

Návrh nebo využití vhodného modelu pro předpověď počasí na základě měření

Bc. Jan Fux

Diplomová práce
2016



Univerzita Tomáše Bati ve Zlíně
Fakulta aplikované informatiky

ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Bc. Jan Fux**

Osobní číslo: **A14459**

Studijní program: **N3902 Inženýrská informatika**

Studijní obor: **Informační technologie**

Forma studia: **kombinovaná**

Téma práce: **Návrh nebo využití vhodného modelu pro předpověď počasí na základě měření**

Téma anglicky: **The Design or Use of an Appropriate Model for Weather Forecasts Based on Measurements**

Zásady pro vypracování:

1. Seznamte se s možnostmi soft computingových nástrojů pro predikování.
2. Zpracujte vhodným způsobem data firmy Navmatix s.r.o. k budoucímu modelování předpovědi počasí.
3. Využijte neuronové sítě a navrhnete vhodný typ, strukturu a natrénujte na zvoleném datasetu.
4. Provedte validaci a testování.
5. Provedte hodnocení míry úspěšnosti předpovědi počasí.

Rozsah diplomové práce:

Rozsah příloh:

Forma zpracování diplomové práce: **tištěná/elektronická**

Seznam odborné literatury:

1. MAQSOOD, I., KHAN, M. R., ABRAHAM, A.: An ensemble of neural networks for weather forecasting, *Neural Computing Application*, roč. 2004 čís. 13. s. 112122. Springer Verlag. 2004. DOI 10.1007/s00521-004-0413-4.
2. BISHOP, Christopher M. *Pattern recognition and machine learning*. New York: Springer, c2006, xx, 738 s. ISBN 0-387-31073-8.
3. MITCHELL, Tom M. *Machine learning*. Boston: WCB/McGraw-Hill, c1997, xvii, 414 s. ISBN 0-07-042807-7.
4. BÍLA J.: *Umělá inteligence a neuronové sítě v aplikacích*, ČVUT, 1996, ISBN 80-01-01275-1.
5. ZELINKA, Ivan. *Umělá inteligence I. Volume 1*. Zlín: Vutium, Brno, 1998, 126 p. ISBN 80-214-1163-5.
6. BOSE N.K., LIANG P.: *Neural Network Fundamentals with Graphs, Algorithms, and Applications*, McGraw-Hill Series in Electrical and Computer Engineering, 1996, ISBN 0-07-006618-3.
7. CULCLASURE, Andrew: *Using Neural Networks to Provide Local Weather Forecasts (2013)*. Electronic Theses & Dissertations. Paper 32. College of Graduate Studies. Georgia Southern University.
8. ŠÍMA J., NERUDA R.: *Teoretické otázky neuronových sítí*. Vyd. 1. Praha, 1996: Matfyzpress, 390 s., ISBN 80-85863-18-9

Vedoucí diplomové práce: **doc. Ing. Zuzana Komínková Oplatková, Ph.D.**
Ústav informatiky a umělé inteligence

Konzultant: **Jiří Hanslian**
Navmatix s.r.o.

Datum zadání diplomové práce: **5. února 2016**

Termín odevzdání diplomové práce: **20. května 2016**

Ve Zlíně dne 5. února 2016



doc. Mgr. Milan Adámek, Ph.D.
děkan



doc. Mgr. Roman Jašek, Ph.D.
ředitel ústavu


Prohlašuji, že

- beru na vědomí, že odevzdáním diplomové/bakalářské práce souhlasím se zveřejněním své práce podle zákona č. 111/1998 Sb. o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách), ve znění pozdějších právních předpisů, bez ohledu na výsledek obhajoby;
- beru na vědomí, že diplomová/bakalářská práce bude uložena v elektronické podobě v univerzitním informačním systému dostupná k prezenčnímu nahlédnutí, že jeden výtisk diplomové/bakalářské práce bude uložen v příruční knihovně Fakulty aplikované informatiky Univerzity Tomáše Bati ve Zlíně a jeden výtisk bude uložen u vedoucího práce;
- byl/a jsem seznámen/a s tím, že na moji diplomovou/bakalářskou práci se plně vztahuje zákon č. 121/2000 Sb. o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon) ve znění pozdějších právních předpisů, zejm. § 35 odst. 3;
- beru na vědomí, že podle § 60 odst. 1 autorského zákona má UTB ve Zlíně právo na uzavření licenční smlouvy o užití školního díla v rozsahu § 12 odst. 4 autorského zákona;
- beru na vědomí, že podle § 60 odst. 2 a 3 autorského zákona mohu užít své dílo – diplomovou/bakalářskou práci nebo poskytnout licenci k jejímu využití jen připouští-li tak licenční smlouva uzavřená mezi mnou a Univerzitou Tomáše Bati ve Zlíně s tím, že vyrovnání případného přiměřeného příspěvku na úhradu nákladů, které byly Univerzitou Tomáše Bati ve Zlíně na vytvoření díla vynaloženy (až do jejich skutečné výše) bude rovněž předmětem této licenční smlouvy;
- beru na vědomí, že pokud bylo k vypracování diplomové/bakalářské práce využito softwaru poskytnutého Univerzitou Tomáše Bati ve Zlíně nebo jinými subjekty pouze ke studijním a výzkumným účelům (tedy pouze k nekomerčnímu využití), nelze výsledky diplomové/bakalářské práce využít ke komerčním účelům;
- beru na vědomí, že pokud je výstupem diplomové/bakalářské práce jakýkoliv softwarový produkt, považují se za součást práce rovněž i zdrojové kódy, popř. soubory, ze kterých se projekt skládá. Neodevzdání této součásti může být důvodem k neobhájení práce.

Prohlašuji,

- že jsem na diplomové/bakalářské práci pracoval samostatně a použitou literaturu jsem citoval. V případě publikace výsledků budu uveden jako spoluautor.
- že odevzdaná verze diplomové práce a verze elektronická nahraná do IS/STAG jsou totožné.

Ve Zlíně, dne 13. 5. 2016


.....
podpis diplomanta

ABSTRAKT

Práce se zabývá krátkodobou předpovědí meteorologických veličin pomocí metod umělé inteligence, konkrétně za pomoci neuronové sítě. Zaměřuje se na různě dlouhé intervaly předpovědi a předkládá několik přístupů a metod, jejichž cílem je dosáhnout co nejlepší možné kvality předpovědi a minimální chyby. Mezi zkoumané metody patří hledání optimální struktury sítě, normalizace dat, využití dat z více meteorologických stanic pro učení sítě a učení daty z dlouhého časového období. V závěru pak práce shrnuje získané výsledky a předkládá doporučení pro optimální přístup k předpovědi zkoumaných meteorologických veličin.

Klíčová slova: předpověď počasí, meteorologie, neuronová síť, neuron, normalizace, Levenberg-Marquardt algoritmus, Back - Propagation

ABSTRACT

Thesis is focused on short-term prediction of the meteorological quantities using methods of the artificial intelligence, specifically neural networks. It is focused on prediction intervals with different sizes and several approaches to gain best possible quality of the prediction and the smallest error is presented. The research in this thesis contains experiments with the network size and structure, data normalization, usage of more data inputs from second meteorological station and network training by the long-term data. In the end, there is a summary of results and recommendations of best approaches to meteorological quantities prediction.

Keywords: weather forecast, meteorology, neural network, neuron, normalization, Levenberg-Marquardt algorithm, Back - Propagation

Tímto chci poděkovat své přítelkyni Petře, za její podporu a toleranci během psaní této práce. Dodávala mi duševní sílu a klid, který jsem při psaní velmi ocenil.

Dále bych rád poděkoval mé vedoucí doc. Ing. Zuzaně Komínkové Oplatkové, Ph.D. za její odpovědi na mé zdlouhavé dotazy během konzultací, její teoretické znalosti a informace, ze kterých jsem během psaní čerpal, mému konzultantovi Ing. Jirímu Hanslianovi a mému kamarádovi Tomášovi, s jejichž pomocí jsem získal meteorologická data, která jsem v práci využil.

Prohlašuji, že odevzdaná verze bakalářské/diplomové práce a verze elektronická nahraná do IS/STAG jsou totožné.

OBSAH

ÚVOD	9
I TEORETICKÁ ČÁST.....	10
1 PŘEDPOVĚĎ POČASÍ.....	11
2 NEURONOVÉ SÍTĚ.....	12
2.1 BIOLOGICKÝ MODEL NEURONU [2].....	12
2.2 MATEMATICKÝ MODEL NEURONU [2].....	13
2.3 NEURONOVÁ SÍŤ [7]	14
2.3.1 Vícevrstvá síť [9].....	16
2.3.2 Back-propagation algoritmus [9].....	16
2.4 VÝBĚR VHODNÉ MODIFIKACE BACK-PROPAGATION ALGORITMU	17
2.5 LEVENBERG-MARQUARDT ALGORITMUS [10].....	17
3 VYUŽITÍ GLOBÁLNÍHO DRUŽICOVÉHO POLOHOVÉHO SYSTÉMU	19
4 POPIS POUŽITÉHO DATASETU	20
4.1 METEOROLOGICKÁ DATA OD NAVMATIX S.R.O.	20
4.2 METEOROLOGICKÁ DATA OD HVĚZDÁRNY A PLANETÁRIA BRNO, P. O.	21
4.3 GNSS DATA OD NAVMATIX S.R.O.	21
4.4 DATA PRO METODU PPP OD ORGANIZACE THE INTERNATIONAL GNSS SERVICE (IGS)	22
II PRAKTICKÁ ČÁST	23
5 PŘEDPOVĚĎ NA ZÁKLADĚ METEOROLOGICKÝCH DAT.....	24
5.1 PŘEDPOVĚĎ VELIČIN 10 MINUT DOPŘEDU	24
5.1.1 Aplikace 10minutové předpovědi na teplotu	24
5.1.2 Aplikace 10minutové předpovědi na tlak	25
5.1.3 Aplikace 10minutové předpovědi na vlhkost.....	27
5.2 PŘEDPOVĚĎ VELIČIN 1 HODINU DOPŘEDU	27
5.2.1 Předpověď teploty po 1 hodině	28
5.2.2 Předpověď tlaku po 1 hodině	29
5.2.3 Předpověď vlhkosti po 1 hodině	31
5.3 PŘEDPOVĚĎ VELIČIN 1 DEN DOPŘEDU.....	31
5.3.1 Předpověď teploty za 1 den.....	32
5.3.2 Předpověď vlhkosti za 1 den.....	33
5.3.3 Předpověď tlaku za 1 den.....	34
6 METODY OPTIMALIZACE PŘEDPOVĚDI DLE METEOROLOGICKÝCH DAT	36
6.1 VLIV VELIKOSTI SÍTĚ NA KVALITU PŘEDPOVĚDI	36
6.1.1 Vliv velikosti sítě na kvalitu předpovědi teploty	36
6.1.2 Vliv velikosti sítě na kvalitu předpovědi tlaku	37
6.1.3 Vliv velikosti sítě na kvalitu předpovědi vlhkosti.....	39

6.2	VLIV NORMALIZACE DAT NA KVALITU PŘEDPOVĚDI.....	41
6.2.1	Vliv normalizace dat na kvalitu předpovědi teploty.....	42
6.2.2	Vliv normalizace dat na kvalitu předpovědi tlaku.....	44
6.2.3	Vliv normalizace dat na kvalitu předpovědi vlhkosti.....	46
6.3	PŘEDPOVĚĎ S VYUŽITÍM DAT Z DRUHÉ METEOROLOGICKÉ STANICE.....	47
6.3.1	Předpověď teploty s využitím dat z druhé meteorologické stanice.....	48
6.3.2	Předpověď tlaku s využitím dat z druhé meteorologické stanice.....	49
6.4	TRÉNOVÁNÍ DATY Z DELŠÍHO ČASOVÉHO HORIZONTU.....	50
6.4.1	Předpověď teploty při použití dat z delšího období.....	51
6.4.2	Předpověď vlhkosti při použití dat z delšího období.....	52
6.4.3	Předpověď tlaku při použití dat z delšího období.....	53
6.5	VYUŽITÍ GNSS DAT JAKO DALŠÍ VSTUP DO SÍTĚ.....	54
6.5.1	Předpověď teploty při použití GNSS dat.....	54
6.5.2	Předpověď vlhkosti při použití GNSS dat.....	55
6.5.3	Předpověď tlaku při použití GNSS dat.....	56
6.6	PŘEDPOVĚĎ VŠECH VELIČIN POMOCÍ 1 SÍTĚ.....	57
7	POPIS POUŽITÉHO VÝPOČETNÍHO SYSTÉMU.....	60
8	MOŽNOSTI BUDOUCÍHO ROZŠÍŘENÍ PRÁCE.....	61
	ZÁVĚR.....	62
	SEZNAM POUŽITÉ LITERATURY.....	64
	SEZNAM POUŽITÝCH SYMBOLŮ A ZKRATEK.....	66
	SEZNAM OBRÁZKŮ.....	67
	SEZNAM TABULEK.....	69
	SEZNAM PŘÍLOH.....	70

ÚVOD

Diplomová práce se zabývá tématem predikce počasí. Byla zpracována na základě zadání od firmy Navmatix s.r.o., která v průběhu vypracování poskytovala konzultace a data potřebná pro experimenty. Předpověď počasí je v dnešní době důležitá výzkumná oblast, která lidem umožňuje předvídat různé chování atmosféry, přírodní jevy a také blížící se katastrofy jako například hurikány. Práce byla zaměřena spíše na předpovídání veličin, které nás u počasí nejvíce zajímají, než na slovní vyjádření typu „jasno“, „polojasno“, „zataženo“, neboť tyto a podobné pojmy je možné stanovit právě až na základě znalosti hodnot konkrétních meteorologických veličin. Jednalo se především o teplotu, vlhkost a atmosférický tlak, což byly také veličiny požadované firmou Navmatix s.r.o.

Pro předpověď bylo nejprve třeba zvolit vhodný model. Na základě publikovaných vědeckých článků a jejich doporučení byla vybrána vícevrstvá neuronová síť, která bude trénována jako síť s učitelem, a to nějakou vhodnou variantou algoritmu typu Back-Propagation. Dále bylo potřeba získat datový set minulých hodnot daných veličin, kterým bude síť učena a které bude zároveň předpovídat. Část těchto experimentálních dat (krátkodobá data) poskytla firma Navmatix s.r.o. z její meteostanice MetPak a dlouhodobější data organizace Hvězdárna a Planetárium Brno p.o. Data byla vhodně rozdělena, upravena a použita k trénování neuronové sítě a posléze i k otestování její úspěšnosti. Byly vyzkoušeny předpovědi na různě dlouhá časová období.

Další část práce se zabývá různými způsoby optimalizace sítě, popřípadě datasetu, aby byla předpověď co nejpřesnější. Popisuje například experimenty s různou architekturou (velikostí) sítě za účelem nalézt nejlepší variantu. Jsou prováděny experimenty s daty z více meteostanic, opět s cílem zpřesnění předpovědi oproti klasickému přístupu jen s jednou meteostanicí. Testuje se zde možnost normalizace dat, jako způsob předzpracování datového souboru před vlastním použitím v síti, ale i využití dlouhodobého datového souboru pro natrénování sítě. Nakonec je zařazen krátký experiment s využitím dat z Globálního družicového polohového systému pro zlepšení předpovědi.

Závěrečná kapitola práce je věnována pokusu o předpovězení všech veličin v rámci jedné sítě. Úspěšnosti jednotlivých experimentů jsou v závěru srovnány a je připojeno doporučení, jak lze z dostupných dat získat co možná nejkvalitnější předpověď.

I. TEORETICKÁ ČÁST

1 PŘEDPOVĚĎ POČASÍ

Vzhledem k rostoucí potřebě člověka znát budoucí vývoj počasí se tímto tématem zabývá velmi mnoho vědeckých článků a výzkumných skupin. K úspěšné předpovědi je zapotřebí mít správný model atmosféry, který, naplněn správnými daty, jako výsledek vrátí budoucí stavy zkoumaných atmosférických veličin. Rozlišujeme v zásadě dva druhy předpovědí, a to *dlouhodobou*, která je zaměřena na předpověď hodnot co nejdále do budoucnosti, což je ovšem vykoupeno hrubším rozlišením získaných hodnot veličin (větší perioda získaných vzorků), a předpověď *krátkodobou*, u které potřebujeme vědět co nejpřesněji vývoj veličin v následujících hodinách či minutách. Ta je ovšem vykoupena menším časovým intervalem, který lze s rozumnou přesností předpovídat. Limitem je jednak výkon výpočetních strojů, které mohou mít problém se zpracováním velkého množství vzorků (se zmenšující se periodou vzorků na stejném předpovídaném časovém úseku roste počet vzorků, které je potřeba dopočítat), ale hlavně fakt, že se zvětšující se délkou předpovídaného úseku klesá kvalita předpovědi. [16]

Pro sestavení předpovědního modelu lze využít dvou hlavních přístupů. První z nich je souhrnem rovnic popisujících jevy v atmosféře. Výstupní data jsou poté dopočítána vyřešením těchto rovnic. Jedná se o složité modely, pro jejichž sestavení je důležitá dokonalá znalost fyzikálních jevů probíhajících v atmosféře. Jedná se o nejstarší přístup, který je ovšem zatím stále nejvíce využíván. Druhou možností je sestavení modelu s pomocí umělé inteligence, například jako sítě prvků, které budeme postupně předkládat historická naměřená data a zároveň očekávané výstupy, které budeme porovnávat se skutečným výstupem ze sítě. Postupně tedy bude docházet k adaptaci takového modelu na požadované hodnoty. Pokud takový adaptovaný model poté naplníme daty, která jsou nejaktuálnější, získáme předpověď nových hodnot. Velkou výhodou tohoto přístupu je, že pro jeho použití není nutná znalost fyzikálních principů, které jsou v atmosféře uplatňovány a které vývoj počasí ovlivňují. V poslední době se touto cestou ubírají mnohé studie, které si dávají za cíl ověřit, zda je možné dosáhnout kvalitní předpovědi. Konkrétně je takovýto model realizován zpravidla tzn. neuronovou sítí. Vzhledem k dobrým výsledkům, které tyto studie často předkládají, byly využity i v této práci.

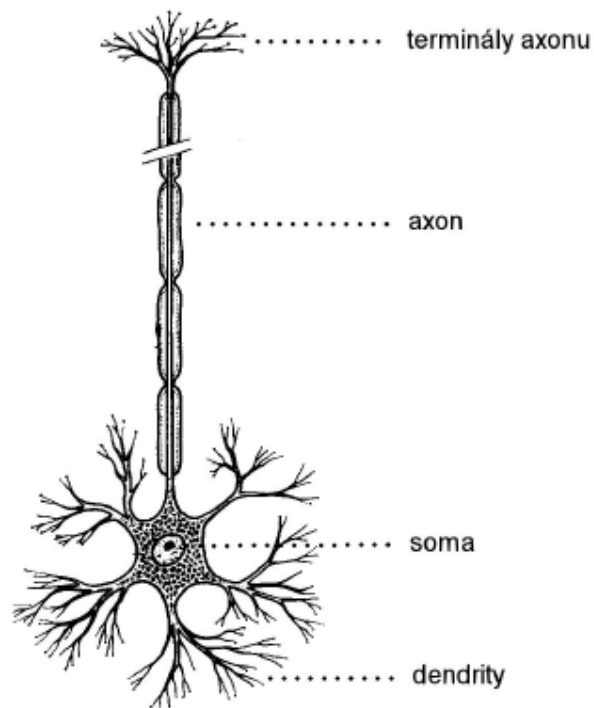
2 NEURONOVÉ SÍTĚ

Neuronová síť je výpočetní model, který spadá do oboru umělé inteligence. Tento model se inspiroval lidským nervovým systémem, který je složen z neuronů. Tyto neurony jsou vzájemně propojeny a jsou schopny si přes tato propojení předávat informace (signály, vzruchy) napříč celým nervovým systémem člověka. Jako první začali o konstrukci modelu takovéto koncepce uvažovat neurovědec Warren McCulloch spolu s logikem Walterem Pittsem, kteří v roce 1943 publikovali ve své práci návod, jak lze pomocí neuronové sítě počítat logaritmické a aritmetické funkce [1]. Praktická realizace byla sice zatím daleko, ale článek posloužil pro další vědce jako podnět ke zkoumání, jak by se těchto sítí dalo využít. Pro účely předpovědi počasí se jedná o poměrně výkonný nástroj.

