

UNIVERZITA PARDUBICE

Fakulta elektrotechniky a informatiky

Využití algoritmů strojového učení

při predikci počasí

Oldřich Svoboda

Diplomová práce

2020

Univerzita Pardubice
Fakulta elektrotechniky a informatiky
Akademický rok: 2019/2020

ZADÁNÍ DIPLOMOVÉ PRÁCE

(projektu, uměleckého díla, uměleckého výkonu)

Jméno a příjmení:	Bc. Oldřich Svoboda
Osobní číslo:	I16240
Studijní program:	N2646 Informační technologie
Studijní obor:	Informační technologie
Téma práce:	Využití algoritmů strojového učení při predikci počasí
Zadávací katedra:	Katedra softwarových technologií

Zásady pro vypracování

V rámci diplomové práce budou aplikovány zadané algoritmy strojového učení na meteorologická data. Získané predikční modely budou verifikovány a bude srovnána jejich přesnost. Práce bude členěna na teoretickou a praktickou část. Teoretická část bude obsahovat deklaraci základních pojmů, přehled metod strojového učení a jejich způsob jejich implementace při predikci z časových řad. Teoretická část bude dále obsahovat úvod do problematiky meteorologie, zpracování a interpretace meteorologických dat. Praktická část bude obsahovat popis implementace zadaných algoritmů, výsledky experimentální verifikace získaných modelů a jejich srovnání.

Rozsah pracovní zprávy: **50-70 stran**
Rozsah grafických prací:
Forma zpracování diplomové práce: **tištěná**

Seznam doporučené literatury:

ALPAYDIN, Ethem a Geoffrey I. WEBB. Introduction to machine learning. Third edition. Cambridge, Massachusetts: The MIT Press, 2014. ISBN 978-0-262-02818-9. DUCHON, Claude a Robert HALE. Time series analysis in meteorology and climatology an introduction. Third edition. Chichester, West Sussex: Wiley-Blackwell, 2012. ISBN 978-111-9953-104. HARPER, Kristine. Weather by the numbers: the genesis of modern meteorology. Cambridge, Mass.: MIT Press, c2008. Transformations (M.I.T. Press). ISBN 978-026-2083-782. SAMMUT, Claude a Geoffrey I. WEBB. Encyclopedia of machine learning. London: Springer, 2010. ISBN 978-038-7345-581.

Vedoucí diplomové práce: **Ing. Soňa Neradová, Ph.D.**
Katedra informačních technologií

Datum zadání diplomové práce: **5. listopadu 2019**
Termín odevzdání diplomové práce: **15. května 2020**



Ing. Zdeněk Němec, Ph.D.
děkan

prof. Ing. Antonín Kavička, Ph.D.
vedoucí katedry

Prohlášení autora

Prohlašuji, že jsem tuto práci vypracoval samostatně. Veškeré literární prameny a informace, které jsem v práci využil, jsou uvedeny v seznamu použité literatury.

Byl jsem seznámen s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorský zákon, zejména se skutečností, že Univerzita Pardubice má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona, a s tím, že pokud dojde k užití této práce mnou nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Pardubice oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

Souhlasím s prezenčním zpřístupněním své práce v Univerzitní knihovně.

V Pardubicích dne

Oldřich Svoboda

Poděkování

Na tomto místě bych rád poděkoval vedoucí této diplomové práce, bez jejíž trpělivosti by tato práce neměla šanci vzniknout. Dále bych moc rád poděkoval své matce, jejíž psychika při psaní této práce utrpěla nejvíce. Děkuji, bez Vás by to nešlo.

Anotace

Diplomová práce se zaměřuje na předpovídání počasí za využití dat z jedné fyzické lokace a algoritmu strojového učení. Předmětem je předpověď v krátkodobém měřítku 24 hodin.

Klíčová slova

Strojové učení, meteorologie, krátkodobá předpověď, teplota, vlhkost, barometrický tlak

Title

Use of machine learning algorithms in weather prediction

Annotation

This diploma thesis focuses on weather forecasting using data from one physical location and a machine learning algorithms. The subject is a short-term forecast for next 24 hours.

Keywords

Machine learning, meteorology, short-term forecast, temperature, humidity, barometric pressure

Obsah

Seznam zkratk	8
Seznam obrázků	9
Úvod	11
1 Základní pojmy v rámci počasí a předpovídání počasí	12
2 Problematika předpovědi počasí	16
2.1 Krátkodobá předpověď počasí	16
2.2 Dlouhodobá předpověď	17
2.3 Obvyklé přístupy k předpovědi počasí	17
2.3.1 GFS	17
2.3.2 WRF Model	18
2.3.3 MPAS.....	19
2.3.4 Amatérské meteostanice	20
3 Základní pojmy strojového učení	22
3.1 Metody strojového učení	23
3.1.1 Regresní metoda	23
3.1.2 Umělé neuronové sítě	23
3.1.2.1 Základní model neuronu	23
3.1.2.2 Neuronová síť	24
3.1.2.3 Lineární perceptron	25
3.1.2.4 Hopfieldova síť	25
3.1.2.5 Dopředná vícevrstvá umělá neuronová síť / vícevrstvý perceptron	26
3.1.2.6 Zpětné šíření chyby – Backpropagation	28
4 Praktická předpověď	29
4.1 Praktické učení a předpověď hodnot časové řady	30
4.1.1 Získání a příprava dat	30

4.1.2	Příklad nastavení a spuštění předpovědního modelu pro jednu vstupní a výstupní hodnotu.....	30
4.2	Předpověď teploty.....	34
4.3	Předpověď vlhkosti.....	37
4.4	Předpověď tlaku.....	39
4.5	Kombinované předpověď teploty a vlhkosti.....	41
4.6	Kombinovaná předpověď teploty a atmosférického tlaku.....	45
4.7	Kombinovaná předpověď vlhkosti a atmosférického tlaku.....	48
4.8	Kombinované předpověď teploty, vlhkosti a tlaku.....	51
4.9	Porovnání se skutečnou předpovědí	56
5	Závěr	58
	Literatura	59
	Seznam příloh.....	61
	Příloha A – Seznam použitého SW.....	62

Seznam zkratek

GFR	Global Forecasting System
MPAS	Model for Prediction Across Scales
RMSE	Root Mean Squared Error
WRF	Weather Research and Forecasting

Seznam obrázků

Obrázek 1 – GFS (National Centers for Environmental Information, 2020)	18
Obrázek 2 - WRF model (UCAR, 2020)	19
Obrázek 3 - MPAS (Los Alamos National Security, LLC (LANS) , 2020).....	20
Obrázek 4 - Jednoduchá domácí meteostanice (Čermák, 2020).....	20
Obrázek 5 - Poloprofesionální meteostanice (Čermák, 2020).....	21
Obrázek 6 - Lineární regrese (Moravec, 2020).....	23
Obrázek 7 - Rovnice lin. regrese	23
Obrázek 8 - Model neuronu (Ing. Petr Doležel, 2016)	24
Obrázek 9 - Hopfieldova síť (Ing. Petr Doležel, 2016)	26
Obrázek 10 - Dopředná vícevrstvá umělá neuronová síť (Ing. Petr Doležel, 2016)	27
Obrázek 11 - Vývoj teploty	34
Obrázek 12 - Předpověď teploty	35
Obrázek 13 - Předpověď teploty na 24h	36
Obrázek 14 - Vývoj vlhkosti.....	37
Obrázek 15 - Předpověď vlhkosti	37
Obrázek 16 - Předpověď vlhkosti na 24 h	38
Obrázek 17 - Vývoj atmosférického tlaku	39
Obrázek 18 - Předpověď vývoje atmosférického tlaku	40
Obrázek 19 - Kombinovaná předpověď teploty a vlhkosti – teplota.....	41
Obrázek 20 - Kombinovaná předpověď teploty a vlhkosti – teplota na 24 h.....	42
Obrázek 21 - Kombinovaná předpověď teploty a vlhkosti – vlhkost.....	43
Obrázek 22 - Kombinovaná předpověď teploty a vlhkosti – vlhkost na 24 h	44
Obrázek 23 - Kombinovaná předpověď teploty a atmosférického tlaku – teplota.....	45
Obrázek 24 - kombinovaná předpověď teploty a atmosférického tlaku – teplota na 24 h	46
Obrázek 25 - Kombinovaná předpověď teploty a atmosférického tlaku – tlak	46
Obrázek 26 - Kombinovaná předpověď teploty a atmosférického tlaku – tlak na 24 h	47
Obrázek 27 - Kombinovaná předpověď vlhkosti a atmosférického tlaku – tlak	48
Obrázek 28 - Kombinovaná předpověď vlhkosti a atmosférického tlaku – tlak na 24 h	49
Obrázek 29 - Kombinovaná předpověď atmosférického tlaku a vlhkosti – vlhkost	49
Obrázek 30 - Kombinovaná předpověď atmosférického tlaku a vlhkosti – vlhkost na 24 h....	50
Obrázek 31 - Kombinovaná předpověď teploty, vlhkosti a tlaku – teplota.....	51
Obrázek 32 - Kombinovaná předpověď teploty, vlhkosti a tlaku – teplota na 24 h.....	52

Obrázek 33 - Kombinovaná předpověď teploty, vlhkosti a tlaku – vlhkost.....	53
Obrázek 34 - Kombinovaná předpověď teploty, vlhkosti a tlaku – vlhkost na 24 h	54
Obrázek 35 - Kombinovaná předpověď teploty, vlhkosti a tlaku – tlak.....	54
Obrázek 36 - Kombinovaná předpověď teploty, vlhkosti a tlaku – tlak na 24 h.....	55
Obrázek 37 - Předpověď vs realita – teplota	56
Obrázek 38 - Předpověď vs realita – vlhkost.....	57
Obrázek 39 - Předpověď vs realita – tlak	57

Úvod

Tato práce si klade za cíl zhodnotit vhodnost použití strojového učení na speciální případ předpovídání počasí pro krátkodobou budoucnost v jedné geografické lokalitě bez možnosti využít data z okolních lokalit a informací o lokální geografii.

Motivací k tvorbě této práce bylo v současné době nedostupnost spolehlivého prostředku pro lokální předpověď počasí počínaje domácími meteostanicemi a poloprofesionálními stanicemi konče. Veškeré dostupné prostředky předpovídají buď ráz počasí dle pohybu atmosférického tlaku či vycházejí z velkých meteorologických modelů, jež publikují výsledky online, a nikoliv vždy je poskytují i pro danou zájmovou lokalitu.

Náplní teoretické části je seznámit čtenáře se základními pojmy z oblasti meteorologie a přiblížit mu problematiku předpovědi počasí jako takové. Dále se seznámí s nejčastěji používanými pojmy, strukturami a algoritmy v rámci strojového učení, které je posléze aplikováno na získaná meteorologická data.

Praktická část se zabývá experimentální předpovědí počasí, respektive hodnot zvolených meteorologických atributů v rámci dané časové řady. Primárním zaměřením práce je krátkodobá předpověď na 24 h, přičemž je i názorně ukázáno proč byla zvolena právě krátkodobá předpověď. Výstupem práce je možnost předpovídat počasí pouze z místně dostupných údajů po zacvičení dané neuronové sítě.

1 Základní pojmy v rámci počasí a předpovídání počasí

Počasí samotné je chápáno jako stav atmosféry určený množinou hodnot všech meteorologických prvků a atmosférickými jevy v konkrétním místě a čase.

Předpověď počasí je vyjádřením budoucího stavu počasí. Vzniká na základě použití meteorologických poznatků, jako jsou například pozorování a měření na meteorologických stanicích, družicová měření, statistické údaje, výstupy numerických modelů a případně i vlastní zkušenosti meteorologa a jeho intuice. Označuje se také jako prognóza.

Podnebí (klima) je definováno jako dlouhodobý průměrný stav atmosféry v určitém místě.

Speciální předpověď jsou předpovědi zaměřené dle konkrétních požadavků, například pro zimní údržbu silnic a dálnic, leteckou dopravu, zemědělství, energetiku a další.

Úspěšnost předpovědi počasí je nejvyšší pro prvních pět dní. Proto je nejčastěji vydávána předpověď právě na následujících pět dní a vyhlídkově na další tři dny. Při předpovědi pro deset a více dní lze předpovědět spíše charakter počasí, nikoliv jeho konkrétní podobu pro jednotlivé dny.

Jako *nebezpečné jevy* se označují všechny hydrometeorologické jevy, jež mohou mít za následek hmotné škody, ohrožení zdraví či života. Jedná se kupříkladu o dlouhotrvající sucha, sněhové jazyky, smogové situace, silné mrazy, ledovku či silné srážky vedoucí k povodním.

Nižší polohy jsou místa s nadmořskou výškou do 400 metrů nad mořem. *Střední polohy* jsou místa od 400 do 600 metrů nad mořem. *Vyšší polohy* jsou od 600 do 800 metrů nad mořem. Konečně jakožto *horské polohy* označujeme místa nad 800 metrů nad mořem.

Synoptická situace označuje charakter či předpokládaný stav frontálních a termobarických útvarů a vzduchových hmot nad velkou určenou konkrétní geografickou oblastí.

Atmosférická fronta je označení pro úzkou přechodovou zónu mezi vzduchovými hmotami různých vlastností, zejména teploty a vlhkosti.

Teplá fronta je fronta, která se přesouvá na stranu studeného vzduchu, po jejím přechodu obvykle dojde k oteplení. Před teplou frontou vzniká charakteristický oblačný systém, začínající vysokou oblačností umístěnou stovky kilometrů před frontou, pokračující střední a dále nízkou oblačností blíže před frontou, kde se vyskytuje pásmo trvalých srážek široké až do čtyř set kilometrů.

Studená fronta je naopak fronta, jež se pohybuje na stranu studeného vzduchu a obvykle způsobuje ochlazení. Typickým znakem je výskyt kupovitých oblaků druhu cumulonimbus, přeháňkovými dešti, často i bouřkami. (Český hydrometeorologický ústav, 2018)

Okluzní fronta vzniká spojením teplé a studené fronty. To znamená, že studená fronta dohnala teplou frontu. Vlastnosti obou front jsou kombinovány.

Zvlněná studená fronta vzniká za zvláštních podmínek na pomalu se pohybující studené frontě, kde se její část změní na frontu teplou. Může být provázeno tlakovou níží spolu s velkou oblačností a srážkami.

Tlaková výše alias anticyklona je oblast s vysokým či zvýšeným tlakem vzduchu, jenž se na synoptické mapě projevuje jako alespoň jedna uzavřená izobara. V zimním období její výskyt znamená mrazivé slunečné počasí nebo inverzní počasí s výskytem mlh a nízké oblačnosti či mrholením. V letním období znamená teplé počasí.

Tlaková níže alias cyklona je naopak oblast se sníženým či nízkým tlakem vzduchu, jenž se též projevuje na synoptické mapě jako alespoň jedna uzavřená izobara, přičemž tlak uvnitř je nižší než v okolí. Obvykle její výskyt znamená velkou oblačnost a srážky.

Množství oblačnosti na obloze se vyjadřuje jejím pokrytí v osminách:

Jasno (0/8) je obloha bez oblačnosti, tj. obloha bez mraků.

Skoro jasno (1/8 až 2/8 pokrytí oblohy) označuje většinu oblohy bez mraků.

Polojasno (3/8 až 4/8) označuje polovinu oblohy pokrytou mraky.

Oblačno (5/8 až 6/8) znamená, že většina oblohy je zakryta mraky.

Skoro zataženo (7/8) vyjadřuje, že téměř celá obloha je zakryta.

Zataženo (8/8) tj., že úplně celá obloha je zakryta mraky.

Atmosférické jevy jsou jevy, jež jsou pozorovány v atmosféře či na zemském povrchu na meteorologických stanicích a v jejich okolí.

Plošná četnost meteorologických jevů označuje, na jakém množství (obvykle v procentech) území se očekává výskyt tohoto jevu:

Bez určení obvykle znamená, že jev se bude vyskytovat nad sedmdesáti či více procenty území.

Na většině území označuje, že jev se bude vyskytovat nad více než padesáti procenty území.

Místa má význam výskytu jevu na třiceti až šedesáti devíti procenty plochy. *Ojediněle* znamená, jev se bude vyskytovat mezi pěti až dvaceti devíti procenty území. *Bez jevu* se použije v případě, že jev se bude vyskytovat maximálně na čtyřech procentech území.

Bouřka je popisný název pro skupinu meteorologických jevů, které se vyskytují při vývoji konvektivních oblaků typu cumulonimbus. Rozlišují se bouřky frontální na atmosférických frontách a nefrontální např. vzniklé z tepla. Jev je spojen s nárazovým větrem, přivalovými srážkami, elektrickými výboji, méně často kroupami nebo tornádem (viditelný, silně rotující „chobot“ vyvíhající ze spodní základny cumulonimbu). Nejčastěji se vyskytují v letním období. (Český hydrometeorologický ústav, 2018)

Mlha je jev označující soustavu velmi malých vodních kapiček rozptýlených ve vzduchu, jenž snižuje vodorovnou dohlednou vzdálenost pod jeden kilometr. Může být i pouze v jednom směru. Relativní vlhkost v tomto jevu je vysoká, dosahuje i sta procent.

Mrznoucí mlha je mlha při teplotách pod bodem mrazu, kde kapičky mají teplotu již pod bodem mrazu. Obvykle vznikají námrazkové jevy.

Vítr je určen primárně dvěma charakteristikami směrem a rychlostí. Směr se udává, jako převládající směr odkud vane vítr, kupříkladu severní, jižní, severozápadní atd. Rychlost mlže být uvedena v m/s či km/h případně slovně. Jakožto náraz větru se považuje krátkodobé zvýšení rychlosti větru oproti průměru o alespoň pět m/s po dobu minimálně jedné a maximálně dvaceti vteřin. Je též stanovena minimální hranice pro rychlost na dvanáct m/s.

Tlak vzduchu se obvykle vyjadřuje jako tlak vzduchového sloupce na hladinu moře v hektopascálech. Podle míry změny tlaku se rozlišujeme tyto kategorie *tlakových tendencí*:

Setrvalý stav, tj. tlak se nebude měnit.

Slabý pokles či vzestup, tlak se změní o maximálně jeden hektopascal za tři hodiny.

Mírný pokles či vzestup, tlak se změní o jeden až tři hektopascal za tři hodiny.

Silný pokles či vzestup, tlak se změní o více než tři hektopascal za tři hodiny.

(Český hydrometeorologický ústav, 2018)

„*Srážky atmosférické* jsou soustava vodních částic vzniklé kondenzací či sublimací vodní páry v ovzduší ve stavu kapalném (děšť, mrznoucí děšť, mrhnutí, mrznoucí mrhnutí) nebo tuhém

(sněžení, sněhové krupky, sněhová zrna, krupky, zmrzlý déšť, kroupy a ledové jehličky) padající nebo vznášející se v atmosféře nebo zdvižené větrem z povrchu země nebo usazené na předmětech na zemi či ve volné atmosféře (rosa, jíní, námraza a ledovka).“ (Český hydrometeorologický ústav, 2018)

Podle *doby* trvání padajících srážek rozlišujeme:

Trvalé jsou charakteristické velkou dobou a stálou intenzitou velkoplošných srážek.

