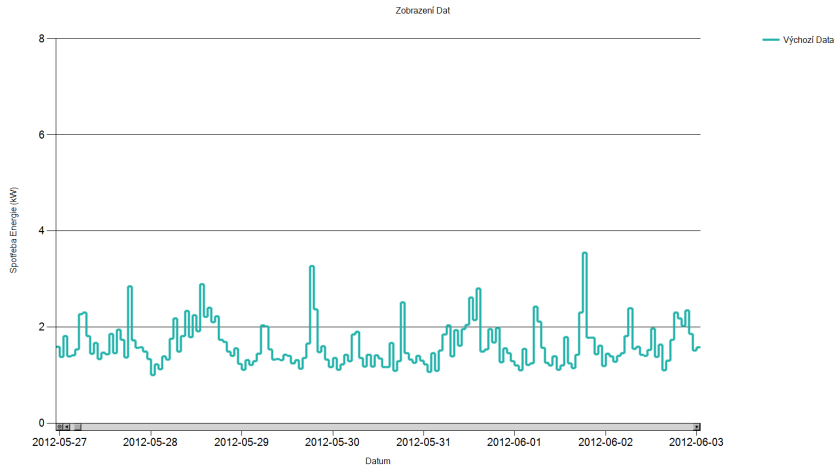
A.2 Používaná data



Obrázek A.1: Zobrazení týdenního intervalu ve francouzských datech

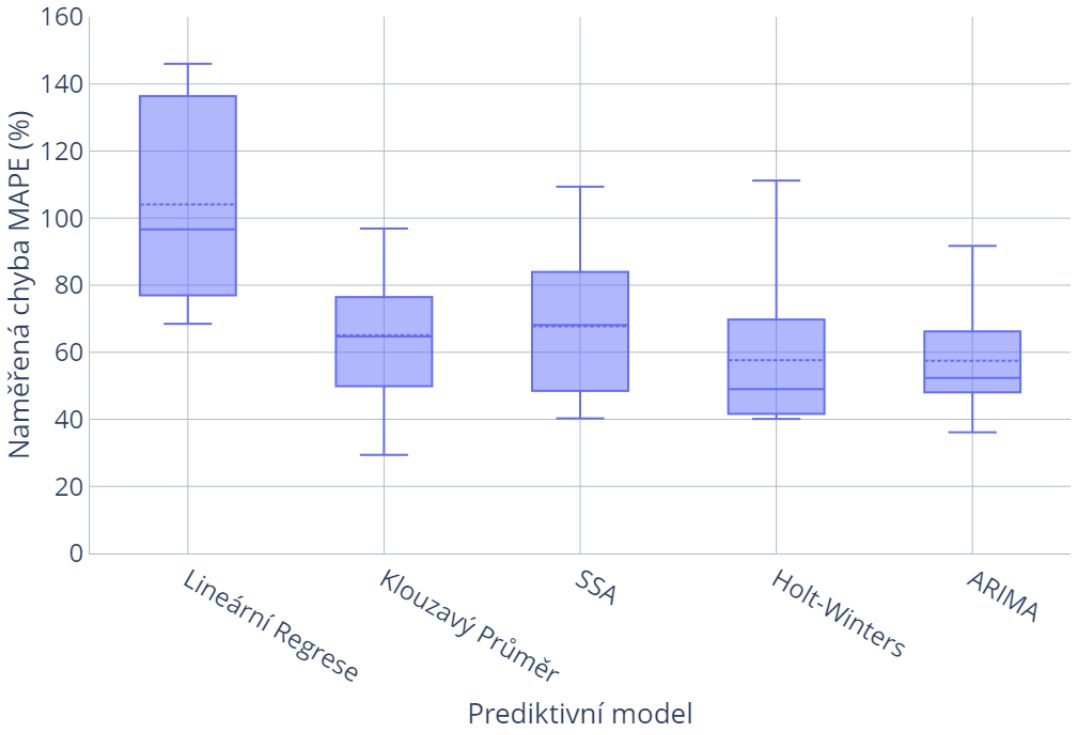


Obrázek A.2: Zobrazení srpnové anomálie ve francouzských datech

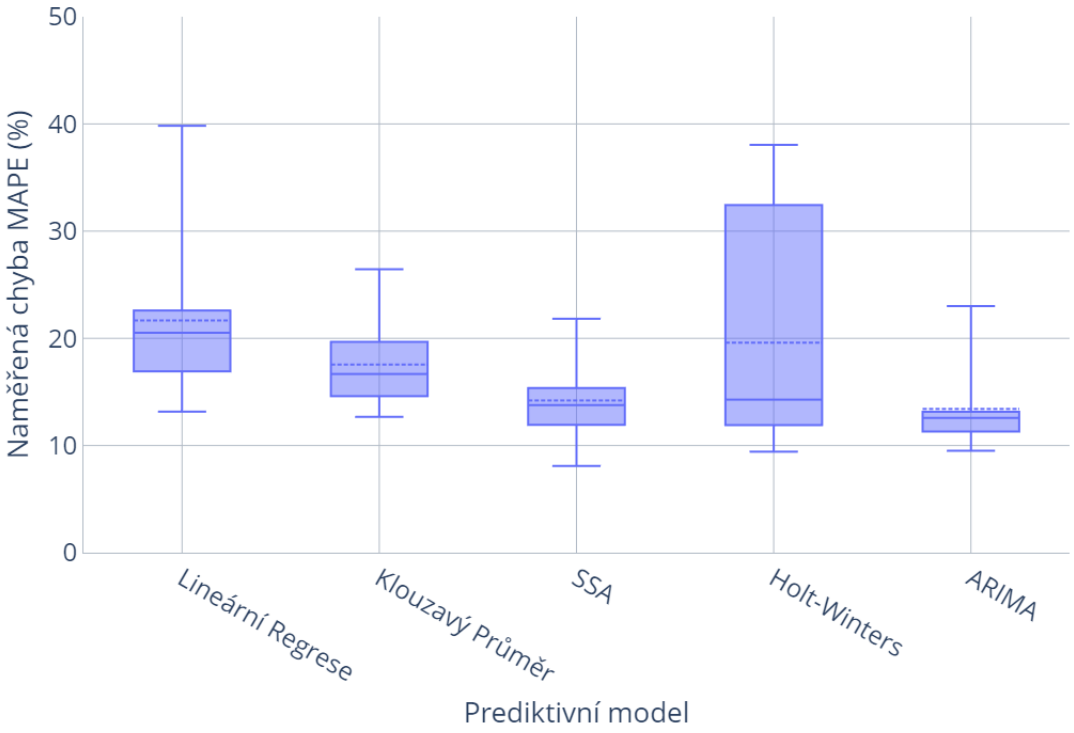


Obrázek A.3: Zobrazení týdenního intervalu v londýnských datech