Neuronové sítě jsou odolné proti zašuměným datům či datům obsahujícím chyby [6], což by v případě této práce, která předpokládá měření meteorologických veličin za pomoci čidel, mohla být výhoda. Jejich vstupem i výstupem mohou být libovolné reálné hodnoty, ačkoli je potřeba volit je rozumně, a také libovolný počet těchto vstupů a výstupů. Nevýhodou oproti jiným metodám umělé inteligence bývá delší čas učení. To však v případě předpovídání hodnot veličin nehraje příliš významnou roli, protože není příliš pravděpodobné neustálé přeučování sítě novými daty. Spíše se předpokládá naučení na historicky významném období obsahujícím pokud možno co nejvíce různých vstupních kombinací bez dalšího doučování sítě. Výhodou naopak je rychlý výsledek, což je u předpovědi především na krátkou dobu dopředu velmi důležité. Pokud by například zvolená metoda 10 minut počítala hodnotu teploty za 10 minut od aktuálního času, pozbyl by výsledek významu. Neuronové sítě se tedy ukázaly jako ideální volba pro tento typ úlohy. To potvrzují i vědecké články [3], [5], [7], [8] a mnohé další, které se právě aplikací neuronových sítí pro úlohu předpovídání počasí zabývají.

2.1 Biologický model neuronu [2]

Jak již bylo řečeno, skládá se nervová soustava živých organismů z tzv. neuronů. Jsou to buňky, které se skládají z několika částí. Vstupy do neuronu jsou nazývány *dendrity*, výstup z neuronu je zvaný *axon*, pomocí něhož je neuron připojen k dendritům dalších neuronů. Třetí část je *soma*, buněčné tělo neuronu. Na obrázku (Obr. 1) je zobrazen biologický neuron.

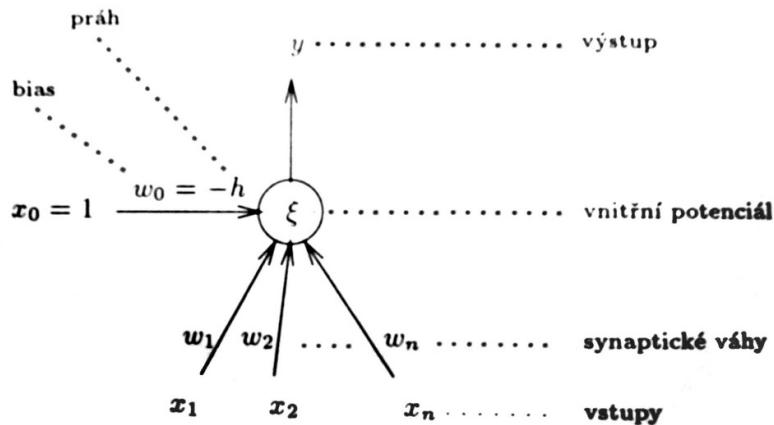


Obr. 1. Biologický neuron [2]

Neurony plní v nervovém systému tři hlavní funkce – přenos, zpracování a uchování informací, které jím procházejí. Těmito informacím se v biologii říká vzruchy. Vzruchy se šíří sítěmi vzájemně pospojovaných neuronů z místa vyslání až k příjemci.

2.2 Matematický model neuronu [2]

Jako je v lidském těle základem nervového systému biologický neuron, pro matematické pojetí neuronové sítě bylo potřeba definovat matematický ekvivalent skutečného neuronu. Formální neuron má stejné vlastnosti jako ten biologický. Má *vstupy* x a *výstup* y . Na vstupech neuronu jsou přítomny *váhy* w , které mohou měnit *vnitřní stav* daného neuronu, který se značí ζ . Po dosažení *prahu* h se určuje *výstup* y jako velikost vnitřního potenciálu neuronu. Na obrázku (Obr. 2) je struktura formálního neuronu ukázána graficky.



Obr. 2. Formální neuron [2]

2.3 Neuronová síť [7]

Struktura neuronové sítě užívané v umělé inteligenci je tedy založená na formálních neuronech, které jsou propojeny. Například výstup jednoho neuronu je vstupem několika dalších. Podle počtu neuronů a jejich vzájemného propojení určujeme tzv. *topologii* (architekturu) sítě. V nervové soustavě živých organismů narazíme na pojmy *receptory*, *efektory* a *projekční dráhy* mezi nimi. Nejedná se o nic jiného, než o *vstupní* a *výstupní* neurony (vrstvy neuronů) a ekvivalent k projekčním drahám jsou *skryté* neurony (vrstva neuronů). Dále lze definovat pojmy *konfigurace sítě*, která je určena váhami všech spojení v síti, a *stav sítě*, který je dán stavem všech neuronů v síti. Síť se během přenosu signálů vyvíjí, mění se hodnoty vah a vnitřní potenciál jednotlivých neuronů. Tento proces se nazývá *učení*.

Postup učení sítě [3]:

1. Shromáždění trénovacích dat

- Volba trénovacích dat je nejzásadnějším krokem v celém procesu učení. Špatně zvolená trénovací data budou zcela degradovat výsledek, ať už budou další kroky jakkoli správné.
- Trénovací data je třeba volit s ohledem na očekávaná výstupní data. Nelze naučit síť na omezeném rozsahu hodnot a očekávat, že bude síť dávat kvalitní výsledky i pro vstupní data z jiného rozsahu, proto je vždy potřeba zabezpečit přítomnost co největšího rozsahu hodnot v trénovací množině.

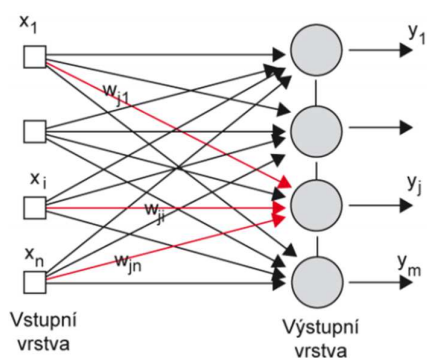
- Pokud je možný výskyt extrémů během používání sítě, je také vhodné tyto extrémy zahrnout do trénovací množiny.
2. Vytvoření sítě
 - Specifikujeme parametry sítě, podle kterých bude poté sestavena. Důležitý je nejen počet vstupů a počet očekávaných výstupů, ale také například její velikost (počet vrstev, počet neuronů v jednotlivých vrstvách), výstupní funkce, učicí algoritmus, rozložení dat (náhodné, lineární,...) a počáteční hodnoty vah.
 - Danou síť sestavíme a inicializujeme.
 3. Učení sítě
 - Síti předkládáme na vstup trénovací data, ke kterým známe požadovaný výstup, a po každém průchodu kontrolujeme, o kolik se skutečný výstup od požadovaného liší. Na základě odlišnosti potom provedeme přenastavení jednotlivých vah v síti.
 - Jakmile je splněna výstupní funkce, učení končí (například chyba výstupu klesne pod určitou hranici).
 4. Testování sítě
 - Síti předložíme na vstup jiná data než trénovací (testovací data), ke kterým ale také známe požadovaný výstup. Ověříme, zda je podobnost výstupu a očekávaného výstupu dostatečná. Pokud ano, ukončíme učení sítě, pokud ne, pokračujeme dále, nebo se vracíme k definici struktury sítě a dalších parametrů

Neuronovou síť lze využít na nejrůznější aplikace, ale především tam, kde neznáme algoritmus řešení, nebo je jeho řešení příliš složité nebo časově náročné. Pro předpověď vývoje počasí v této práci byla zvolena právě z důvodu jednoduchosti a také skutečnosti, že již byla v mnoha vědeckých článcích označena za dobrého kandidáta pro tuto úlohu. Není nutné definovat složité rovnice pro popis fyzikálních dějů v atmosféře, postačí historická data z meteorologické stanice, která budou využita pro natrénování sítě.

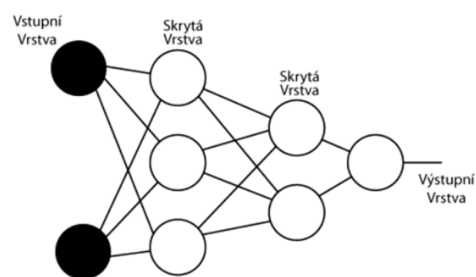
V práci byla využita vícevrstvá síť s back-propagation algoritmem (algoritmus zpětného šíření chyby), čímž bylo navázáno například na článek [3], který tuto síť doporučuje jako dobrého kandidáta.

2.3.1 Vícevrstvá síť [9]

Nejjednodušší (jednovrstvá) síť by měla vstupní neurony spojeny přímo s výstupními. Tato struktura by však neumožňovala síti potřebnou variabilitu. Vícevrstvá síť přidává k předešlé struktuře ještě minimálně jednu vrstvu, která je zařazena mezi vstupní a výstupní neurony. Takovým vrstvám se potom říká „skryté vrstvy“. V práci byly využívány sítě výhradně s jednou skrytou vrstvou, které poskytují dobrou variabilitu a zároveň nejsou zbytečně složité a náročné na trénování. Navíc přidáním dalších skrytých vrstev většinou nedochází k výraznějšímu zlepšení.



Obr. 3. Jednovrstvá síť [9]



Obr. 4. Vícevrstvá (třívrstvá) síť [9]

2.3.2 Back-propagation algoritmus [9]

V překladu *Algoritmus zpětného šíření chyby* je jeden z možných přístupů k učení sítě *s učitelem*. Z výstupu sítě a očekávaného výstupu je vypočítána chyba, která se postupně šíří sítí opačným směrem, od výstupních neuronů ke vstupním. Tato chyba je pro každý neuron upravena derivací přenosové funkce výstupní vrstvy a vahou, která k přenosové funkci vede. Tím je získána chyba, kterou způsobil konkrétní neuron v rámci chyby celkové. Na základě této chyby se poté provede upravení vah. [12]

Postup šíření chyby:

1. Inicializace vah v síti
2. Výběr vstupního vektoru z trénovací množiny a vložení na vstup sítě
3. Zjištění výstupu sítě
4. Vypočtení chyby výstupu oproti očekávanému výstupu
5. Vypočtení chyby na předchozích vrstvách
6. Modifikace vah dle vypočtených chyb
7. Pokračuj bodem 2, dokud není splněna ukončovací podmínka

Základní back-propagation algoritmus je ovšem málo variabilní. Neumožňuje měnit parametry přenosových funkcí a omezuje se na modifikaci vah a prahů v neuronové síti. Aby byla síť více variabilní, vznikalo mnoho modifikací základního back-propagation algoritmu, které umožňovaly měnit během učení více parametrů sítě. Tím je potom dosažena rychlejší konvergence – rychlejší nalezení extrému, a tedy řešení, a také vyšší kvalita řešení [11].

2.4 Výběr vhodné modifikace back-propagation algoritmu

Vzhledem k tomu, že se tímto tématem zabývalo již několik vědeckých článků, bylo při výběru vhodného trénovacího algoritmu využito jejich výsledků k inspiraci.

Například v článku [4], který porovnává algoritmy Levenberg-Marquardt, conjugate gradient a resilient back-propagation, byly tyto postupně využívány pro předpověď průtoku vody na vodních tocích a pro zjištění půdního pnutí, což je důležitým aspektem například při stavbě tunelů [4]. Algoritmus Levenberg-Marquardt dosáhl nejlepší výkonnosti i rychlosti oproti ostatním [4].

Levenberg-Marquardt algoritmus byl již také úspěšně aplikován pro podobnou úlohu, jakou se zabývá tato práce, a sice pro předpovídání teploty, jak popisuje článek [5]. Úspěšnost dosaženou při experimentech přirovnávají autoři ke konvenčním přístupům k předpovídání počasí, kterou přesáhli [5].

V článku [3] shrnuli autoři výsledky výzkumů v oblasti neuronových sítí se zaměřením na předpovídání počasí. Mimo jiné zde uvádějí závěry práce kolektivu Arti R. Naik, Prof. S. K. Pathan a další z roku 2012, kteří vyzdvihují Levenberg-Marquardt algoritmus především pro jeho rychlost a i sami autoři článku [3] jej doporučují jako algoritmus s nejlepší rychlostí učení z dostupných back-propagation algoritmů.

Dle předložených článků lze tedy usuzovat, že tento algoritmus je skutečně dobrým kandidátem pro účely předpovědi teploty, případně i dalších veličin, které jsou pro meteorologii významné. Proto byl použit také v této práci.

2.5 Levenberg-Marquardt algoritmus [10]

Jedná se o nelineární metodu druhého řádu (využívá druhé derivace při výpočtu nových hodnot vah), která je založená na metodě nejmenších čtverců. Tato metoda spojuje dvě

jednodušší metody - Gauss-Newtonovu metodu a Gradientní optimalizaci klasického back-propagation algoritmu.

$$w_{k+1} = w_k - (J_k^T J_k + \mu I)^{-1} J_k e_k \quad (1)$$

Metoda je založena na rovnici (1), která popisuje aktualizaci vah na základě znalosti vah předchozích w_k , vypočtené Jacobiánské matice J , matice identit (jednotková matice) I a kombinačního koeficientu μ (vždy kladné číslo).

$$J = \begin{bmatrix} \frac{\partial e_{1,1}}{\partial w_1} & \frac{\partial e_{1,1}}{\partial w_2} & \dots & \frac{\partial e_{1,1}}{\partial w_N} \\ \frac{\partial e_{1,2}}{\partial w_1} & \frac{\partial e_{1,2}}{\partial w_2} & \dots & \frac{\partial e_{1,2}}{\partial w_N} \\ \dots & \dots & \dots & \dots \\ \frac{\partial e_{1,M}}{\partial w_1} & \frac{\partial e_{1,M}}{\partial w_2} & \dots & \frac{\partial e_{1,M}}{\partial w_N} \\ \dots & \dots & \dots & \dots \\ \frac{\partial e_{p,1}}{\partial w_1} & \frac{\partial e_{p,1}}{\partial w_2} & \dots & \frac{\partial e_{p,1}}{\partial w_N} \\ \frac{\partial e_{p,2}}{\partial w_1} & \frac{\partial e_{p,2}}{\partial w_2} & \dots & \frac{\partial e_{p,2}}{\partial w_N} \\ \dots & \dots & \dots & \dots \\ \frac{\partial e_{p,M}}{\partial w_1} & \frac{\partial e_{p,M}}{\partial w_2} & \dots & \frac{\partial e_{p,M}}{\partial w_N} \end{bmatrix} \quad (2)$$

Jacobiánskou matici J lze vypočítat dle vztahu (2), který využívá derivace váhového vektoru w_k a derivace hodnot chyby m -tého výstupu při použití p -tého vstupu během trénování $e_{p,m}$. Aktuální chyba $e_{p,m}$ se dá vypočítat jako rozdíl požadovaného výstupu a aktuálního výstupu v daném kroku (p , m). P značí počet trénovacích vstupů, M počet výstupů a N je počet vah. e_k potom značí chybový vektor zachycený vztahem (3).

$$J = \begin{bmatrix} e_{1,1} \\ e_{1,2} \\ \dots \\ e_{1,M} \\ \dots \\ e_{p,1} \\ e_{p,2} \\ \dots \\ e_{p,M} \end{bmatrix} \quad (3)$$

3 VYUŽITÍ GLOBÁLNÍHO DRUŽICOVÉHO POLOHOVÉHO SYSTÉMU

Dosti diskutovaná je v poslední době možnost využití Globálního družicového polohového systému (GNSS) pro zpřesnění předpovědi počasí. V této práci byla tato metoda také okrajově prozkoumána, jako jeden z možných přístupů ke zlepšení předpovědi získané pouze na základě standardních meteorologických veličin. V případě neuronové sítě se jednalo o další pomocný parametr, který by byl získán ze systému GNSS, a který by posloužil jako další vstup do neuronové sítě. Jako vhodný byl zvolen parametr Zenith Total Delay (ZTD), který udává zpoždění satelitního signálu vlivem troposféry. Tento parametr však bylo nejprve nutné získat. Pro tyto účely byl využit software RTKLib 2.4.2 [17], který na základě souborů RINEX (The Receiver Independent Exchange Format) umí ZTD vypočítat. RINEX soubory se využívají pro ukládání a výměnu satelitních dat. Pro výpočet byla použita metoda Precise Point Positioning (PPP), která vyžadovala 4 druhy RINEX souborů. *Navigation Data* soubory (*.N) a *Observation Data* soubory (*.O) dodala firma Navmatix s.r.o. a soubory *Precise Clocks* (*.CLK) a *Orbits* (*.SP3) jsou volně poskytovány například organizací The International GNSS Service (IGS), od které byly také využity. U metody PPP je ZTD parametr vlastně vedlejším produktem. Primárním účelem u této metody je přesné zjištění pozice daného bodu. Dosahuje přesnosti až několika centimetrů. [19], [20]

4 POPIS POUŽITÉHO DATASETU

Data využitá v této práci pocházejí z více zdrojů. Nejdůležitějším zdrojem byla zadávající firma Navmatix s.r.o., která dodala krátkodobá data meteorologická a data z Globálního družicového polohového systému (GNSS) a dále příspěvková organizace Hvězdárna a planetárium Brno p. o., která poskytla dlouhodobá data meteorologická.

4.1 Meteorologická data od Navmatix s.r.o.

Krátkodobý meteorologický dataset, který byl hlavním zdrojem pro experimenty v rámci této práce byl získán z měření za pomoci lokální meteostanice firmy Navmatix s.r.o. Jedná se o meteostanici MetPak od firmy Gill Instruments. Tato stanice slouží pro lokální měření okamžitých hodnot atmosférických veličin, které jsou pro předpověď počasí významné. Data byla zaznamenávána do textového souboru jako hodnoty oddělené čárkami. Během zpracovávání byly firmou Navmatix s.r.o. dodány 2 datasety. Oba měly stejný formát, ale lišily se počtem dní, ze kterých byly data nasbírány a také periodou vzorkování. První dataset (kratší) byl dodán před začátkem práce, druhý až v průběhu. V obou datasetech byly zastoupeny tyto veličiny: *směr větru*, *rychlost větru*, *atmosférický tlak* (nepřepočtený na hladinu moře), *vlhkost*, *teplota*, *rosný bod*, *sklon v ose X*, *sklon v ose Y*, *hodnota digitálního vstupu* a *napětí*. Dataset číslo 2 obsahoval navíc ještě *datum* a *UTC čas*. Ne všechny hodnoty však byly identifikovány jako využitelné pro účely experimentů s predikcí vývoje počasí. Prvních šest zmíněných veličin by bylo pro předpověď vhodných, první dataset však neobsahoval validní hodnoty vlhkosti a tím pádem i rosného bodu, který je na vlhkosti závislý. Problém byl charakterizován jako chyba vlhkosního čidla, která se projevovala nulovými hodnotami vlhkosti v prvním datasetu. Vzhledem k tomuto problému byly nakonec pro všechny experimenty s teplotou a tlakem, které byly zpracovávány dříve, využity jen veličiny *rychlost větru*, *směr větru*, *atmosférický tlak* a *teplota*. Pro experimenty s vlhkostí už byl použit dataset 2, u kterého bylo možno využít všech 5 vstupních veličin, tedy *rychlost větru*, *směr větru*, *atmosférický tlak*, *vlhkost* a *teplota*.

První datový soubor byl využíván především na experimenty s teplotou a tlakem. Obsahoval naměřené hodnoty z více než 3 dní v měsíci lednu 2016 v době od 19. do 22. ledna. Celkově se jednalo o 264 195 vzorků. Pro rovnoměrnější rozložení počtu ranních i odpoledních měření byl dataset ořezán tak, aby obsahoval právě 3 dny naměřených hodnot, tedy 259 200 vzorků. Jednotlivé vzorky dat byly zaznamenávány po 1 sekundě.

Druhý datový soubor byl využit především pro experimenty s vlhkostí a obsahoval naměřené hodnoty z období 17. 2. 2016 až 12. 3. 2016. Jednalo se o 186 357 vzorků. Vzorky ze začátku datového souboru však stále neměly validní vlhkost, proto byl datový soubor shora ořezán na 140 111 vzorků, které již obsahovaly validní hodnoty všech veličin a zároveň tak bylo dosaženo zarovnání dat na celé dny (jednalo se o data za 18 dní). Vzorky byly v datech ukládány po 11 sekundách.