Občasné se vyskytují opakovaně s přestávkami v řádech hodin. Používá se jak pro velkoplošné, tak i pro konvektivní srážky (přeháňky).

Přeháňky je označení, pokud jsou přestávky mezi opakováním v řádech minut a intenzita srážek rychle kolísá. Mezi srážkami může dojít i k vyjasnění.

Četné srážky jsou srážky, které se opakují v řádech minut.

Bez určení doby trvání se obvykle využívá při přechodu fronty.

(Český hydrometeorologický ústav, 2018)

„*Déšť* jsou vodní srážky vypadávající z oblaků ve tvaru kapek větším než 0,5 mm nebo i menším, pokud jsou velmi rozptýlené. Vypadávají zpravidla z vrstevnaté oblačnosti.“ (Český hydrometeorologický ústav, 2018)

Přeháňky definujeme jako srážky s rapidním začátkem i koncem, krátkým trváním a velmi rychlým střídáním intenzity. Obvykle dochází ke krátkému vyjasnění. Lze obsáhnout i sněhové krupky.

(Český hydrometeorologický ústav, 2018)

Mrholení lze určit jako kapalné srážky s vysokou hustotou skládající se z kapiček o průměru menším než 0,5 mm vypadávající z nízké oblačnosti typu stratus nebo z mlhy.

Sněžení jsou srážky tvořené z ledových krystalků či ze sněhových vloček, vznikají obvykle z vrstevnaté oblačnosti. (Český hydrometeorologický ústav, 2018)

„*Sněhové přeháňky* jsou sněhové srážky vypadávající z kupovité (konvektivní) oblačnosti.“ (Český hydrometeorologický ústav, 2018)

Intenzita srážek je množství srážek za jednotku času. Nejčastěji určené výškou vodního či sněhového sloupce za jednu hodinu. Rozlišujeme:

Velmi slabou intenzitu při neměřitelném množství srážek.

Slabou intenzitu, kde déšť je v rozmezí 0,1 až 2,5 mm/h a sníh do 0,5 cm, čímž neovlivňuje dohlednost.

Mírná intenzita, kdy déšť je v rozmezí 2,6 až 8 mm/h a sníh od 0,6 do 4 cm, čímž je dohlednost mírně zhoršená.

Silná intenzita, kdy déšť je v rozmezí 8 až 40 mm/h a sníh více než 4 cm, čímž je dohlednost snižena na 500 m.

Velmi silná intenzita, kdy déšť přesahuje 40 mm/h a sníh může krátkodobými intenzivními přeháňkami snížit dohlednost pod 500 m.

(Český hydrometeorologický ústav, 2018)

2 Problematika předpovědi počasí

Předpověď počasí se ve většině případů rozděluje na krátkodobou a dlouhodobou, přičemž hranice mezi nimi je určena na přibližně deset dní, kdy předpověď již ze své samotné podstaty nemůže dosahovat přiměřené přesnosti.

Nepřesnost předpovědi je způsobována jednak charakterem atmosféry, množstvím vstupujících faktorů (přesnost a počet měření, reálné množství údajů atd.) tak hlavní měrou i skutečností, že stále plně nerozumíme všem zahrnutým přírodním procesům, které jej ovlivňují, a to bez ohledu na náročnost jejich simulování.

(InMeteo, 2017)

2.1 Krátkodobá předpověď počasí

Krátkodobá předpověď počasí je obvykle určena časovým intervalem do přibližně deseti dní, přičemž na prvních pět dní se obvykle provádí podrobná předpověď a na následující dny pouze výhledová, orientační předpověď.

Podrobná předpověď je vytvářena pro jednotlivé části území a krátké časové úseky (například hodinová předpověď), kdežto výhledová předpověď je obvykle větší úseky či celé území v rámci dní. Např. pro celou republiku jsou určovány přepokládané denní teploty. (InMeteo, 2017)

2.2 Dlouhodobá předpověď

„Na rozdíl od klasických předpovědí (tj. do cca 10 dnů dopředu) není účelem dlouhodobých předpovědí určit počasí v konkrétní den. Není tedy možné od nich chtít vědět, zdali třeba za 4 týdny v daný den bude teplo nebo pršet, ale spíš jaký bude převažující charakter daného období, tedy průměrné podmínky v tom několikadenním až měsíčním období. Současně to ale neznamená, že tyto podmínky musí v daném období převažovat kontinuálně, naopak, často má počasí různé podoby, které se někdy více někdy méně liší od tohoto průměru.“ (InMeteo, 2017)

Velké oblibě se dostává předpovědi počasí za využití vyjádření pravděpodobnosti výskytu určitých meteorologických jevů, jenž už několik let využívá i ČHMÚ. Příkladem takovéto předpovědi může být údaj, že dané období bude s pravděpodobností 50% srážkově nadprůměrné, s pravděpodobností 35% blízko průměru a s pravděpodobností 15% srážkově podprůměrné.

Dlouhodobá předpověď však nemusí mít pouze nám obvykle známou podobu, ale může zahrnovat i kupříkladu vývoj teploty a proudění v oceánu či vývoj klimatu jako takového.

Dlouhodobá předpověď však nemusí vždy přinášet použitelné údaje. Kupříkladu v rámci střední Evropy je předpověď od jednoho měsíce již tak nespolehlivá, že většina organizací ji vůbec nezveřejňuje. Naproti tomu v některých částech Severní Ameriky pomohla dlouhodobá předpověď pomoci zmírnit dopad velkých období sucha. (InMeteo, 2017)

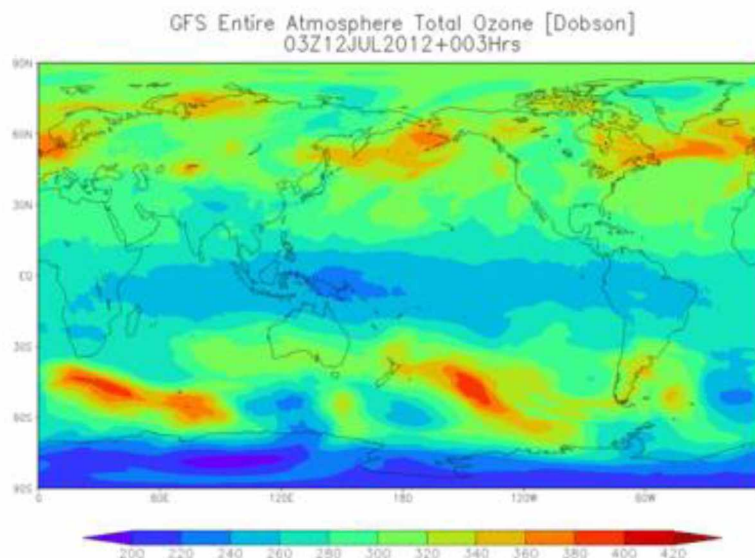
2.3 Obvyklé přístupy k předpovědi počasí

Nejčastějším přístupem k předpovědi počasí na profesionální úrovni je využití numerických modelů, které mají simulovat vývoj počasí. Tyto modely jsou obvykle založeny na vyjádření přírodních jevů, které jsou meteorologickými jevy či jejich působením tyto jevy vznikají. (InMeteo, 2017)

2.3.1 GFS

Global Forecasting System (tedy globální předpovědní systém) je numerický model provozovaný americkým národním centrem pro predikci životního prostředí. Předpovídá vývoj desítek parametrů na šestnáct dnů tj. 384 h. Model GFS vzniká spojením modelu atmosféry, oceánu, půdy a mořského ledu pro co nejpřesnější popis počasí. (InMeteo, 2017)

V současné době je rozlišovací schopnost tohoto systému třináct kilometrů a stále probíhá jeho vývoj a zlepšování. (InMeteo, 2017)



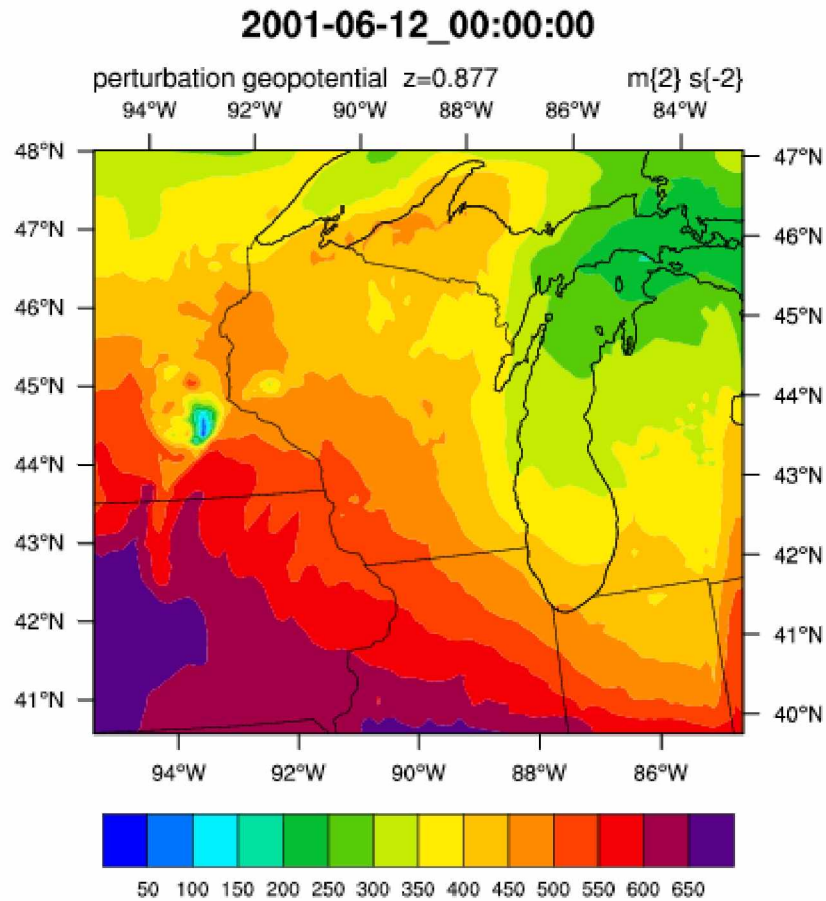
Obrázek 1 – GFS (National Centers for Environmental Information, 2020)

2.3.2 WRF Model

Weather Research and Forecasting (WRF) model je model numerické přípovědi počasí navržený pro výzkum atmosférických jevů a jejich předpovídání. Model je používán na meteorologické prvky v rozsahu od desítek metrů do tisíců kilometrů. Tento model je vyvíjen od pozdních let devadesátých společnou prací Národního střediska pro výzkum atmosféry

(NCAR), Národního správa oceánů a atmosféry, amerického letectva, námořní výzkumná laboratoře, univerzity v Oklahomě a Federální letecká správy (FAA).

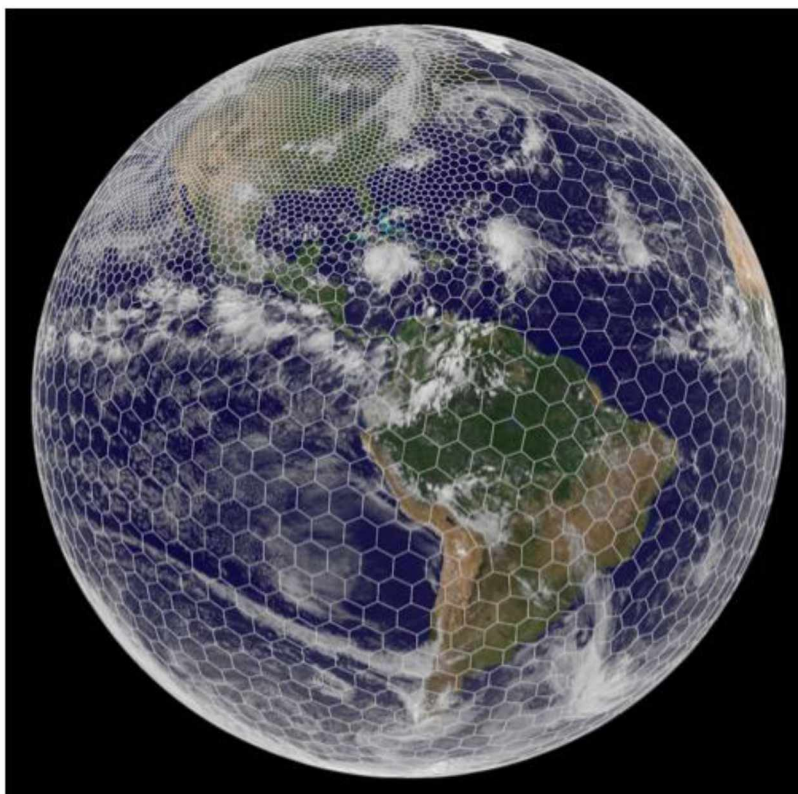
Tento model může poskytovat simulace buď na základě skutečných atmosférických podmínek (získaných pozorování) či za využití takzvaných ideálních podmínek. WRF má poměrně širokou uživatelskou základnu složenou převážně z výzkumných stanic, univerzit, a dokonce i komerčních společností. (UCAR , 2020)



Obrázek 2 - WRF model (UCAR, 2020)

2.3.3 MPAS

Model for Prediction Across Scales (MPAS) je kolaborativní projekt zaměřený na vývoj atmosféry, oceánu a dalších simulačních systémů Země pro použití v klimatických, regionálních klimatických a povětrnostních studiích. Primárními partnery vývoje je Národní laboratoř Los Alamos a Národní středisko pro výzkum atmosféry. (Los Alamos National Security, LLC (LANS) , 2020)



Obrázek 3 - MPAS (Los Alamos National Security, LLC (LANS) , 2020)

2.3.4 Amatérské meteostanice

Amatérské meteostanice lze dále rozdělit na dva druhy:

- Bez venkovního čidla či s jednoduchým čidlem teploty/vlhkosti ovzduší. Takovéto stanice se obvykle nacházejí v domácnostech, jakožto náhrada dříve hojně se vyskytujícího nástěnného barometru.



Obrázek 4 - Jednoduchá domácí meteostanice (Čermák, 2020)

Tyto stanice však předpovídají vývoj počasí pouze na základě pohybu tlaku, za předpokladu, že jsou vybaveny příslušným čidlem. Stanice pouze zobrazující internetovou předpověď jsou též jednou z variant.

- S venkovním čidlem na měření většina běžně měřených veličin (teplota, vlhkost, tlak, směr a rychlost větru, množství srážek)



Obrázek 5 - Poloprofesionální meteostanice (Čermák, 2020)

Tyto stanice se běžně vyskytují pouze u nadšených amatérských meteorologů, případně v profesionálních variantách na význačných místech, kde jsou zapojeny do komunitní sítě meteostanic. Předpoklad k předpovědi počasí je zde velmi dobrý, avšak obvykle se omezuje na odhad vývoje počasí opětovně dle vývoje změny barometrického tlaku či na hromadném zapojení dat z více stanic jakožto součást většího výpočetního modelu.

3 Základní pojmy strojového učení

„*Strojové učení* je oblastí umělé inteligence, zabývající se algoritmy a technikami, které umožňují počítačovému systému 'učit se'.“ (Matoušek)

Učením se v tomto kontextu chápe jako přizpůsobení vnitřního fungování systému tak, aby efektivněji přizpůsoboval změnám svého okolí.

Algoritmy strojového učení lze rozdělit dle druhu učení na následující kategorie:

Učení s učitelem je učení je za předpokladu že pro vstupní data máme k dispozici správná výstupní data na porovnání.

Učení bez učitele je oproti tomu učení systému bez znalosti správných výstupních dat. Tyto dva přístupy je možno kombinovat do takzvaného *kombinovaného učení* s učitelem a bez učitele.

Učení se zpětnou vazbou je učení, kde vstupem je zpětná vazba z výstupu existujícího systému či jeho simulace. Příkladem může být řízení chemické reakci pomocí fyzikálního modelu či řízení auta v reálném prostředí či simulátoru.

Dále lze rozdělit metody strojového učení dle způsobu zpracování dat na následující:

Dávkové, kde jsou všechna data dostupná již před začátkem výpočtu.

Inkrementální, jež jsou schopné zpracovat nové informace z nových dat bez nutnosti začít celý výpočet znovu od začátku.

Řešené úlohy lze rozdělit na následující druhy:

Klasifikace je úloha, kde je cílem rozdělit vstupní data do dvou či více tříd.

Regrese je úloha, kde je účelem odhadnout číselnou na výstupu podle hodnoty na vstupu.

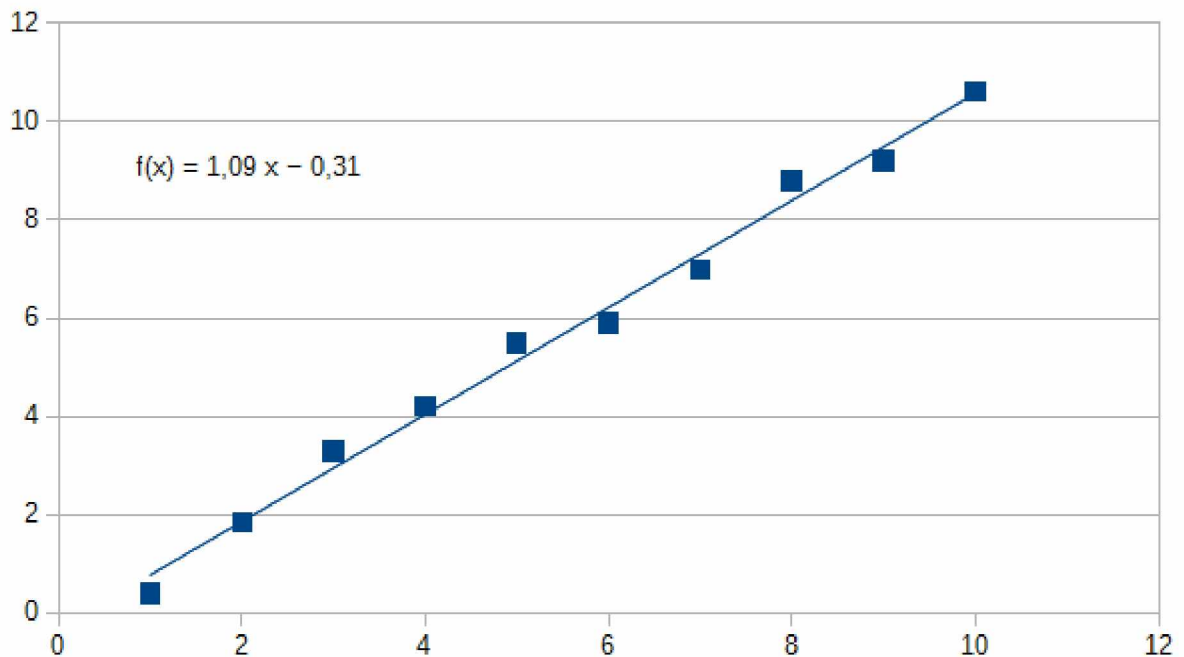
Shlukování je úloha, kde je účelem vytvořit skupiny (shluky) objektu s podobnými ználostmi.