A.3 Dodatečné grafy



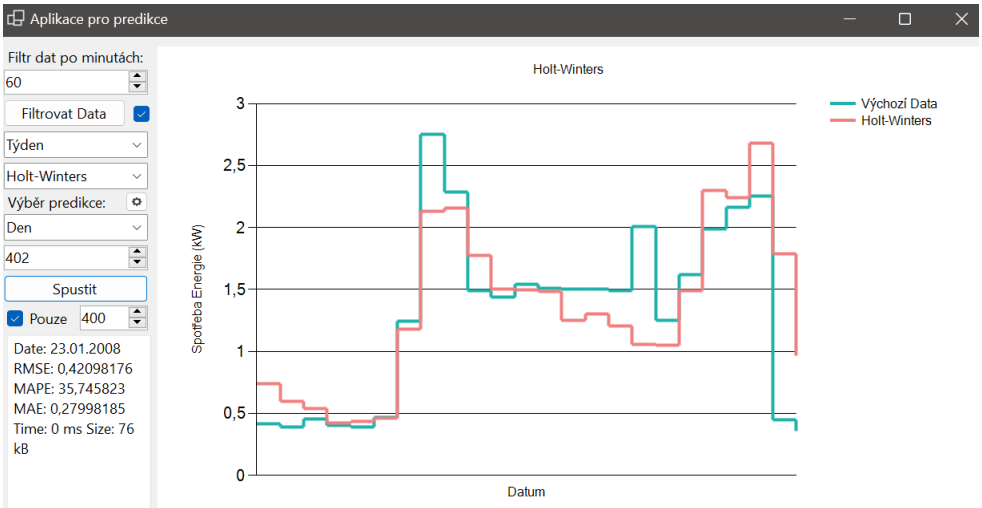
Obrázek A.4: Porovnání chyby MAPE jednotlivých metod ve francouzských datech



Obrázek A.5: Porovnání chyby MAPE jednotlivých metod v londýnských datech

A.4 Testovací aplikace

Metody byly zprovozněny v testovací aplikaci s jednoduchým uživatelským rozhraním.



Obrázek A.6: Zobrazení vytvořené aplikace