4.2 Meteorologická data od Hvězdárny a planetária Brno, p. o.

Vzhledem k tomu, že krátkodobá data se ukázala být značně omezujícími při experimentech a firma Navmatix s.r.o. neměla k dispozici data z delšího časového horizontu, byla požádána tato příspěvková organizace o nějaká dlouhodobější. Obdržená data byla ve formátu CSV jako hodnoty oddělené středníky. Obsahovala následující hodnoty: *datum a čas vložení, datum změření, čas změření, vnitřní teplota, venkovní teplota, rosný bod, vnitřní vlhkost, venkovní vlhkost, rychlost větru, směr větru, směr větru textově, pocitová teplota, srážky 1h, srážky 24h, srážky celkově, tlak, tlaková tendence a slovní předpověď aktuálního stavu počasí*. Z dostupných veličin však byly využity pouze ty hodnoty, které byly dostupné také v datasetu od firmy Navmatix s.r.o., tedy *venkovní teplota, tlak, rychlost větru a směr větru*.

Datový soubor byl složen z dat v období od 29. 3. 2012 do 23. 2. 2016 a obsahoval celkem 461 247 vzorků. Vzorky byly od sebe vzdáleny o periodu 4 minuty. Data bohužel obsahovaly občasné výpadky v průběhu celého datasetu, které však byly velmi krátkodobé – většinou výpadek jen 1 vzorku, tedy žádné kontinuální přerušení nebylo v datech identifikováno.

4.3 GNSS data od Navmatix s.r.o.

Tento dataset byl použit pouze pro experimenty v rámci kapitoly 6.5 a jejích podkapitol. Pocházel z vícefrekvenčního GNSS přijímače firmy Navmatix s.r.o. a obsahoval data ze 13 dnů ve formátu RINEX z období od 28. 2. 2016 do 11. 3. 2016. Pro každý den byly obsaženy soubory *Observation Data, GPS Navigation Data a Glonass Navmess Data*. Poslední zmíněné nebyly v práci využity.

4.4 Data pro metodu PPP od organizace The International GNSS Service (IGS)

Organizace The International GNSS Service (IGS) poskytuje na svých webových stránkách [18] volně k dispozici soubory *Precise Clocks* a *Orbits*. Z tohoto zdroje bylo čerpáno pro získání těchto dvou typů souborů pro období od 28. 2. 2016 do 11. 3. 2016, aby mohly být přidány k souborům z kapitoly 4.3 a využity pro metodu Precise Point Positioning (PPP). Tento dataset byl také použit pouze pro experimenty v rámci kapitoly 6.5 a jejích podkapitol.

II. PRAKTICKÁ ČÁST

5 PŘEDPOVĚĎ NA ZÁKLADĚ METEOROLOGICKÝCH DAT

První část práce je zaměřena na sestavení a otestování úspěšnosti modelů pouze na základě historických meteorologických dat. Vzhledem k tomu, že dle požadavků firmy Navmatix s.r.o. měla být práce zaměřena na prozkoumání možností předpovědi teploty, atmosférického tlaku a vlhkosti, jsou jednotlivé podkapitoly členěny na pod-části, které se vždy postupně věnují všem těmto veličinám v rámci jednoho přístupu. Postupně byl vyzkoušen různě dlouhý interval předpovědi (10 minut, 1 hodina, 1 den), poté byl zvolen interval 1 hodiny jako referenční a byly prováděny pokusy o optimalizaci této předpovědi. Bylo provedeno několik sestavení sítě s rozdílným počtem neuronů za účelem nalézt optimální velikost sítě. Bylo experimentováno s normalizací dat a také vlivem dat z více meteorologických stanic na kvalitu předpovědi. Poté byl trénovací datový soubor zaměněn za dlouhodobý a byly prováděny experimenty se sítěmi natrénovanými dlouhodobými daty. Nakonec byl otestován druhý možný přístup k sestavení sítě, a sice vytvoření sítě, která bude předpovídat všechny meteorologické veličiny naráz (1 síť na vše). Dá se předpokládat, že každá ze zkoumaných veličin bude vyžadovat jinak nastavenou a velkou síť. [16]

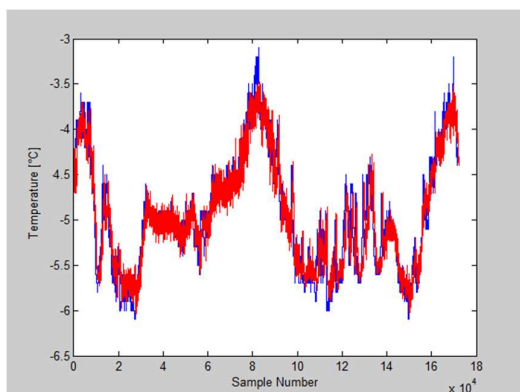
5.1 Předpověď veličin 10 minut dopředu

Pro první část experimentů byl zvolen interval 10 minut. Trénovací a cílová (target) data měla tedy vzorky vzájemně posunuty o 10 minut, tak aby síť dávala pro aktuální hodnotu vstupu předpovězenou hodnotu výstupu, který je vzdálen 10 minut do budoucna. Zdrojová data byla ponechána beze změny, tedy s periodou 1 sekundy mezi samplly. Data ze dvou dnů byla použita pro trénování a data ze třetího dne poté k otestování předpovědi. V praxi by tedy taková předpověď fungovala tak, že by síti bylo předloženo nejaktuálnějších 10 minut známých hodnot (samozřejmě s periodou menší než 10 minut), a síť by předpověděla následující hodnoty pro 0 – 10 minut dopředu.

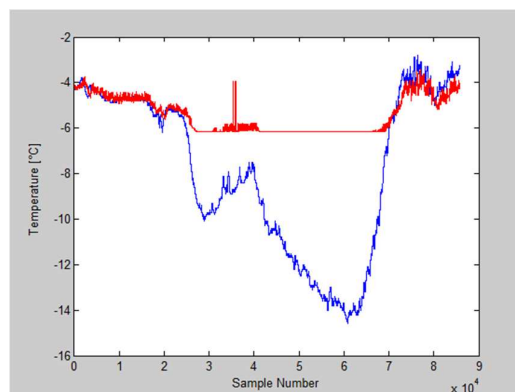
5.1.1 Aplikace 10minutové předpovědi na teplotu

Pro první experimenty byla zvolena předpověď teploty, protože tato veličina je obecně nejvíce sledovaná a lze tedy předpokládat její největší důležitost i zde. Lze vidět, že křivka trénování poměrně pěkně kopíruje požadovanou hodnotu. U křivky testování je ale patrné, že výsledný vývoj teploty byl předpovědí kopírován jen do zhruba jedné třetiny dne. Další špatný vývoj grafu má na svědomí skutečnost, že data třetího dne obsahovala nižší teploty,

než první dva dny použité pro trénování. V prvních dvou dnech se teplota dostala na minimum -6 stupňů Celsia, kdežto třetí den klesla až na -14. Rozhodně lze proto doporučit pro trénování vybrat více rozdílná data namísto souvislého intervalu pár dní. Spíše lze doporučit různé úseky, které budou obsahovat co nejvíce různých situací (rozdílné hodnoty teplot).



Obr. 5. Trénování 10minutové předpovědi teploty



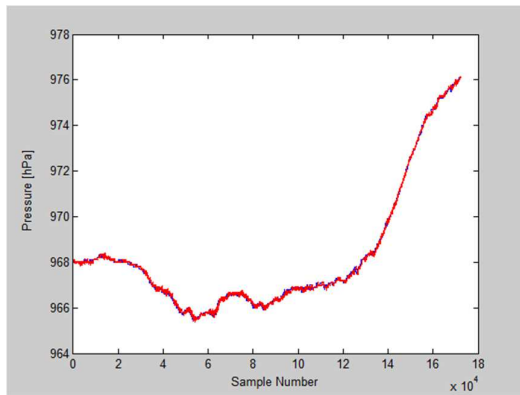
Obr. 6. Testování 10minutové předpovědi teploty

Během testování bylo dosaženo maximální chyby 8,45 °C. Průměrná chyba byla 2,6 °C a MSE byla 14,0620.

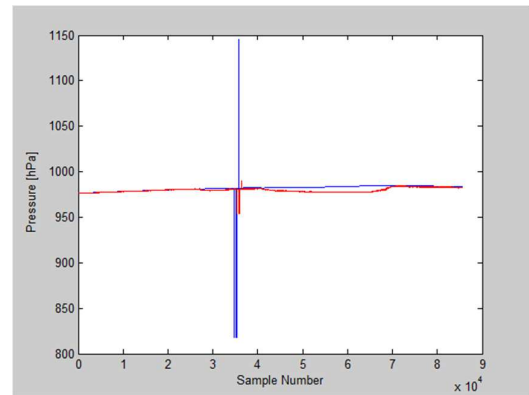
5.1.2 Aplikace 10minutové předpovědi na tlak

Druhou veličinou, jejíž předpovězená hodnota byla bodem zájmu této práce, je atmosférický tlak. Neuronová síť byla trénována stejnými daty, ale jako výstupní známá veličina byl namísto teploty použit tlak. Graf (Obr. 7) zachycující průběh trénování ukazuje velmi dobrý výsledek. Křivky jsou téměř totožné. U testování je už vidět malá odlišnost ve druhé třetině grafu (Obr. 8). Také je zde patrných několik zubů, které značí úskok hodnoty tlaku poměrně výrazně nahoru i dolů. To také způsobilo zakolísání předpovědi bezprostředně po této události. Byla provedena analýza dat a vzhledem k tomu, že na ostatních veličinách nebyla v těchto několika málo vzorcích pozorována žádná výrazná změna, která by naznačovala, že tyto úskoky hodnoty tlaku jsou opodstatněné, byly tyto úskoky přičteny chybě měřidla a tedy dat. V takovém případě je samozřejmě ovlivněna přesnost předpovědi, a proto, vzhledem k tomu, že takto „vadných“ vzorků nebylo v datovém souboru mnoho, byl datový soubor ručně přefiltrován. Rozdíl mezi skutečnou hodnotou a předpovězenou byl bez přefiltrování v průměru 2,14 hPa, což by samo o sobě značilo dobrý výsledek, ale maximální rozdíl byl 165,20 hPa, což už je

naopak skutečně špatný výsledek. Motivace k přefiltrování dat proto byla o to větší. Hodnota MSE je v tomto případě 10,8566.

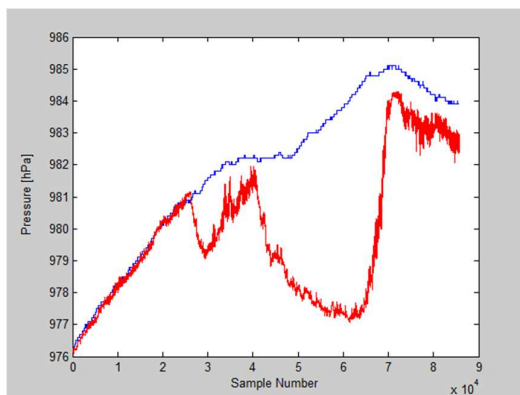


Obr. 7. Trénování 10minutové předpovědi tlaku



Obr. 8. Testování 10minutové předpovědi tlaku

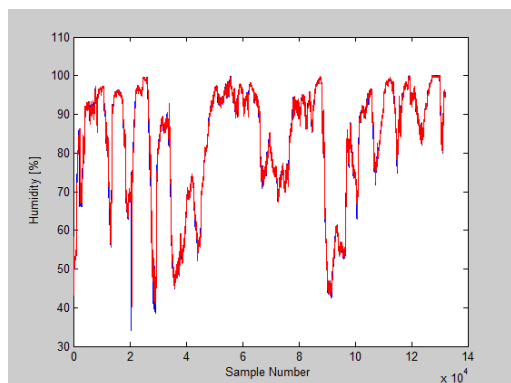
Po přefiltrování dat (odstranění 4 vadných hodnot), byla testovací množina opět předložena síti. Výsledek je zobrazen na grafu (Obr. 9) níže. Ačkoliv se na první pohled může zdát, že předpověď je mnohem horší, než u předchozího případu, není tomu tak. Nyní je totiž maximální rozdíl 7,31 hPa, což je oproti předchozímu stavu velké zlepšení. Průměrná chyba byla nyní opět 2,14 hPa a hodnota MSE 9,5858. Ke zlepšení tedy došlo, minimálně v úseku, kde byly filtrované data. Výsledek nicméně není úplně optimální a bude tedy na místě další prověření, jak lze síť optimalizovat (velikost sítě, trénovací data,...), aby více dodržovala trend vývoje dat (prostřední část grafu).



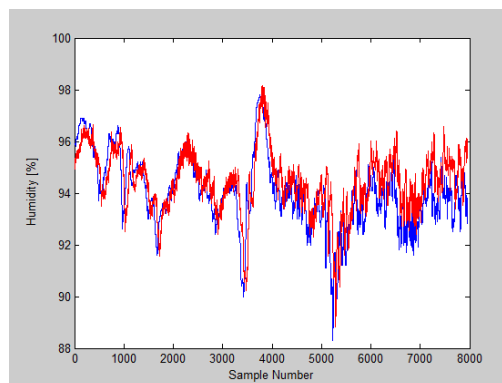
Obr. 9. Testování 10minutové předpovědi tlaku (filtrovaná data)

5.1.3 Aplikace 10minutové předpovědi na vlhkost

Protože dříve využívaná data neměla k dispozici validní hodnotu vlhkosti, byly experimenty s touto veličinou ponechány na konec. Mezitím proběhla oprava vlhkostního čidla a byla získána nová data, kde již byly všechny veličiny validní. Datový soubor tedy obsahoval 5 veličin, a to *směr větru*, *rychlost větru*, *tlak*, *vlhkost* a *teplotu*. Všechny byly použity jako vstup do sestavené neuronové sítě. Datový soubor nyní ovšem neměl vzorky vzdálené o periodu vzorkování 1 sekunda, jako tomu bylo u teploty a tlaku, ale o 11 sekund. To má za výsledek menší počet dat jak pro trénování, tak i pro testování. Tento nedostatek je ale kompenzován delším časovým horizontem, ze kterého byla data k dispozici. Data byla sesbírána od 22. 2. 2016 14:22:39 do 11. 3. 2016 14:22:37 a jednalo se o 140 111 vzorků. Pro testování byla použita data z posledních 24 hodin (tedy 8 015 vzorků) a zbylá data (132 096 vzorků) zůstala jako trénovací. Velikost sítě byla volena standardních 10 neuronů ve skryté vrstvě. Volba počtu neuronů bude blíže popsána v kapitole 6.1.



Obr. 10. Trénování 10minutové předpovědi vlhkosti



Obr. 11. Testování 10minutové předpovědi vlhkosti

Výsledek předpovědi je zobrazen na grafech výše. Křivka výstupu trénování opět velmi přesně kopíruje očekávanou hodnotu. Při předložení testovacích dat jsou vidět drobné rozdíly, ovšem je třeba vzít v úvahu měřítko osy y. V průběhu testování byl rozdíl mezi skutečnou hodnotou a předpovězenou v průměru 0,83 % a maximální rozdíl byl 4,24 %. Hodnota MSE je v tomto případě 1,1891.

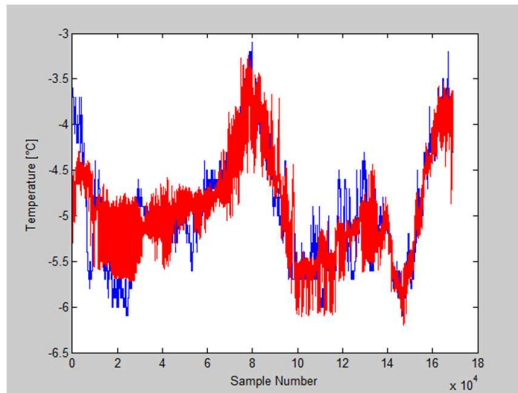
5.2 Předpověď veličin 1 hodinu dopředu

Druhá část experimentů byla zaměřena na zvětšený interval předpovědi (zvětšenou vzdálenost vzorku předkládaného a předpovídaného) na 1 hodinu, stále však bylo počítáno

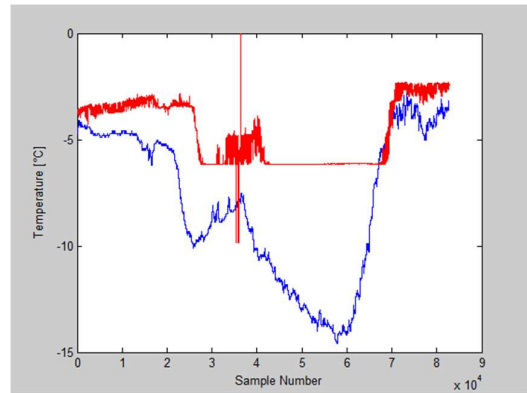
se stejným rozložením datového souboru na 2 dny trénování a 1 den testování. Na základě každého vzorku má tedy síť dát hodnotu předpokládané veličiny za 1 hodinu.

5.2.1 Předpověď teploty po 1 hodině

Z grafů lze vyzorovat, že předpověď je značně horší, než když síť předpovídala jen 10 minut dopředu, a to i v první třetině grafu, kde předchozí desetiminutová předpověď fungovala dobře. Se zvětšující se vzdáleností předpovídané veličiny od aktuálního vzorku můžeme tedy předpokládat zhoršení přesnosti. Opět je také patrné oříznutí grafu hodnotou -6 stupňů celsia, což byla v trénovací množině minimální hodnota. Je to tedy chyba plynoucí z nedostatečného trénovacího souboru. Maximální chyba během testování byla $8,46$ °C, průměrná chyba byla $3,38$ °C a hodnota MSE $17,2296$.



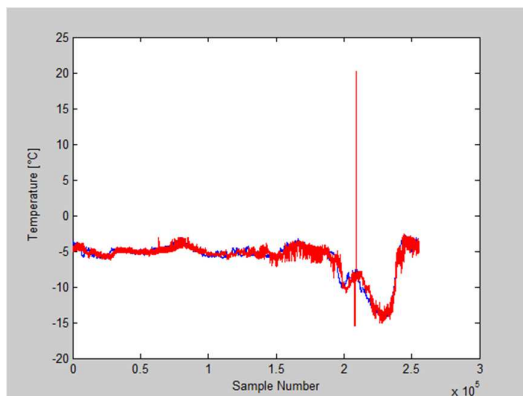
Obr. 12. Trénování hodinové předpovědi teploty (2 dny)



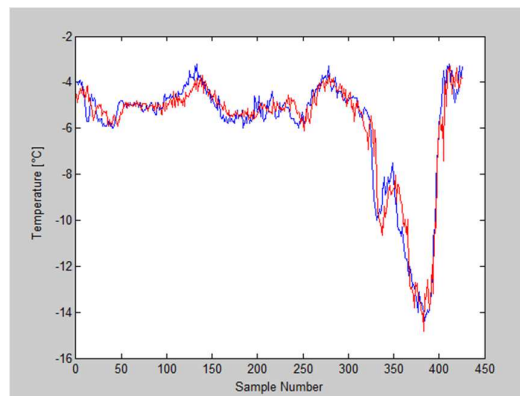
Obr. 13. Testování hodinové předpovědi teploty (1 den)

Jelikož výsledek nebyl úplně přesvědčivý, rozhodl jsem se pro změnu datasetu. Rozdělení 2 dny trénování versus 1 den testování se ukázal jako nevhodný. Proto jsem provedl přerozdělení dat. Nyní jsem použil jako trénovací data všechny 3 dostupné dny. Vybral jsem z nich však každý šestistý vzorek, což znamená hodnotu jednou za 10 minut, a tyto vybrané hodnoty byly uloženy do testovací množiny. Z trénovací množiny byly tyto hodnoty odstraněny, aby byla zachována věrohodnost předpovědi, tedy že síť daná data skutečně při trénování neviděla. Na druhou stranu lze předpokládat, že tímto rozdělením dat bude simulován případ, kdy je síť natrénována dostatečným množstvím různých dat, které pokrývají využitelný rozsah hodnot teploty.

Výsledky předpovědi při tomto rozdělení vstupních dat lze vidět na grafech níže.