(Matoušek)

Root Mean Squared Error (RMSE) alias odmocnina ze střední kvadratické chyby je nejznámějším a nejčastější metrikou chyby odhadu. Zde bude využívána pro znázornění chyby predikce oproti skutečně naměřeným hodnotám. (Havlík, 2014)

3.1 Metody strojového učení

3.1.1 Regresní metoda



Obrázek 6 - Lineární regrese (Moravec, 2020)

Lineární regrese je statistická metoda umožňující předvídat hodnoty funkce.

$$y = X\beta + \epsilon$$

Obrázek 7 - Rovnice lin. regrese

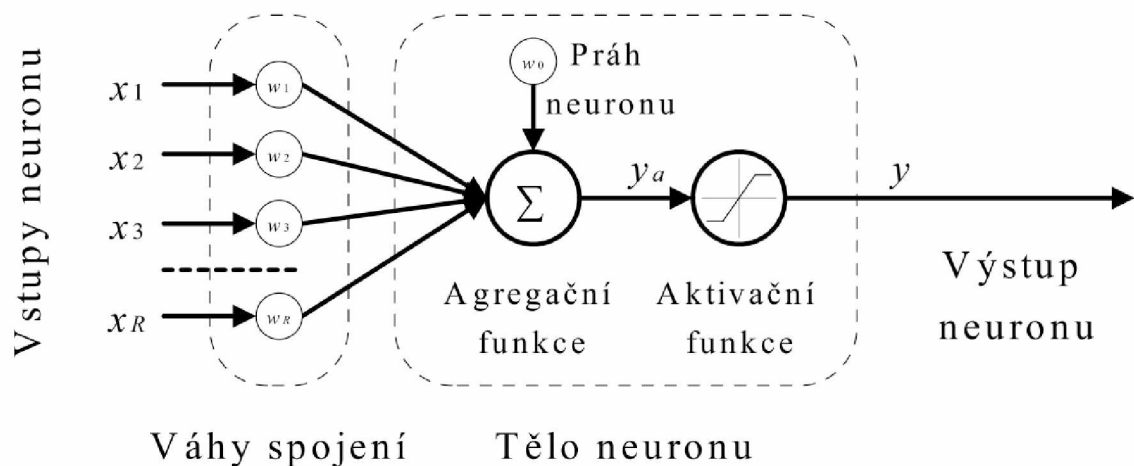
Ve většině případů se popisuje jakožto model, jehož cílem je se, co nejvíce přiblížit skutečné funkci, jež je znázorněna daty za pomoci sumy nejmenších čtverců. Lineární regrese je nejjednodušší variantou, vložená data lze samozřejmě prokládat téměř libovolnou funkcí. (Doc. Dr. Vladimír Homola, 2020)

3.1.2 Umělé neuronové sítě

Umělé neuronové sítě jsou odvozené od skutečných neuronových sítí, tj. mozku, kde vykonávají všechny rozhodovací a učící se procesy. Jejich model je tedy stvořen z formálních neuronů, které jsou přibližným obrazem reálných neuronů. (Ing. Petr Doležel, 2016)

3.1.2.1 Základní model neuronu

V rámci formálního neuronu jsou všechny biologické procesy nahrazeny matematickými funkcemi. (Ing. Petr Doležel, 2016)



Obrázek 8 - Model neuronu (Ing. Petr Doležel, 2016)

„Vstupy neuronu $x_1; x_2; \dots; x_R$ modelují dendrity, váhy spojení $w_1; w_2; \dots; w_R$ modelují synapse, výstupy simuluje činnost axonu. Agregace vstupních signálů, jejich porovnání s prahovou hodnotou w_0 a následně jejich nelineární zobrazení představují model těla neuronu, tj. vyhodnocení celkového vstupního aktivačního potenciálu a jeho transformaci na výstupní signál. Každý umělý neuron obsahuje konečný počet R vstupů a jediný výstup. Tento jediný výstup je samozřejmě možno rozmnožit do potřebného počtu kopií – vstupů do následných neuronů. V každém neuronu (popisovaný základní model nevyjímaje) se vstupní hodnoty transformují na výstup pomocí minimálně dvou výpočetních procedur. Konkrétně se jedná o výpočet vstupního potenciálu y_a a o tzv. aktivační funkci $f(\cdot)$.“ (Ing. Petr Doležel, 2016)

3.1.2.2 Neuronová síť

Neuronová síť je založena na množině spojených jednotek nebo uzlů nazývaných umělé neurony (analogické k biologickým neuronům v mozku). Každé spojení (synapse) mezi umělými neurony může přenášet signál z jednoho do druhého. Umělý neuron, který přijímá signál, dokáže zpracovat a poté předat výsledný signál do dalších umělých neuronů připojených k němu pomocí svých vlastních synapsí. (Ing. Petr Doležel, 2016)

U běžných neuronových sítí je signál na spojení mezi umělými neurony reálným číslem a výstup každého umělého neuronu je vypočítán nelineární funkcí součtu jeho vstupů. Umělé neurony a spojení obvykle mají váhu, která se přizpůsobuje, jak probíhá učení. Váha zvyšuje nebo snižuje sílu signálu. Umělé neurony mohou mít takovou hranici, že pouze pokud agregátní signál překročí tuto hranici, je odeslán signál. Typicky jsou umělé neurony uspořádány ve vrstvách. Různé vrstvy mohou na svých vstupech provádět různé druhy transformací. Signály

se pohybují od první (vstupní) do poslední (výstupní) vrstvy, případně po překročení vrstev vícekrát. (Ing. Petr Doležel, 2016)

Původním cílem bylo vyřešit problémy stejným způsobem, jakým by mohl lidský mozek. Postupem času se pozornost soustředila na přizpůsobení specifickým mentálním schopnostem, což vedlo k odchylkám od biologie. Umělé neuronové sítě byly použity na různé úkoly, včetně počítačového vidění, rozpoznávání řeči, strojového překladu, filtrování sociálních sítí a lékařské diagnostiky. (Alpaydin, 2014)

V rámci topologie obvykle rozeznáváme následující vrstvy a jim příslušné neurony:

Vstupní vrstvu, jež zajišťuje vstup signálů z okolí.

Skryté vrstvy, které však nemusí existovat, pokud si model vystačí s neurony z ostatních vrstev.

Výstupní vrstvu, jež předá signál do okolí.

Též lze sítě dělit podle toku signálu *dopředné*, kde signál postupuje pouze dopředu (jedním směrem) a *zpětnovazební*, kde signál může cestovat všemi směry. (Ing. Petr Doležel, 2016)

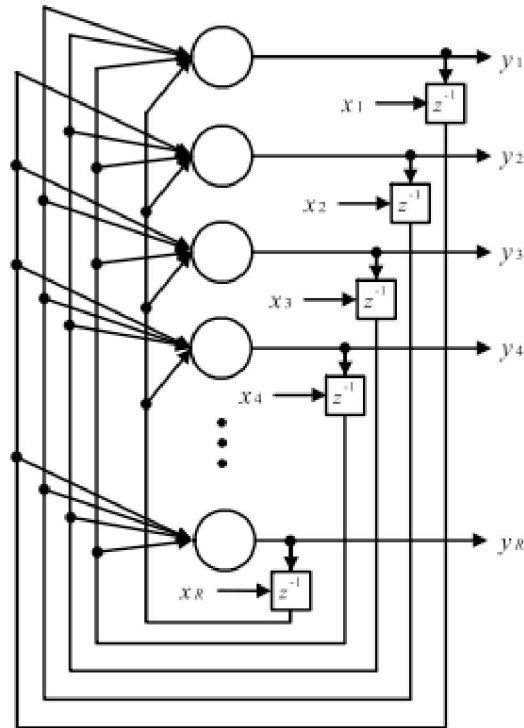
3.1.2.3 Lineární perceptron

Perceptron je jednou z historicky neznámějších variant neuronových sítí. Rozeznáváme jednoduchý perceptron (tvořen jedním neuronem) a vícevrstvý perceptron (tvořen více neurony).

Perceptron je algoritmus pro učení binárních klasifikátorů (funkce, které mohou rozhodnout, zda vstup, reprezentovaný číslem vektoru, patří do určité třídy nebo nikoliv). Jedná se o druh takzvaného klasifikačního algoritmu. Tento algoritmus umožňuje klasifikaci lineárně separovatelných dat. (Ing. Petr Doležel, 2016)

3.1.2.4 Hopfieldova síť

Je neuronová síť s neurony v jedné vrstvě a úplnými váženými spojeními mezi všemi neurony vyjma sebe sama. Výstup z této sítě tedy není závislý nejen na současném stavu vstupů do sítě, ale také na stavu všech neuronů obsažených v této síti. Hopfieldova síť obsahuje právě tolik neuronů, kolik je vstupů, respektive výstupů z neuronové sítě. Každý neuron je tedy vstupním i výstupním neuronem. Vzhledem k variantě propojení „každý s každým“ zde vzniká zpětnovazební smyčka, jde tedy o variantu zpětnovazební neuronové sítě. Hopfieldova síť je autoasociativní paměť, jejíž vybavování probíhá postupně v jednotlivých krocích, po kterých se zpracovávají vstupy v rámci této sítě. (Ing. Petr Doležel, 2016)



Obrázek 9 - Hopfieldova síť (Ing. Petr Doležel, 2016)

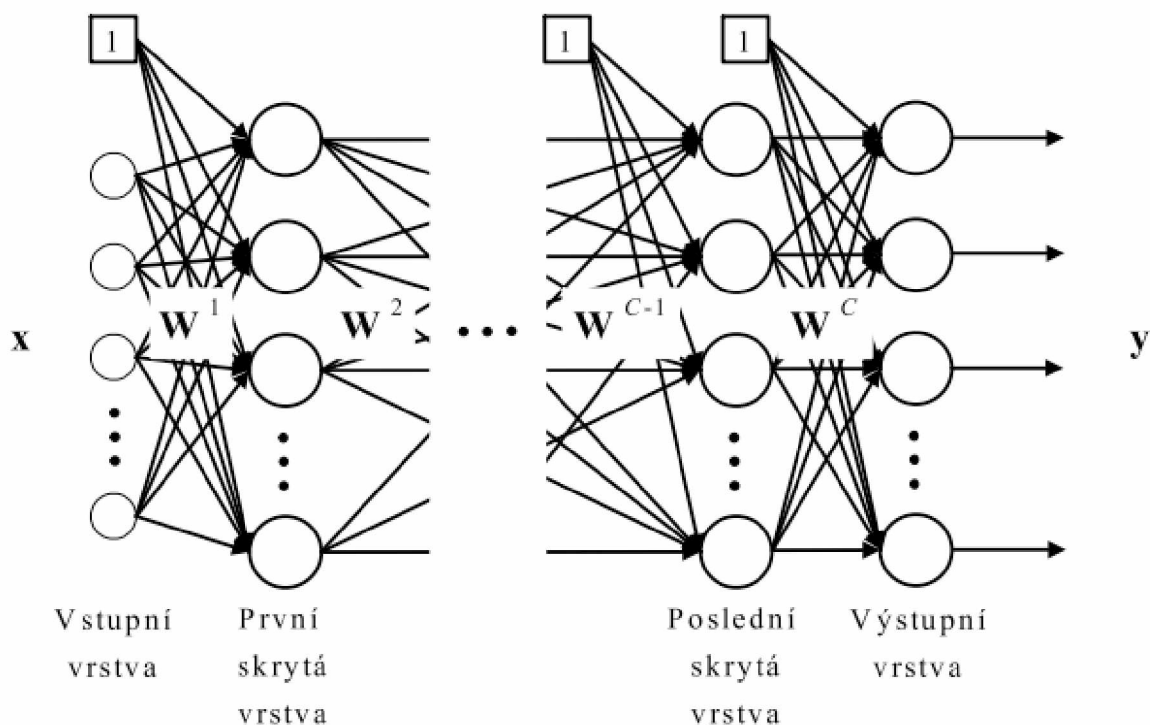
3.1.2.5 Dopředná vícevrstvá umělá neuronová síť / vícevrstvý perceptron

Dopředná vícevrstvá umělá neuronová síť je nejpoužívanějším typem neuronové sítě. Využívá se jako univerzální aproximátor, prediktor, rozhodovací člen a rozpoznávač vzorů pro kryptografii a podobně. Stejně jako ostatní zde zmíněné sítě je založena na formálním neuronu.

Často bývá dopředná vícevrstvá umělá neuronová síť označována jakožto vícevrstvý perceptron, přičemž zde však existuje rozdíl mezi těmito sítěmi. Oproti vícevrstvému perceptronu je u dopředné vícevrstvé umělé neuronové sítě možno použít libovolnou aktivační funkci neuronů. (Ing. Petr Doležel, 2016)

„Platí, že výstup z jednoho neuronu vstupem do několika dalších neuronů, podobně jako terminály axonu biologického neuronu jsou přes synaptické vazby spojeny s dendrity jiných neuronů.“ (Ing. Petr Doležel, 2016)

Základní charakteristikou je, že propojení neuronů existuje jenom mezi neurony přilehlých vrstev a je jednosměrné důsledkem tohoto je skutečnost, že neurony ve vlastní vrstvě mezi sebou a mezi dalšími vrstvami neurony propojeny nejsou a mezi neurony neexistují ani zpětné vazby.



Obrázek 10 - Dopředná vícevrstvá umělá neuronová síť (Ing. Petr Doležel, 2016)

„Pro návrh topologie neuronové sítě neexistuje analytický postup. V literatuře je uvedena řada různých doporučení, ale ta jsou vždy poplatná určité řešené úloze a mají tak značně subjektivní charakter. Většinou vycházejí ze série simulačních výpočtů nebo experimentů. Počet skrytých vrstev se volí podle složitosti řešené úlohy. Ze zkušeností vyplývá, že v řadě případů dosáhneme vyhovujícího chování za použití jediné skryté vrstvy. Pro odhad počtu neuronů p v této skryté vrstvě je někdy možno použít empirický vztah

$$p = \sqrt{RQ}.$$

Pro síť s dvěma skrytými vrstvami je dále např. doporučeno volit počty neuronů v těchto vrstvách p_1 a p_2 podle vztahů

$$p_1 = Q^{3/2} \sqrt{\frac{R}{Q}}, \quad p_2 = Q^3 \sqrt{\frac{R}{Q}}.$$

“ (Ing. Petr Doležel, 2016)

Ve většině případů se postupuje experimentálně, a to dvěma způsoby. Metodou postupného odebírání, kde se na začátku vytvoří přehnaně rozsáhlá neuronová síť (větší počet vrstev a neuronů) a postupně se odebírají vrstvy a neurony, dokud síť vyhovuje úloze. Nebo metodou

přírůstků, kde se vytvoří jednodušší síť a postupně se do ní přidávají neurony a vrstvy, dokud výstup nebude vyhovovat dané úloze. (Ing. Petr Doležel, 2016)

3.1.2.6 Zpětné šíření chyby – Backpropagation

Zpětné šíření chyby je algoritmus učení umělých neuronových sítí s učitelem pomocí gradientu sestupu. Pro danou umělou neuronovou síť a její chybovou funkci vypočítává tato metoda gradient chybové funkce s ohledem na váhy v dané neuronové síti. (Brilliant.org, 2020)

Název „zpětné“ se odkazuje na skutečnost, že k výpočtu gradientu postupuje postupně od výstupních vrstev až po vrstvy vstupní, čímž je zajištěno, že váhy výstupních neuronů jsou započítány jakožto první a váhy vstupních neuronů jakožto poslední, přičemž část výpočtu se opětovně použije při výpočtu pro vrstvu následující. Tento specifický přístup k výpočtu gradientu zajišťuje efektivnější výpočet oproti výpočtu pro jednotlivé vrstvy samostatně (bez započítání zde popsaného „přelivu“ chyby mezi jednotlivými vrstvami) (Brilliant.org, 2020)

Tento algoritmus je považován za efektivní algoritmus, obzvláště při řešení rozpoznávání zvuku a obrazu při využití grafických karet pro zvýšení výkonu. (Brilliant.org, 2020)

Detailní popis chování a výpočtu v rámci tohoto algoritmu a jeho případné porovnání s konkurenčními algoritmy vysoce přesahuje rozsah této práce, je tedy na čtenáři, aby v případě zájmu využil například zde citovanou literaturu.

4 Praktická předpověď

Pro účely této práce byly vybrány následující ukazatele stavu počasí v našem okolí:

1. Teplota ovzduší – jde o nejčastěji udávanou číselnou hodnotu popisující stav atmosféry okolo nás, je také nejčastěji sledována širokou veřejností.
2. Atmosférický tlak – též velmi často udáván a sledován. I v domácích podmínkách umožňuje předpovídat vývoj počasí v dané lokalitě.
3. Relativní vlhkost ovzduší – zvolen z důvodu jednoduchosti měření v domácích podmínkách a návaznosti hodnot na jinak obtížně vyčíslitelné atmosférické jevy jako je například mlha a déšť.

Uživatelsky velmi oblíbené předpovědi pomocí piktogramů označujících deštivo, slunečno, sněžení a podobně jsou bohužel těžko vyjádřitelné v absolutních číslech a proto se zde s nimi nebudeme zabývat.

Rozsah vstupních dat z hlediska času je stanoven na interval osmi let v rozsahu od 1. dubna roku 2010 do 31. března roku 2018 včetně. Před tímto datem jsou bohužel v dostupných údajích mezery trvající i několik dní, proto byl zvolen úsek bez těchto výpadků pro omezení možného nežádoucího ovlivnění výsledků.

Lokalita získaných dat je hlavní město Praha, zde vybrané pro jednoduchost získání historických údajů i konzistenci v samotných měřeních.

Data zde uvedená jsou s jednotlivými pozorování po celé jedné hodině, výstupem všech pokusů budou data s časovým odstupem též jednu hodinu. Tímto získáváme více než 67 tisíc záznamů pro provádění experimentů.

Historické údaje byly získány ze serveru timeanddate.com, jenž umožňuje relativně jednoduché získání dat z vystavených scriptů. (Time and Date AS, 2020)

Teplota je udávána ve stupních celsia, vlhkost v procentech a barometrický tlak v hPa případně mbar, pokud není uvedeno jinak.

Pro výpočty a předpovědi bylo využito softwaru Matlab, jenž poskytuje dobře zdokumentované prostředí pro práci s množinami dat a neuronovými sítěmi.

4.1 Praktické učení a předpověď hodnot časové řady

4.1.1 Získání a příprava dat

Pro získání (extrakci) potřebných historických dat bylo nutné vytvořit miniaplikaci, která se postarala o odeslání potřebných parametrů na server protistrany a následné uložení odpovědi do přijatelného formátu v lokálním úložišti.

Dále bylo nutné data předzpracovat, tj. odfiltrovat pouze chtěné hodnoty, ověřit, zda není přerušena posloupnost dat, či se zde nevyskytuje neočekávaná hodnota. Náprava výskytů odlehlých/chybějících hodnot byla provedena ručně, jejich výskyt byl ve většině případů ohlášen chybou scriptu či aplikace.

Též byl vytvořen automatický cron job a skript jenž obstarává stažení aktuální předpovědi pro následující den, a to vždy právě o půlnoci, aby bylo dodrženo předpovídání na následujících 24 h. Tuto akci bylo nutná provést záhy, jelikož jednotlivé předpovědi se nearchivují a pokud ano pak nejsou dostupné veřejnosti.

Veškeré znovupoužitelné scripty/aplikace jsou v příloze této práce v adresáři *scripts*. Veškeré datové soubory, které byly získány ať již došlo k jejich využití či zpracování jsou v příloze této práce v adresáři *data*, kde jsou dále rozděleny do dvou adresářů, a to *historic* a *forecast*, dle toho, zda jde o data z předpovědního modulu yr.no či o historická data/pozorování.