Obr. 14. Trénování hodinové předpovědi teploty (3 dny)

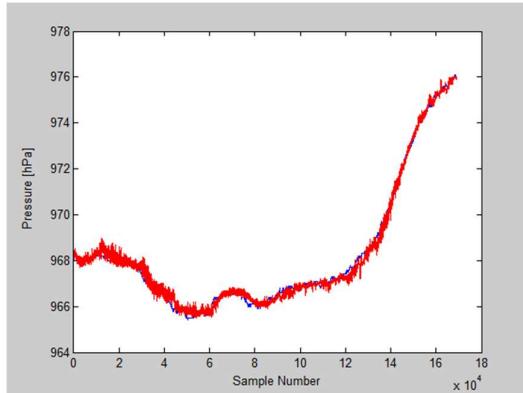


Obr. 15. Testování hodinové předpovědi teploty (každý 600. vzorek)

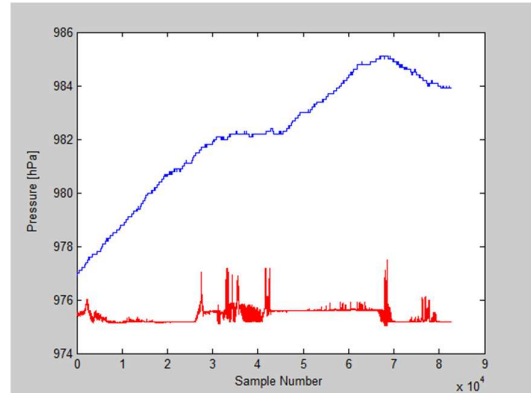
Výsledná předpověď je nyní velmi přesná. Tím je tedy podpořena důležitost výběru správných trénovacích dat. Pro trénování je nutné zvolit data, která obsáhnou co nejvíce možných hodnot teploty, aby byla síť připravena na co nejvíce situací vstupních parametrů. Po úpravě dat bylo dosaženo maximální chyby 3,91 °C, v průměru byla chyba ale jen 0,47 °C a hodnota MSE byla 0,4398, což je také velmi dobrá hodnota. Zlepšení tedy bylo poměrně razantní.

5.2.2 Předpověď tlaku po 1 hodině

Stejně jako u předpovědi teploty je další část věnována předpovědi hodnoty za 1 hodinu, tentokrát ale atmosférického tlaku. Předpoklad je, že přesnost předpovědi bude horší než v případě předpovědi na 10 minut. Jak lze vidět z grafu (Obr. 17), předpověď je velmi špatná. Křivka předpovědi tlaku v podstatě nijak nereaguje na změnu vstupních hodnot. V tomto stavu je předpověď vlastně nepoužitelná. Pokud ale porovnáme hodnoty grafu (Obr. 16) a (Obr. 17) na ose y, lze vidět, že hodnoty tlaku jsou v testovací množině v úplně jiném intervalu, než je tomu u hodnot trénovacích. Je tedy velmi pravděpodobné, že se zde opět projevil problém ze sekce 5.2.1, a sice, že má síť problém se vstupními hodnotami, na které nebyla natrénovaná. Je to samozřejmě důsledek malého časového intervalu obsaženého v trénovací množině, který nepokryl všechny možné vstupy, které při testování mohou nastat. Maximální chyba v rámci testovacího datasetu byla 10,08 hPa, průměrně byla chyba 6,67 hPa a hodnota MSE byla 49,2279.



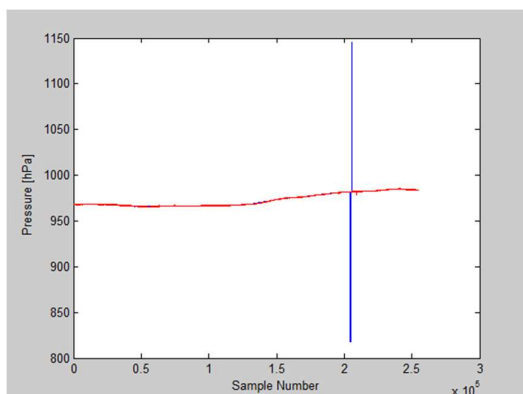
Obr. 16. Trénování hodinové předpovědi tlaku (2 dny)



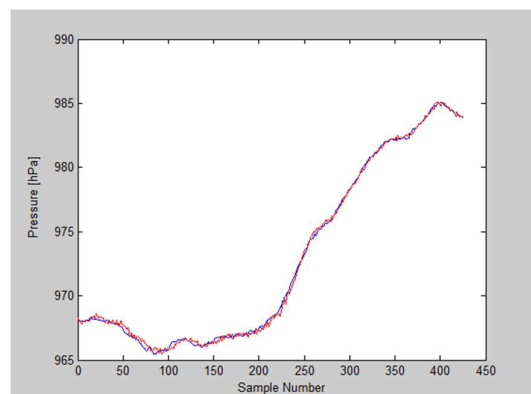
Obr. 17. Testování hodinové předpovědi tlaku (1 den)

Bylo tedy stejně jako v sekci 5.2.1 provedeno přerozdělení vstupních dat. Trénovacími daty byly nyní všechny 3 dostupné dny a každý šestistý vzorek (1x za 10 minut) byl z této trénovací množiny vyjmut. Z těchto vyňatých vzorků byla sestavena nová testovací množina, která bude nyní zaručovat, že testovací hodnoty předkládané síti budou nyní v rozsahu hodnot, které síť zná, ale zároveň to nebudou přímo vzorky, které by byly k trénování použité a síť je tedy znala.

Graf (Obr. 19) ukazuje velmi dobrý průběh křivky předpovězených hodnot oproti křivce skutečného vývoje. Je tedy viditelný rozdíl ve výsledku, pokud síť správně natrénujeme co největším rozsahem hodnot, jaký máme k dispozici. Nyní po úpravě je maximální absolutní chyba 0,68 hPa, v průměru chyba činila 0,16 hPa a hodnota MSE dosáhla dokonce 0,0449.



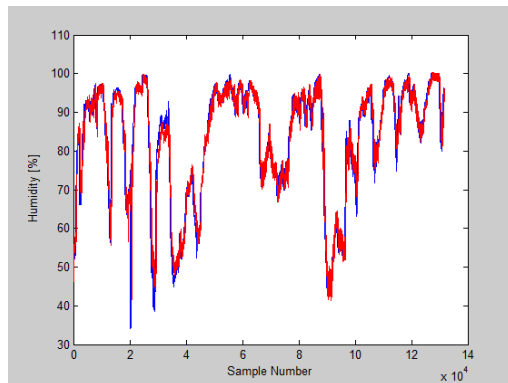
Obr. 18. Trénování hodinové předpovědi tlaku (3 dny)



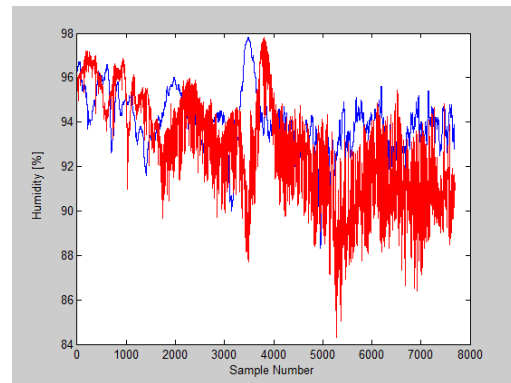
Obr. 19. Testování hodinové předpovědi tlaku (každý 600. vzorek)

5.2.3 Předpověď vlhkosti po 1 hodině

Podobně jako v kapitole 5.1.3, byl pro vlhkost využit dataset se vzorky po 11 sekundách. Tentokrát však z těchto dat byla vygenerována trénovací a cílová data se vzorky vzájemně posunutými o 1 hodinu, čímž je dosaženo, že výstup ze sítě bude hodnota vzdálená od vstupních parametrů o 1 hodinu na časové ose. V datech je to tedy vzdálenost o 327 vzorků.



Obr. 20. Trénování hodinové předpovědi vlhkosti (17 dní)



Obr. 21. Testování hodinové předpovědi vlhkosti (1 den)

Výsledky jsou poněkud překvapivé. Především na grafu testování lze zpozorovat, že výsledná předpovězená hodnota (červená křivka) je vždy posunutá doprava oproti očekávané. Vypadá to tedy, jako by byla předpověď zpožděná. Vzhledem k tomu, že toto posunutí se však velmi blíží počtu vzorků 327, což je posunutí trénovací a cílové množiny, je zde pravděpodobně problém, že síť nefunguje správně. Jinými slovy předává na výstup to, co dostala na vstupu, jen s malou úpravou. Bylo provedeno několik experimentů s velikostí sítě a spousta opětovných trénování. Grafy (Obr. 20) a (Obr. 21) jsou nejlepší výsledek, který se podařil díky těmto experimentům získat při použití 5 neuronů ve skryté vrstvě. Je ovšem jasné, že takto sestavená síť nemůže být nasazena v reálném provozu, protože by selhávala. V dalších částech bude snaha o optimalizaci ať už velikostí sítě, nebo optimalizací. V tomto případě je největší chyba 10,01 % a průměrná chyba je 2,10 %. MSE je potom 7,0441.

5.3 Předpověď veličin 1 den dopředu

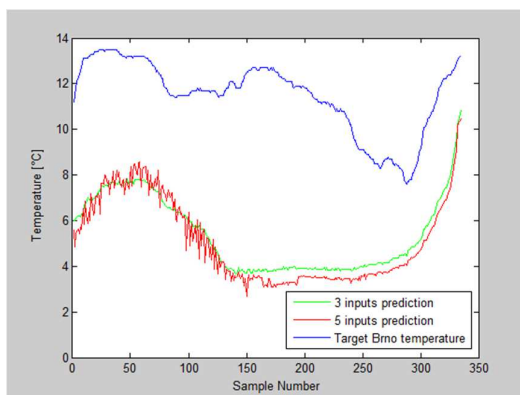
Třetí část experimentů se věnuje poslednímu požadavku na vzdálenost předpovězené hodnoty od hodnoty aktuální, což byl 1 den. Vzhledem k tomu, že první dataset, který byl

pro experimenty k dispozici, obsahoval pouhé 3 dny, bylo jeho využití pro tyto experimenty zavrženo. Nabízelo se tudíž použít buď druhý dataset z dat od firmy Navmatix s.r.o., nebo využít dlouhodobá data z datasetu Hvězdárny a Planetária Brno p.o. Na otázku, který dataset bude lépe zvolit, nebyla jednoznačná odpověď, proto byl proveden experiment s oběma datasety a výsledky byly bez zkoumání ideální velikosti sítě a dalších možných optimalizací porovnány mezi sebou.

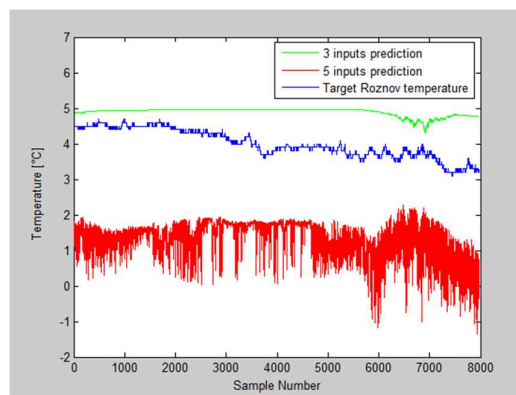
Dataset od firmy Navmatix s.r.o. byl rozdělen na trénovací množinu, která zahrnovala vzorky ze 14 dnů se vzorkovací periodou 11 sekund a testovací množinu obsahující vzorky za 1 den, taktéž s periodou 11 sekund. Dataset od organizace Hvězdárna a Planetárium Brno p.o. byl zase rozdělen na trénovací množinu obsahující data za 4 roky a testovací množinu obsahující data za 1 den. Vzorky byly od sebe vzdáleny o periodu 5 minut.

5.3.1 Předpověď teploty za 1 den

Předpověď byla provedena dvakrát za použití obou dlouhých datasetů (Rožnov a Brno). Nejprve byly obě sítě sestaveny za použití všech 5 vstupních veličin a poté ještě jednou při použití pouze 3 vstupních veličin (síla a směr větru byly vynechány). Obě varianty z obou datasetů byly poté porovnávány. K sestavování sítí bylo využito blokové rozdělení dat a počet neuronů ve skryté vrstvě byl volen jako aritmetický průměr součtu počtu vstupů a výstupů – tedy 3 neurony pro 5 vstupních veličin a 2 neurony pro 3 vstupní veličiny.



Obr. 22. Předpověď teploty za 1 den – dataset Brno



Obr. 23. Předpověď teploty za 1 den – dataset Rožnov

Graf (Obr. 23) ukazuje, že blíže ke skutečnému průběhu je výstup ze sítě, která nemá vstupy pro směr a rychlost větru. Také lze vidět, že červená křivka je mnohem více zašuměná. Vynecháním obou veličin jsme tedy získali přesnější předpověď a ještě navíc méně zašuměnou. Graf (Obr. 22) potvrzuje pozitivní vliv vynechání veličin síly a směru

větru, protože zelená křivka je také mnohem méně zašuměná, ale rozdíl mezi oběma předpovězenými teplotami a skutečnou je poměrně velký.

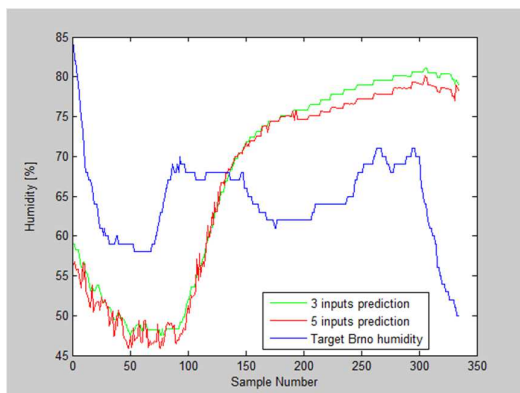
Dataset	Počet vstupů	Největší chyba	Průměrná chyba	MSE
Brno	5	9,82 °C	6,38 °C	43,6265
	3	9,00 °C	6,09 °C	39,6121
Rožnov	5	4,90 °C	2,71 °C	7,6015
	3	1,70 °C	0,90 °C	0,9364

Tab. 1. Chyby predikce teploty na 1 den

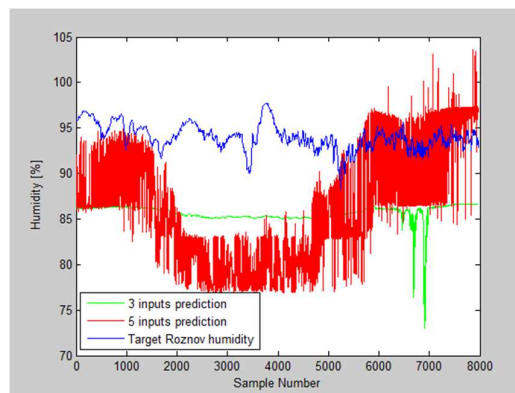
Z tabulky (Tab. 1) je jasné, že mnohem lépe fungovala předpověď při použití datasetu od firmy Navmatix s.r.o. Průměrná chyba pod 1 stupeň Celsia slibuje dobrou přesnost předpovědi při použití pouze tlaku, vlhkosti a teploty jako vstupy do sítě. Horší přesnost u datasetu z Brna může být dána buď skutečností, že dataset obsahuje občasné výpadky, ale spíše délkou trénovacího souboru, který obsahoval 4 roky trénovacích dat, což bylo pravděpodobně příliš mnoho. Dataset z Rožnova obsahoval pouze 14 dní, přesto bylo jeho použitím dosaženo menší chyby předpovědi.

5.3.2 Předpověď vlhkosti za 1 den

Další veličinou, pro kterou je předpověď zkoumána, je vlhkost. Opět byly využité stejné datasety se stejným rozložením testovacích a trénovacích hodnot a opět za použití i bez použití rychlosti a směru větru jako vstupu do sítě. Srovnání výsledků následuje v grafech níže.



Obr. 24. Předpověď vlhkosti za 1 den – dataset Brno



Obr. 25. Předpověď vlhkosti za 1 den – dataset Rožnov

U datasetu z Brna je patrný problém s dodržáním trendu a odchylka od skutečného vývoje je opět poměrně velká, což lze vidět v grafu (Obr. 24). Dataset z Rožnova opět fungoval lépe, i když odchylka od skutečného vývoje není malá. Navíc je patrné velmi silné zašumění při použití všech 5 vstupních veličin. Toto je opět eliminováno při použití pouze 3 vstupních veličin.

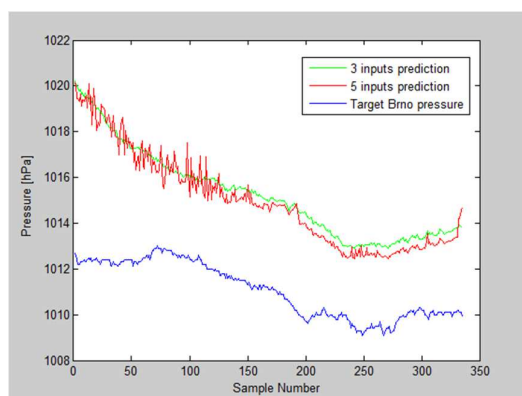
Dataset	Počet vstupů	Největší chyba	Průměrná chyba	MSE
Brno	5	28,90 %	12,23 %	182,1115
	3	29,53 %	12,58 %	190,1224
Rožnov	5	20,47 %	8,53 %	98,5408
	3	19,73 %	8,33 %	71,7866

Tab. 2. Chyby predikce vlhkosti na 1 den

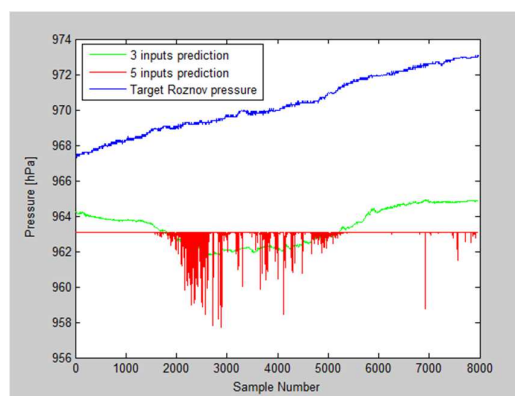
Tabulka (Tab. 2) opět shrnuje chyby, které použitím obou datasetů vznikly. Dataset z Rožnova dosáhl průměrné absolutní chyby kolem 8,55 %. Opět se tedy ukazuje jako výhodnější kratší dataset s menší periodou. Zlepšení při vyřazení rychlosti a směru větru tentokrát není tak patrné, nicméně z grafů bylo vidět, že dojde alespoň k redukci šumu, což je při reálném použití výhodné.

5.3.3 Předpověď tlaku za 1 den

Poslední zkoumanou veličinou je tlak. Jeho předpověď na vzdálenost 1 dne opět zahrnuje sestavení, trénování a otestování sítě jak za pomoci datasetu Brněnského, tak i Rožnovského a jejich porovnání.



Obr. 26. Předpověď tlaku za 1 den – dataset Brno



Obr. 27. Předpověď tlaku za 1 den – dataset Rožnov

Oba grafy mají podobný trend jako jejich skutečné předlohy průběhu tlaku. Trend je u obou porušen jenom na začátku asi do pětiny průběhu. Oba grafy vykazují určitý offset od skutečného průběhu a červená křivka v grafu (Obr. 27), která reprezentuje předpověď při použití Rožnovského datasetu a všech 5 vstupních veličin, má téměř konstantní průběh až na pár míst s šumem. Vzhledem k tomu, že skutečný vývoj tlaku byl sice lineární, ale nikoliv konstantní, lze tento přístup k předpovědi tlaku vyloučit už na základě grafu.

Dataset	Počet vstupů	Největší chyba	Průměrná chyba	MSE
Brno	5	7,60 hPa	3,90 hPa	16,5658
	3	7,74 hPa	4,18 hPa	18,5315
Rožnov	5	13,73 hPa	7,32 hPa	56,2906
	3	8,24 hPa	6,94 hPa	50,3060

Tab. 3. Chyby predikce tlaku na 1 den

Tentokrát je výsledek ovšem jiný než u teploty a vlhkosti. Jako lepší se jeví síť natrénovaná daty dlouhodobými, tedy z několika let. Dalším překvapením je lepší výsledek při využití rychlosti a směru větru (tedy 5 vstupů do sítě), než při jejich absenci. Není to rozdíl markantní, ale každé zlepšení je ku prospěchu. Pro předpovědi tlaku by tedy stálo za to považovat o delších datasetech, kterými budeme síť trénovat. Pokud už by byl k dispozici pouze dataset z několika dní, je dle výsledků lepší vítr do sítě nepřidávat, protože z krátkodobého hlediska způsobuje problémy.