V Příloze A, která je součástí této práce je seznam použitého softwaru využitého při tvorbě této práce.

4.1.2 Příklad nastavení a spuštění předpovědního modelu pro jednu vstupní a výstupní hodnotu

Pro simulaci umělé neuronové sítě byla využita funkce:

```
trainNetwork(XTrain, YTrain, layers, options);
```

jenž je dostupná v novějších verzích aplikace Matlab.

Jakožto algoritmičtí řešitelé aplikovatelní v této funkci jsou k dispozici *sgdm* (využívá klesající gradient s optimalizátorem hybnosti), *rmsprop* (udržuje klouzavý průměr čtverce gradientů, dělí gradient odmocninou tohoto průměru) a nejoblíbenější *adam* (metoda stochastického gradientu sestupu, která je založena na adaptivním odhadu momentů prvního a druhého řádu). (MathWorks, 2020)

V rámci prováděných experimentů jsou vyzkoušeny všechny tyto algoritmy, přičemž výsledky se pro přehlednost použije pouze z nejúspěšnějšího případně nejrychlejšího z kandidátů. Pokud není uvedeno jinak pak použitý model obsahuje řešitelský algoritmus *adam*.

Pro předpovídání hodnot za využití již naučené sítě pomocí předchozího příkladu bylo využito následující funkce:

```
predictAndUpdateState(net,YPred(:,i-1),'ExecutionEnvironment','cpu');
```

Množství vstupu a výstupů bylo zvoleno v závislosti na právě prováděném experimentu. Množství neuronů ve skrytých vrstvách bylo pro první spuštění nastaveno na hodnotu 200 a posléze bylo provedeno několik pokusů, zda změna počtu způsobí zlepšení kvality výstupů. Množství iterací bylo standardně ponecháno na 250, přičemž prvotní úroveň rychlosti učení byla nastavena na 0.005, tato hodnota po uplynutí 125 tj. poloviny iterací klesne na 0.001 pro zjemnění a zpřesnění výsledků.

Vstupem a výstupem do neuronové sítě byla vždy standardizovaná hodnota pro zajištění stejného výkonu sítě při použití libovolných jednotek. To samozřejmě znamená nutnost „odstandardizovat“ výstup z neuronové sítě před jeho použitím např. v grafu.

Následuje příklad zdrojového kódu aplikace Matlab pro trénink a předpověď teploty za využití výše zmíněných funkcí a hodnot, zde pro rozložení tréninkových a testovacích dat v poměru 9:1. Doba běhu tohoto skriptu je závislá: na množství použitých dat, vybavení a výkonu zařízení na němž je spuštěn.

```
A = importdata('history.txt',';');
temp = A.data(:,1);
data = temp';

figure
plot(data)
xlabel("Hodiny")
ylabel("Teplota [°C]")
title("Vývoj teploty")

numTimeStepsTrain0 = floor(0.9*numel(data));

dataTrain = data(1:numTimeStepsTrain0+1);
dataTest = data(numTimeStepsTrain0+1:end);

mu = mean(dataTrain);
sig = std(dataTrain);

dataTrainStandardized = (dataTrain - mu) / sig;
```



```

XTrain = dataTrainStandardized(1:end-1);
YTrain = dataTrainStandardized(2:end);

numFeatures = 1;
numResponses = 1;
numHiddenUnits = 200;

layers = [ ...
    sequenceInputLayer(numFeatures)
    lstmLayer(numHiddenUnits)
    fullyConnectedLayer(numResponses)
    regressionLayer];

options = trainingOptions('adam', ...
    'MaxEpochs',250, ...
    'GradientThreshold',1, ...
    'InitialLearnRate',0.005, ...
    'LearnRateSchedule','piecewise', ...
    'LearnRateDropPeriod',125, ...
    'LearnRateDropFactor',0.2, ...
    'Verbose',0, ...
    'Plots','training-progress');

net = trainNetwork(XTrain,YTrain,layers,options);

dataTestStandardized = (dataTest - mu) / sig;
XTest = dataTestStandardized(1:end-1);
net = predictAndUpdateState(net,XTrain);
[net,YPred] = predictAndUpdateState(net,YTrain(end));

numTimeStepsTest = numel(XTest);
for i = 2:numTimeStepsTest
    [net,YPred(:,i)] = predictAndUpdateState(net,YPred(:,i-1),'ExecutionEn-
vironment','cpu');
end
YPred = sig*YPred + mu;

YTest = dataTest(2:end);
rmse = sqrt(mean((YPred-YTest).^2))

figure
plot(dataTrain(1:end-1))
hold on
idx = numTimeStepsTrain:(numTimeStepsTrain+numTimeStepsTest);
plot(idx,[data(numTimeStepsTrain) YPred],'.-')
hold off
legend(["Skutečnost" "Předpověď"])
figure
subplot(2,1,1)
plot(YTest)
hold on
plot(YPred,'.-')
hold off
legend(["Skutečnost" "Předpověď"])
ylabel("Teplota [°C]")
title("Předpověď")

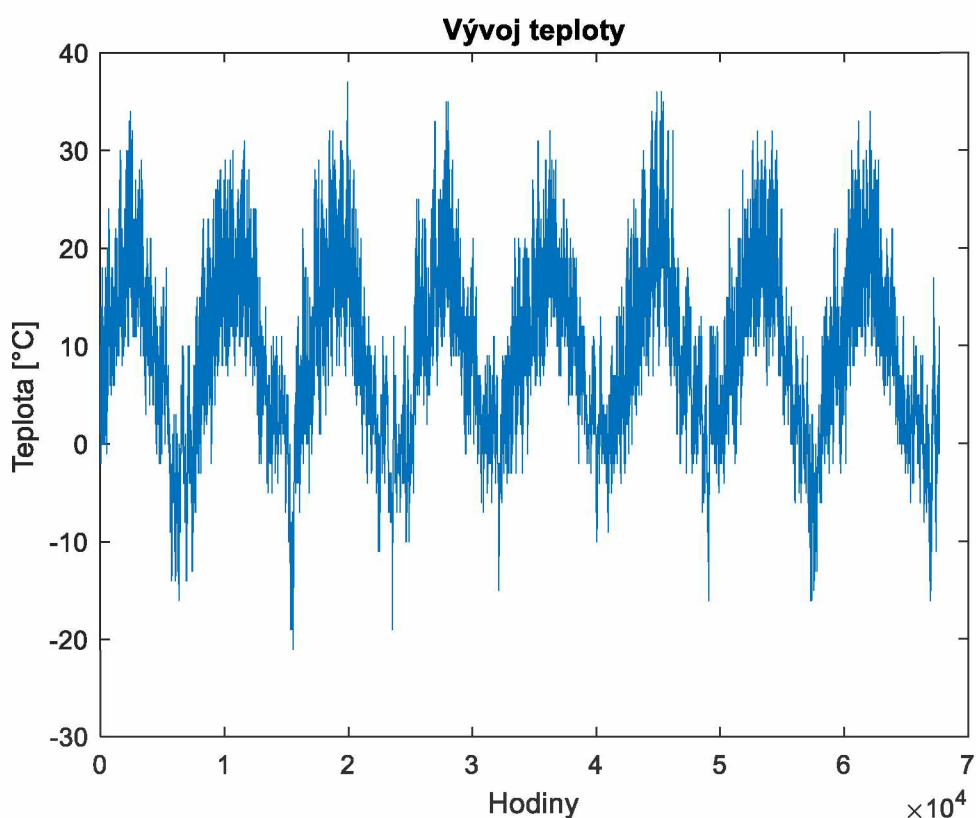
subplot(2,1,2)
stem(YPred - YTest)

```

```
xlabel("Hodiny")  
ylabel("Error")  
title("RMSE = " + rmse)
```

4.2 Předpověď teploty

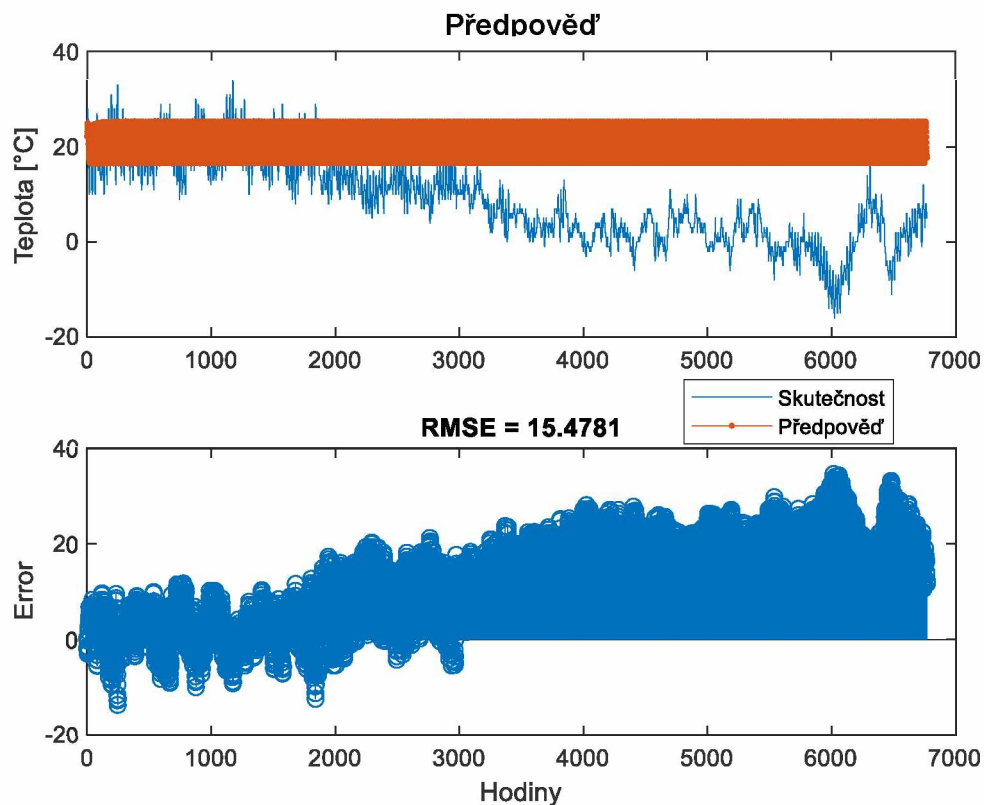
Jakožto první experiment je zadána možnost předpovídat teplotu pouze na základě znalosti předchozí teploty a ničeho jiného. Je známým faktem, že teplota ovzduší se pohybuje v cyklech s tím, jak se postupně mění roční období a též v jednotlivých denních cyklech. Oba cykly jsou způsobeny vzájemným pohybem a postavením Země oproti Slunci, což má za následek právě toto chování.



Obrázek 11 - Vývoj teploty

Na výše uvedeném obrázku je cykličnost v rámci ročního cyklu krásně viditelná, je zde rozpoznatelných osm nosných křivek, každá pro jednotlivý rok.

Pokud zvolíme obvyklý poměr 90 % dat pro učení sítě a 10 % na ověření pak při využití 200 skrytých neuronů získáme následující předpověď.

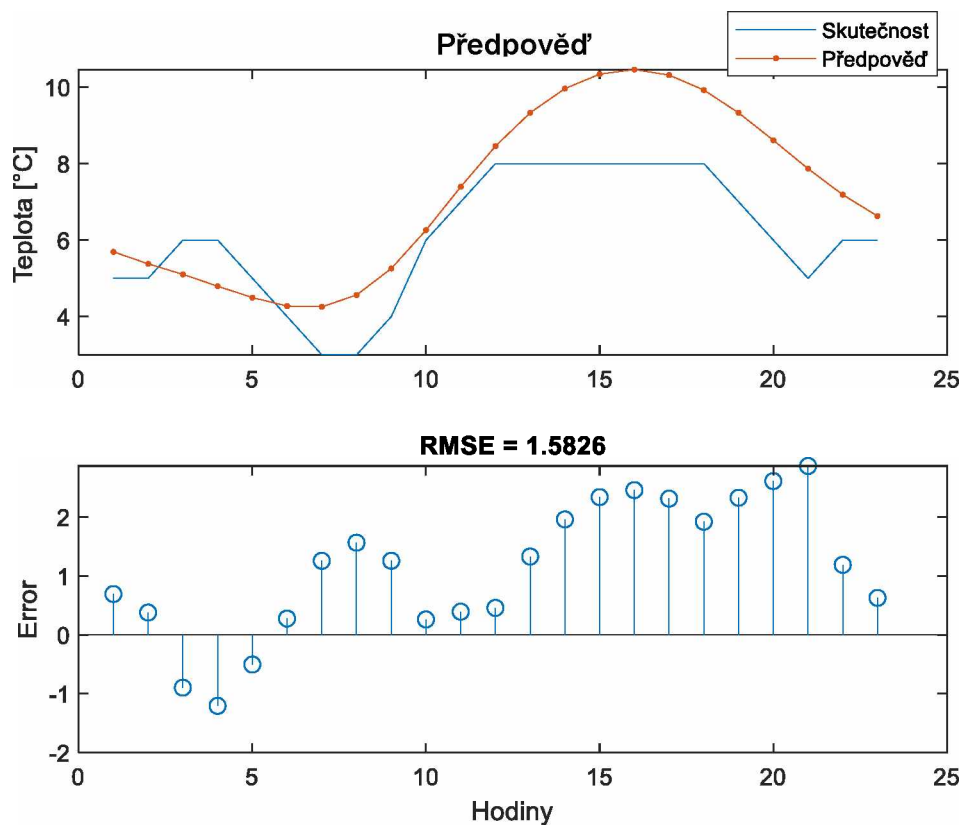


Obrázek 12 - Předpověď teploty

Zde uvedených 10 % odpovídá přibližně jedné osmině roku. Šlo by tedy o dlouhodobou předpověď. Bohužel jak je již evidentní ze znázorněných dat, pro dlouhodobou předpověď je tento model zcela nepoužitelný. Rozdíl 15 °C mezi předpovědí a skutečností je pro praktické použití příliš velký.

Zvýšení počtu neuronů ve skrytých vrstvách nepřináší lepší výsledky, stejně tak zvýšení množství iterací se neprojevilo kladným účinkem.

Jelikož předmětem této práce je však krátkodobá předpověď, pojďme se tedy podívat na výsledek krátkodobé předpovědi na následujících 24 h.



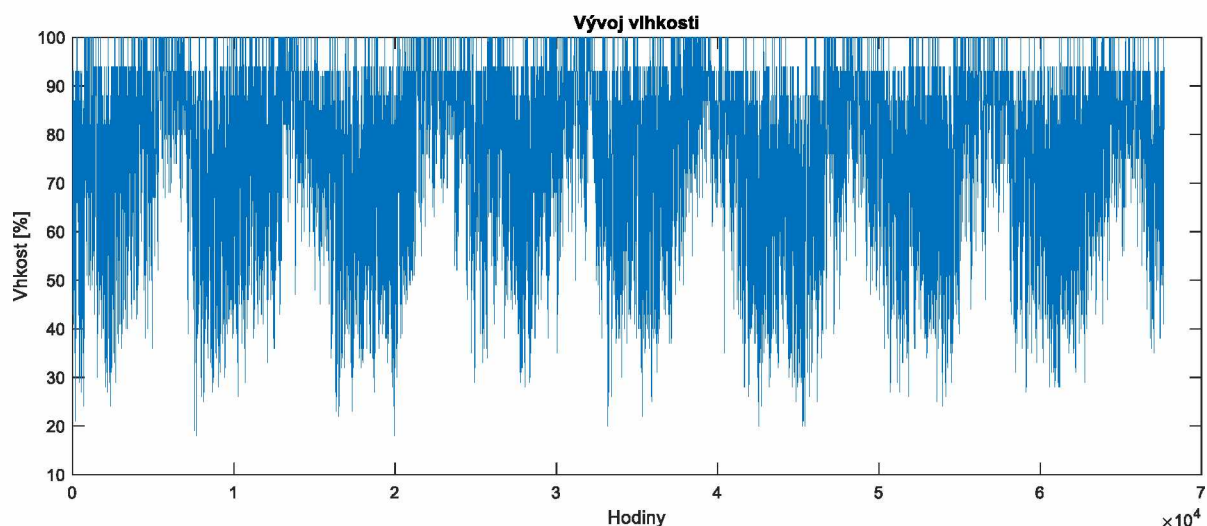
Obrázek 13 - Předpověď teploty na 24h

Zde je již patrné, že námi vytrénovaná neuronová síť relativně dobře pochytila chování teploty v denním cyklu oproti cyklu ročnímu, jak jsme se přesvědčili výše.

Rozdíl skutečnosti oproti předpovědi okolo 2 °C lze považovat za velice dobrý a pro praktické použití za dostačující. Zde tedy naše neuronová síť splnila svou funkci. Přidáním neuronů či iterací se zde opětovně nepodařilo dosáhnout lepších výsledků.

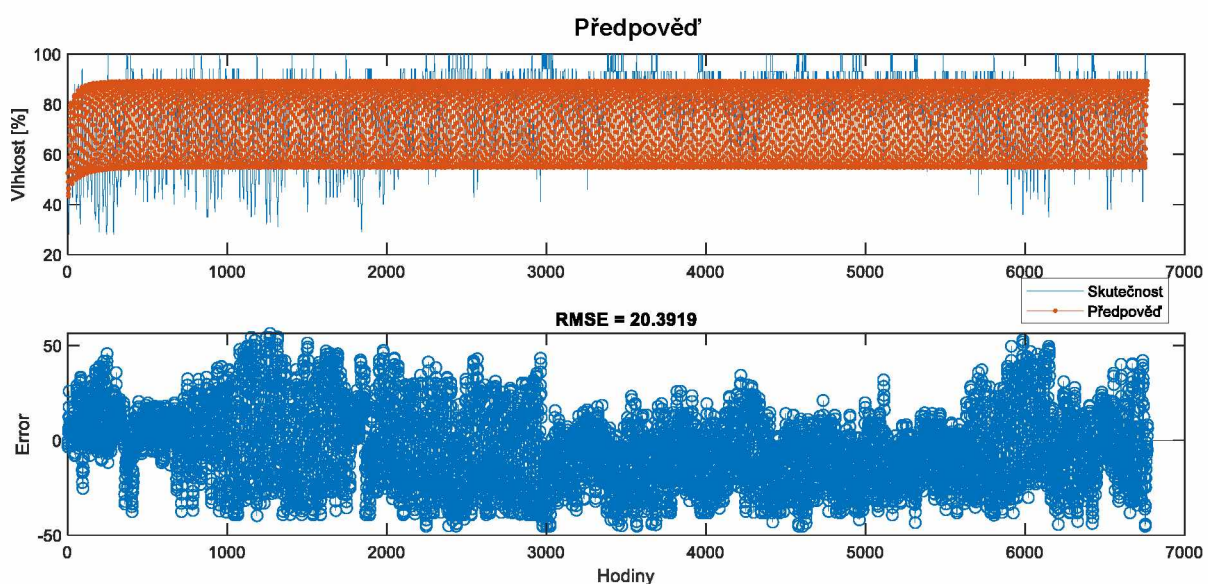
4.3 Předpověď vlhkosti

Podívejme se tedy na vývoj vlhkosti v našem období.



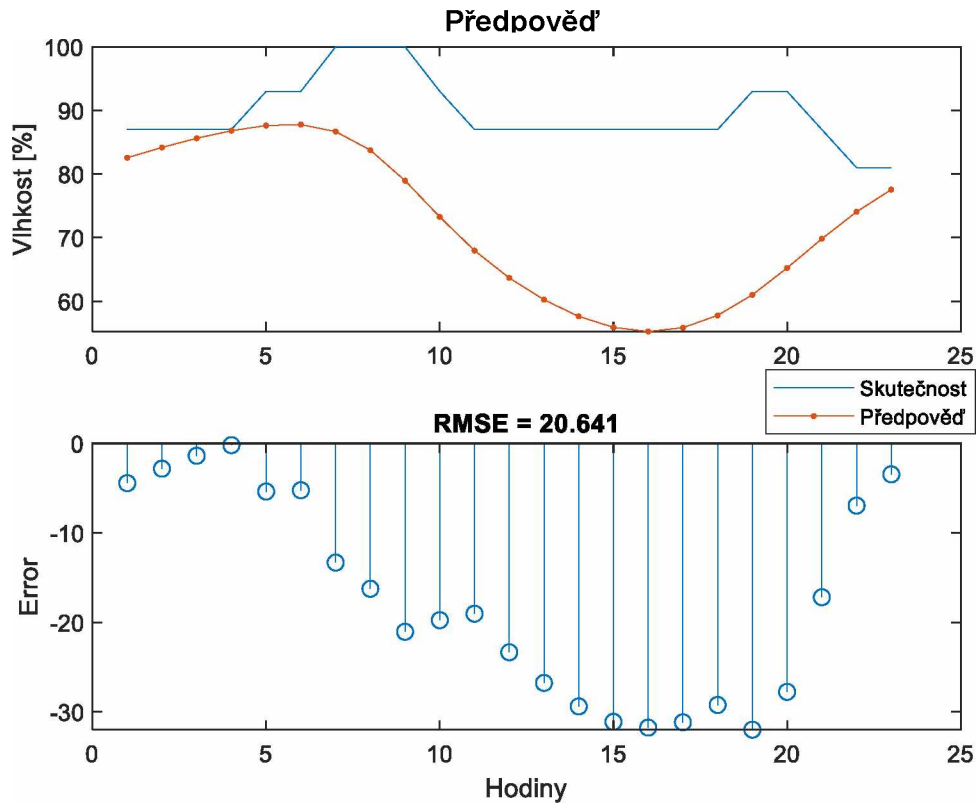
Obrázek 14 - Vývoj vlhkosti

I zde je vidět jistá periodicita vlhkých a suchých období. Bohužel i zde je výsledek nevalný při pokusu o dlouhodobou předpověď.



Obrázek 15 - Předpověď vlhkosti

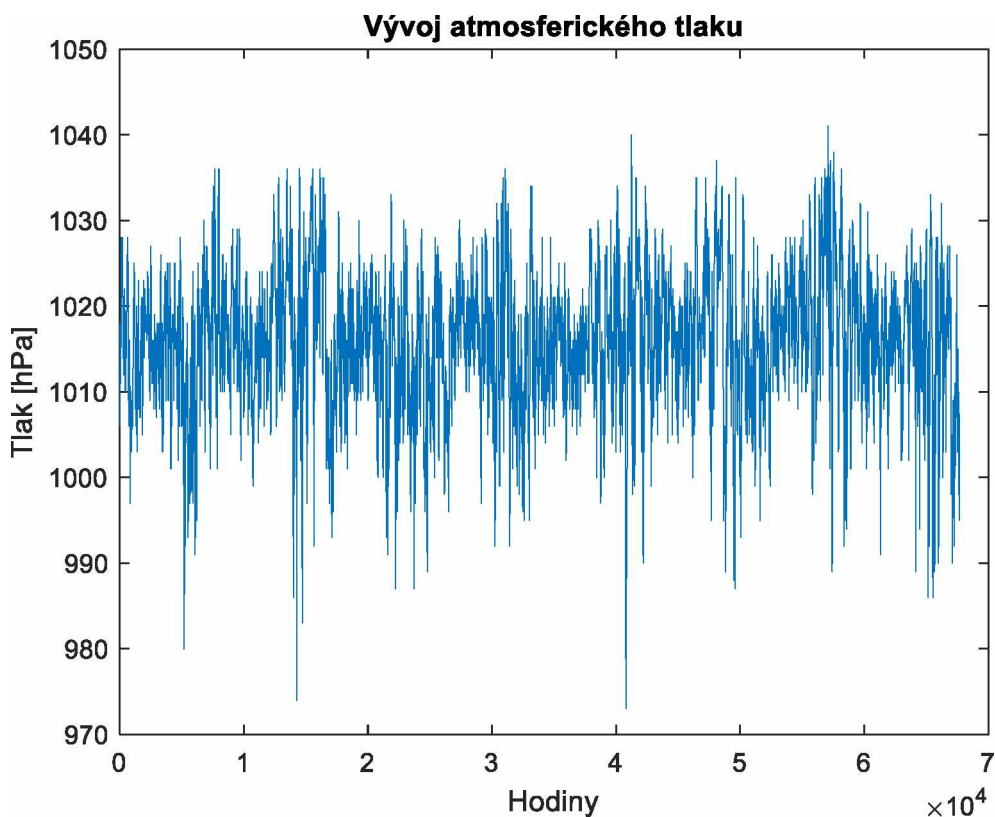
Chyba pohybující se okolo 20procentních stupňů je již mimo akceptovatelnou oblast, obzvláště pokud vezmeme v úvahu, že výsledku není opět patrná cyklická složka založená na střídání ročních období. Podívejme se tedy opět na krátkodobou předpověď na 24 h.



Obrázek 16 - Předpověď vlhkosti na 24 h

Zde je chybovost prakticky stejná, nelze tedy daný model prakticky využít pro předpověď vývoje vlhkosti, a to ani pro krátkodobou předpověď. S největší pravděpodobností je zde chybějící veličina, která by mohla ovlivnit výsledky předpovědi. Přidáním neuronů či iterací se zde opětovně nepodařilo dosáhnout lepších výsledků.

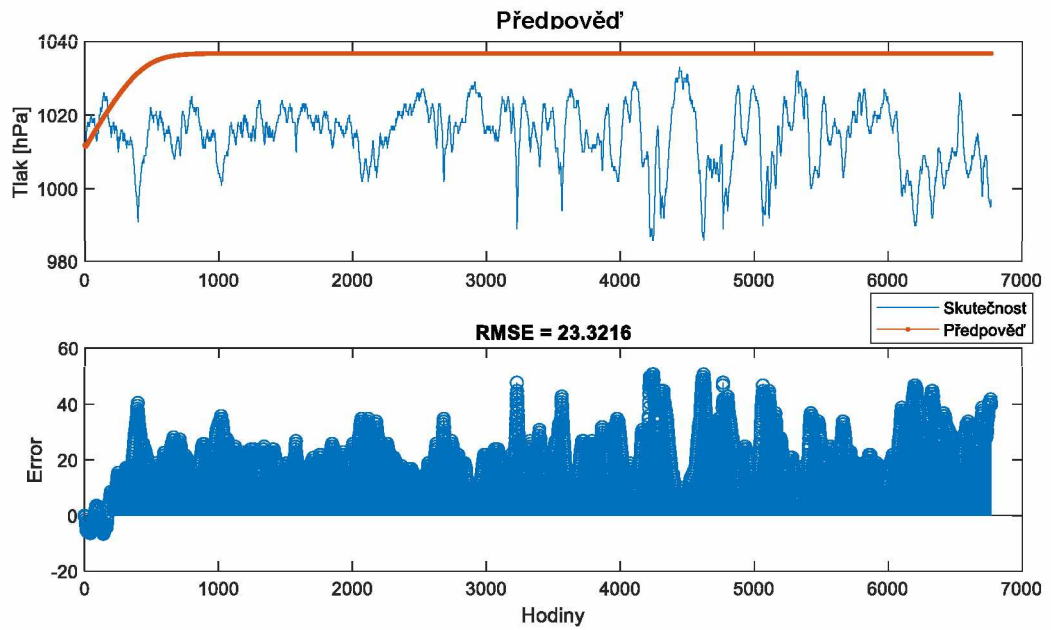
4.4 Předpověď tlaku



Obrázek 17 - Vývoj atmosférického tlaku

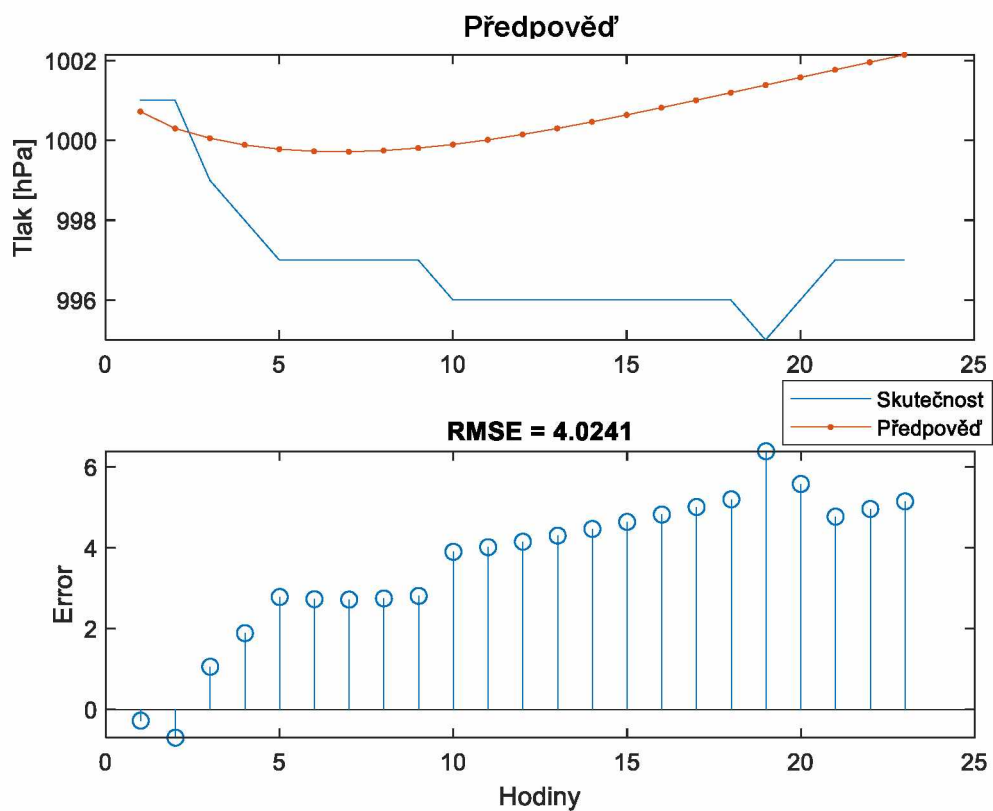
Při pohledu na vývoj atmosférického tlaku v našem období, lze zahlédnout též náznaky periodicity, ač již nikoliv tak zřetelně. Z pohledu meteorologie jsou velice zajímavá minima, která se zde nacházejí pod hodnotou 980 hPa. Jedno z těchto minim pravděpodobně bude mít souvislost s povodněmi v roce 2011.

Při pohledu na naši dlouhodobou předpověď je jasné, že naše síť se chopila pouze jednoho trendu a následně kompletně „přestřelila“ většinu hodnot. Z pohledu meteorologie toto znamená, že naše předpověď vždy hlásila hezké počasí, ačkoliv právě v oblasti zuřila bouřka. Nutno však podotknout, že tento výsledek byl očekáván, dlouhodobá předpověď je obecně považována za velmi složitou a extrémně nepřesnou.



Obrázek 18 - Předpověď vývoje atmosférického tlaku

Podívejme se tedy na naši krátkodobou předpověď.

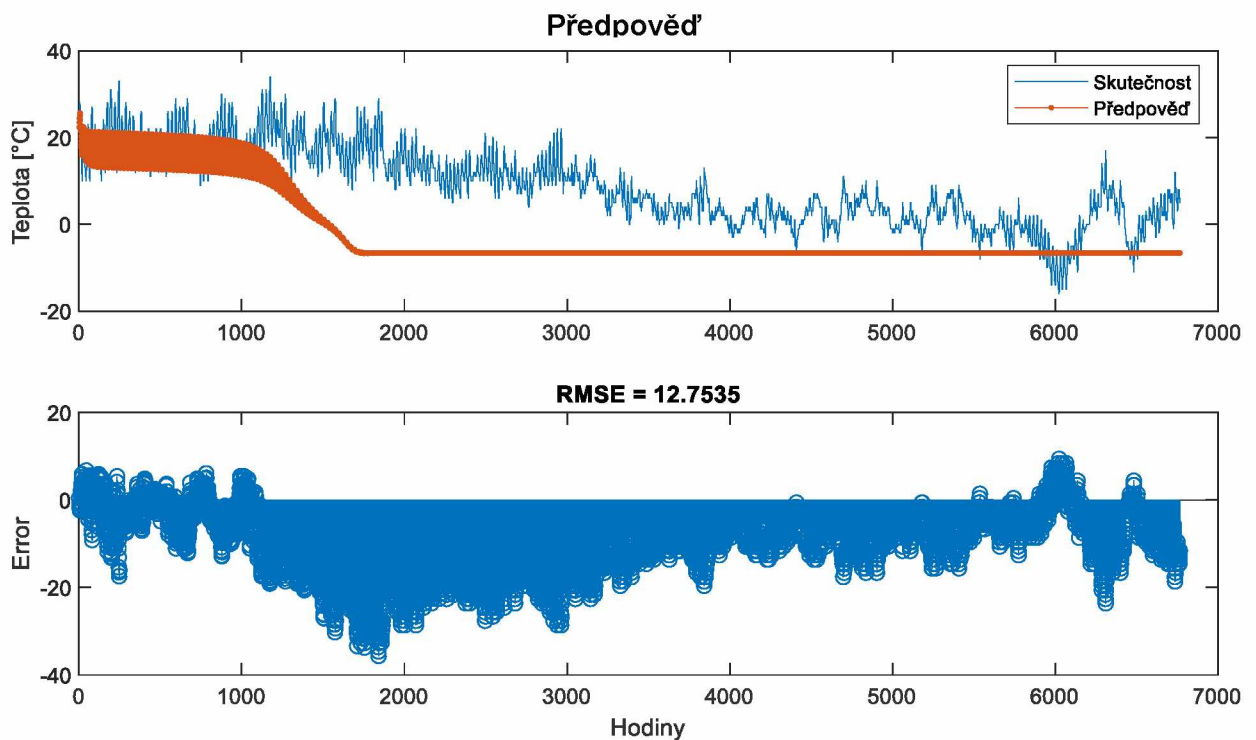


4.5 Kombinované předpověď teploty a vlhkosti

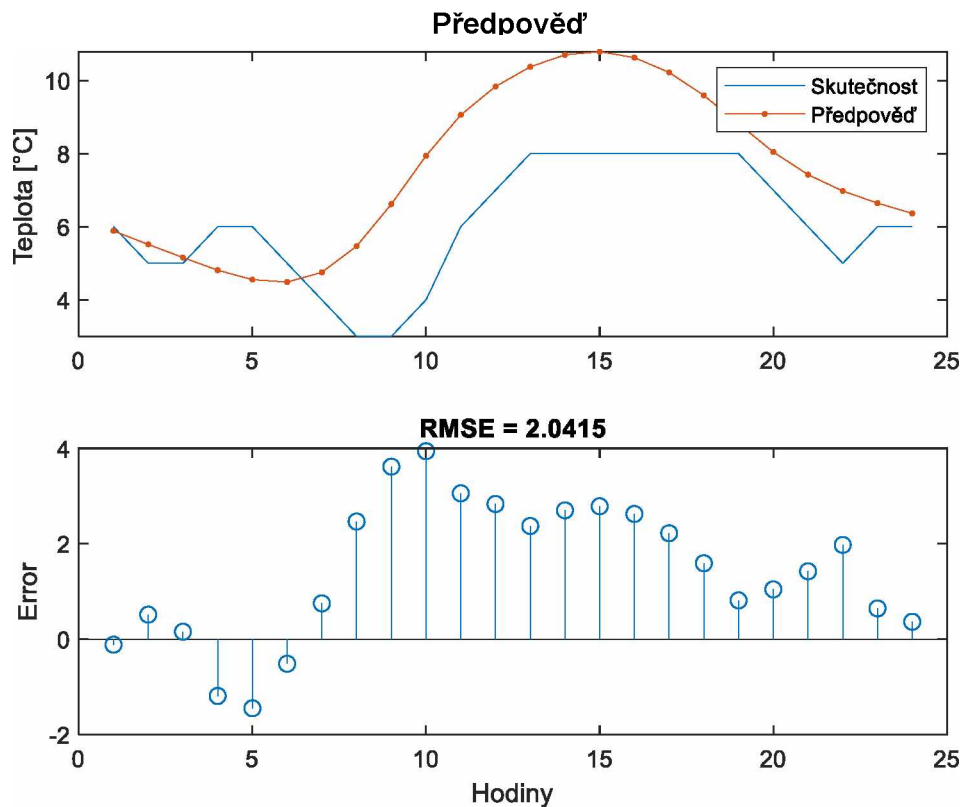
Pro další experiment se pokusme provést předpověď ze dvou vstupních veličin na dvě výstupní veličiny. Zde teploty a vlhkosti a sledujme, zda se předpověď zlepší oproti použití predikce na samostatné veličiny.

Překvapivě je v rámci dlouhodobé předpovědi vidět zlepšení, které však nemůže přesáhnout přibližně 1000 hodin, což je přibližně 1,3 měsíce.

O to zajímavější je pohled na krátkodobou předpověď.