6 METODY OPTIMALIZACE PŘEDPOVĚDI DLE METEOROLOGICKÝCH DAT

V následujících podkapitolách budou předloženy a testovány různé přístupy optimalizace předpovědi. Bude se jednat zejména o experimentování s velikostí sítě, využití většího objemu dat, popřípadě dat z delšího časového období, i když bude jejich objem podobný. Dále například experimenty s normalizací dat před sestavením sítě a také spojení dat z více stanic do jedné sítě opět ve snaze zpřesnit základní předpověď. Poslední část se navíc věnuje pokusu o předpovězení hodnot všech 3 veličin pouze pomocí 1 neuronové sítě. V rámci následujících experimentů byly prováděny především předpovědi meteorologických veličin na vzdálenost 1 hodiny, pokud není v konkrétní podkapitole stanoveno jinak.

6.1 Vliv velikosti sítě na kvalitu předpovědi

V předchozí části byly sestavovány sítě o 4 vstupních neuronech, 1 výstupním a o standardním počtu 10 neuronů ve skryté vrstvě (vyjma kapitoly 5.2.3, kde bylo použito 5 neuronů, jelikož síť s deseti neurony nefungovala dobře, a kromě všech podkapitol kapitoly 5.3, kde bylo také nutné optimalizovat již během prvotního vývoje, protože při standardním nastavení nebyly získány dobré výsledky). Protože je snahou najít co nejoptimálnější výsledky, byl proveden test, jak se bude kvalita předpovědi měnit při použití různého počtu neuronů, tedy různé velikosti sítě. Změna velikosti má vliv na počet vah v síti, do kterých se mohou změny sítě během trénování promítnout. Experimenty byly prováděny se sítěmi s počtem 8, 10, 15 a 20 neuronů. U výsledků byla sledována hodnota MSE, pomocí které lze získat představu o rozložení chyb. Jelikož byly nejlepší výsledky získány z hodinové předpovědi, kde byla testovací data vybrána jako každý šestistý vzorek naměřených dat, byly experimenty s velikostí sítě prováděny povětšinou pro tento případ. Případné výjimky jsou uvedeny.

6.1.1 Vliv velikosti sítě na kvalitu předpovědi teploty

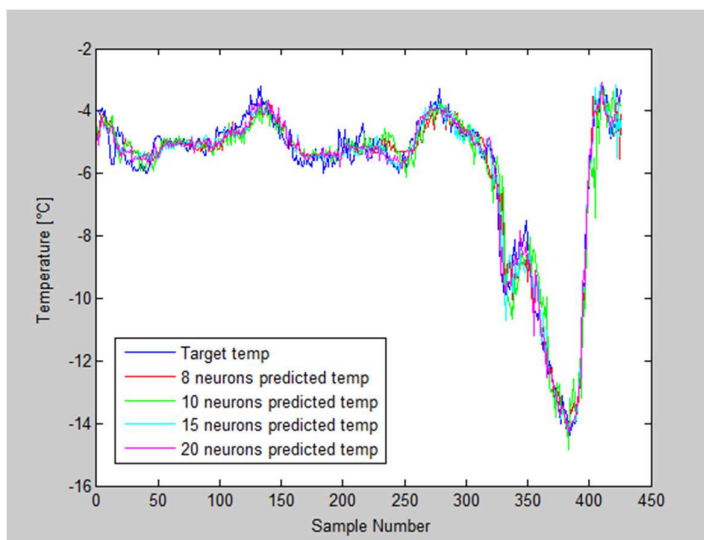
Trénování všech případů trvalo poměrně dlouho (v řádu hodin). To je dáno velikostí trénovacího datasetu, který obsahoval 255 174 hodnot každé ze 4 vstupních veličin a rychlostí výpočetního stroje, který výpočet prováděl. Z tabulky (Tab. 4) lze vidět, že výsledek s nejmenší chybou (nejmenší MSE) byl získán při použití 20 neuronů ve skryté vrstvě neuronové sítě. To ovšem není jediný ukazatel kvality natrénované sítě, neboť síť

může být *přetrénována*. To je stav, kdy je síť již natolik adaptována na trénovací data, že není schopná reagovat na data jiná, která při trénování neviděla (testovací data).

Počet neuronů	Trénování	Testování		
	MSE	MSE	Absolutní chyba (MAX)	Absolutní chyba (Prům.)
8	0,2495	0,2573	2,88 °C	0,39 °C
10	0,4159	0,4398	3,91 °C	0,47 °C
15	0,1780	0,1924	2,20 °C	0,32 °C
20	0,1314	0,1284	1,30 °C	0,28 °C

Tab. 4. Závislost hodnoty chyby teploty na velikosti sítě

Graf (Obr. 28) potom ukazuje křivky všech předpovědí spolu se skutečným průběhem předpovězených hodnot (modrá křivka). Je patrné, že například síť s 10 neurony vykazovala občas větší odchylku od skutečné hodnoty (jedná se o drobné ústřely od modré čáry). Nejlépe naopak kopíruje modrou křivku fialová křivka (síť s 20 neurony).



Obr. 28. Srovnání předpovědí teploty při použití různého počtu neuronů

6.1.2 Vliv velikosti sítě na kvalitu předpovědi tlaku

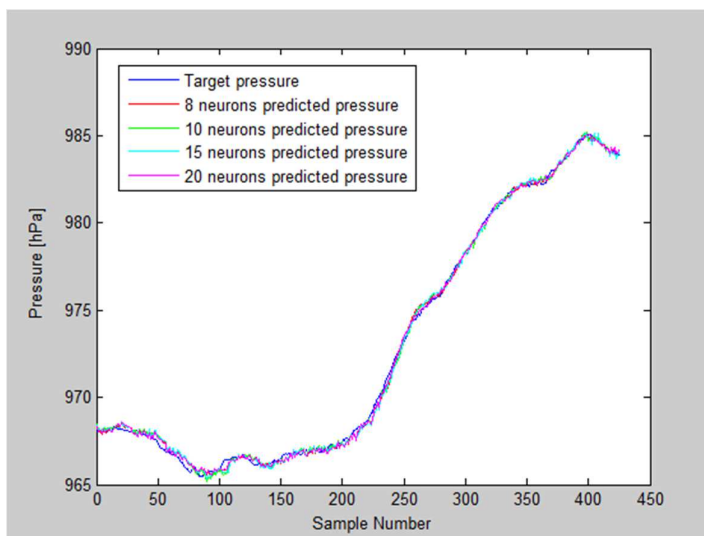
Doporučení na velikost sítě z předchozí kapitoly je platné pro teplotu. Nemusí ale nutně platit i pro tlak. Proto bylo experimentováno i s velikostmi sítí pro tuto veličinu. Bude

zajímavé zjistit, zda lze velikost sítě doporučit obecně pro všechny veličiny, nebo zda budou vyžadovat různý přístup. Experiment byl založen na hodinové předpovědi se třemi dny trénovacích dat a vybranými vzorky pro množinu testovacích.

Počet neuronů	Trénování	Testování		
	MSE	MSE	Absolutní chyba (MAX)	Absolutní chyba (Prům.)
8	0,4726	0,0506	0,69 hPa	0,18 hPa
10	0,4890	0,0629	0,79 hPa	0,20 hPa
15	0,4933	0,0628	0,76 hPa	0,20 hPa
20	0,4633	0,0419	0,66 hPa	0,16 hPa

Tab. 5. Závislost hodnoty chyby tlaku na velikosti sítě

Délka trénování byla opět poměrně dlouhá, natrénovat všechny 4 varianty sítě trvalo cca 6 – 8 hodin. Bohužel přesný čas není v tomto případě k dispozici. Dataset obsahoval 255 174 hodnot každé ze 4 vstupních veličin. Tabulka (Tab. 5) ukazuje hodnoty absolutní chyby a MSE pro jednotlivé varianty sítě. Je vidět, že s rostoucím počtem neuronů se chyba zvětšuje, ale nakonec pro 20 neuronů je nejlepší (nejmenší). Opět, stejně jako u teploty, je to pravděpodobně dáno počtem vstupních parametrů, které do sítě vstupují. Pokud je hodnota počtu neuronů dělitelná počtem vstupů, dostáváme nejlepší možné výsledky. 20neuronová síť navíc ještě přinesla mírné zlepšení chyby oproti 8neuronové a lze ji tedy v tomto případě doporučit nejvíce.



Obr. 29. Srovnání předpovědí tlaku při použití různého počtu neuronů

Graf (Obr. 29) zobrazuje průběh předpovědi tlaku a je vidět, že rozdíly jsou opravdu velmi malé. Přesto si lze všimnout drobného kolísání například 10neuronové křivky kolem stého vzorku, nebo 15neuronové kolem čtyřstého vzorku. Další dvě sítě s menšími MSE tyto kolísání neobsahují.

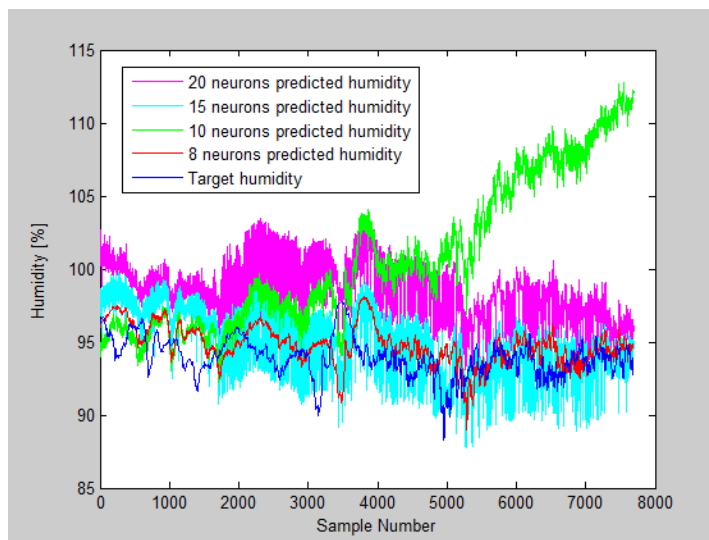
6.1.3 Vliv velikosti sítě na kvalitu předpovědi vlhkosti

Další veličinou, která zaslouží samostatný průzkum optimální velikosti sítě, je vlhkost. Pro experimenty byla použita data se 17 trénovacími dny a jedním testovacím a předpověď vlhkosti za 1 hodinu. Tentokrát nebylo využito přerozdělení dat tak, aby testovací data byly vybrané vzorky z trénovací množiny, protože to nebylo třeba. Trénovací data (17 dní) obsahovala dostatečný rozsah hodnot, jinými slovy v testovací množině (1 den) nebyly žádné hodnoty vlhkosti, které by síť před trénováním neviděla.

Počet neuronů	Trénování	Testování		
	MSE	MSE	Absolutní chyba (MAX)	Absolutní chyba (Prům.)
8	14,2374	3,7325	6,90 %	1,52 %
10	13,9768	78,2740	19,55 %	7,10 %
15	14,6879	6,3284	8,35 %	2,08 %
20	9,8103	26,1567	11,85 %	4,67 %

Tab. 6. Závislost hodnoty chyby vlhkosti na velikosti sítě

Z tabulky jde vidět, že výsledky si poněkud odporují. Při testování se zdá, že nejlepší výsledky podává síť s 8 a 15 neurony. Pokud se ale podíváme na jejich MSE během trénování, jedná se o nejhorší výsledky. Z grafu lze zase pozorovat, že větší počet neuronů způsobuje větší oscilaci v předpovědi. Poměrně pěknou křivku generuje síť s 8 neurony, nicméně všechny křivky včetně této jsou posunuté vpravo od očekávaného průběhu a kopírují spíše průběh vstupních hodnot. Jedná se o problém, který se projevil již v kapitole 5.2.3, a ani zde se nám jej nepodařilo zbavit. Zdá se tedy, že optimalizace pomocí změny velikosti sítě příliš úspěšná nebyla. Z výsledků lze vyvodit, že je lepší se vyhnout síti s 10 neurony, protože v druhé části dne přestává fungovat a vrací hodnoty vlhkosti dokonce větší než 100 %, což je samozřejmě nesmysl. Ani 20 neuronů nemá příliš přesvědčivý průběh. Jednak je hodně zašuměný a také má poměrně velké posunutí od předpokládaných hodnot, a to v celé délce.



Obr. 30. Srovnání předpovědí vlhkosti při použití různého počtu neuronů

6.2 Vliv normalizace dat na kvalitu předpovědi

Normalizací rozumíme proces transformace hodnot vstupních i výstupních veličin do jiného rozsahu, tradičně 0 až 1 nebo -1 až 1. V této práci byl využíván převod do rozsahu 0 až 1. Přepočtení je uvedeno v definičním vztahu (4). Důvodem k provádění normalizace nad používanými daty je skutečnost, že síť mívá lepší a věrohodnější odezvu, pokud vstupní proměnné mají hodnoty ve stejném rozsahu.

$$NormValue = \frac{Value - minValue}{maxValue - minValue} \quad (4)$$

Nicméně tato úprava s sebou nese určitá negativa. Například v případě nasazení bude nutno jakákoliv vstupní data před vpuštěním do sítě za účelem získání předpovědi nejprve normalizovat. Ze sítě navíc opět získáme data, která jsou v intervalu 0 až 1, takže je třeba provést tzv. de-normalizaci, abychom získali data opět v rozsahu, kterému rozumíme. De-normalizace výstupů je prováděna dle definičního vztahu (5).

$$Value = NormValue * (maxValue - minValue) + minValue \quad (5)$$

Normalizace může za určitých okolností předpověď vylepšit, ale může mít i negativní vliv. Záleží vždy na konkrétní situaci a nejlépe se to ověří experimentem.

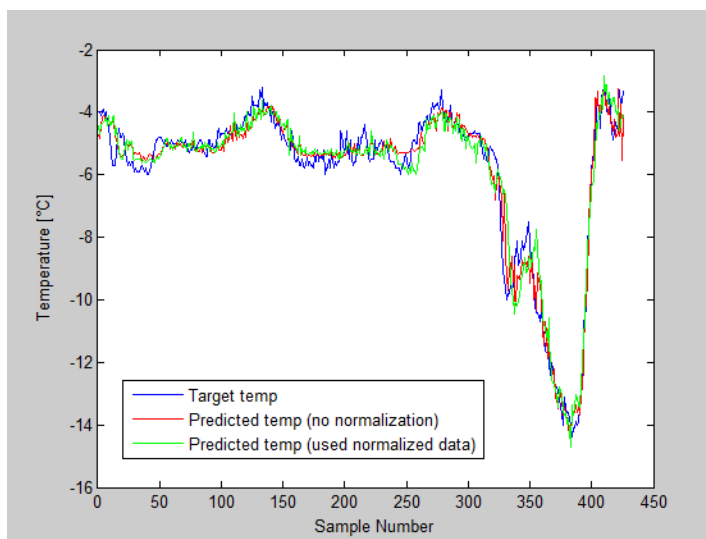
Bylo opět využito předpovědi teploty za 1 hodinu z druhé části sekce 5.2.1. Pro normalizaci byly zvoleny tyto konstanty pro maximum a minimum daných veličin:

	Teplota [°C]	Atmosf. tlak [%]	Rychlost větru [m/s]	Směr větru [°]
MAX	70	1100	60	359
MIN	-35	600	0	0

Tab. 7. Konstanty normalizace použitých veličin

6.2.1 Vliv normalizace dat na kvalitu předpovědi teploty

Opět je první část experimentů věnována teplotě. Výsledná křivka předpovězené teploty (po de-normalizaci výstupu ze sítě) je porovnána s křivkou získanou pomocí nenormalizovaných dat, obě při použití 8 neuronů ve skryté vrstvě.

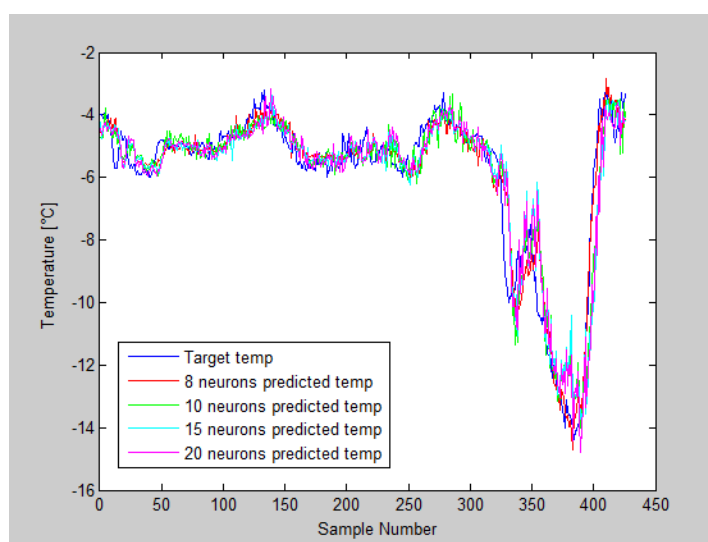


Obr. 31. Srovnání předpovědi teploty s/bez normalizace dat

Z grafu (Obr. 31) lze vidět, že rozdíl není nijak markantní. V některých částech grafu je lepší předpověď z normalizovaných dat, v některých naopak nenormalizovaných. Lze vypořadovat, že předpověď bez normalizace je více vyhlazená, což na jednu stranu umožní větší odolnost vůči chybám, ale reakce na prudkou změnu bude pomalejší. To ale bude s největší pravděpodobností způsobeno délkou trénování. Zatímco síti

s nenormalizovanými daty trvalo trénování v řádu hodin, natrénování normalizovanými daty bylo jen několik minut. Díky tomu může být v praxi umožněno využít pro trénování větší dataset (pokud bude k dispozici) než pro trénování nenormalizovanými daty, a tím dosažení větší přesnosti předpovědi.

Pro úplnost ještě následuje graf předpovědí při použití rozdílných počtů neuronů ve skryté vrstvě sítě - (Obr. 32). Situace je obdobná, jako u předpovědi bez normalizace, tedy že rozdíly nejsou markantní, ale nejvíce kopíruje reálný vývoj teploty předpověď osmi neuronovou sítí. Z tabulky (Tab. 8) jdou vyvodit stejné závěry. Nejlépe fungovala jednoznačně předpověď založená na 8 neuronech ve skryté vrstvě.



Obr. 32. Srovnání předpovědí teploty při použití různého počtu neuronů a normalizovaných dat

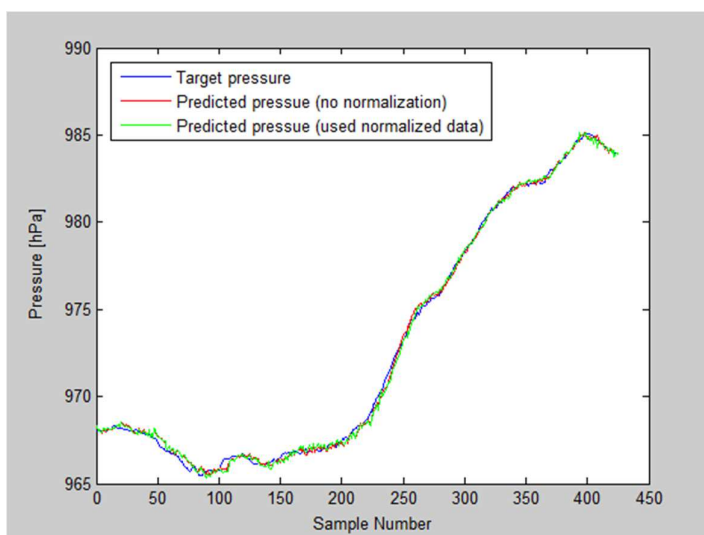
Počet neuronů	Absolutní chyba (MAX)	Absolutní chyba (Prům.)	MSE
8	2,94 °C	0,46 °C	0,2495
10	3,66 °C	0,58 °C	0,4159
15	4,39 °C	0,59 °C	0,1780
20	4,27 °C	0,58 °C	0,1314

Tab. 8. Závislost chyby teploty na velikosti sítě zpracovávající normalizovaná data

V porovnání s výsledky bez normalizace z kapitoly 6.1.1 jsou nicméně aktuální výsledky horší, takže normalizace pravděpodobně nebude u teploty krok správným směrem.