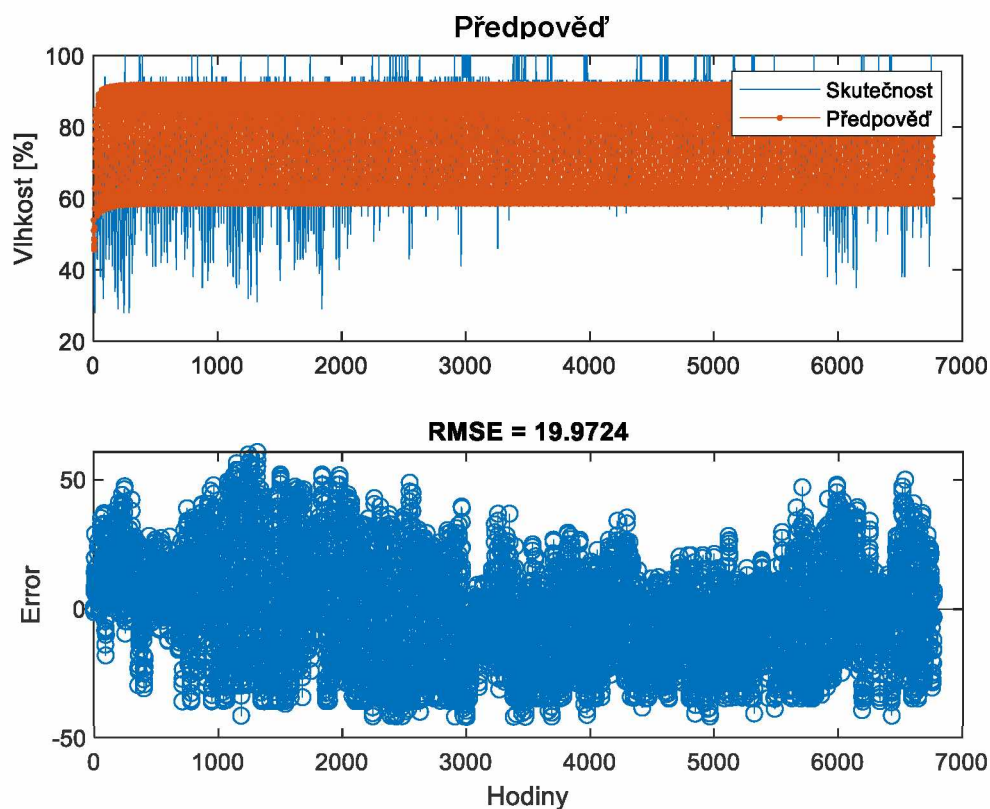
Obrázek 19 - Kombinovaná předpověď teploty a vlhkosti – teplota



Obrázek 20 - Kombinovaná předpověď teploty a vlhkosti – teplota na 24 h

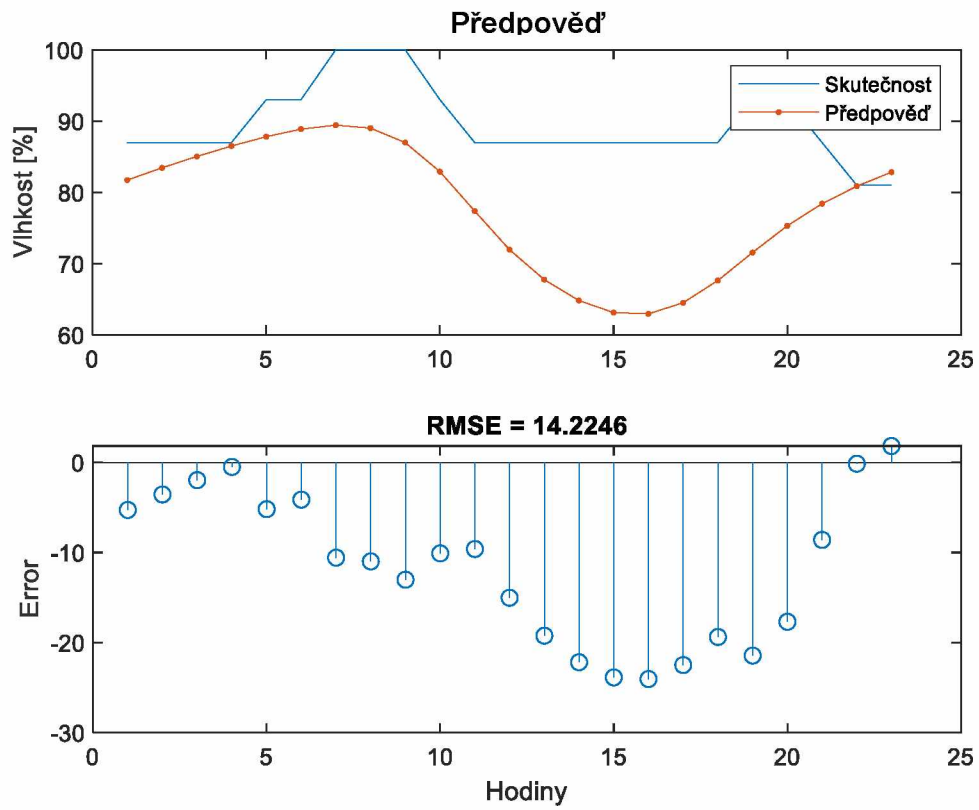
Zde je překvapením, že došlo naopak ke zhoršení přesnosti předpovědi. Je zde tedy možné, že přidání vlhkosti jakožto další veličiny vstupující do neuronové sítě zvýšilo hladinu šumu, jenž se takto projevuje na výstupu, tj. předpovědi samotné.

Porovnejme ještě tedy výsledky predikce vlhkosti:



Obrázek 21 - Kombinovaná předpověď teploty a vlhkosti – vlhkost

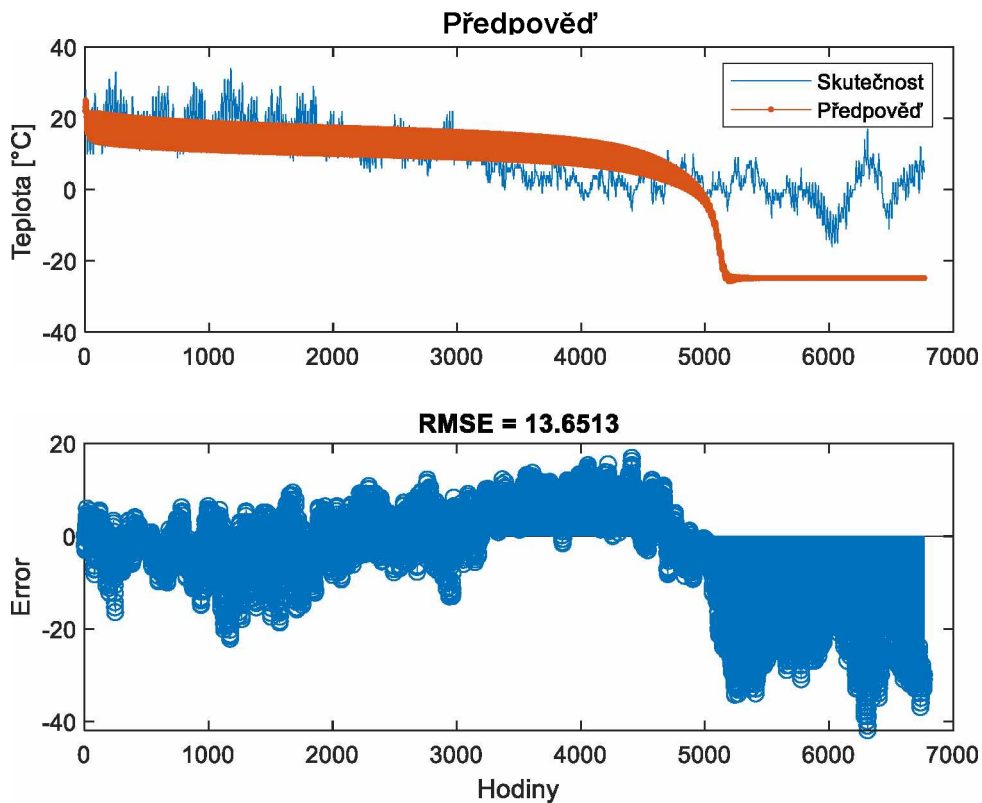
I zde je vidět drobné zlepšení v rámci dlouhodobé předpovědi. Při pohledu na krátkodobou předpověď viz níže, je vidět, že se přesnost předpovědi opět zlepšila. Jedinou oblastí, v které tedy došlo ke zhoršení výsledků, byla krátkodobá předpověď teploty.



Obrázek 22 - Kombinovaná předpověď teploty a vlhkosti – vlhkost na 24 h

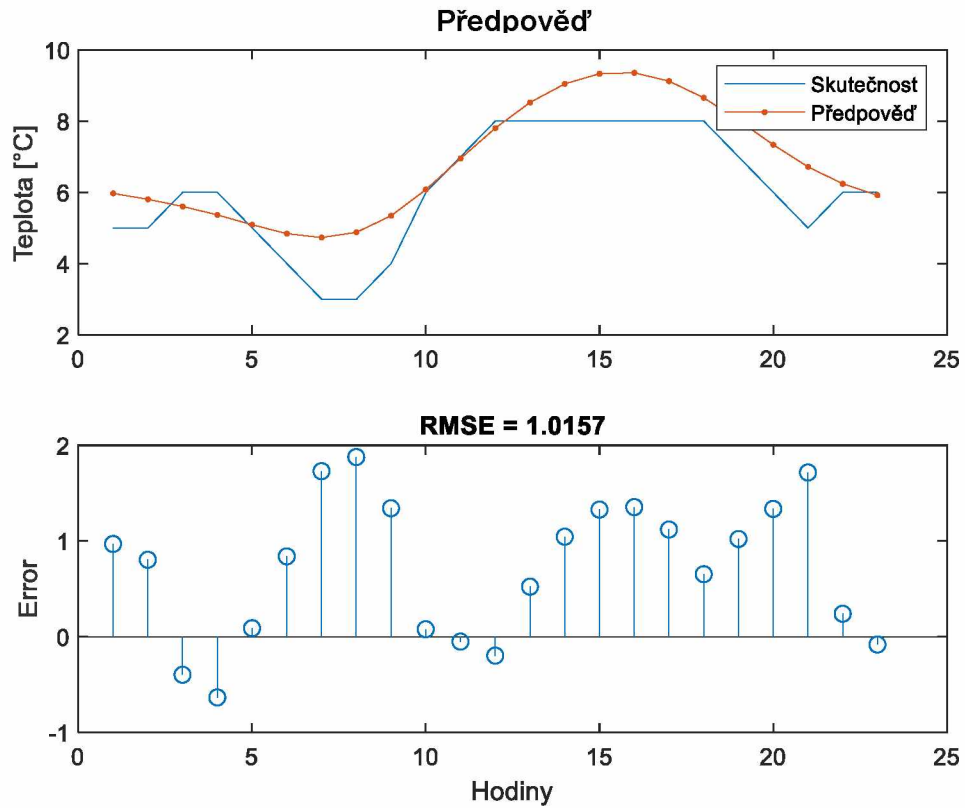
4.6 Kombinovaná předpověď teploty a atmosférického tlaku

Nyní se podíváme na kombinaci dvou nejčastěji používaných ukazatelů v případech, když se obvykle mluví o počasí.

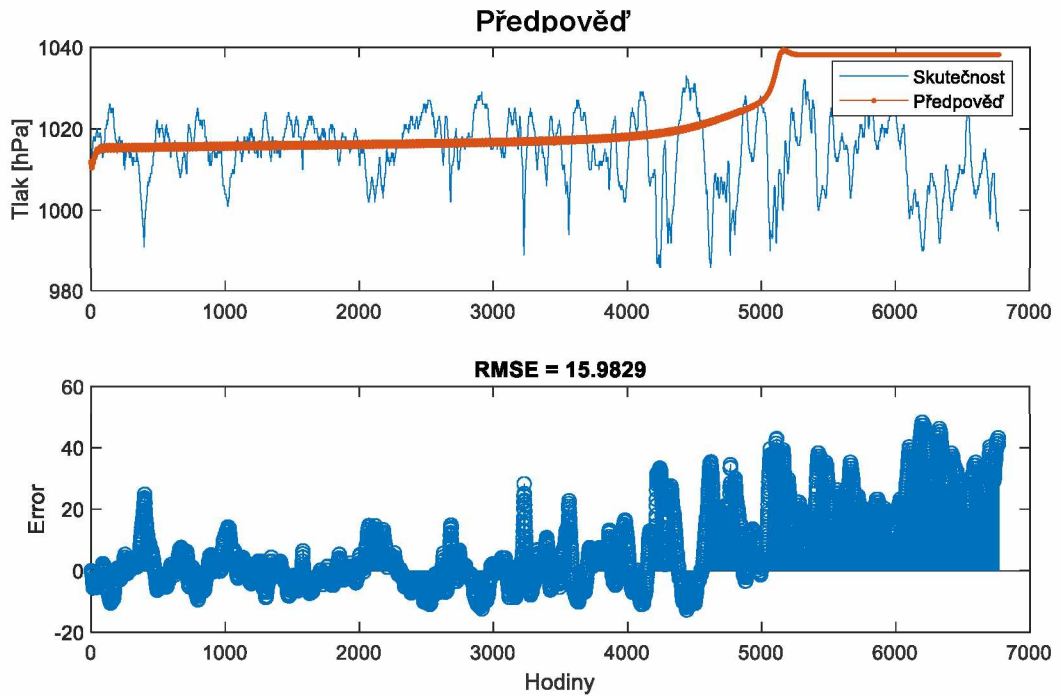


Obrázek 23 - Kombinovaná předpověď teploty a atmosférického tlaku – teplota

Při pohledu na dlouhodobou předpověď teploty zjišťujeme, že oproti předpovědi teploty samotné se přesnost opět zvýšila, nicméně nikoliv o tolik jako v případě kombinace teplota + vlhkost. Oproti tomu krátkodobá předpověď teploty níže ukazuje, že došlo ke zmenšení chyby na téměř celou 1 a křivka svým tvarem a hodnotami téměř kopíruje skutečné hodnoty.



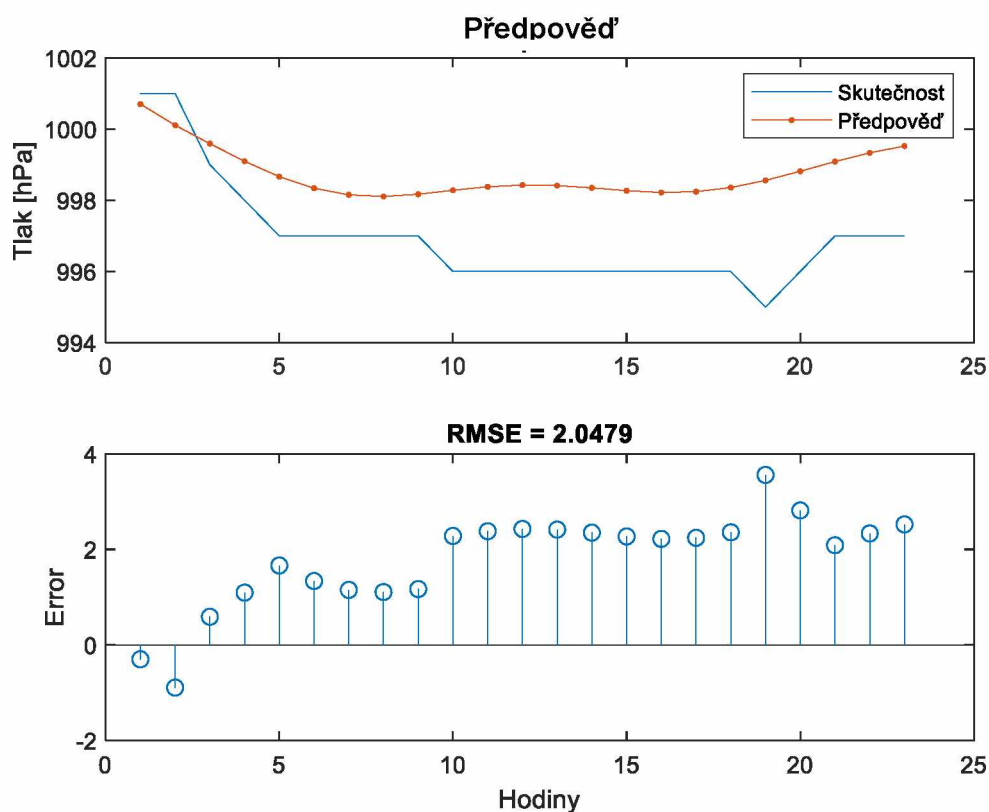
Obrázek 24 - kombinovaná předpověď teploty a atmosférického tlaku – teplota na 24 h



Obrázek 25 - Kombinovaná předpověď teploty a atmosférického tlaku – tlak

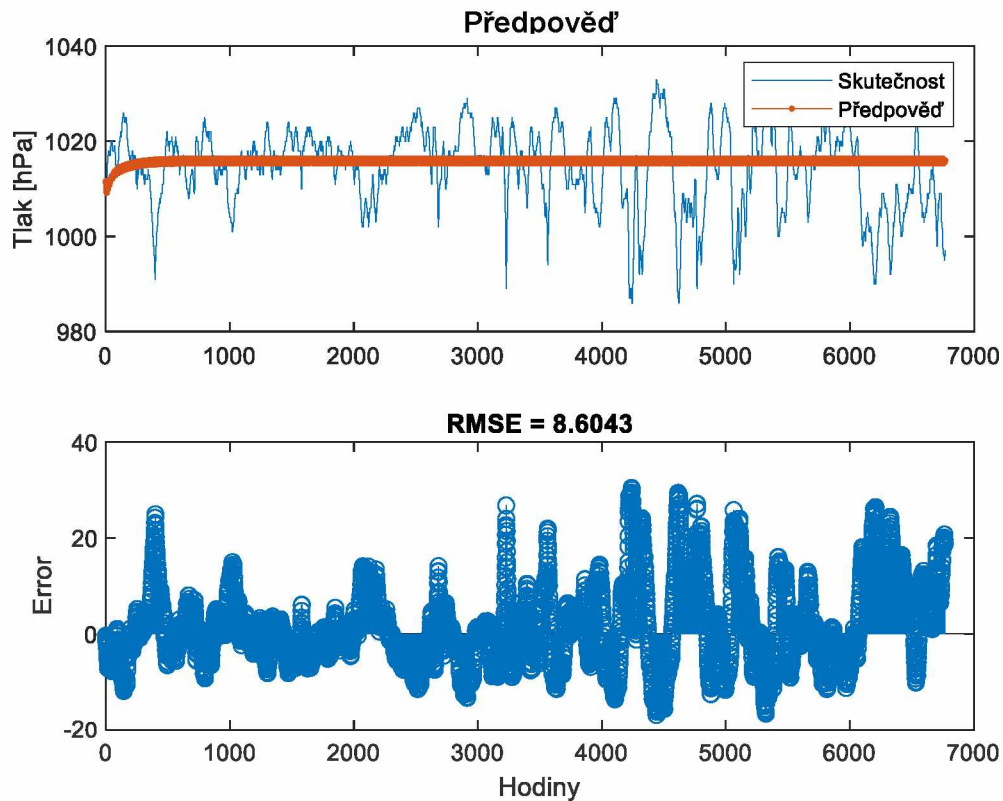
Při prozkoumání dlouhodobé předpovědi tlaku, docházíme k závěru, že i zde došlo ke zlepšení přesnosti oproti samotné předpovědi tlaku. Nicméně již z letmého pohledu je jasné, že tato předpověď stále není použitelná pro praktické účely.

Při porovnání hodnot krátkodobé předpovědi tlaku, zjišťujeme, že i zde došlo ke zmenšení chyby na přibližně polovinu, přičemž křivka předpovědi již svým tvarem odpovídá přibližnému tvaru křivky skutečných hodnot.