6.2.2 Vliv normalizace dat na kvalitu předpovědi tlaku

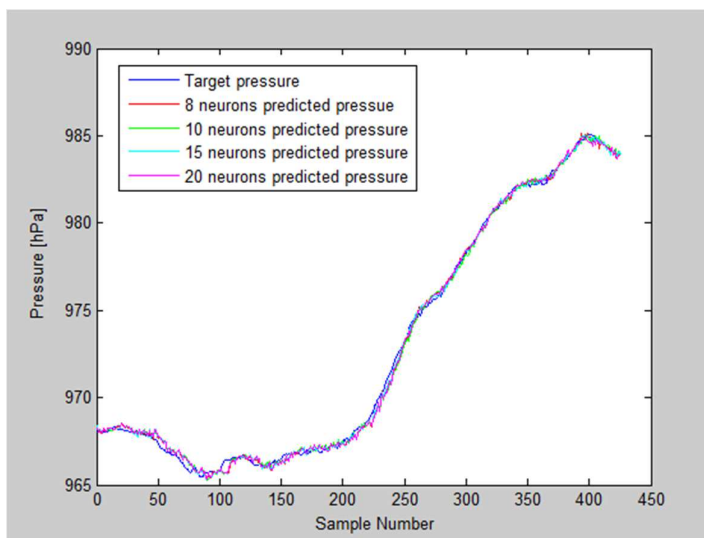
Při predikci hodnot tlaku byl u výsledných křivek poměrně dobrý výsledek již při predikci bez normalizace. Byl tedy předpoklad, že z grafu nebudou rozdíly tak viditelné a zajímavější tedy bude číselné vyjádření chyb předpovědi. Rozložení hodnot bylo opět převzato z druhé části sekce 5.2.2, stejně jako u teploty. Síť využívají 8 neuronů ve skryté vrstvě.



Obr. 33. Srovnání předpovědi tlaku s/bez normalizace dat

Obě křivky z grafu (Obr. 33) (s normalizací dat i bez normalizace) kopírují poměrně dobře skutečný vývoj veličin. Na první pohled se ale nedá říct, jestli je lepší výsledek jeden nebo druhý. Hlavní rozdíl byl opět v délce trénování sítě. Zatímco u sítě bez normalizace trvalo trénování několik hodin, u normalizovaných dat byla délka trénování několik minut, stejně jako u teploty.

Graf použití různého počtu neuronů u nenormalizovaných dat následuje pro doplnění také. Uvádím jej hlavně z toho důvodu, aby bylo možné zjistit, zda má normalizace dat vliv na to, jak velká síť dá nejlepší předpovědi veličin.



Obr. 34. Srovnání předpovědí tlaku při použití různého počtu neuronů a normalizovaných dat

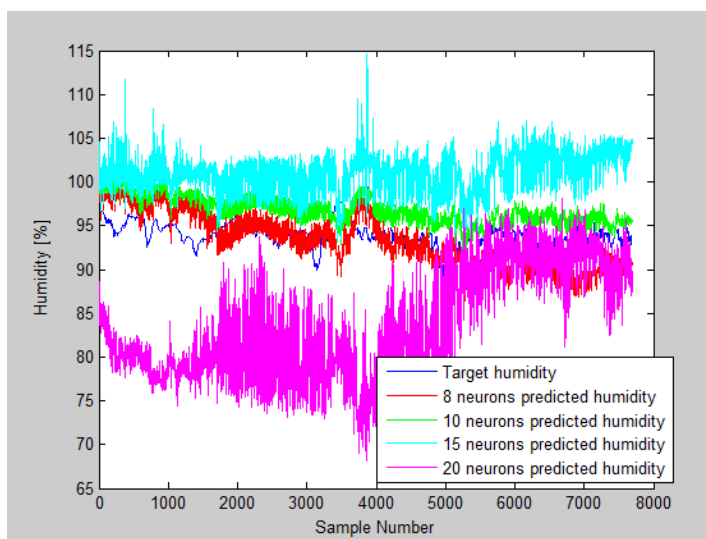
Počet neuronů	Absolutní chyba (MAX)	Absolutní chyba (Prům.)	MSE
8	0,80 hPa	0,20 hPa	1.9653e-006
10	0,76 hPa	0,21 hPa	2.1236e-006
15	0,79 hPa	0,18 hPa	1.9170e-006
20	0,79 hPa	0,20 hPa	1.9500e-006

Tab. 9. Závislost chyby tlaku na velikosti sítě zpracovávající normalizovaná data

Rozdíly opět nejsou z grafu (Obr. 34) téměř viditelné, takže se spíše zaměřím na tabulku. Co se týče chyb, tak jsou v podstatě srovnatelné. Žádná ze sestavených sítí není špatná a všechny by pro danou úlohu fungovaly dobře. Pokud porovnáme výsledky s tabulkou (Tab. 5), zjistíme, že výsledky jsou velmi podobné, a dokonce například pro 20neuronovou síť horší, než byly u nenormalizovaných dat. Hlavní přínos by tedy v tomto případě byl tedy pouze v rychlosti trénování sítě. Co se přesnosti předpovědi týče, normalizace by zde neměla přílišný smysl.

6.2.3 Vliv normalizace dat na kvalitu předpovědi vlhkosti

Pro vlhkost byla také sestavena neuronová síť, která předpokládá normalizované hodnoty. Byly převedeny do intervalu 0 až 1 podle stejného pravidla jako u teploty, pouze zde navíc přibyla vlhkost, která byla převedena tak, že hodnota 1 znamená 100% a 0 znamená 0% vlhkost.



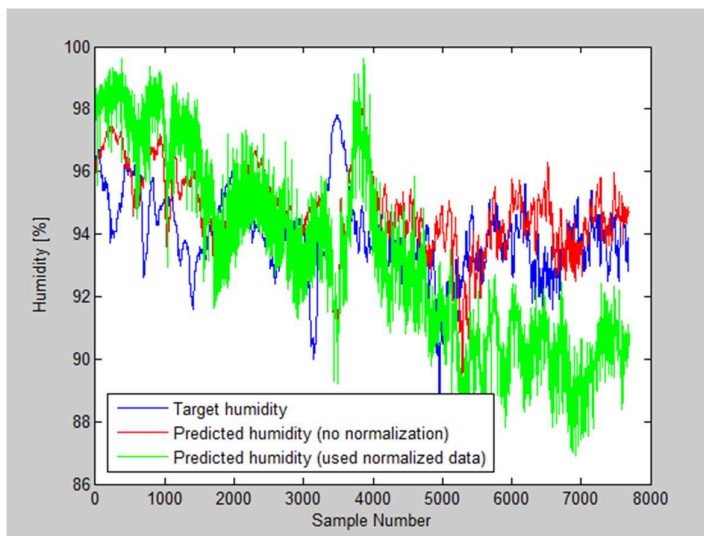
Obr. 35. Srovnání předpovědí vlhkosti při použití různého počtu neuronů a normalizovaných dat

Počet neuronů	Absolutní chyba (MAX)	Absolutní chyba (Prům.)	MSE
8	8,49 %	2,44 %	0,0015
10	7,74 %	2,98 %	0,0015
15	20,38 %	7,11 %	0,0013
20	26,40 %	10,65 %	0,0010

Tab. 10. Závislost chyby vlhkosti na velikosti sítě zpracovávající normalizovaná data

Také bylo otestováno, jak velká síť bude lepší v případě normalizovaných hodnot. Byly natrénovány sítě s 8, 10, 15 a 20 neurony. Z grafu (Obr. 35) lze vidět, že lepší výsledky dostáváme ze sítě s menším počtem neuronů, tedy 8 v tomto případě. Síť s 20 neurony je zcela špatná a síť s 15 neurony se sice liší méně, ale zase předpovídá vlhkost nad logickou

mezi 100 %. V tomto případě je tedy lépe využít menší síť. Průměrné absolutní chyby z tabulky (Tab. 10) ukazují také osmi neuronovou síť jako nejúspěšnější, přestože maximální chybu má větší než například síť s 10 neurony.



Obr. 36. Srovnání předpovědi vlhkosti s/bez normalizace dat

Graf (Obr. 36) ukazuje porovnání obou sítí s nejlepším dosaženým průběhem (síť s 8 neurony bez normalizace a stejně velká síť s normalizací). Křivku skutečného vývoje vlhkosti lépe kopíruje síť bez normalizace dat. Vzhledem k tomu, že ke zlepšení normalizací dat nedochází (spíše ke zhoršení), a že normalizace znamená další zpracování dat, které je tím pádem zbytečná operace navíc, nelze pro vlhkost normalizaci dat doporučit.

6.3 Předpověď s využitím dat z druhé meteorologické stanice

Tato sekce zkoumá další možný přístup, který by mohl kvalitu předpovědi vylepšit. Je to přidání dalších vstupů do sítě v podobě dalších naměřených hodnot. Tyto hodnoty však již nebudou pocházet ze stanice, v jejímž okolí nás předpověď zajímá, nýbrž ze stanice pomocné, která je v určité vzdálenosti od našeho bodu zájmu. I když je pomocná stanice vzdálena asi 116,6 km kilometrů vzdušnou čarou, můžou z ní naměřená data síti s přesností předpovědi pomoci.

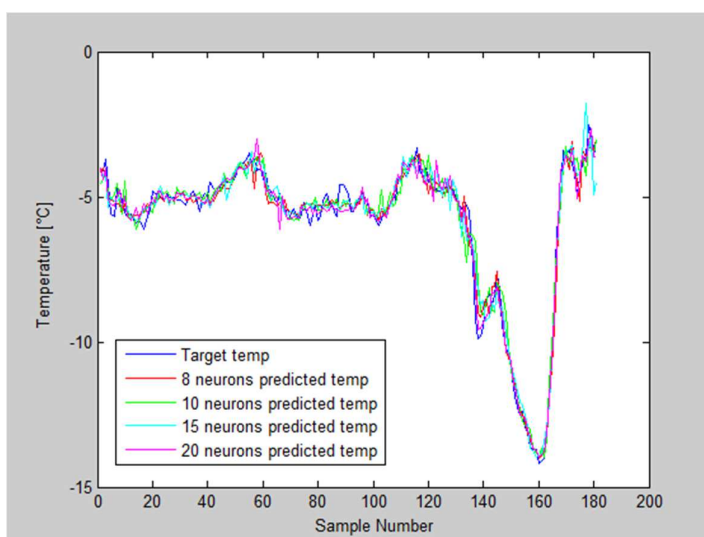
Pro tento experiment byl zvolen datový soubor od firmy Hvězdárna a Planetárium Brno p. o., která data pro účely této práce poskytla. V první řadě však bylo nutné oba datové soubory srovnat tak, aby byly vzorkovány se stejnou periodou (4 minuty, což je

perioda získaných dat z pomocné stanice) a aby obsahovaly data ze stejné doby. Předpověď veličin byla prováděna na 1 hodinu dopředu.

Experimenty byly prováděny pouze s veličinami teplota a atmosférický tlak, protože v době jejich provádění ještě nebyl k dispozici druhý dataset od firmy Navmatix s.r.o., ze kterého by bylo možné použít validní hodnotu vlhkosti. Z tohoto důvodu nemohly experimenty s vlhkostí v této části proběhnout.

6.3.1 Předpověď teploty s využitím dat z druhé meteorologické stanice

Do sítě tedy vstupovalo 8 parametrů (teplota, tlak, síla a směr větru ze dvou stanic) a výstupem je předpovězená teplota za 1 hodinu. Menší „roztřepanost“ sítě oproti grafu (Obr. 28) lze přisoudit menšímu počtu hodnot (konkrétně 181 hodnot testovacích a 905 trénovacích). K redukci hodnot došlo kvůli delší vzdálenosti mezi vzorky (4 minuty).



Obr. 37. Srovnání předpovědí teploty při použití různého počtu neuronů a dat ze 2 stanic

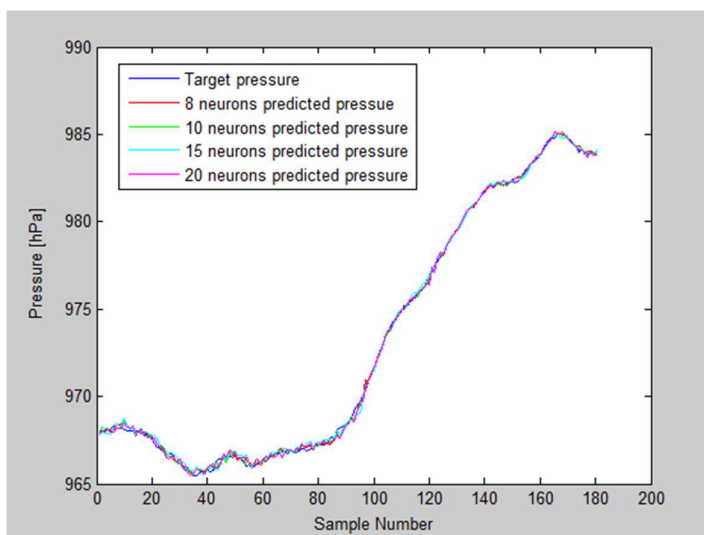
Z tabulky (Tab. 11) jde vidět, že hodnoty chyb i MSE jsou nižší než při použití dat pouze z 1 meteostanice, není to ovšem rozdíl až tak markantní. V případě, že by předpověď dle 1 stanice byla s horší úspěšností, lze očekávat, že zlepšení by bylo větší. Lze také vidět, že sítě s 8 a 20 neurony poměrně dobře kopírují skutečnou křivku. Například síť s 10 neurony obsahuje větší množství špiček, které ale přestřelují skutečnou hodnotu. Z toho lze vyvodit, že optimální počet neuronů ve skryté vrstvě bude dobré přizpůsobit počtu vstupů, jako jejich násobky, nebo alespoň půlnásobky.

Počet neuronů	2 stanice			1 stanice (pro srovnání)		
	Abs. chyba (MAX)	Abs. chyba (Prům.)	MSE	Abs. chyba (MAX)	Abs. chyba (Prům.)	MSE
8	1,36 °C	0,30 °C	0,1324	2,88 °C	0,39 °C	0,2573
10	2,65 °C	0,35 °C	0,2132	3,91 °C	0,47 °C	0,4398
15	1,85 °C	0,35 °C	0,1772	2,20 °C	0,32 °C	0,1924
20	1,21 °C	0,28 °C	0,1067	1,30 °C	0,28 °C	0,1284

Tab. 11. Závislost chyby teploty na velikosti sítě – srovnání 2 stanic s 1 stanicí

6.3.2 Předpověď tlaku s využitím dat z druhé meteorologické stanice

Další veličinou, která byla podrobena zkoumání kvality předpovědi s využitím další pomocné stanice byl atmosférický tlak. Sestavené sítě měly opět 8 vstupů a 1 výstup, tentokrát předpovězenou hodnotu tlaku za následující hodinu. Počty a rozložení vzorků byly využity z experimentu s teplotou z kapitoly 6.3.1.



Obr. 38. Srovnání předpovědí tlaku při použití různého počtu neuronů a dat ze 2 stanic

V grafu (Obr. 38) je vidět, že všechny sítě kopírují křivku skutečného vývoje s velmi dobrou přesností a rozdíl oproti grafu (Obr. 29) není na první pohled příliš zřetelný. Zajímavé proto bude především číselné vyjádření absolutních chyb.

Počet neuronů	2 stanice			1 stanice (pro srovnání)		
	Abs. chyba (MAX)	Abs. chyba (Prům.)	MSE	Abs. chyba (MAX)	Abs. chyba (Prům.)	MSE
8	0,59 hPa	0,14 hPa	0,0235	0,69 hPa	0,18 hPa	0,0506
10	0,51 hPa	0,11 hPa	0,0166	0,79 hPa	0,20 hPa	0,0629
15	0,70 hPa	0,17 hPa	0,0401	0,76 hPa	0,20 hPa	0,0628
20	0,59 hPa	0,14 hPa	0,0180	0,66 hPa	0,16 hPa	0,0419

Tab. 12. Závislost chyby tlaku na velikosti sítě – srovnání 2 stanic s 1 stanicí

Pokud porovnáme výsledky s experimentem v kapitole 6.1.2, který využíval stejná data, ale pouze z jedné stanice, zlepšení jsme dosáhli. Není sice radikální – v řádu setin, ale vzhledem k nízké chybě už při použití 1 stanice byl prostor pro zlepšení v tomto případě malý. Pokud bychom ale měli s předpovědí dle jedné stanice větší problémy, zlepšení by se projevilo značněji. Samozřejmě by se v takovém případě dalo uvažovat o více referenčních stanicích jako vstupech neuronové sítě, aby bylo dosaženo přesnosti ještě lepší.

6.4 Trénování daty z delšího časového horizontu

Jelikož nejdříve získaný datový soubor firmy Navmatix s.r.o. neobsahoval příliš dlouhý časový úsek (pouze 3 dny dat), bylo doposud učení sítě založeno na znalosti krátkodobé historie meteorologických veličin. Rozhodně však stojí za zkoušku zjistit, jak se bude chovat síť, která bude natrénována daty z delšího časového horizontu. Pro tento pokus byl použit dataset od organizace Hvězdárna a Planetárium Brno p.o. Pro trénování sítě byla použita data z prvních 2 let datasetu a pro testování byla určena data z dalších 2 let. Pro zobrazení se ovšem tak dlouhá testovací data ukázala jako nevhodná, takže byl jejich počet zmenšen na 1000 vzorků, což je při periodě dat 4 minuty něco přes 1 den. Vzhledem k získané zkušenosti, že je dobré volit počet neuronů jako násobky vstupních dat, byly sestaveny 2 sítě s 8 a 12 neurony ve skryté vrstvě pro předpověď teploty (4 vstupní veličiny). Pro předpověď vlhkosti a tlaku již bylo využito všech 5 veličin (přidána vlhkost), takže sítě byly poté sestavovány s 5 a 10 neurony ve skryté vrstvě. Předpovědi veličin byly prováděny na 1 hodinu dopředu.

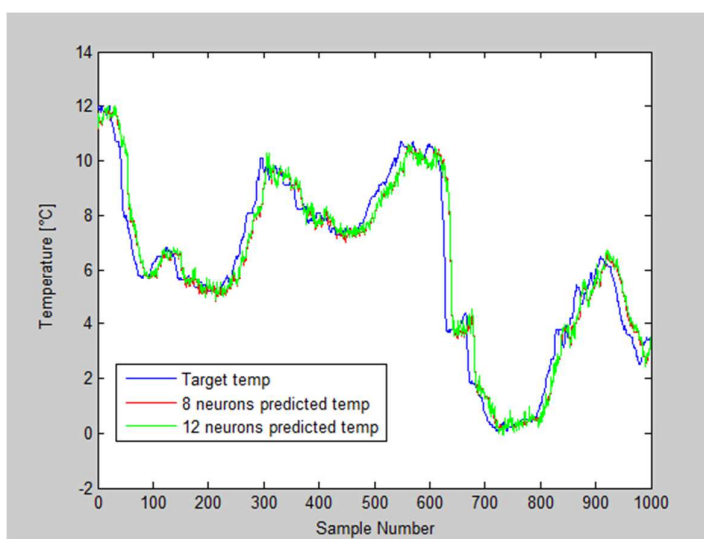
6.4.1 Předpověď teploty při použití dat z delšího období

Z grafu (Obr. 39) jde vidět drobné posunutí předpovědi o několik vzorků vpravo oproti očekávané hodnotě. Toto posunutí je pravděpodobně způsobeno výpadky, které jsou v datasetu přítomny. To dokládá i fakt, že při použití krátkého datasetu bez výpadků v předchozích kapitolách nebylo posunutí patrné, ačkoli je využit stejný přístup. Kromě této skutečnosti je ale získaný průběh teploty velmi dobrý a v porovnání s grafy zachycující průběhy teplot při trénování krátkým datasetem věrněji kopíruje průběh očekávaný. Vzhledem k posunutí jsou hodnoty MSE v tabulce (Tab. 13) trochu zavádějící, a pokud by křivky lícovaly, byly by mnohem nižší.

Co se týče počtu neuronů, jde vidět, že zvětšení jejich počtu zde nemá příliš výrazný vliv. Lze se tedy spokojit s 8 neurony a získat díky tomu kratší čas trénování.

Počet neuronů	Trénování	Testování		
	MSE	MSE	Absolutní chyba (MAX)	Absolutní chyba (Prům.)
8	1,2716	0,7186	5,63 °C	0,55 °C
12	1,2604	0,7318	5,55 °C	0,56 °C

Tab. 13. Chyby teploty u sítě trénované daty z delšího časového horizontu



Obr. 39. Průběh předpovědi teploty na 1 h dopředu při použití 2 let trénovacích dat

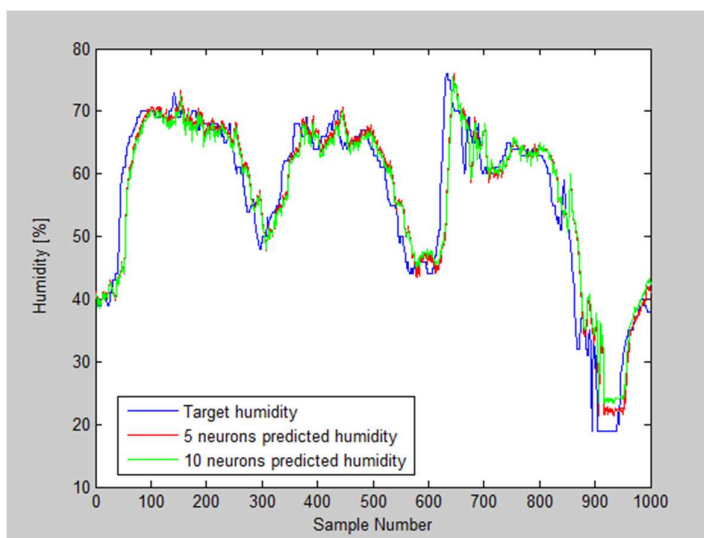
6.4.2 Předpověď vlhkosti při použití dat z delšího období

V grafu (Obr. 40) můžeme opět vidět určité posunutí, což podporuje domněnku z kapitoly 6.4.1, že posunutí může být způsobeno výpadky, které jsou v datasetu přítomny, protože jednotlivé vzorky jsou u vlhkosti i teploty stejné (sejmuté ve stejný čas), a tudíž by i posunutí mělo v takovém případě být stejné. Jinak ale oba výstupy kopírují skutečný průběh vlhkosti velmi dobře a rozdíly mezi oběma sítěmi s různým počtem neuronů také nejsou nijak významné.

Maximální velikosti chyb zachycené v tabulce (Tab. 14) jsou poměrně velké, což způsobují části prudkého stoupaní vlhkosti v grafu, kde posunutím vzorků vzniká v konkrétním čase velký rozdíl. Průměrná chyba je ovšem poměrně dobrá. Dataset bez výpadků (nebo vhodně upravený) měl vyloučit posunutí a tím pádem aktuální extrémní chyb. I průměrné chyby by tak mohly být ještě mnohem menší.

Počet neuronů	Trénování	Testování		
	MSE	MSE	Absolutní chyba (MAX)	Absolutní chyba (Prům.)
5	19,4188	22,7123	23,96 %	3,13 %
10	19,0658	23,4613	23,74 %	3,25 %

Tab. 14. Chyby vlhkosti u sítě trénované daty z delšího časového horizontu



Obr. 40. Průběh předpovědi vlhkosti na 1 h dopředu při použití 2 let trénovacích dat

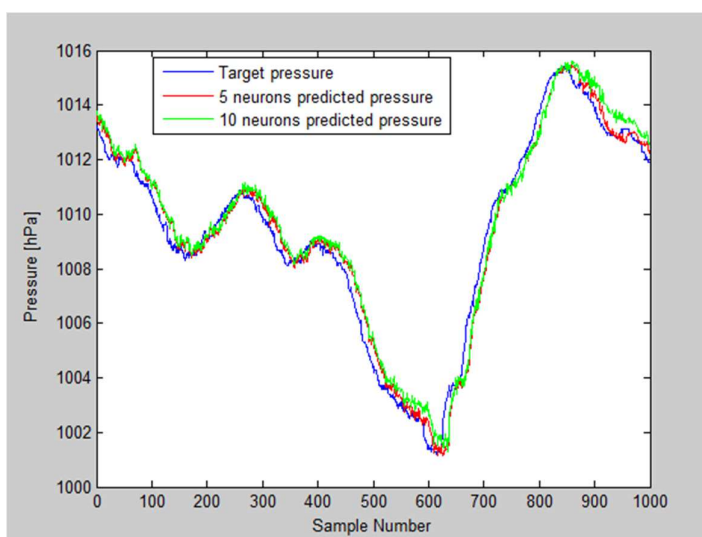
6.4.3 Předpověď tlaku při použití dat z delšího období

Průběhy předpovědí tlaku jsou zachyceny v grafu (Obr. 41). Posunutí předpovězených hodnot oproti skutečným je zde, stejně jako u teploty a vlhkosti, viditelné. Opět je podpořena domněnka o výpadcích v rámci datového použitého souboru z kapitoly 6.4.1. Kromě tohoto posunutí průběhy opět dobře kopírují tvar křivky skutečného vývoje tlaku. Rozdíly mezi výstupy sítě s 5 neurony a s 10 neurony jsou malé až na poslední klesající úsek grafu, kde se drobně odchýlily.

Z tabulky (Tab. 15) jde vidět, že obě sítě dávají chybu v podstatě stejnou. Jelikož je důležitější průměrná chyba, lze se přiklonit spíše k variantě s 5 neurony, kterou podporuje také kratší čas trénování sítě. Opět by se dalo hovořit o možném zlepšení v případě datasetu bez výpadků, to však již bylo dostatečně probráno v předchozích kapitolách.

Počet neuronů	Trénování	Testování		
	MSE	MSE	Absolutní chyba (MAX)	Absolutní chyba (Prům.)
5	0,2175	0,3062	2,09 hPa	0,43 hPa
10	0,2121	0,4136	2,03 hPa	0,54 hPa

Tab. 15. Chyby tlaku u sítě trénované daty z delšího časového horizontu



Obr. 41. Průběh předpovědi tlaku na 1 h dopředu při použití 2 let trénovacích dat

6.5 Využití GNSS dat jako další vstup do sítě

GNSS data byla předzpracována softwarem RTKLIB pro získání parametru ZTD. Předpověď byla prováděna opět na 1 hodinu dopředu nejprve pouze s meteorologickými daty a následně i s využitím ZTD parametru. Následovalo porovnání obou získaných výsledků se skutečným vývojem hodnot. Síť byla sestavována se 3 neurony ve skryté vrstvě pro předpověď bez ZTD parametru a se 4 neurony při jeho použití. Rozdílný počet u obou předpovědí byl zvolen z toho důvodu, aby byl dodržen počet neuronů shodný s počtem vstupů sítě.

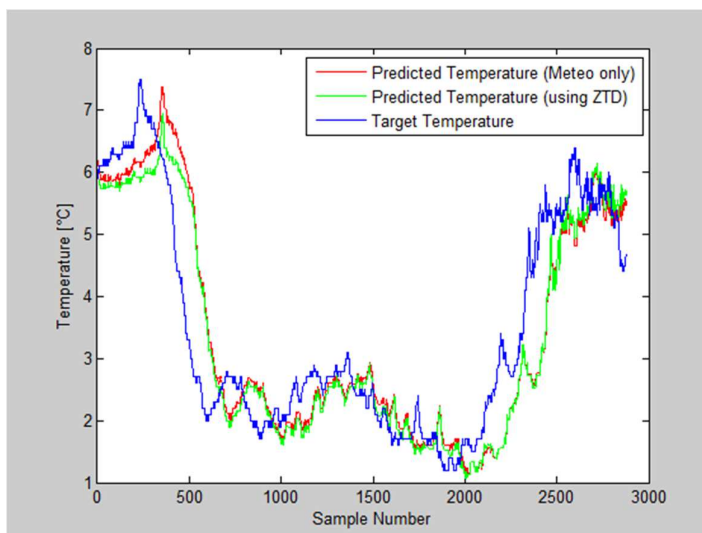
6.5.1 Předpověď teploty při použití GNSS dat

Srovnání průběhů předpovědí s využitím parametru ZTD a bez něj je v grafu (Obr. 42). Opět jde vidět posunutí oproti skutečným hodnotám, nicméně pro účely porovnání dvou předpovědí nemusí být posunutí až takový problém. Naopak, lze zkoumat, jak bude na takovýto problém reagovat síť s přidáním ZTD parametrem a která síť bude odolnější proti tomuto posunutí.

Průběh obou předpovědí je velmi podobný skutečnému, i když je vidět, že křivka předpovědi se ZTD parametrem má pozvolnější růst teploty k lokálnímu maximu v první části grafu a teplota v maximu je menší, než tomu je u skutečného vývoje. Pokud porovnáme hodnoty v tabulce (Tab. 16), je vidět zlepšení všech zkoumaných parametrů. MSE a absolutní chyba klesly o jednu až dvě desetiny, ale průměrná chyba jen o setinu. Přínos tedy využití parametru ZTD mělo, ale nikoliv zásadní.

	MSE	Absolutní chyba (MAX)	Absolutní chyba (Prům.)
Se ZTD	0,1559	2,68 °C	0,63 °C
Bez ZTD	0,2273	2,86 °C	0,64 °C

Tab. 16. Chyby teploty u sítě trénované bez / s GNSS daty



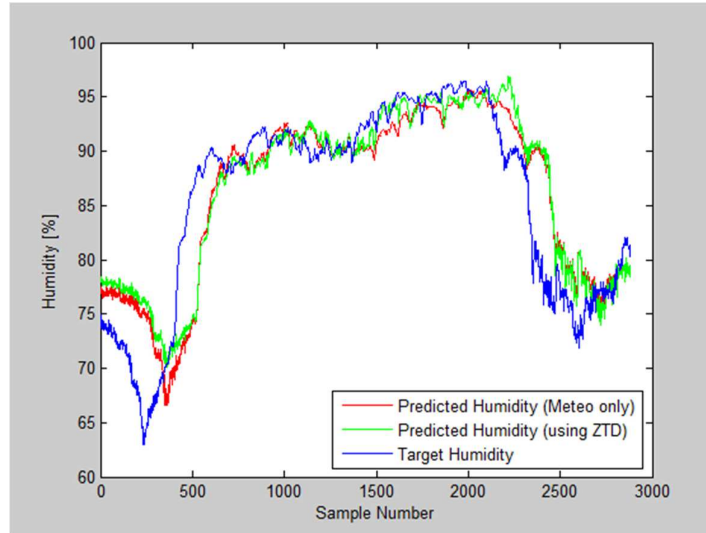
Obr. 42. Průběh předpovědi teploty na 1 h dopředu při použití GNSS dat

6.5.2 Předpověď vlhkosti při použití GNSS dat

Následuje pokus s předpovědí hodnot vlhkosti. Opět můžeme v grafu na obrázku (Obr. 43) pozorovat posunutí oproti hodnotám skutečným, ale hlavně jdou vidět drobné rozdíly obou předpovědí v extrémech. Na první pohled to tedy vypadá, že budou výsledky stejné, jako u předpovědi teploty z předchozí kapitoly, z tabulky (Tab. 17) ovšem plyne, že výsledky s využitím parametru ZTD jsou spíše horší. Došlo ke zhoršení všech parametrů a u maximální absolutní chyby dokonce téměř o jedno procento.

	MSE	Absolutní chyba (MAX)	Absolutní chyba (Prům.)
Se ZTD	9,7266	14,46 %	3,10 %
Bez ZTD	10,8315	13,69 %	3,04 %

Tab. 17. Chyby vlhkosti u sítě trénované bez / s GNSS daty



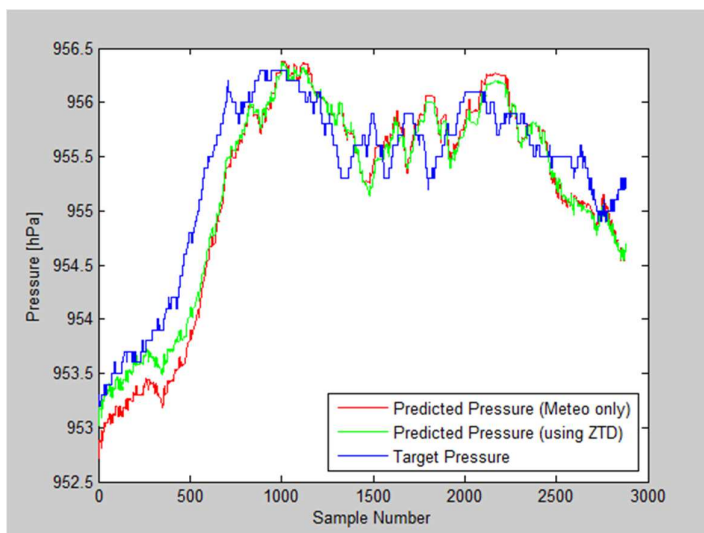
Obr. 43. Průběh předpovědi vlhkosti na 1 h dopředu při použití GNSS dat

6.5.3 Předpověď tlaku při použití GNSS dat

Poslední část byla věnována tlaku. Opět je z grafu (Obr. 44) patrné posunutí hodnot od skutečného vývoje podrobněji zmiňované v předchozích kapitolách. Zvláštností oproti grafům teploty a vlhkosti je nyní malý rozdíl předpovězených hodnot při použití parametru ZTD a bez něj. Pouze na začátku grafu je znát lepší průběh křivky předpovědi se ZTD parametrem, ale poté se průběhy obou předpovědí víceméně srovnaly. V tabulce (Tab. 18) je vidět, že parametr ZTD ke zlepšení pomohl, ale stejně jako v kapitole 6.5.1 to nebylo zlepšení velké, nýbrž v řádu setin až desetin.

	MSE	Absolutní chyba (MAX)	Absolutní chyba (Prům.)
Se ZTD	0,0776	0,99 hPa	0,29 hPa
Bez ZTD	0,1032	1,11 hPa	0,34 hPa

Tab. 18. Chyby tlaku u sítě trénované bez / s GNSS daty

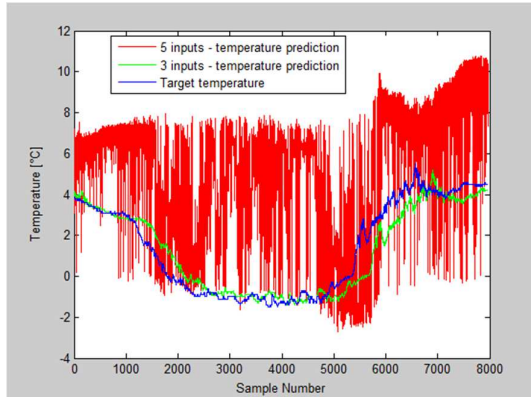


Obr. 44. Průběh předpovědi tlaku na 1 h dopředu při použití GNSS dat

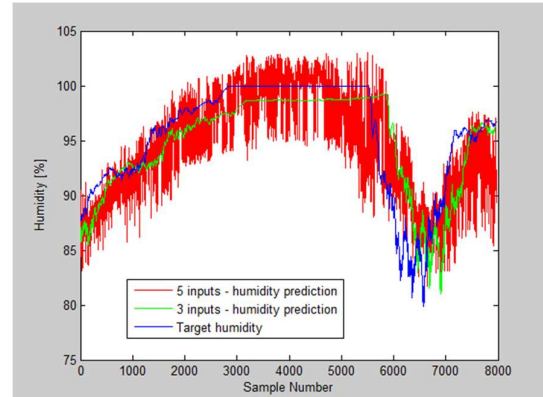
6.6 Předpověď všech veličin pomocí 1 sítě

Tento přístup se zdá jako nejjednodušší řešení, proto bylo také na místě vyzkoušet, jak bude fungovat. Jako vstupní data byly použity hodnoty 5 známých veličin (rychlost větru, směr větru, atmosférický tlak, vlhkost a teplota). Byl použit druhý (delší) dataset firmy Navmatix s.r.o. a předpověď byla prováděna na vzdálenost 1 hodiny dopředu. Pro testování byl použit 1 den z datasetu a pro trénování předcházejících 18 dní. Jednotlivé vzorky měli mezi sebou vzdálenost (periodu) cca 11 sekund.

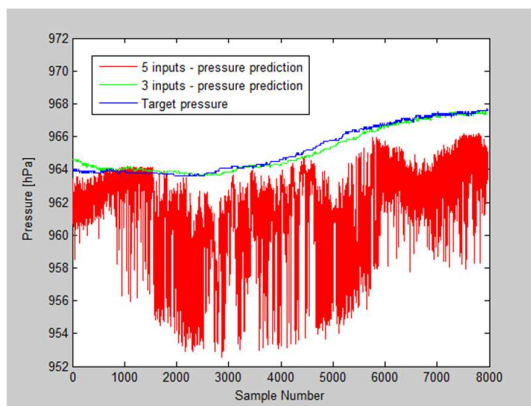
Předpovědi všech veličin byly provedeny dvakrát, pokaždé s takovou velikostí sítě, jaký byl počet vstupů a zároveň výstupů (tedy 3 a 5 skrytých neuronů). Nejdříve byly jako vstupy i výstupy využity všechny diskutované veličiny, tedy 5 vstupů i výstupů sítě, poté byly vyřazeny *rychlost a směr větru*, síť tedy měla 3 vstupy a 3 výstupy, a síť byla sestavena a natrénována znovu. Obě tyto varianty jsou v grafech (Obr. 45), (Obr. 46) a (Obr. 47) odlišeny jinou barvou křivky. Vzhledem k tomu, že u sítě s 3 vstupy a výstupy nemohly být na výstupu získávány hodnoty rychlosti a směru větru, proto je v grafech (Obr. 48) a (Obr. 49) těchto veličin zachycen pouze výstup ze sítě s 5 veličinami.



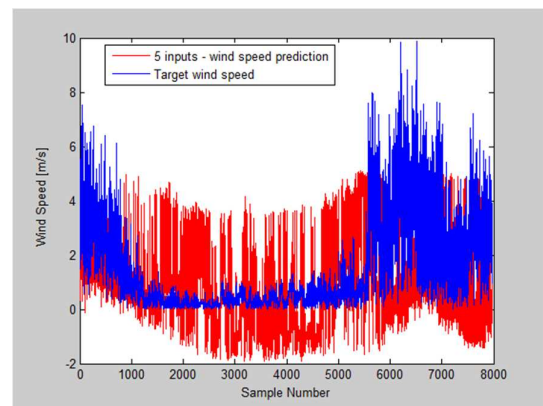
Obr. 45. Předpověď všech veličin pomocí 1 sítě – křivka teploty



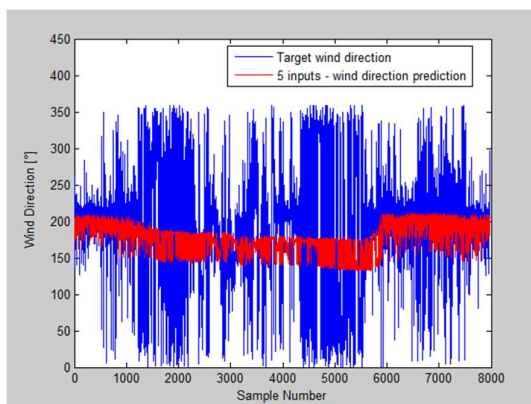
Obr. 46. Předpověď všech veličin pomocí 1 sítě – křivka vlhkosti



Obr. 47. Předpověď všech veličin pomocí 1 sítě – křivka tlaku



Obr. 48. Předpověď všech veličin pomocí 1 sítě – křivka rychlosti větru



Obr. 49. Předpověď všech veličin pomocí 1 sítě – křivka směru větru

Z předložených grafů lze v první řadě usuzovat, že skutečně není dobrý nápad využívat rychlost a směr větru jako vstup do sítě, a to ani co se týče předpovězení vlastního chování těchto veličin, ani pro vylepšení předpovědi veličin jiných. Z prvních 3 grafů jde jasně porovnat, že výsledný průběh u sítě s 5 vstupy není dobrý, a i když se například u vlhkosti snaží kopírovat průběh skutečného vývoje veličiny, tak se především v zašumění nemůže rovnat výstupu ze sítě o 3 vstupech i výstupech. Nicméně jinak průběhy ze sítě o 3 veličinách velmi pěkně kopírují skutečné průběhy.