Obrázek 26 - Kombinovaná předpověď teploty a atmosférického tlaku – tlak na 24 h

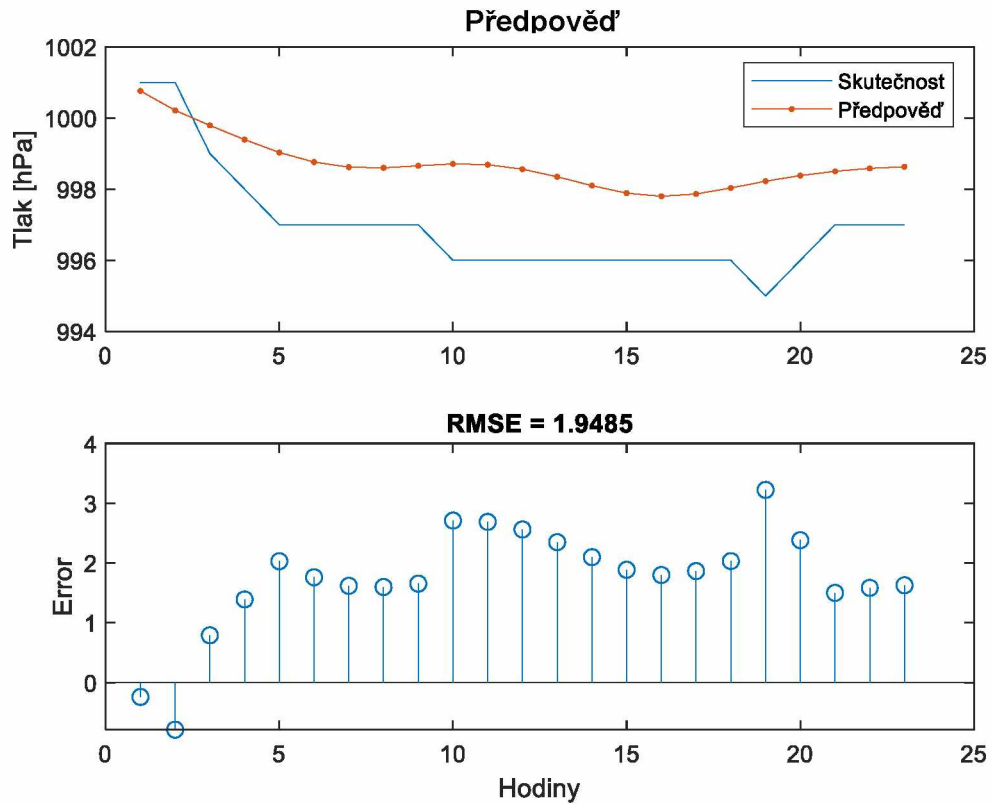
4.7 Kombinovaná předpověď vlhkosti a atmosférického tlaku



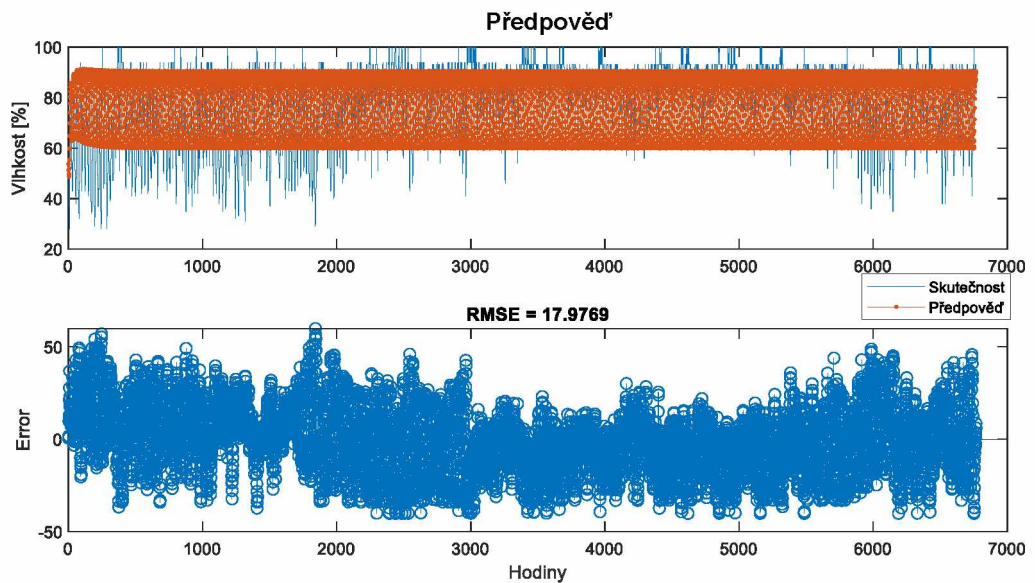
Obrázek 27 - Kombinovaná předpověď vlhkosti a atmosférického tlaku – tlak

Z dat více zjišťujeme, že dlouhodobá předpověď tlaku je v tomto případě se zatím nejmenší chybou, nicméně dle rozptření křivky předpovědi oproti křivce se skutečnými hodnotami zjišťujeme, že výsledná data stále nejsou prakticky použitelná. U krátkodobé předpovědi jsme se již dostali k chybě pod 2 celé a křivka svým tvarem kopíruje tvar křivky se skutečnými hodnotami.

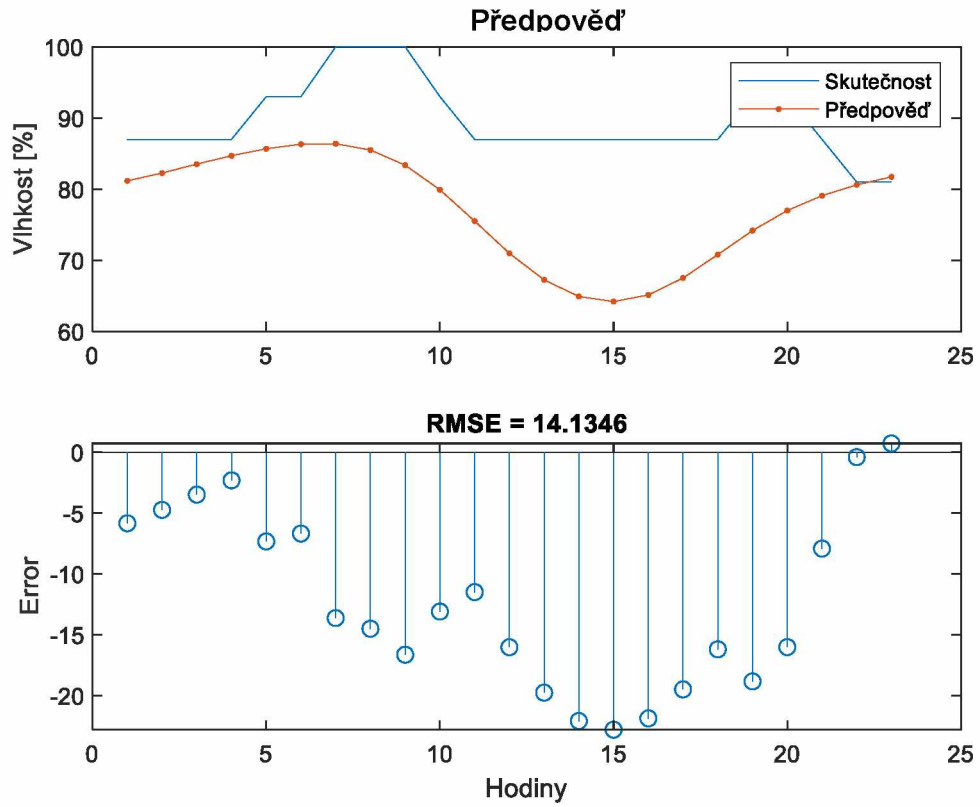
Při pohledu na graf dlouhodobé předpovědi vlhkosti opětovně dostáváme nejnižší zatím dosaženou chybu



Obrázek 28 - Kombinovaná předpověď vlhkosti a atmosférického tlaku – tlak na 24 h

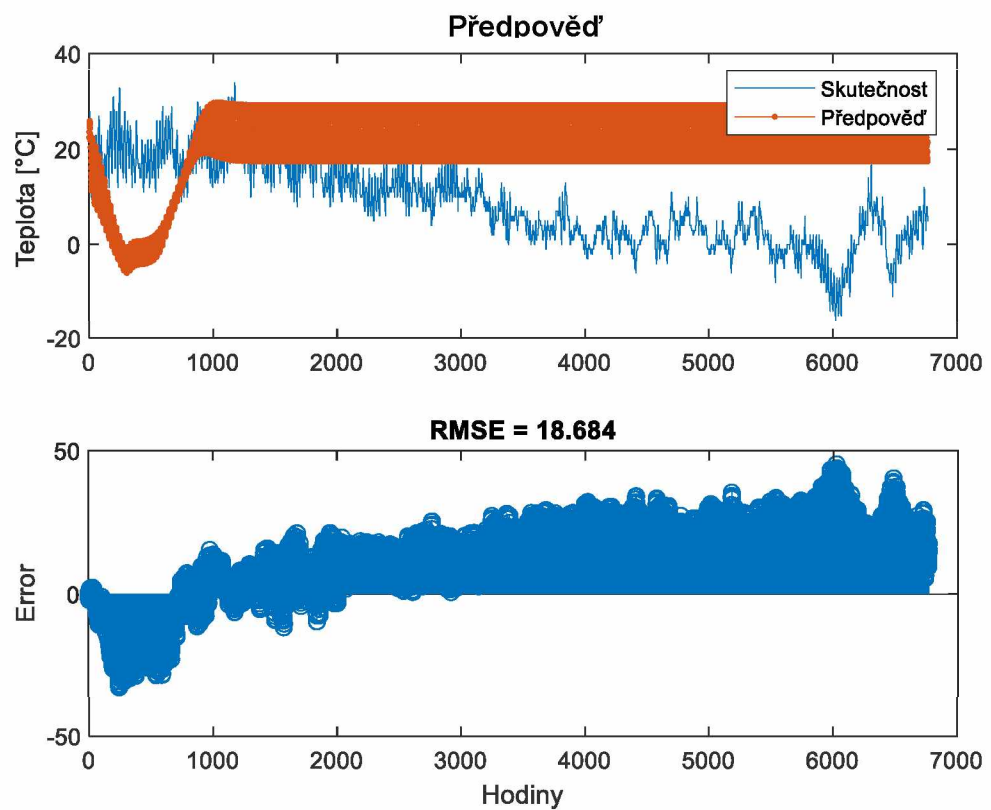


Obrázek 29 - Kombinovaná předpověď atmosférického tlaku a vlhkosti – vlhkost



Obrázek 30 - Kombinovaná předpověď atmosférického tlaku a vlhkosti – vlhkost na 24 h

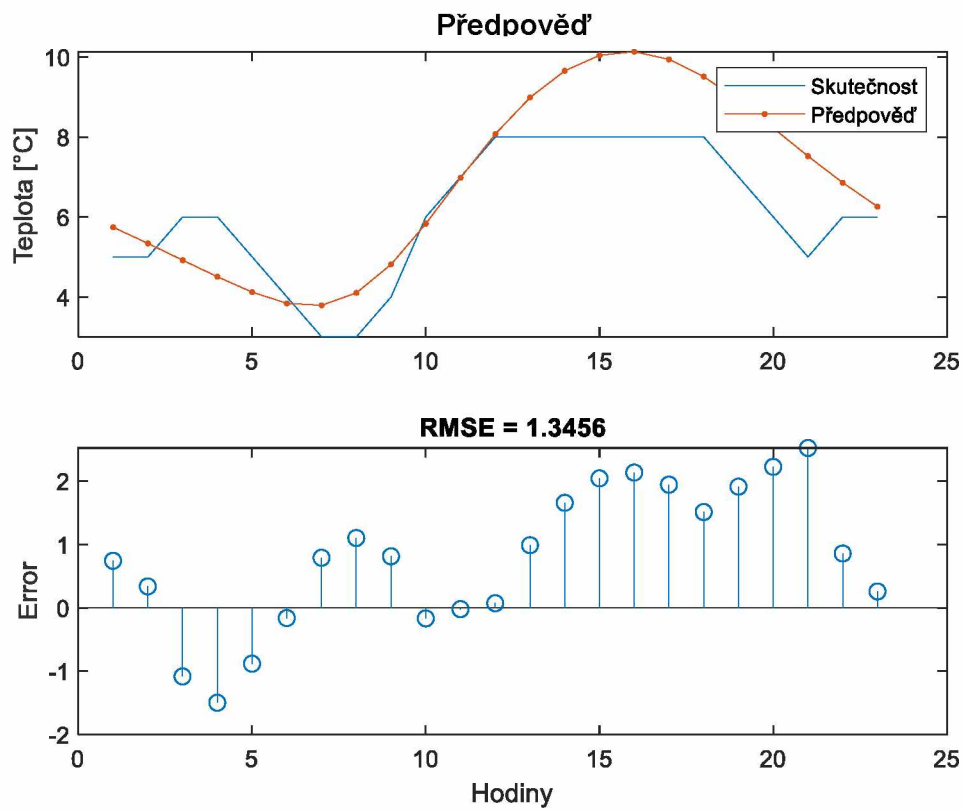
4.8 Kombinované předpověď teploty, vlhkosti a tlaku



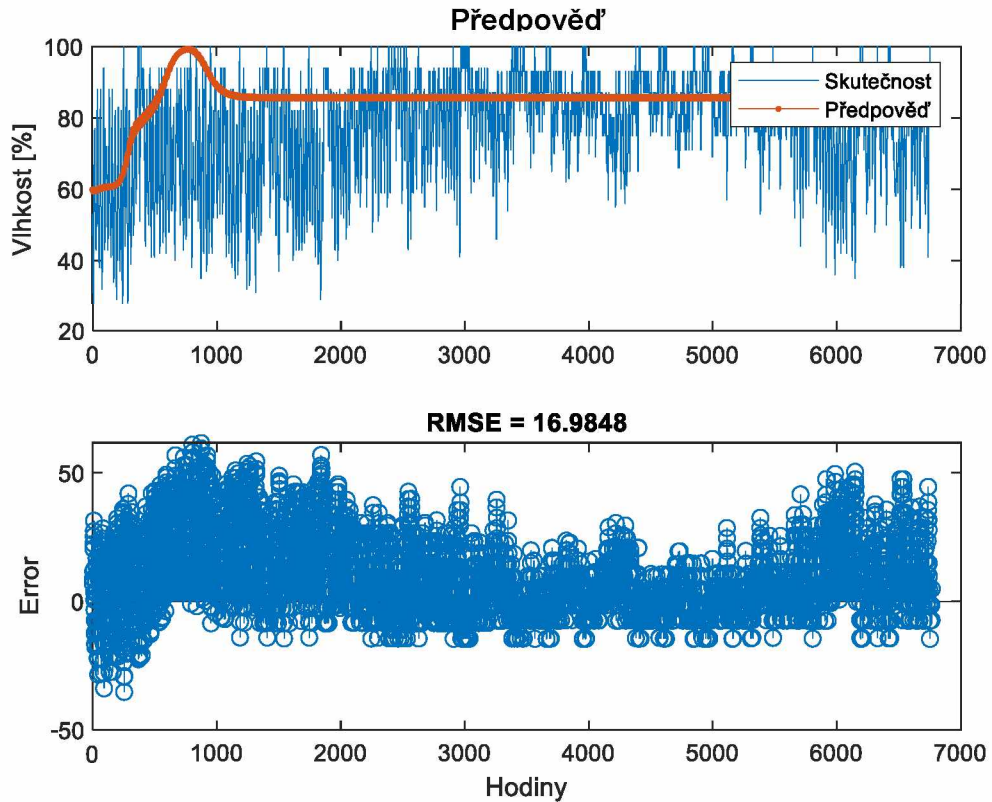
Obrázek 31 - Kombinovaná předpověď teploty, vlhkosti a tlaku – teplota

Dlouhodobá předpověď teploty se v této kombinaci vyznačuje větší chybou než při samotné predikci pouze na základě teploty.

Naproti tomu krátkodobá předpověď má jen o přibližně tři desetiny druhou nejlepší přesnost. V několika bodech se již křivky protínají a jejich tvar přibližně odpovídá.



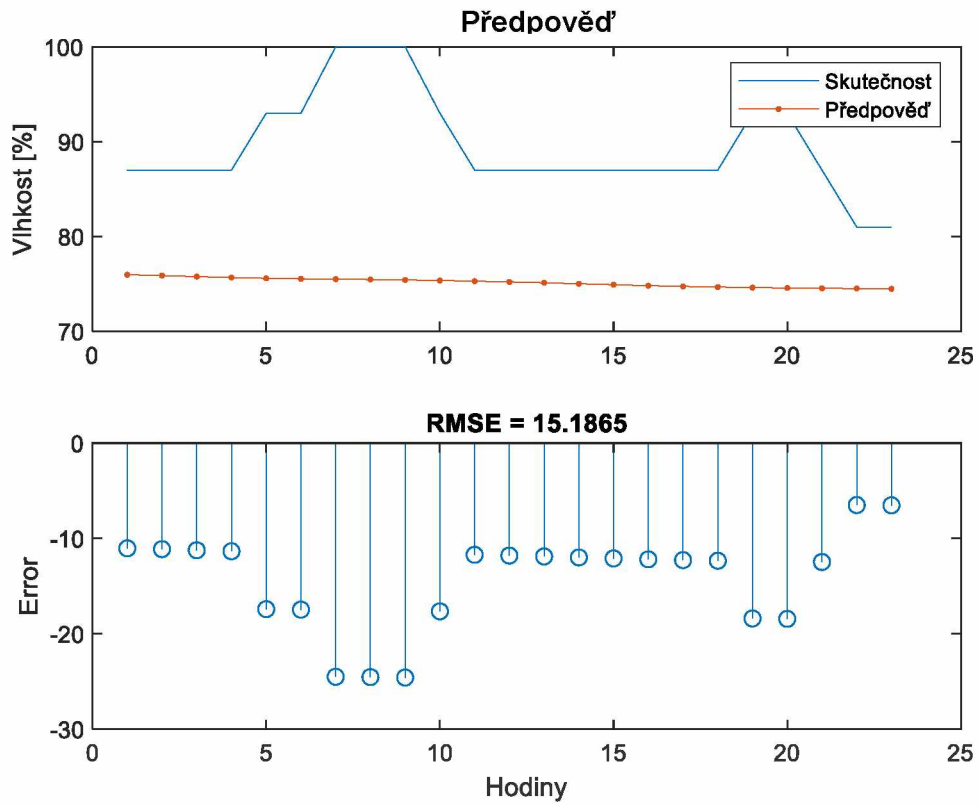
Obrázek 32 - Kombinovaná předpověď teploty, vlhkosti a tlaku – teplota na 24 h



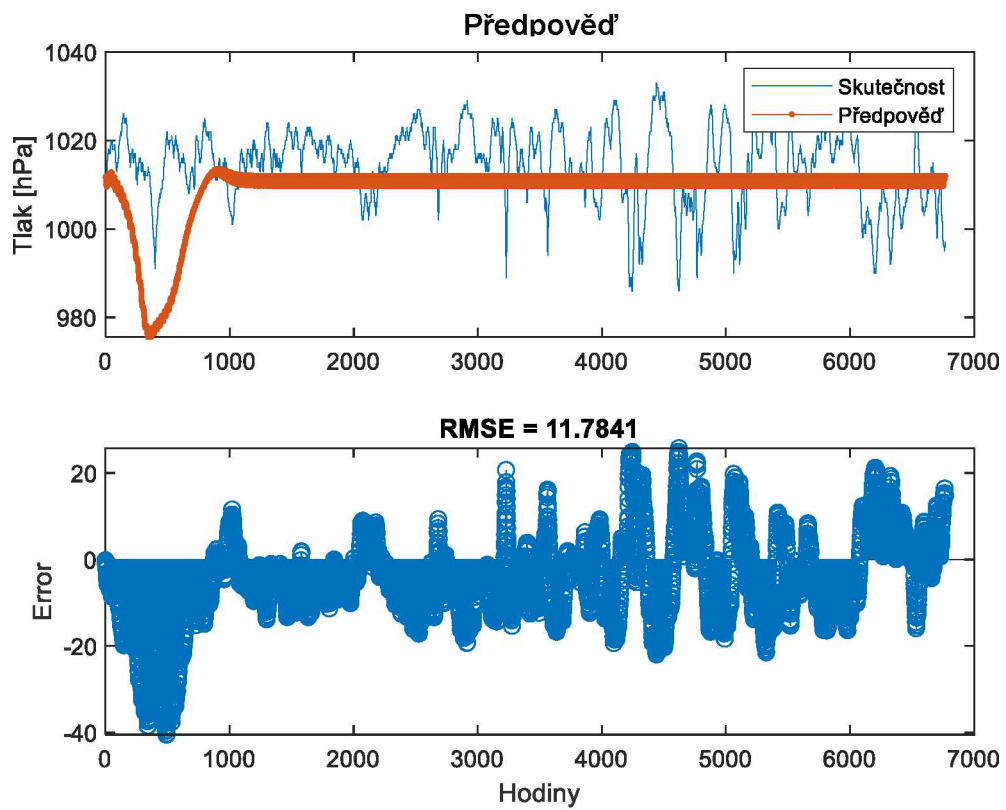
Obrázek 33 - Kombinovaná předpověď teploty, vlhkosti a tlaku – vlhkost

Dlouhodobá předpověď vlhkosti v této variantě též dosáhla své druhé největší chyby, a to v porovnání se všemi předchozími pokusy.