Veličina	Počet veličin v síti	Absolutní chyba (MAX)	Absolutní chyba (Prům.)	MSE - učení
Teplota	5	9,37 °C	4,17 °C	5001,3
	3	2,85 °C	0,54 °C	8,6537
Vlhkost	5	13,04 %	2,66 %	3942,9
	3	10,16 %	2,19 %	17,2073
Tlak	5	12,02 hPa	3,59 hPa	4814,6
	3	0,70 hPa	0,21 hPa	44,7594
Směr větru	5	225,47 °	60,27 °	7220,6
Rychlost větru	5	8,97 m/s	1,83 m/s	4857,0

Tab. 19. Chyby veličin u sítě predikující všechny veličiny najednou

Výsledky jsou poměrně dobré a dokonce i srovnatelné s výsledky sítí, které predikují zvlášť jednotlivé veličiny, jak je popsáno v předchozích kapitolách. Při pohledu na hodnoty MSE při učení sítě lze vidět obrovské rozdíly u sítí s 3 veličinami a s 5 veličinami. To je dáno šumem, který s sebou nesou veličiny rychlost a směr větru.

7 POPIS POUŽITÉHO VÝPOČETNÍHO SYSTÉMU

Pro všechny experimenty prováděné v rámci této práce byl použit notebook MSI GX623 s procesorem Intel Core2Duo P7350 s pracovní frekvencí 2.00GHz, vybavený 4GB RAM pamětí. Počítač byl vybaven operačním systémem Microsoft Windows 10 Professional 64bit. Pro experimenty s neuronovými sítěmi byl využit software MATLAB 7.11.0 (R2010b) od firmy MathWorks vybavený Neural Networks Toolboxem. Pro předzpracování dat byl využit z části software Microsoft Excel 2013 z kancelářského balíku Microsoft Office 365 ProPlus a z části linuxová konzole Cygwin pro MS Windows od Red Hat, Inc. doplněná o balíky SED a AWK.

8 MOŽNOSTI BUDOUCÍHO ROZŠÍŘENÍ PRÁCE

Možným rozšířením práce může být například hlubší prozkoumání využitelnosti dat ze systému GNSS. V práci využitý parametr ZTD je jedna z možných variant zpracování GNSS dat, lze ovšem využít i Integrated Water Vapour (IWV) parametr, který udává množství vodní páry v troposféře. Lze předpokládat, že by tento parametr mohl být vhodnější pro předpověď vlhkosti, než využitý ZTD.

ZÁVĚR

Předpověď počasí je v dnešní době stále významnější vědeckou oblastí, která lidem poskytuje cenné informace. V této práci byl prezentován pohled na předpověď poněkud nekonvenční cestou, a to za pomoci umělé inteligence. Na základě prezentovaných vědeckých článků byly vybrány Neuronové sítě jako dobrý kandidát na reprezentaci předpovědního modelu pro meteorologicky významné veličiny jako *atmosférický tlak*, *teplota* a *vlhkost*. Pro všechny prezentované metody byly využity všechny tyto veličiny jako vstup do sítě, aby se síť naučila na závislosti mezi těmito veličinami. Pro učení sítě byla využita optimalizovaná varianta Back-Propagation algoritmu, konkrétně Levenberg-Marquardt algoritmus. Sítě byly sestaveny pro různé intervaly předpovědi (10 minut, 1 hodina, 1 den) a následně byl vybrán interval 1 hodiny, na kterém byly zkoumány možné optimalizace předpovědi. U ostatních intervalů lze očekávat podobné závěry. Výsledky neoptimalizovaných metod lze nalézt v příslušných kapitolách.

Pro předpověď teploty bylo shledáno jako nejlepší řešení využít sítě s jedním výstupem (pouze teplota) a s 20 neurony ve skryté vrstvě neuronové sítě. Neosvědčila se normalizace dat ani využití datového souboru z dlouhého časového období (síť naučená daty z 3 dnů fungovala lépe než síť naučená daty z 2 minulých let). Osvědčilo se případné využití druhé referenční meteostanice pro zpřesnění předpovědi. V obou případech bylo dosaženo průměrné absolutní chyby 0,28 °C. Metoda se dvěma stanicemi však měla o něco menší maximální chybu (1,21 °C) oproti klasické 20neuronové síti bez optimalizace (1,30 °C). Využití 1 sítě pro předpověď všech 3 veličin dalo chybu teploty přibližně dvojnásobnou, také se tedy nejedná o optimální řešení.

Atmosférický tlak je veličina, která se mění poměrně pozvolna a tomu odpovídaly i průběhy v grafech. Nejlépe se osvědčilo využití hodnot z druhé referenční stanice jako další vstupy pro neuronovou síť. Průměrná absolutní chyba byla poté 0,11 hPa. Velký rozdíl ovšem nebyl ani při využití dat jen z 1 stanice. V takovém případě bylo dosaženo průměrné chyby 0,16 hPa. Obě zmíněné chyby byly dosaženy při použití 20 neuronů ve skryté vrstvě neuronové sítě, což byl nejlepší výsledek. Normalizace dat způsobila zhoršení výsledku, i když téměř zanedbatelné, nicméně vzhledem k tomu, že motivací k jejímu využití byla právě snaha o zlepšení, nelze ji doporučit. Totéž jde říct o metodě učení daty z dlouhého časového horizontu (chyba asi 4 krát větší). Při využití 1 sítě pro

předpověď všech veličin najednou byla chyba pouze dvojnásobná oproti metodě využití přídatné referenční stanice.

Poslední důležitou veličinou, na kterou byla práce zaměřena, je vlhkost. Vzhledem k chybějícím datům nemohla být podrobena testu na optimalizaci pomocí druhé referenční stanice, takže předložené výsledky jsou prezentovány bez této varianty. Nejlepší výsledek (průměrná absolutní chyba 1,52 %) byl získán při použití 8 neuronů ve skryté vrstvě neuronové sítě bez jakékoli úpravy dat. Data pro učení byla využita sice krátkodobá, ale z delšího období než u teploty a tlaku (17 dní). Všechny ostatní optimalizace (dlouhý datový soubor pro učení, normalizace dat, výstup všech veličin v rámci 1 sítě) poskytly horší výsledek. Vzhledem k dobrým výsledkům u predikce tlaku a teploty při využití meteorologických dat z druhé referenční stanice se dá předpokládat mírné zlepšení i u vlhkosti.

Během provádění experimentů byly také vyzorovány určité zákonitosti, které je vhodné při aplikaci sítě na meteorologická data vzít v úvahu. Mohou například vzniknout problémy, v případě zanesení veličin *rychlost větru* a *směr větru* do sítě jako další vstupy. Zahnutí těchto veličin působí problémy na výstupu sítě, kde většinou vzniká poměrně velký šum. To je s největší pravděpodobností způsobeno velmi rychlými, až nepředvídatelnými, změnami hodnot těchto veličin. Pro účely predikce proto doporučuji se těmto veličinám spíše vyhnout, nebo alespoň hodnoty před použitím vhodně upravit. Například před využitím proložit okamžité hodnoty křivkou trendu (vyššího řádu než lineárního) a využít pak těchto hodnot na křivce. Další důležitá věc je vhodné sestavení trénovací množiny. Je důležité zajistit, aby v ní byly pokud možno obsaženy hodnoty napříč celým intervalem, ve kterém se má pohybovat výsledná předpovězená hodnota. V opačném případě může dojít k tomu, že výstupní hodnoty, které by měly ležet mimo naučený rozsah, sít' ořízne právě rozsahem, který při učení „viděla“.

S využitím GNSS dat pro zpřesnění předpovědi bylo v práci experimentováno jen okrajově, nicméně získané výsledky nejsou příliš přesvědčivé. Ke zlepšení předpovědi došlo u teploty a atmosférického tlaku, zlepšení ale nebylo nikterak významné. Metoda příliš nefungovala pro předpověď vlhkosti, u které způsobila spíše zhoršení výsledků. Vzhledem k rozsáhlosti problematiky by si metoda ovšem zasloužila podrobnější výzkum v samostatné práci.

SEZNAM POUŽITÉ LITERATURY

- [1] MCCULLOCH, Warren S. a Walter PITTS. A Logical Calculus of the Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics*. 1943, **5**, 115-133.
- [2] ŠÍMA, Jiří a Roman NERUDA. *Teoretické otázky neuronových sítí*. Praha: MATFYZPRESS, 1996. ISBN 80-85863-18-9.
- [3] MALIK, Pooja, Saranjeet SINGH a Binni ARORA. A Review on Weather Forecasting Using Neural Network. *International Journal of Innovative Research & Studies* [online]. 2014, **2014**(Vol 3 5), 543-548 [cit. 2016-01-27]. Dostupné z: http://www.ijirs.com/vol3_issue-5/39.pdf
- [4] KIŞI, Özgür a Erdal UNCUOĞLU. Comparison of three back-propagation training algorithms for two case studies. *Indian Journal of Engineering & Materials Sciences* [online]. 2005, **2005**(Vol. 12), 434-442 [cit. 2016-02-23]. Dostupné z: [http://nopr.niscair.res.in/bitstream/123456789/8460/1/IJEMS%2012\(5\)%20434-442.pdf](http://nopr.niscair.res.in/bitstream/123456789/8460/1/IJEMS%2012(5)%20434-442.pdf)
- [5] SHEREEF, Kadar a Baboo SANTHOSH. A New Weather Forecasting Technique using Back Propagation Neural Network with Modified Levenberg-Marquardt Algorithm for Learning. *IJCSI International Journal of Computer Science Issues*[online]. 2011, **2011**(Vol. 8, 6, No 2), 153-160 [cit. 2016-02-23]. ISSN 1694-0814. Dostupné z: <http://ijcsi.org/papers/IJCSI-8-6-2-153-160.pdf>
- [6] MITCHELL, Tom M. *Machine learning*. Boston: McGraw-Hill, c1997. ISBN 0-07-042807-7.
- [7] MALIK, Pooja, Saranjeet SINGH a Binni ARORA. An Effective Weather Forecasting Using Neural Network. *International Journal of Emerging Engineering Research and Technology*. 2014, **2**(2), 209-212.
- [8] MAQSOOD, Imran, Muhammad Riaz KHAN a Ajith ABRAHAM. An ensemble of neural networks for weather forecasting. *Neural Computing & Applications*. 2004, **13**(2), 112–122. DOI: 10.1007/s00521-004-0413-4.
- [9] HOLČÍK, Jiří, KOMENDA, Martin. *Matematická biologie: e-learningová učebnice*. 1. vydání. Brno: Masarykova univerzita, 2015. ISBN 978-80-210-8095-9.

- [10] Levenberg–Marquardt Training. YU, Hao a Bogdan M. WILAMOWSKI. The Industrial Electronics Handbook: Intelligent systems. 2nd ed. Boca Raton: CRC Press, c2011, 12-1 to 12-16. ISBN 9781439802830.
- [11] TUČKOVÁ, Jana. *Vybrané aplikace umělých neuronových sítí při zpracování signálů*. Vyd. 1. Praha: České vysoké učení technické v Praze, 2009. ISBN 978-80-01-04229-8.
- [12] BISHOP, Christopher M. *Pattern recognition and machine learning*. New York: Springer, c2006. Information science and statistics. ISBN 0-387-31073-8.
- [13] BÍLA, Jiří. *Umělá inteligence a neuronové sítě v aplikacích*. Dot. 1. vyd. Praha: České vysoké učení technické, 1996. ISBN 80-01-01275-1.
- [14] ZELINKA, Ivan. *Umělá inteligence I: neuronové sítě a genetické algoritmy*. V Brně: VUTUM, 1998. ISBN 80-214-1163-5.
- [15] BOSE, N. K. a Ping LIANG. *Neural network fundamentals with graphs, algorithms, and applications*. N. Y.: McGraw-Hill, c1996. ISBN 0070066183.
- [16] CULCLASURE, Andrew. *Using Neural Networks to Provide Local Weather Forecasts*. Jack N. Averitt College of Graduate Studies, 2013. Paper 32. Electronic Theses & Dissertations. Georgia Southern University.
- [17] *RTKLIB: An Open Source Program Package for GNSS Positioning* [online]. GitHub: Tomoji Takasu, 2013 [cit. 2016-05-10]. Dostupné z: <http://www.rtklib.com/>
- [18] *International GNSS Service* [online]. IGS Central Bureau: IGS, 2016 [cit. 2016-05-10]. Dostupné z: https://igsceb.jpl.nasa.gov/components/prods_cb.html
- [19] GUEROVA, G., A. STOYCHEVA a I. MANAFOV. Use of GNSS Tropospheric Products And Numerical Weather Prediction Model to Study Fog in Bulgaria. *ESA Conference Bureau* [online]. 2015, , 1-4 [cit. 2016-05-10]. Dostupné z: http://congrexprojects.com/docs/default-source/15a08_session1/057_guerova.pdf
- [20] YUAN, Yubin, Witold ROHM, Suqin WU, Kefei ZHANG a Suelynn CHOY. Advanced Weather Prediction and Severe Weather Monitoring Using National GNSS CORS Infrastructure – Preliminary Results. *Geospatial Science Research 2* [online]. 2012, **2012**(2), 1-8 [cit. 2016-05-10]. Dostupné z: http://eurws.org/Vol-1328/GSR2_Yuan.pdf

SEZNAM POUŽITÝCH SYMBOLŮ A ZKRATEK

MSE	Mean Squared Error (Střední chyba umocněná na druhou)
CSV	Comma-separated values
SED	Stream Editor
AWK	Interpretovaný programovací jazyk
GNSS	Globální družicový polohový systém
RINEX	Receiver Independent Exchange Format
ZTD	Zenith Total Delay
IWV	Integrated Water Vapor
GPS	Global Positioning System
PPP	Precise Point Positioning

SEZNAM OBRÁZKŮ

Obr. 1. Biologický neuron [2].....	13
Obr. 2. Formální neuron [2].....	14
Obr. 3. Jednovrstvá síť [9].....	16
Obr. 4. Vícevrstvá (třívrstvá) síť [9]	16
Obr. 5. Trénování 10minutové předpovědi teploty	25
Obr. 6. Testování 10minutové předpovědi teploty.....	25
Obr. 7. Trénování 10minutové předpovědi tlaku	26
Obr. 8. Testování 10minutové předpovědi tlaku.....	26
Obr. 9. Testování 10minutové předpovědi tlaku (filtrovaná data).....	26
Obr. 10. Trénování 10minutové předpovědi vlhkosti	27
Obr. 11. Testování 10minutové předpovědi vlhkosti	27
Obr. 12. Trénování hodinové předpovědi teploty (2 dny).....	28
Obr. 13. Testování hodinové předpovědi teploty (1 den).....	28
Obr. 14. Trénování hodinové předpovědi teploty (3 dny).....	29
Obr. 15. Testování hodinové předpovědi teploty (každý 600. vzorek).....	29
Obr. 16. Trénování hodinové předpovědi tlaku (2 dny).....	30
Obr. 17. Testování hodinové předpovědi tlaku (1 den).....	30
Obr. 18. Trénování hodinové předpovědi tlaku (3 dny).....	30
Obr. 19. Testování hodinové předpovědi tlaku (každý 600. vzorek).....	30
Obr. 20. Trénování hodinové předpovědi vlhkosti (17 dní)	31
Obr. 21. Testování hodinové předpovědi vlhkosti (1 den).....	31
Obr. 22. Předpověď teploty za 1 den – dataset Brno	32
Obr. 23. Předpověď teploty za 1 den – dataset Rožnov	32
Obr. 24. Předpověď vlhkosti za 1 den – dataset Brno.....	33
Obr. 25. Předpověď vlhkosti za 1 den – dataset Rožnov	33
Obr. 26. Předpověď tlaku za 1 den – dataset Brno	34
Obr. 27. Předpověď tlaku za 1 den – dataset Rožnov	34
Obr. 28. Srovnání předpovědí teploty při použití různého počtu neuronů	37
Obr. 29. Srovnání předpovědí tlaku při použití různého počtu neuronů.....	39
Obr. 30. Srovnání předpovědí vlhkosti při použití různého počtu neuronů	41
Obr. 31. Srovnání předpovědi teploty s/bez normalizace dat	42

Obr. 32. Srovnání předpovědí teploty při použití různého počtu neuronů a normalizovaných dat	43
Obr. 33. Srovnání předpovědi tlaku s/bez normalizace dat	44
Obr. 34. Srovnání předpovědí tlaku při použití různého počtu neuronů a normalizovaných dat	45
Obr. 35. Srovnání předpovědí vlhkosti při použití různého počtu neuronů a normalizovaných dat	46
Obr. 36. Srovnání předpovědi vlhkosti s/bez normalizace dat	47
Obr. 37. Srovnání předpovědí teploty při použití různého počtu neuronů a dat ze 2 stanic.....	48
Obr. 38. Srovnání předpovědí tlaku při použití různého počtu neuronů a dat ze 2 stanic.....	49
Obr. 39. Průběh předpovědi teploty na 1 h dopředu při použití 2 let trénovacích dat	51
Obr. 40. Průběh předpovědi vlhkosti na 1 h dopředu při použití 2 let trénovacích dat	52
Obr. 41. Průběh předpovědi tlaku na 1 h dopředu při použití 2 let trénovacích dat	53
Obr. 42. Průběh předpovědi teploty na 1 h dopředu při použití GNSS dat.....	55
Obr. 43. Průběh předpovědi vlhkosti na 1 h dopředu při použití GNSS dat	56
Obr. 44. Průběh předpovědi tlaku na 1 h dopředu při použití GNSS dat	57
Obr. 45. Předpověď všech veličin pomocí 1 sítě – křivka teploty	58
Obr. 46. Předpověď všech veličin pomocí 1 sítě – křivka vlhkosti	58
Obr. 47. Předpověď všech veličin pomocí 1 sítě – křivka tlaku	58
Obr. 48. Předpověď všech veličin pomocí 1 sítě – křivka rychlosti větru	58
Obr. 49. Předpověď všech veličin pomocí 1 sítě – křivka směru větru	58

SEZNAM TABULEK

Tab. 1. Chyby predikce teploty na 1 den.....	33
Tab. 2. Chyby predikce vlhkosti na 1 den	34
Tab. 3. Chyby predikce tlaku na 1 den.....	35
Tab. 4. Závislost hodnoty chyby teploty na velikosti sítě	37
Tab. 5. Závislost hodnoty chyby tlaku na velikosti sítě	38
Tab. 6. Závislost hodnoty chyby vlhkosti na velikosti sítě	40
Tab. 7. Konstanty normalizace použitých veličin.....	42
Tab. 8. Závislost chyby teploty na velikosti sítě zpracovávající normalizovaná data	43
Tab. 9. Závislost chyby tlaku na velikosti sítě zpracovávající normalizovaná data	45
Tab. 10. Závislost chyby vlhkosti na velikosti sítě zpracovávající normalizovaná data.....	46
Tab. 11. Závislost chyby teploty na velikosti sítě – srovnání 2 stanic s 1 stanicí.....	49
Tab. 12. Závislost chyby tlaku na velikosti sítě – srovnání 2 stanic s 1 stanicí.....	50
Tab. 13. Chyby teploty u sítě trénované daty z delšího časového horizontu.....	51
Tab. 14. Chyby vlhkosti u sítě trénované daty z delšího časového horizontu.....	52
Tab. 15. Chyby tlaku u sítě trénované daty z delšího časového horizontu.....	53
Tab. 16. Chyby teploty u sítě trénované bez / s GNSS daty.....	54
Tab. 17. Chyby vlhkosti u sítě trénované bez / s GNSS daty	55
Tab. 18. Chyby tlaku u sítě trénované bez / s GNSS daty.....	56
Tab. 19. Chyby veličin u sítě predikující všechny veličiny najednou.....	59

SEZNAM PŘÍLOH

- PI Ukázkový skript předpovědi teploty s GUI pro MATLAB software a Neural Network toolbox na CD.