Krátkodobá předpověď vlhkosti též dosáhla své druhé nejhorší velikosti, přičemž křivky si svým tvarem neodpovídají.

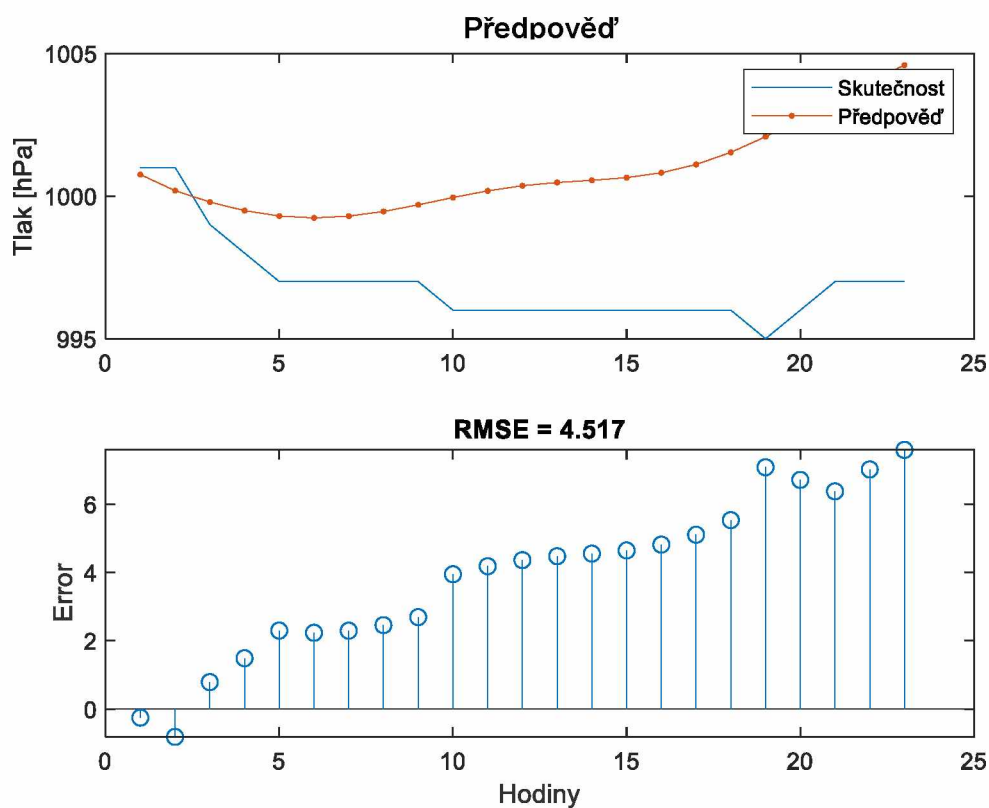


Obrázek 34 - Kombinovaná předpověď teploty, vlhkosti a tlaku – vlhkost na 24 h



Obrázek 35 - Kombinovaná předpověď teploty, vlhkosti a tlaku – tlak

Dlouhodobá předpověď tlaku dosáhla na druhou nejlepší velikost své chyby. Přičemž krátkodobá předpověď má chybu nejhorší ze všech prováděných experimentů.

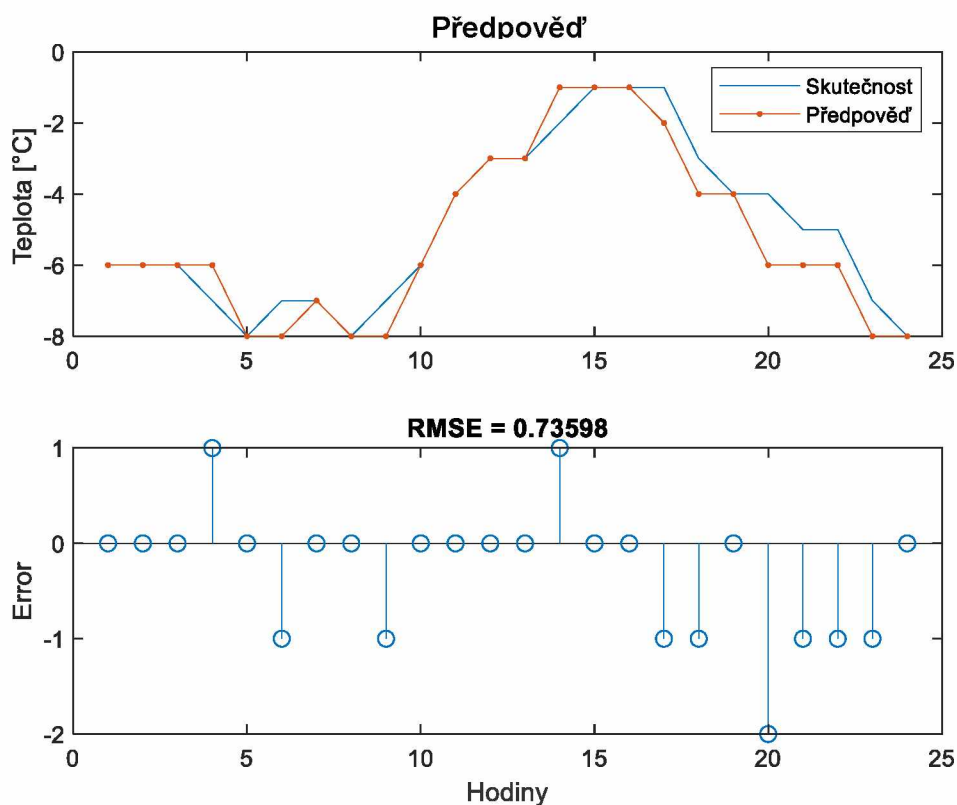


Obrázek 36 - Kombinovaná předpověď teploty, vlhkosti a tlaku – tlak na 24 h

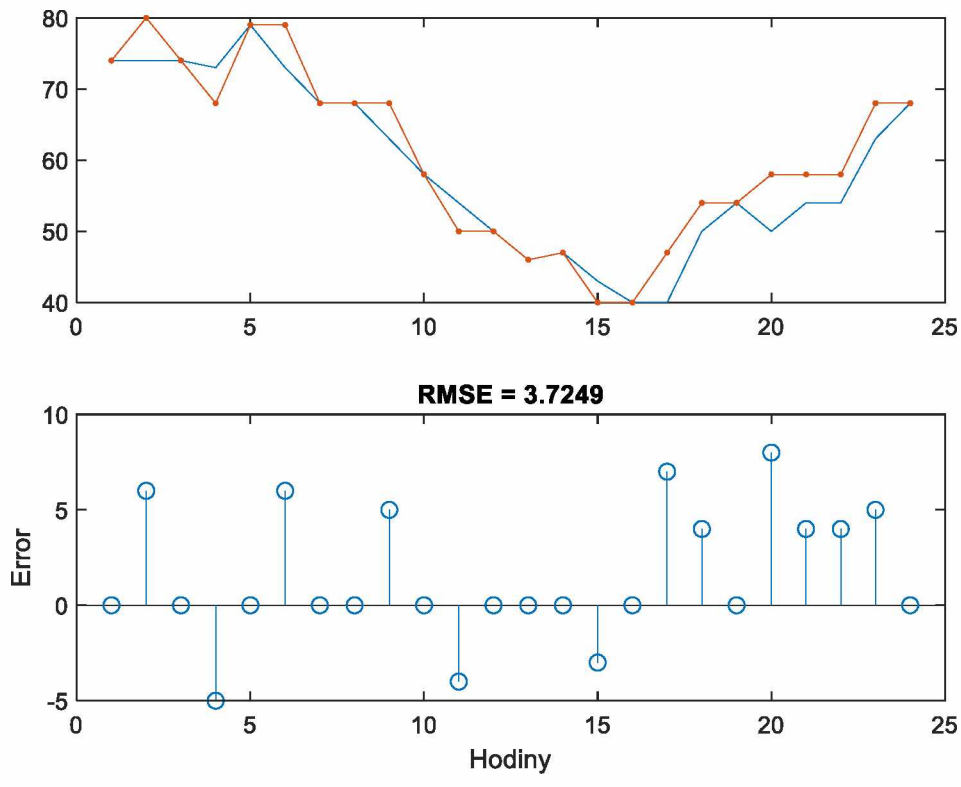
4.9 Porovnání se skutečnou předpovědí

Níže lze nalézt porovnání naměřených hodnot se zveřejněnou předpovědí na serveru www.yr.no. (Norwegian Meteorological Institute, 2020)

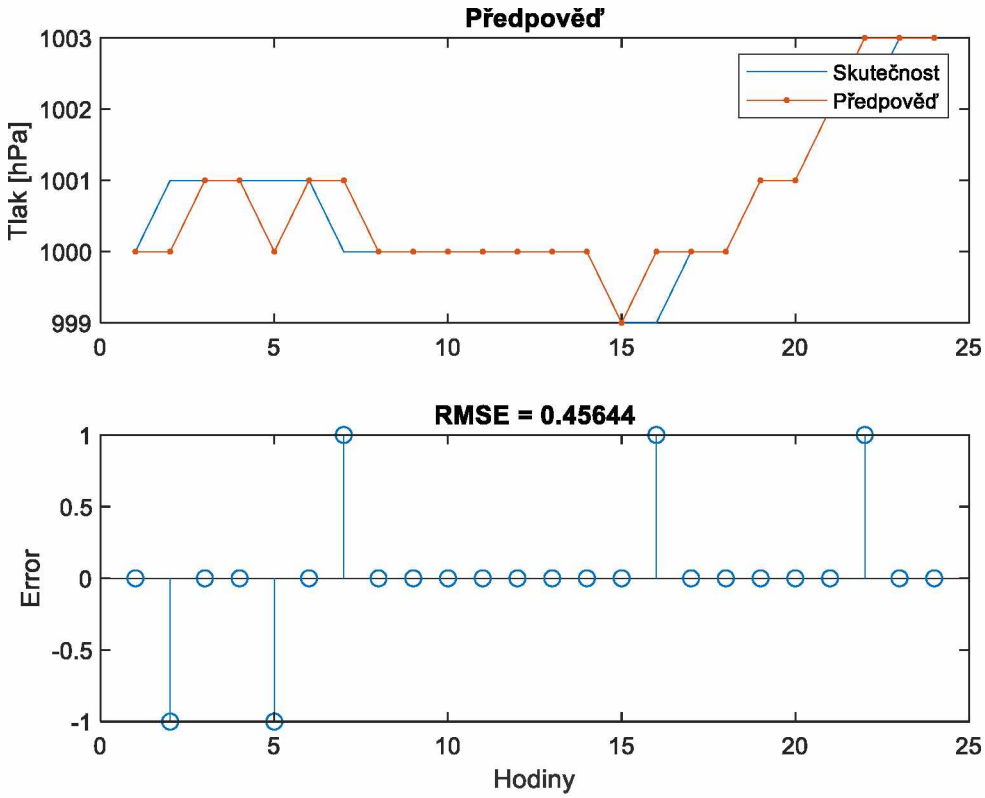
Předpověď z profesionálního zdroje ve všech ohledech poráží všechny zde provedené experimenty. Je tedy vhodné použít jakožto zdroj předpovědi tuto variantu, za předpokladu, že je pro danou lokaci k dispozici. Pokrytí jednotlivých vesnic i měst by v případě citovaného yr.no neměl být problém, jelikož zahrnují do svých předpovědí i vesnice o velikosti 230 obyvatel, což je právě případ autorova bydliště.



Obrázek 37 - Předpověď vs realita – teplota



Obrázek 38 - Předpověď vs realita – vlhkost



Obrázek 39 - Předpověď vs realita – tlak

5 Závěr

Stanovených cílů se podařilo dosáhnout částečně. Za stanovených podmínek je predikce nepoužitelná pro dlouhodobou předpověď kteréhokoliv ze sledovaných parametrů.

Naproti tomu v rámci krátkodobé předpovědi se ukázalo, že je v praxi použitelný model, kde se předpovídá krátkodobá budoucnost teploty pouze na základě předchozí teploty.

Kombinované předpovědi přinesly smíšené výsledky, kde docházelo ke zlepšení méně důležitých parametrů (vlhkost) a zároveň vznikala větší chyba u více sledovaného parametru (teplota).

Kombinace všech vstupních dat se ukázala ve výsledku jakožto neuspokojující, vzhledem k tomu, že jednotlivé dřívější modely nabízeli přesnější výsledky. Zde je však možné, že s nárůstem dat by se tato nevýhoda odstranila a modely by byli alespoň rovnocenné.

Porovnání přesnosti předpovědi s profesionální službou yr.no dopadlo ve všech směrech ku prospěchu yr.no. Toto však může být dáno skutečností, že se k předpovědi přistupuje z rozdílných stran (lokální vs globální předpověď).

Na základní otázku, zda je možné předpovídat počasí pouze na základě lokálních dat, je odpověď ano, ale je zapotřebí brát na zřetel, že takováto předpověď dosahuje i řádově horší přesnosti než předpověď klasická.

Z algoritmů řešitelů umělých neuronových sítí se ukázal nejvhodnější pro zadanou úlohu algoritmus *adam*. Ostatní dosáhli podobných výsledků v delším čase či s větší měřenou chybou.

Autor pozitivně kvituje možnost spouštění trénování v neuronových sítích za pomoci grafických karet využívaných programem Matlab, což úspěšně zkrátilo dobu trénování z řádů až desítek hodin na pouhé desítky minut.

Literatura

Alpaydm, Ethem. 2014. *Introduction to Machine Learning*. Massachusetts Institute of Technology : The MIT Press Cambridge, Massachusetts , 2014. ISBN 978-0-262-02818-9 .

Brilliant.org. 2020. Backpropagation. *Brilliant*. [Online] 2020. <https://brilliant.org/wiki/backpropagation/>.

Claude Sammut, Geoffrey I. Webb (Eds.). 2011. *Encyclopedia of Machine Learning*. New York : Springer Science+Business Media, 2011. ISBN 978-0-387-34558-1.

Čermák, Ondřej. 2020. Meteostanice. *Meteostanice.com*. [Online] 2020. <https://www.meteostanice.com/>.

Český hydrometeorologický ústav. 2018. Meteorologická terminologie. *Portál ČHMÚ*. [Online] 2018. [Citace: 01. 02 2018.] <http://portal.chmi.cz/informace-pro-vas/meteorologicka-terminologie/zakladni-pojmy>.

Doc. Dr. Vladimír Homola, Ph.D. 2020. Lineární regrese. *Lineární regrese*. [Online] 2020. <http://homel.vsb.cz/~hom50/SLBSTATS/LR2/GS04.HTM>.

Duchon, Claude a Hale, Robert. 2012. *Time Series Analysis in Meteorology and Climatology An Introduction*. místo neznámé : John Wiley & Sons, 2012. ISBN 978-0-470-97199-4.

Havlík, Jindřich. 2014. *Hodnocení kvality odhadu stavu stochastických systémů - diplomová práce*. [Pdf] Plzeň : Západočeská univerzita v Plzni, Fakulta aplikovaných věd, 2014.

Ing. Petr Doležel, Ph.D. 2016. *Úvod do umělých neuronových sítí pro studenty technických vysokých škol*. Pardubice : Univerzita Pardubice, 2016. ISBN 978-80-7560-022-6.

InMeteo. 2017. Encyklopedie počasí. *In-pocasi*. [Online] 2017. [Citace: 01. 02 2018.] <https://www.in-pocasi.cz/encyklopedie-pocasi/>.

Los Alamos National Security, LLC (LANS) . 2020. MPAS Overview. *MPAS Model for Prediction Across Scales*. [Online] Los Alamos National Security, LLC (LANS) , 2020. <https://mpas-dev.github.io/>.

MathWorks. 2020. trainingOptions. *Help Center*. [Online] MathWorks, 2020. https://www.mathworks.com/help/deeplearning/ref/trainingoptions.html#bu80qkw-3_head.

Matoušek, Václav. Strojové učení. *KIV ZČU.* [Online] https://www.kiv.zcu.cz/studies/predmety/uzi/Folie_ZS/Stroj_uceni.pdf.

Moravec, Zdeněk. 2020. Lineární regrese. *Web o chemii, elektronice a programování.* [Online] 2020. <http://z-moravec.net/chemie/zaklady-chemie/linearni-regrese/>.

National Centers for Environmental Information. 2020. Global Forecast System (GFS). *National Centers for Environmental Information (NCEI).* [Online] National Centers for Environmental Information, 2020. <https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forecast-system-gfs>.

Norwegian Meteorological Institute. 2020. Yr. *Hourly forecast for Prague (Czech Republic).* [Online] Norwegian Meteorological Institute and the Norwegian Broadcasting Corporation, 2020. https://www.yr.no/place/Czech_Republic/Prague/Prague/hour_by_hour.html.

Time and Date AS. 2020. Past Weather in Prague, Czechia. *timeanddate.com.* [Online] Time and Date AS, 2020. <https://www.timeanddate.com/weather/czech-republic/prague/historic>.

UCAR. 2020. WEATHER RESEARCH AND FORECASTING MODEL. *National Center for Atmospheric Research.* [Online] UCAR , 2020. <https://www.mmm.ucar.edu/weather-research-and-forecasting-model>.

UCAR. 2020. WEATHER RESEARCH AND FORECASTING MODEL (WRF). *Research applications laboratory.* [Online] NCAR, 2020. <https://ral.ucar.edu/solutions/products/weather-research-and-forecasting-model-wrf>.

Seznam příloh

- Příloha A – Seznam použitého SW

Příloha A – Seznam použitého SW

- Oracle Java 8
- Apache NetBeand 8.2
- IntelliJ IDEA 2020.1
- MathWorks Matlab R2020a
- Debian 10
- Cron
- Bash
- Microsoft Excel 2016